Successive Wyner–Ziv Coding Scheme and Its Application to the Quadratic Gaussian CEO Problem

Jun Chen, *Member, IEEE*, and Toby Berger, *Fellow, IEEE*

***Abstract*—In this paper, we introduce a distributed source coding scheme called successive Wyner–Ziv coding. We show that every point in the rate region of the quadratic Gaussian CEO problem can be achieved via successive Wyner–Ziv coding. The concept of successive refinement in single source coding is generalized to the distributed source coding scenario, which we refer to as distributed successive refinement. For the quadratic Gaussian CEO problem, we establish a necessary and sufficient condition for distributed successive refinement, where the successive Wyner–Ziv coding scheme plays an important role.**

***Index Terms*—CEO problem, contra-polymatroid, rate splitting, source splitting, successive refinement, Wyner–Ziv coding.**

I. Introduction

The problem of distributed source coding has assumed renewed interest in recent years. Many practical compression schemes have been proposed for Slepian–Wolf coding (e.g., [1], [2] and the reference therein) and Wyner–Ziv coding (e.g., [3] and the reference therein), whose performances are close to the fundamental theoretical bounds [4][5]. Therefore it is of interest to reduce the general distributed source coding problem to these well-studied cases.

Given L independent and identically distributed (i.i.d.) discrete sources $X_1, X_2, \ldots, X_L$, the Slepian–Wolf rate region is the union of all the rate vectors $(R_1, R_2, \ldots, R_L)$ satisfying

$$\sum_{i\in\mathcal{A}} R_i \geq H\left(X_{\mathcal{A}} \,|\, X_{\mathcal{I}_L\setminus\mathcal{A}}\right) \qquad \forall \text{ nonempty set } \mathcal{A}\subseteq\mathcal{I}_L$$

where $\mathcal{I}_L = \{1,2,\ldots,L\}$ and $X_{\mathcal{A}} = (X_i)_{i\in\mathcal{A}}$. The Slepian–Wolf reigon is a contra-polymatroid[1] with $L!$ vertices [7][8]. Specifically, if π is a permutation on $\mathcal{I}_L$, define the vector $(R_1(\pi), R_2(\pi), \ldots, R_L(\pi))$ by

$$R_{\pi(i)}(\pi) = H\left(X_{\pi(i)} \,|\, X_{\pi(i+1)}, \ldots, X_{\pi(L)}\right), \quad i = 1, \ldots, L-1$$

$$R_{\pi(L)}(\pi) = H\left(X_{\pi(L)}\right).$$

Then $(R_1(\pi), R_2(\pi), \ldots, R_L(\pi))$ is a vertex of the Slepian–Wolf region for every permutation π. It is known that vertices of the Slepian–Wolf region can be achieved with complexity which is significantly lower than that of a general point. It was observed in [9] that by splitting a source into two virtual sources one can reduce the problem of coding an arbitrary point in an L-dimensional Slepian–Wolf region to that of coding a vertex of a $(2L-1)$-dimensional Slepian–Wolf region. The source-splitting approach was also adopted in distributed lossy source coding [10]. In the distributed lossy source coding scenario, we shall refer to source splitting as quantization splitting, since it is the quantization output, not the source, that gets split. Finally, we point out that the source-splitting idea has a dual in the problem of coding for multiple access channels, that is referred to as rate-splitting [11]–[14].

The portion of this paper following this introductory section is divided into three sections numbered II, III, and IV. In Section II, we introduce a low complexity successive Wyner–Ziv coding scheme and prove that any point in the rate region of the quadratic Gaussian CEO problem can be achieved via this scheme. The duality between superposition coding in multiaccess communication and successive Wyner–Ziv coding is briefly discussed. The concept of distributed successive refinement is introduced in Section III. The quadratic Gaussian CEO problem is used as an example, and the necessary and sufficient condition for distributed successive refinement is established. We conclude the paper in Section IV.

We use boldfaced letters to indicate (n-dimensional) vectors, capital letters for random objects, and small letters for their realizations. For example, we let $\mathbf{X} = (X(1), \ldots, X(n))^T$ and $\mathbf{x} = (x(1), \ldots, x(n))^T$. Calligraphic letters are used to indicate a set (say, $\mathcal{A}$). We use $U_{\mathcal{A}}$ to denote the vector $(U_i)_{i\in\mathcal{A}}$ with index i in increasing order and use $U_{\mathcal{A},\mathcal{B}}$ to denote $(U_{\mathcal{A},j})_{j\in\mathcal{B}}$.[2] For example, if $\mathcal{A} = \mathcal{B} = \{1,2\}$, then $U_{\mathcal{A}} = (U_1, U_2)$ and $U_{\mathcal{A},\mathcal{B}} = (U_{1,1}, U_{2,1}, U_{1,2}, U_{2,2})$. Here U_i and $U_{i,j}$ can be random variables, constants or functions. We let $U_{\mathcal{A}}$ be a constant if $\mathcal{A}$ is an empty set. We use $\mathcal{I}_K$ to denote the set $\{1,2,\ldots,K\}$ for any positive integer K, and use $\mathbb{R}_+^K$ to denote the set of K-dimensional vectors with nonnegative entries. Throughout this paper, the logarithm function is to the base e unless specified otherwise.

II. Successive Wyner–Ziv Coding Scheme

A. The CEO Problem and Successive Wyner–Ziv Coding

We adopt the model of the CEO problem which has been studied for many years [15]–[17]. However, some of our results also hold for many other distributed source coding models. Here is a brief description of the CEO problem (also see Fig. 1).

Manuscript received April 16, 2006; revised September 12, 2007.

J. Chen is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (e-mail: junchen@ece.mcmaster.ca).

T. Berger is with the Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22904 USA (e-mail: tb6n@virginia.edu).

Communicated by M. Effros, Associate Editor for Source Coding.

Digital Object Identifier 10.1109/TIT.2008.917687

[1]See [6] for the definition of polymatroid and contra-polymatroid.

[2]Here the elements of $\mathcal{A}$ and $\mathcal{B}$ are assumed to be nonnegative integers.

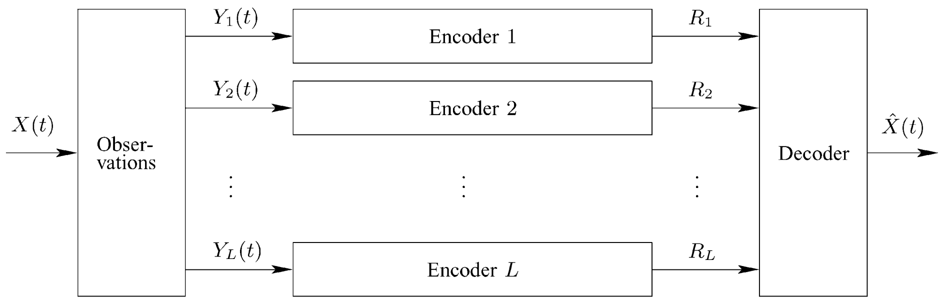

Fig. 1. The CEO problem.

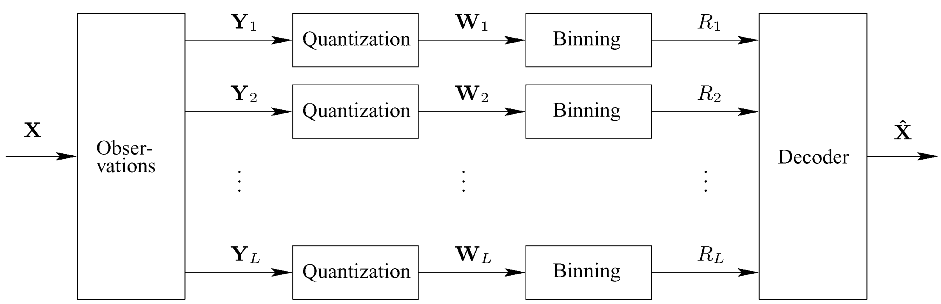

Fig. 2. Berger–Tung coding.

Let $\{X(t), Y_1(t), \ldots, Y_L(t)\}_{t=1}^{\infty}$ be a temporally memoryless source with instantaneous joint probability distribution $p(x, y_1, \ldots, y_L)$ on $\mathcal{X} \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_L$, where $\mathcal{X}$ is the common alphabet of the random variables $X(t)$ for $t = 1, 2, \ldots$, and $\mathcal{Y}_i$ $(i = 1, 2, \ldots, L)$ is the common alphabet of the random variables $Y_i(t)$ for $t = 1, 2, \ldots$. $\{X(t)\}_{t=1}^{\infty}$ is the target data sequence that the decoder is interested in. This data sequence cannot be observed directly. L encoders are deployed, where encoder i observes $\{Y_i(t)\}_{t=1}^{\infty}, i = 1, 2, \ldots, L$. The data rate at which encoder i $(i = 1, 2, \ldots, L)$ may communicate information about its observations to the decoder is limited to R_i nats per second. The encoders are not permitted to communicate with each other. Finally, the decision $\{\hat{X}(t)\}_{t=1}^{\infty}$ is computed from the combined data at the decoder so that a desired fidelity constraint can be satisfied.

Definition 2.1: An L-tuple of rates $R_{\mathcal{I}_L}$ is said to be D-admissible if for each $\epsilon > 0$, there exists an n_0 such that for all $n > n_0$ there exist encoders

$$f_i^{(n)} : \mathcal{Y}_i^n \to \mathcal{C}_i^{(n)}, \qquad i = 1, 2, \ldots, L$$

and a decoder

$$g^{(n)} : \prod_{i=1}^{L} \mathcal{C}_i^{(n)} \to \mathcal{X}^n$$

such that

$$\frac{1}{n} \log \left| \mathcal{C}_i^{(n)} \right| \leq R_i + \epsilon, \qquad i = 1, 2, \ldots, L$$
$$\frac{1}{n} \mathbb{E} \left[\sum_{t=1}^{n} d(X(t), \hat{X}(t)) \right] \leq D + \epsilon$$

where $\hat{\mathbf{X}} = g^{(n)}(f_1^{(n)}(\mathbf{Y}_1), \ldots, f_L^{(n)}(\mathbf{Y}_L))$ and $d(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \to [0, d_{\max}]$ is a given distortion measure. We use $\mathcal{R}(D)$ to denote the set of all D-admissible rate tuples.

Definition 2.2 (Berger–Tung Rate Region): Let

$$\mathcal{R}(W_{\mathcal{I}_L}) = \left\{ R_{\mathcal{I}_L} : \sum_{i \in \mathcal{A}} R_i \geq I(Y_{\mathcal{A}}; W_{\mathcal{A}} | W_{\mathcal{I}_L \setminus \mathcal{A}}), \right.$$
$$\left. \forall \text{ nonempty set } \mathcal{A} \subseteq \mathcal{I}_L \right\} \quad (1)$$

where $W_i \to Y_i \to (X, Y_{\mathcal{I}_L \setminus \{i\}}, W_{\mathcal{I}_L \setminus \{i\}})$ form a Markov chain for all $i \in \mathcal{I}_L$. The Berger-Tung rate region with respect to distortion D is

$$\mathcal{R}_{\mathrm{BT}}(D) = \operatorname{conv} \left(\bigcup_{W_{\mathcal{I}_L} \in \mathcal{W}(D)} \mathcal{R}(W_{\mathcal{I}_L}) \right)$$

where $\mathcal{W}(D)$ is the set of all $W_{\mathcal{I}_L}$ satisfying the following properties:

1) $W_i \to Y_i \to (X, Y_{\mathcal{I}_L \setminus \{i\}}, W_{\mathcal{I}_L \setminus \{i\}})$ form a Markov chain 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 $\mathbb{E} d(X, \hat{X}) \leq D$, where $\hat{X} = f(W_{\mathcal{I}_L})$.

Remark: The auxiliary random variable W_i can be interpreted as a quantized version (or a description) of Y_i, $i = 1, 2, \ldots, L$. The coding scheme associated with the Berger–Tung rate region is depicted in Fig. 2.

It was shown in [18]–[20] that $\mathcal{R}_{\mathrm{BT}}(D) \subseteq \mathcal{R}(D)$. The Berger–Tung rate region is the largest known achievable rate region for the general CEO problem although it was shown by

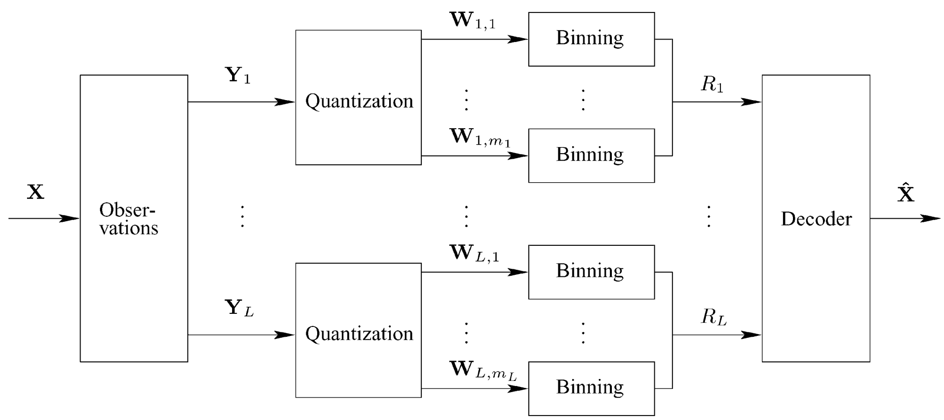

Fig. 3. Successive Wyner–Ziv coding.

Körner and Marton [21] that it is not always tight. Computing the Berger–Tung rate region involves complicated optimization and convexification. Hence, we shall focus only on $\mathcal{R}(W_{\mathcal{I}_L})$. We will see that for the quadratic Gaussian CEO problem, the properties of the Berger-Tung rate region are determined completely by those of $\mathcal{R}(W_{\mathcal{I}_L})$.

It was proved in [22][23] that $\mathcal{R}(W_{\mathcal{I}_L})$ is a contra-polymatroid with $L!$ vertices. Specifically, if π is a permutation on $\mathcal{I}_L$, define the vector $R_{\mathcal{I}_L}(\pi)$ by

$$\begin{aligned} R_{\pi(i)}(\pi) &= I(Y_{\pi(i)}; W_{\pi(i)} \,|\, W_{\pi(i+1)}, \dots, W_{\pi(L)}), \\ &\qquad\qquad\qquad\qquad i = 1, \dots, L-1 \\ R_{\pi(L)}(\pi) &= I(Y_{\pi(L)}; W_{\pi(L)}). \end{aligned}$$

Then $R_{\mathcal{I}_L}(\pi)$ is a vertex of $\mathcal{R}(W_{\mathcal{I}_L})$ for every permutation π. The dominant face of $\mathcal{R}(W_{\mathcal{I}_L})$ is the convex polytope consisting of all points $R_{\mathcal{I}_L} \in \mathcal{R}(W_{\mathcal{I}_L})$ such that $\sum_{i=1}^{L} R_i = I(Y_{\mathcal{I}_L}; W_{\mathcal{I}_L})$. Every rate tuple $R_{\mathcal{I}_L}$ on the dominant face of $\mathcal{R}(W_{\mathcal{I}_L})$ has the property that

$$R'_{\mathcal{I}_L} \le R_{\mathcal{I}_L} \Rightarrow R'_{\mathcal{I}_L} = R_{\mathcal{I}_L} \qquad \forall\, R'_{\mathcal{I}_L} \in \mathcal{R}\left(W_{\mathcal{I}_L}\right)$$

where $R'_{\mathcal{I}_L} \le R_{\mathcal{I}_L}$ means $R'_i \le R_i$ for all $i \in \mathcal{I}_L$. It is easy to verify that the vertices of $\mathcal{R}(W_{\mathcal{I}_L})$ are on the dominant face. For each vertex $R_{\mathcal{I}_L}(\pi)$, there exists a low-complexity successive Wyner–Ziv coding scheme which can be roughly described as follows.

1) Encoder $\pi(L)$ employs conventional lossy source coding. Encoder $\pi(i)$ $(i = L-1, L-2, \dots, 1)$ employs Wyner–Ziv coding with side information $\mathbf{W}_{\pi(i+1)}, \dots, \mathbf{W}_{\pi(L)}$ at the decoder.
2) The decoder first decodes the codeword $\mathbf{W}_{\pi(L)}$ from encoder $\pi(L)$, then successively decodes the codeword $\mathbf{W}_{\pi(i)}$ $(i = L-1, L-2, \dots, 1)$ from encoder $\pi(i)$ with side information $\mathbf{W}_{\pi(i+1)}, \dots, \mathbf{W}_{\pi(L)}$.

Rate tuples on the dominant face other than these $L!$ vertices were previously known to be attainable only by one of two methods. The first method known to achieve these difficult rate tuples is time sharing between vertices. This approach can require as many as L successive decoding schemes,[3] each scheme requiring L decoding steps. The second approach to achieve these rate tuples is joint decoding. This is quite difficult to implement in practice since random codes have a decoding complexity of the order of $2^{n\left[\sum_{i=1}^{L} I(Y_i;W_i)-I(Y_{\mathcal{I}_L};W_{\mathcal{I}_L})\right]}$, where n is the block length.

We will show that any rate tuple in $\mathcal{R}(W_{\mathcal{I}_L})$ can be achieved by a low-complexity successive Wyner–Ziv coding scheme with at most $2L-1$ steps. Without loss of generality, we only need to consider rate tuples on the dominant face of $\mathcal{R}(W_{\mathcal{I}_L})$. Before proceeding to prove this result, we shall first give a formal description of the general successive Wyner–Ziv coding scheme (see Fig. 3). The main idea of successive Wyner–Ziv coding is as follows: Encoder i forms several descriptions of Y_i and bins these descriptions separately; the bin index of each description is sent to the decoder, and the decoder recovers the descriptions from all the encoders successively according to a prescribed decoding order.

Let $(W_{1,\mathcal{I}_{m_1}}, W_{2,\mathcal{I}_{m_2}}, \dots, W_{L,\mathcal{I}_{m_L}})$ be jointly distributed with the generic source variables $(X, Y_{\mathcal{I}_L})$ such that $W_{i,\mathcal{I}_{m_i}} \to Y_i \to (X, Y_{\mathcal{I}_L \backslash \{i\}}, W_{j,\mathcal{I}_{m_j}}, j \in \mathcal{I}_L \backslash \{i\})$ form a Markov chain for all $i \in \mathcal{I}_L$. Let σ be a permutation on $\{W_{1,\mathcal{I}_1}, \dots, W_{L,\mathcal{I}_{m_L}}\}$ (where $\{W_{1,\mathcal{I}_1}, \dots, W_{L,\mathcal{I}_{m_L}}\} = \{W_{i,j} : i = 1, \dots, L; j = 1, \dots, m_i\}$) such that for all $i \in \mathcal{I}_L$, $W_{i,j}$ is placed before $W_{i,k}$ if $j < k$ (we refer to this type of permutation as the well-ordered permutation). Let $\{W_{i,j}\}_{\sigma}^{-}$ denote the set of random variables that appear before $W_{i,j}$ in the permutation σ.

Random Binning at Encoder i: In what follows we shall adopt the notation and conventions of [25]. Let n-vectors $\mathbf{W}_{i,1}(1), \dots, \mathbf{W}_{i,1}(M_{i,1})$ be drawn independently according to a uniform distribution over the set $T_{\epsilon}(W_{i,1})$ of ϵ-typical $W_{i,1}$ n-vectors, where $M_{i,1} = \lfloor 2^{n(I(Y_i,W_{i,1})+\epsilon'_{i,1})} \rfloor$. That is, $\Pr(\mathbf{W}_{i,1}(k) = \mathbf{w}_{i,1}) = 1/|T_{\epsilon}(W_{i,1})|$, if $\mathbf{w}_{i,1} \in T_{\epsilon}(W_{i,1})$,

[3]By Carathéodory's fundamental theorem [24], any point in the convex closure of a connected compact set $\mathcal{A}$ in a d-dimensional Euclidean space can be represented as a convex combination of $d+1$ or fewer points in the original set $\mathcal{A}$.

and $=0$ otherwise. Distribute these vectors into $N_{i,1}$ bins: $B_{i,1}(1),\ldots,B_{i,1}(N_{i,1})$, such that

$$\left\lfloor \frac{M_{i,1}}{N_{i,1}} \right\rfloor \le |B_{i,1}(b)|_{W_{i,1}} \le \left\lceil \frac{M_{i,1}}{N_{i,1}} \right\rceil, \qquad b=1,2,\ldots,N_{i,1}$$

where $N_{i,1} = \lfloor 2^{(nI(Y_i,W_{i,1}|\{W_{i,1}\}_\sigma^-)+\epsilon_{i,1})} \rfloor$ and $|B_{i,1}(b)|_{W_{i,1}}$ denotes the number of $\mathbf{W}_{i,1}$-vectors in $B_{i,1}(b)$.

Successively from $j=2, j=3,\ldots$, to $j=m_i$, for each vector $(k_1,\ldots,k_{j-1})$ with $k_s \in \{1,2,\ldots,M_{i,s}\}$, $s=1,\ldots,j-1$, let

$$\mathbf{W}_{i,j}(k_1,\ldots,k_{j-1},1),\ldots,\mathbf{W}_{i,j}(k_1,\ldots,k_{j-1},M_{i,j})$$

be drawn i.i.d. according to a uniform distribution over the set $T_\epsilon(W_{i,j}\,|\,\mathbf{w}_{i,1}(k_1),\ldots,\mathbf{w}_{i,j-1}(k_1,\ldots,k_{j-1}))$ of conditionally ϵ-typical $W_{i,j}$ n-vectors, conditioned on $\mathbf{w}_{i,1}(k_1),\ldots,\mathbf{w}_{i,j-1}(k_1,\ldots,k_{j-1})$, and distribute them uniformly into $N_{i,j}$ bins: $B_{i,j}(1),\ldots,B_{i,j}(N_{i,j})$ such that

$$\left\lfloor \frac{M_{i,j}}{N_{i,j}} \right\rfloor \le |B_{i,j}(b)|_{W_{i,j}} \le \left\lceil \frac{M_{i,j}}{N_{i,j}} \right\rceil, \quad b=1,2,\ldots,N_{i,j}.$$

Here $M_{i,j} = \lfloor 2^{n(I(Y_i,W_{i,j}|W_{i,\mathcal{I}_{j-1}})+\epsilon'_{i,j})} \rfloor, N_{i,j} = \lfloor 2^{n(I(Y_i,W_{i,j}|\{W_{i,j}\}_\sigma^-)+\epsilon_{i,j})} \rfloor$. Note: $\epsilon_{i,j},\epsilon'_{i,j} (i\in\mathcal{I}_L, j\in\mathcal{I}_{m_i})$ are positive numbers of the same order as ϵ which can be made arbitrarily close to zero as $n\to\infty$. Furthermore, we require $\epsilon_{i,j} > \epsilon'_{i,j}$ for all $i\in\mathcal{I}_L, j\in\mathcal{I}_{m_i}$.

Encoding at Encoder i: Given a $\mathbf{y}_i\in\mathcal{Y}_i^n$, find, if possible, a vector $(k^*_{i,1},\ldots,k^*_{i,m_i})$ such that

$$\left(\mathbf{y}_i,\mathbf{w}_{i,1}(k^*_{i,1}),\mathbf{w}_{i,2}(k^*_{i,1},k^*_{i,2}),\ldots,\mathbf{w}_{i,m_i}\left(k^*_{i,1},\ldots,k^*_{i,m_i}\right)\right) \in T_\epsilon\left(Y_i,W_{i,1},W_{i,2},\ldots,W_{i,m_i}\right).$$

Then find bins $B_{i,1}(b^*_{i,1}), B_{i,2}(b^*_{i,2}),\ldots,B_{i,m_i}(b^*_{i,m_i})$ such that $B_{i,j}(b^*_{i,j})$ contains $\mathbf{w}_{i,j}(k^*_{i,1},\ldots,k^*_{i,j}), j=1,2,\ldots,m_i$. Send $(b^*_{i,1},\ldots,b^*_{i,m_i})$ to the decoder. If no such $(k^*_{i,1},\ldots,k^*_{i,m_i})$ exists, simply send $(0,\ldots,0)$.

We can see the resulting transmission rate of encoder i is

$$\begin{aligned} R_i &= \frac{1}{n}\log\left(\prod_{j=1}^{m_i} N_{i,j}+1\right) \\ &\le \sum_{j=1}^{m_i} I(Y_i,W_{i,j}\,|\,\{W_{i,j}\}_\sigma^-) + \sum_{j=1}^{m_i}\epsilon_{i,j} + \frac{1}{n}. \end{aligned} \qquad (2)$$

Decoding: Given $(b^*_{i,1},\ldots,b^*_{i,m_i})$ for all $i\in\mathcal{I}_L$, if $(b^*_{i,1},\ldots,b^*_{i,m_i}) = (0,\ldots,0)$ for some i, declare a decoding failure. Otherwise decode as follows:

Let $\sigma(j)$ denote the jth element in permutation σ. Let $s_1(j), s_2(j)$ be the first and second subscript of $\sigma(j)$, respectively. For example, if $\sigma(j)=W_{3,2}$, then $s_1(j)=3, s_2(j)=2$. The decoder first finds $\mathbf{w}_{s_1(1),s_2(1)}(\hat{k}_{s_1(1),s_2(1)})$ in $B_{s_1(1),s_2(1)}(b^*_{s_1(1),s_2(1)})$. Note: $s_2(1)=1$. Since $B_{s_1(1),s_2(1)}(b^*_{s_1(1),s_2(1)})$ contains at most one vector, we have $\hat{k}_{s_1(1),s_2(1)} = k^*_{s_1(1),s_2(1)}$. Successively from $j=2, j=3,\ldots$, to $j=\sum_{i=1}^L m_i$, if in $B_{s_1(j),s_2(j)}(b^*_{s_1(j),s_2(j)})$, there exists a unique $\hat{k}_{s_1(j),s_2(j)}$ such that

$$\left(\mathbf{w}_{s_1(i),s_2(i)}\left(\hat{k}_{s_1(i),1},\hat{k}_{s_1(i),2},\ldots,\hat{k}_{s_1(i),s_2(i)}\right), i\in\mathcal{I}_j\right) \in T_{\epsilon'}\left(W_{s_1(i),s_2(i)}, i\in\mathcal{I}_j\right)$$

decode $\mathbf{w}_{s_1(j),s_2(j)}(\hat{k}_{s_1(j),1},\hat{k}_{s_1(j),2},\ldots,\hat{k}_{s_1(j),s_2(j)})$; otherwise declare a decoding failure. Note: ϵ' is of the same order as ϵ which can be made arbitrarily close to zero as $n\to\infty$.

By the standard technique, it can be shown that $\Pr(\hat{k}_{i,j} = k^*_{i,j}, \forall i\in\mathcal{I}_L, j\in\mathcal{I}_{m_i}) \to 1$ as $n\to\infty$. Furthermore, by the Markov Lemma [18], we have

$$\Pr\left(\left(\mathbf{X},\mathbf{W}_{i,j}(k^*_{i,1},k^*_{i,2},\ldots,k^*_{i,j}), i\in\mathcal{I}_L, j\in\mathcal{I}_{m_i}\right) \in T_{\epsilon'}\left(X,W_{i,j}, i\in\mathcal{I}_L, j\in\mathcal{I}_{m_i}\right)\right) \to 1$$

as $n\to\infty$. Hence for any function $g:\prod_{i=1}^L\prod_{j=1}^{m_i}\mathcal{W}_{i,j}\to\mathcal{X}$, we have

$$\begin{aligned} &\frac{1}{n}\mathbb{E}\left[\sum_{t=1}^n d(X(t), g(W_{i,j}(k^*_{i,1},k^*_{i,2},\ldots,k^*_{i,j},t), i\in\mathcal{I}_L, j\in\mathcal{I}_{m_i}))\right] \\ &\quad \le \mathbb{E}d\left(X, g(W_{i,j}, i\in\mathcal{I}_L, j\in\mathcal{I}_{m_i})\right) + \epsilon'' d_{\max} \end{aligned}$$

where $W_{i,j}(k^*_{i,1},k^*_{i,2},\ldots,k^*_{i,j},t)$ is the tth entry of $\mathbf{W}_{i,j}(k^*_{i,1},k^*_{i,2},\ldots,k^*_{i,j})$ and ϵ'' is of the same order as ϵ which can be made arbitrarily close to zero as $n\to\infty$.

It is easy to see that if we let $W'_{i,j} = W_{i,\mathcal{I}_j} (\forall\, i\in\mathcal{I}_L, j\in\mathcal{I}_{m_i})$, and replace $W_{i,j}$ by $W'_{i,j}$ in (2), R_i is unaffected. Hence there is no loss of generality to assume that $W_{i,1}\to W_{i,2}\to\cdots\to W_{i,m_i}\to Y_i$ form a Markov chain for all $i\in\mathcal{I}_L$. We can view $W_{i,1},W_{i,2},\ldots,W_{i,m_i}$ as descriptions of Y_i; moreover, along the direction specified by the Markov chain, the description gets finer and finer.

The above coding scheme has the following intuitive interpretation:

Encoder i first splits R_i into m_i pieces: $r_{i,j} = I(Y_i;W_{i,j}|\{W_{i,j}\}_\sigma^-), \forall j\in\mathcal{I}_{m_i}$. Then successively from $j=1, j=2,\ldots$, to $j=m_i$, it uses a Wyner–Ziv code with rate $r_{i,j}$ to convey $\mathbf{W}_{i,j}$ to the decoder which has the side information $\{\mathbf{W}_{i,j}\}_\sigma^-$. The decoder recovers $\{\mathbf{W}_{i,j}, i\in\mathcal{I}_L, j\in\mathcal{I}_{m_i}\}$ successively according to the order in the permutation σ. We can see that this scheme requires $\sum_{i=1}^L m_i$ Wyner–Ziv coding steps. Thus we call it a $\sum_{i=1}^L m_i$-successive Wyner–Ziv coding scheme. A similar successive coding strategy was developed in [26] for tree-structured sensor networks.

The successive Wyner–Ziv encoding and decoding structure of the above scheme significantly reduces the coding complexity compared with joint decoding or time sharing and makes the existing practical Wyner–Ziv coding techniques directly applicable to the more general distributed source coding scenarios. Furthermore, the successive Wyner–Ziv coding scheme possesses a certain robust property which is especially attractive in some applications. In the successive Wyner–Ziv coding scheme, encoder i essentially transmits its codeword in m_i packets. Each packet contains a sub-codeword

$\mathbf{W}_{i,j}(j \in \mathcal{I}_{m_i})$. If a packet, say packet $\mathbf{W}_{i,k}$ is lost in transmission, the decoder is still able to decode packets $\{\mathbf{W}_{i,k}\}_{\sigma}^{-}$. In contrast, the joint decoding scheme does not have this robust property since any corruption in the transmitted codewords may cause a complete failure in decoding.

We need to introduce another definition before giving a formal statement of our first theorem.

Definition 2.3: For any disjoint sets $\mathcal{A}, \mathcal{B} \subseteq \mathcal{I}_L$ ($\mathcal{A}$ is nonempty), let

$$\mathcal{R}(W_{\mathcal{A}} \mid W_{\mathcal{B}}, Z_{\mathcal{I}_L}) = \left\{ R_{\mathcal{A}} : \sum_{i \in \mathcal{S}} R_i \geq I(Y_{\mathcal{S}}; W_{\mathcal{S}} | W_{\mathcal{A} \setminus \mathcal{S}}, W_{\mathcal{B}}, Z_{\mathcal{I}_L}), \quad \forall \text{ nonempty set } \mathcal{S} \subseteq \mathcal{A} \right\}$$

where $Z_i \to W_i \to Y_i \to (X, Y_{\mathcal{I}_L \setminus \{i\}}, W_{\mathcal{I}_L \setminus \{i\}}, Z_{\mathcal{I}_L \setminus \{i\}})$ form a Markov chain for all $i \in \mathcal{I}_L$.

It is easy to verify that $\mathcal{R}(W_{\mathcal{A}} | W_{\mathcal{B}}, Z_{\mathcal{I}_L})$ is a contra-polymatroid with $|\mathcal{A}|!$ vertices. Specifically, if π is a permutation on $\mathcal{A}$, define the vector $R_{\mathcal{A}}(\pi)$ by

$$R_{\pi(i)}(\pi) = I\left(Y_{\pi(i)}; W_{\pi(i)} | W_{\pi(i+1)}, \ldots, W_{\pi(|\mathcal{A}|)}, W_{\mathcal{B}}, Z_{\mathcal{I}_L}\right), \quad i = 1, \ldots, |\mathcal{A}| - 1$$
$$R_{\pi(|\mathcal{A}|)}(\pi) = I\left(Y_{\pi(|\mathcal{A}|)}; W_{\pi(|\mathcal{A}|)} | W_{\mathcal{B}}, Z_{\mathcal{I}_L}\right).$$

Then $R_{\mathcal{A}}(\pi)$ is a vertex of $\mathcal{R}(W_{\mathcal{A}} | W_{\mathcal{B}}, Z_{\mathcal{I}_L})$ for every permutation π. The dominant face $\mathcal{D}(W_{\mathcal{A}} | W_{\mathcal{B}}, Z_{\mathcal{I}_L})$ of $\mathcal{R}(W_{\mathcal{A}} | W_{\mathcal{B}}, Z_{\mathcal{I}_L})$ is the convex polytope consisting of all points $R_{\mathcal{A}} \in \mathcal{R}(W_{\mathcal{A}} \mid W_{\mathcal{B}}, Z_{\mathcal{I}_L})$ such that $\sum_{i \in \mathcal{A}} R_i = I(Y_{\mathcal{A}}; W_{\mathcal{A}} \mid W_{\mathcal{B}}, Z_{\mathcal{I}_L})$. We have $\dim[\mathcal{D}(W_{\mathcal{A}} | W_{\mathcal{B}}, Z_{\mathcal{I}_L})] \leq |\mathcal{A}| - 1$, where $\dim[\mathcal{D}(W_{\mathcal{A}} | W_{\mathcal{B}}, Z_{\mathcal{I}_L})]$ is the dimension of $\mathcal{D}(W_{\mathcal{A}} \mid W_{\mathcal{B}}, Z)$. The equality holds when the $|\mathcal{A}|!$ vertices are all distinct. Every rate tuple $R_{\mathcal{A}} \in \mathcal{D}(W_{\mathcal{A}} \mid W_{\mathcal{B}}, Z_{\mathcal{I}_L})$ has the property that

$$R'_{\mathcal{A}} \leq R_{\mathcal{A}} \Rightarrow R'_{\mathcal{A}} = R_{\mathcal{A}} \qquad \forall\, R'_{\mathcal{A}} \in \mathcal{R}(W_{\mathcal{A}} | W_{\mathcal{B}}, Z_{\mathcal{I}_L}).$$

Theorem 2.1: For any rate tuple $R_{\mathcal{A}} \in \mathcal{D}(W_{\mathcal{A}} | W_{\mathcal{B}}, Z_{\mathcal{I}_L})$, there exist random variables $(W'_{i,1}, \ldots, W'_{i,m_i})_{i \in \mathcal{A}}$ jointly distributed with $(X, Y_{\mathcal{I}_L}, W_{\mathcal{I}_L}, Z_{\mathcal{I}_L})$ satisfying the following:

1) $(W'_{i,m_i})_{i \in \mathcal{A}} = W_{\mathcal{A}}$ (i.e., $(W'_{i,m_i})_{i \in \mathcal{A}}$ and $W_{\mathcal{A}}$ are just two different names for the same random vector);
2) $\sum_{i \in \mathcal{A}} m_i \leq |\mathcal{A}| + \dim[\mathcal{D}(W_{\mathcal{A}} | W_{\mathcal{B}}, Z_{\mathcal{I}_L})]$ and $m_i \leq 2$ for all $i \in \mathcal{A}$;
3) $Z_i \to W'_{i,1} \to W'_{i,m_i} \to Y_i \to (X, Y_{\mathcal{I}_L \setminus \{i\}}, W_{\mathcal{I}_L \setminus \{i\}}, Z_{\mathcal{I}_L \setminus \{i\}}, W'_{j,\mathcal{I}_{m_j}}, j \in \mathcal{A} \setminus \{i\})$ form a Markov chain for all $i \in \mathcal{I}_L$;

and a well-ordered permutation σ on $\{W'_{i,\mathcal{I}_{m_i}}, i \in \mathcal{A}\}$ such that

$$R_i = \sum_{j=1}^{m_i} I\left(Y_i; W'_{i,j} | \{W'_{i,j}\}_{\sigma}^{-}, W_{\mathcal{B}}, Z_{\mathcal{I}_L}\right) \qquad \forall\, i \in \mathcal{A}. \quad (3)$$

Proof: The theorem can be proved in a similar manner as in [13]. The details are omitted. □

When $\mathcal{A} = \mathcal{I}_L$ and $\mathcal{B} = \emptyset$, Theorem 2.1 says that if $\mathbf{Z}_{\mathcal{I}_L}$ is available at the decoder, then encoders $1, 2, \ldots, L$ can convey $\mathbf{W}_{\mathcal{I}_L}$ to the decoder via a $(2L - 1)$-successive Wyner–Ziv coding scheme as long as $R_{\mathcal{I}_L} \in \mathcal{R}(W_{\mathcal{I}_L} | Z_{\mathcal{I}_L})$.

It is noteworthy that $2L - 1$ is just an upper bound; for a rate tuple on the boundary of $\mathcal{D}(W_{\mathcal{I}_L} \mid Z_{\mathcal{I}_L})$, the coding complexity can be further reduced. For example, consider the case where $L = 3$. Let V_1 be the vertex corresponding to permutation $\pi_1 = (1, 2, 3)$, i.e.

$$V_1 = (I(Y_1; W_1 \mid Z_{\mathcal{I}_3}, W_2, W_3), I(Y_2; W_2 \mid Z_{\mathcal{I}_3}, W_3), I(Y_3; W_3 \mid Z_{\mathcal{I}_3})).$$

Let V_2 be the vertex corresponding to permutation $\pi_2 = (1, 3, 2)$, i.e.,

$$V_2 = (I(Y_1; W_1 \mid Z_{\mathcal{I}_3}, W_2, W_3) \; I(Y_2; W_2 \mid Z_{\mathcal{I}_3}), I(Y_3; W_3 \mid Z_{\mathcal{I}_3}, W_2)).$$

For any rate tuple $R_{\mathcal{I}_3}$ on the edge connecting V_1 and V_2, we have $R_1 = I(Y_1; W_1 \mid Z_{\mathcal{I}_3}, W_2, W_3)$. Hence encoder 1 can use a Wyner–Ziv code to convey $\mathbf{W}_1$ to the decoder if $(\mathbf{Z}_{\mathcal{I}_3}, \mathbf{W}_2, \mathbf{W}_3)$ are already available at the decoder. Since (R_2, R_3) is on the dominant face of $\mathcal{R}((W_2, W_3) | Z_{\mathcal{I}_3})$, by Theorem 2.1, encoder 2 and encoder 3 can convey $(\mathbf{W}_2, \mathbf{W}_3)$ to the decoder via a 3-successive Wyner–Ziv coding scheme if $\mathbf{Z}_{\mathcal{I}_3}$ is available to the decoder. Thus, overall it is a 4-successive Wyner–Ziv coding scheme as opposed to a 5-successive one.

In general we can imitate the approach in [27]. For $\emptyset \subset \mathcal{A} \subset \mathcal{I}_L$, define the hyperplane

$$\mathcal{H}(\mathcal{A}) = \left\{ \mathcal{R}_{\mathcal{I}_L} \in \mathcal{R}^L : \sum_{i \in \mathcal{A}} R_i = I(Y_{\mathcal{A}}; W_{\mathcal{A}} \mid Z_{\mathcal{I}_L}) \right\}$$

and let $\mathcal{F}_{\mathcal{A}} = \mathcal{H}(\mathcal{A}) \cap \mathcal{D}(W_{\mathcal{I}_L} | Z_{\mathcal{I}_L})$. If $\emptyset \subset \mathcal{A}_1 \subset \mathcal{A}_2 \subset \cdots \subset \mathcal{A}_k \subset \mathcal{I}_L$ is a telescopic sequence of subsets, then $\mathcal{F}_{\mathcal{A}_1} \cap \mathcal{F}_{\mathcal{A}_2} \cap \cdots \cap \mathcal{F}_{\mathcal{A}_k}$ is a face of $\mathcal{D}(W_{\mathcal{I}_L} | Z_{\mathcal{I}_L})$. Conversely, every face of $\mathcal{D}(W_{\mathcal{I}_L} | Z_{\mathcal{I}_L})$ can be written in this form. Let $\mathcal{B}_i = \mathcal{A}_i \setminus \mathcal{A}_{i-1}$, $i = 1, 2, \ldots, k+1$, where we set $\mathcal{A}_0 = \emptyset$ and $\mathcal{A}_{k+1} = \mathcal{I}_L$. Let Ξ be the set of permutations π on $\mathcal{I}_L$ such that

$$\left\{ \pi\left(\sum_{j=0}^{k-i} |\mathcal{B}_{k+1-j}| + 1 \right), \ldots, \pi\left(\sum_{j=0}^{k+1-i} |\mathcal{B}_{k+1-j}| \right) \right\} = \mathcal{B}_i, \quad i = 1, 2, \ldots, k+1.$$

Each permutation $\pi \in \Xi$ is associated with a vertex of $\mathcal{F}_{\mathcal{A}_1} \cap \mathcal{F}_{\mathcal{A}_2} \cap \cdots \cap \mathcal{F}_{\mathcal{A}_k}$ and vice versa. Hence, $\mathcal{F}_{\mathcal{A}_1} \cap \mathcal{F}_{\mathcal{A}_2} \cap \cdots \cap \mathcal{F}_{\mathcal{A}_k}$ has totally $|\Xi| = \prod_{i=1}^{k+1} (|\mathcal{B}_i|!)$ vertices. Moreover, we have $\dim(\mathcal{F}_{\mathcal{A}_1} \cap \mathcal{F}_{\mathcal{A}_2} \cap \cdots \cap \mathcal{F}_{\mathcal{A}_k}) \leq L - k - 1$, where the equality holds if these $|\Xi|$ vertices are all distinct. For any rate tuple $R_{\mathcal{I}_L} \in \mathcal{F}_{\mathcal{A}_1} \cap \mathcal{F}_{\mathcal{A}_2} \cap \cdots \cap \mathcal{F}_{\mathcal{A}_k}$, it is easy to verify that $\mathcal{R}_{\mathcal{B}_i}$ is on the dominant face of $\mathcal{R}(W_{\mathcal{B}_i} | W_{\bigcup_{j=1}^{i-1} \mathcal{B}_j}, Z_{\mathcal{I}_L})$, $i = 1, 2, \ldots, k+1$. Hence by successively applying Theorem 2.1, we can conclude that an $(L + \bar{L})$-successive Wyner–Ziv coding scheme is sufficient for conveying $\mathbf{W}_{\mathcal{I}_L}$ to the decoder if it has the side information $\mathbf{Z}_{\mathcal{I}_L}$, where

$$\bar{L} = \sum_{i=1}^{k+1} \dim\left[\mathcal{D}\left(W_{\mathcal{B}_i} \mid W_{\bigcup_{j=1}^{i-1} \mathcal{B}_j}, Z_{\mathcal{I}_L} \right) \right] = \dim(\mathcal{F}_{\mathcal{A}_1} \cap \mathcal{F}_{\mathcal{A}_2} \cap \cdots \cap \mathcal{F}_{\mathcal{A}_k}).$$

Corollary 2.1: Every rate tuple $R_{\mathcal{I}_L}$ on the dominant face of $\mathcal{R}(W_{\mathcal{I}_L})$ can be achieved via a K-successive Wyner–Ziv coding scheme for some $K \le 2L-1$.

Proof: Apply Theorem 2.1 with $Z_{\mathcal{I}_L}$ being a deterministic vector. □

B. Duality With Successive Superposition Coding

The successive Wyner–Ziv coding scheme has a dual in multiple access channel coding, which we refer to as the successive superposition coding scheme.

Consider an L-user discrete memoryless multiple access channel. This is defined in terms of a stochastic matrix

$$W : \mathcal{X}_1 \times \cdots \times \mathcal{X}_L \to \mathcal{Y}$$

with entries $W(y|x_1,\ldots,x_L)$ describing the probability that the channel output is y when the inputs are $x_1,\ldots,x_L$.

Now we give a brief description of the successive superposition coding scheme. Let $X_{1,\mathcal{I}_{m_1}}, X_{2,\mathcal{I}_{m_2}},\ldots,X_{L,\mathcal{I}_{m_L}}$ be independent random vectors, i.e.,

$$p\left(x_{1,\mathcal{I}_{m_1}}, x_{2,\mathcal{I}_{m_2}},\ldots,x_{L,\mathcal{I}_{m_L}}\right) = p\left(x_{1,\mathcal{I}_{m_1}}\right) p\left(x_{2,\mathcal{I}_{m_2}}\right)\cdots p\left(x_{L,\mathcal{I}_{m_L}}\right)$$

where $x_{i,j} \in \mathcal{X}_i$ for all $i \in \mathcal{I}_L, j \in \mathcal{I}_{m_i}$. Let σ be a well-ordered permutation on the set $\{X_{1,\mathcal{I}_{m_1}}, X_{2,\mathcal{I}_{m_2}},\ldots,X_{L,\mathcal{I}_{m_L}}\}$.

Encoder i: Let n-vectors $\mathbf{X}_{i,1}(1),\ldots,\mathbf{X}_{i,1}(M_{i,1})$ be drawn independently according to the marginal distribution $p(x_{i,1})$, where $M_{i,1} = \lceil 2^{n(I(X_{i,1};Y\,|\,\{X_{i,1}\}_\sigma^-)-\epsilon_{i,1})}\rceil$. Successively from $j=2, j=3,\ldots$, to $j=m_i$, for each vector $(k_1,\ldots,k_{j-1})$ with $k_s \in \{1,2,\ldots,M_{i,s}\}$ $(s=1,\ldots,j-1)$, let $\mathbf{X}_{i,j}(k_1,\ldots,k_{j-1},1),\ldots,\mathbf{X}_{i,j}(k_1,\ldots,k_{j-1},M_{i,j})$ be drawn i.i.d. according to the marginal conditional distribution $p(x_{i,j}|x_{i,1},\ldots,x_{i,j-1})$, conditioned on $\mathbf{x}_{i,1}(k_1),\ldots, \mathbf{x}_{i,j-1}(k_1,\ldots,k_{j-1})$. Here $M_{i,j} = \lceil 2^{n(I(X_{i,j};Y\,|\,\{X_{i,j}\}_\sigma^-)-\epsilon_{i,j})}\rceil$. Only $\mathbf{x}_{i,m_i}(k_1,\ldots,k_{m_i})$'s will be transmitted. Hence the resulting rate for encoder i is

$$\begin{aligned} R_i &= \frac{1}{n}\log\left(\prod_{j=1}^{m_i} M_{i,j}\right) \\ &\ge \sum_{j=1}^{m_i} I(X_{i,j};Y\,|\,\{X_{i,j}\}_\sigma^-)) - \sum_{j=1}^{m_i}\epsilon_{i,j}. \end{aligned} \quad (4)$$

Decoder: Suppose $\mathbf{x}_{i,m_i}(k_1^*,\ldots,k_{m_i}^*)$, $i \in \mathcal{I}_L$, are transmitted, which generate channel output $\mathbf{y} \in \mathcal{Y}^n$. Decoder first finds a $\hat{k}_{s_1(1),s_2(1)}$ such that $\mathbf{y}$ and $\mathbf{x}_{s_1(1),s_2(1)}(\hat{k}_{s_1(1),s_2(1)})$ are jointly typical. If there is none or more than one such $\hat{k}_{s_1(1),s_2(1)}$, declare a decoding failure. Otherwise, proceed as follows.

Successively from $j=2, j=3,\ldots$ to $j=\sum_{i=1}^L m_i$, if there exists a unique $\hat{k}_{s_1(j),s_2(j)}$ such that

$$\left(\mathbf{y}, \mathbf{x}_{s_1(i),s_2(i)}\left(\hat{k}_{s_1(i),1},\hat{k}_{s_1(i),2},\ldots,\hat{k}_{s_1(i),s_2(i)}\right), i \in \mathcal{I}_j\right) \in T_\epsilon\left(Y, X_{s_1(i),s_2(i)}, i \in \mathcal{I}_j\right)$$

decode $\mathbf{x}_{s_1(j),s_2(j)}(\hat{k}_{s_1(j),1},\hat{k}_{s_1(j),2},\ldots,\hat{k}_{s_1(j),s_2(j)})$, otherwise declare a decoding failure.

By the standard technique, it can be shown that $\Pr(\hat{k}_{i,j} = k_{i,j}^*, \forall i \in \mathcal{I}_L, j \in \mathcal{I}_{m_i}) \to 1$ as $n \to \infty$.

It is easy to see that if we let $X'_{i,j} = X_{i,\mathcal{I}_j} (\forall i \in \mathcal{I}_L, j \in \mathcal{I}_{m_i})$, and replace $X_{i,j}$ by $X'_{i,j}$ in (4), R_i is unaffected. Hence there is no loss of generality to assume that $X_{i,1} \to X_{i,2} \to \cdots \to X_{i,m_i} \to (Y, X_{j,\mathcal{I}_{m_j}}, j \in \mathcal{I}_L\backslash\{i\})$ form a Markov chain for all $i \in \mathcal{I}_L$. Intuitively, along each link of this Markov chain, a higher rate codebook is successively generated via superposition on a lower rate codebook. We refer to the above coding scheme as $\sum_{i=1}^L m_i$-successively superposition coding.

Our successive superposition coding scheme is similar to the rate-splitting scheme introduced in [13]. Actually every rate-splitting scheme can be converted into a successive superposition scheme. To see this, for each user i, let f_i be a splitting function such that $X_i = f_i(U_{i,1},U_{i,2},\ldots,U_{i,m_i})$ and let $X_{i,m_i} = X_i, X_{i,j} = U_{i,\mathcal{I}_j}, j=1,2,\ldots,m_i-1, i \in \mathcal{I}_L$. Then $X_{i,1} \to X_{i,2} \to \cdots \to X_{i,m_i} \to (Y, X_{j,\mathcal{I}_{m_j}}, j \in \mathcal{I}_L\backslash\{i\})$ form a Markov chain for all $i \in \mathcal{I}_L$. In [13] $U_{i,1},U_{i,2},\ldots,U_{i,m_i}$ are required to be independent[4]; if we remove this condition, then every successive superposition coding scheme can also be converted into a rate-splitting scheme simply by setting $U_{i,j} = X_{i,j}, \forall j \in \mathcal{I}_{m_i}$ and $f_i(U_{i,1},U_{i,2},\ldots,U_{i,m_i}) = U_{i,m_i}$.

Let

$$\mathcal{R}(X_{\mathcal{I}_L}) = \left\{ R_{\mathcal{I}_L} \in \mathbb{R}_+^L : \sum_{i\in\mathcal{A}} R_i \le I\left(X_\mathcal{A};Y|X_{\mathcal{I}_L\backslash\mathcal{A}}\right) \quad \forall \text{ nonempty set } \mathcal{A} \subseteq \mathcal{I}_L \right\}.$$

Ahlswede [28] and Liao [29] proved that

$$\mathcal{C} = \mathrm{conv}\left(\bigcup_{p(x_1)p(x_2)\cdots p(x_L)} \mathcal{R}\left(X_{\mathcal{I}_L}\right)\right)$$

where $\mathcal{C}$ is the capacity region of the synchronous multiple access channel.

It can be shown that if $p(x_1,x_2\ldots,x_L) = p(x_1) p(x_2)\cdots p(x_L)$, then $\mathcal{R}(X_{\mathcal{I}_L})$ is a polymatroid with $L!$ vertices [30], [31]. Specifically, if π is a permutation on $\mathcal{I}_L$, define the vector $R_{\mathcal{I}_L}(\pi)$ by

$$\begin{aligned} R_{\pi(i)}(\pi) &= I\left(X_{\pi(i)};Y|X_{\pi(i+1)},\ldots,X_{\pi(L)}\right), \\ &\qquad i=1,\ldots,L-1 \\ R_{\pi(L)}(\pi) &= I\left(X_{\pi(L)};Y\right). \end{aligned}$$

Then $R_{\mathcal{I}_L}(\pi)$ is a vertex of $\mathcal{R}(X_{\mathcal{I}_L})$ for every permutation π. The dominant face of $\mathcal{R}(X_{\mathcal{I}_L})$ is the convex polytope consisting of all points $R_{\mathcal{I}_L} \in \mathcal{R}(X_{\mathcal{I}_L})$ such that $\sum_{i=1}^L R_i = I(X_{\mathcal{I}_L};Y)$. Every rate tuple $R_{\mathcal{I}_L}$ on the dominant face of $\mathcal{R}(X_{\mathcal{I}_L})$ has the property that

$$R'_{\mathcal{I}_L} \ge R_{\mathcal{I}_L} \Rightarrow R'_{\mathcal{I}_L} = R_{\mathcal{I}_L} \qquad \forall R'_{\mathcal{I}_L} \in \mathcal{R}\left(X_{\mathcal{I}_L}\right).$$

[4]This independence condition is unnecessary since $U_{i,1},U_{i,2},\ldots,U_{i,m_i}$ are all controlled by user i. But this condition facilitates the codebook construction and storage, since now the high-rate codebook at each user is essentially a product of low-rate codebooks.

The following corollary is a dual result of Corollary 2.1. The proof is similar to that of Corollary 2.1 and thus omitted.

Corollary 2.2: Every rate tuple $R_{\mathcal{I}_L}$ on the dominant face of $\mathcal{R}(X_{\mathcal{I}_L})$ can be achieved via a K-successive superposition coding scheme for some $K \leq 2L-1$.

C. Application to the Quadratic Gaussian CEO Problem

Although we assumed discrete-alphabet sources and bounded distortion measure in the previous discussion, all our results can be extended to the Gaussian case with squared distortion measure along the lines of [32]–[34]. Now we proceed to study the quadratic Gaussian CEO problem [35], for which some stronger conclusions can be drawn. Let $\{X(t)\}_{t=1}^{\infty}$ be i.i.d. Gaussian random variables with mean zero and variance σ_X^2. Let $\{Y_i(t)\}_{t=1}^{\infty} = \{X(t)+N_i(t)\}_{t=1}^{\infty}$ for all $i \in \mathcal{I}_L$, where $\{N_i(t)\}_{t=1}^{\infty}$ are i.i.d. Gaussian random variables independent of $\{X(t)\}_{t=1}^{\infty}$ with mean zero and variance $\sigma_{N_i}^2$. Also, the random processes $\{N_j(t)\}_{t=1}^{\infty}$ and $\{N_k(t)\}_{t=1}^{\infty}$ are independent for $j \neq k$. For each $i \in \mathcal{I}_L$, let $W_i = Y_i + T_i$, where $T_i \sim \mathcal{N}(0, \sigma_{T_i}^2)$ is independent of $(X, Y_{I_L}, T_{\mathcal{I}_L \setminus \{i\}})$. Moreover, let

$$r_i = I(Y_i, W_i \,|\, X) = \frac{1}{2}\log\frac{\sigma_{N_i}^2 + \sigma_{T_i}^2}{\sigma_{T_i}^2} \qquad \forall\, i \in \mathcal{I}_L. \quad (5)$$

It was computed in [23] and [37] that

$$\begin{aligned}\mathcal{R}(W_{\mathcal{I}_L}) = \Bigg\{ R_{\mathcal{I}_L} : \sum_{i\in\mathcal{A}} R_i \geq \frac{1}{2} & \\ \times \log\left(\frac{\frac{1}{\sigma_X^2} + \sum_{i=1}^{L}\frac{1-\exp(-2r_i)}{\sigma_{N_i}^2}}{\frac{1}{\sigma_X^2} + \sum_{i\in\mathcal{I}_L\setminus\mathcal{A}}\frac{1-\exp(-2r_i)}{\sigma_{N_i}^2}}\right) & \\ + \sum_{i\in\mathcal{A}} r_i, \forall \text{ nonempty set } \mathcal{A} \subseteq \mathcal{I}_L \Bigg\} & \\ \triangleq \mathcal{R}(r_{\mathcal{I}_L}). & \quad (6)\end{aligned}$$

Furthermore, it was shown in [36][37] that

$$\mathcal{R}(D) = \bigcup_{r_{\mathcal{I}_L}\in\mathcal{F}(D)} \mathcal{R}(r_{\mathcal{I}_L}) \quad (7)$$

where

$$\mathcal{F}(D) = \left\{ r_{\mathcal{I}_L} \in \mathbb{R}_+^L : \frac{1}{\sigma_X^2} + \sum_{i=1}^{L}\frac{1-\exp(-2r_i)}{\sigma_{N_i}^2} \geq \frac{1}{D} \right\}. \quad (8)$$

Definition 2.4: Let $\partial\mathcal{R}(D)$ denote the boundary of $\mathcal{R}(D)$, i.e.

$$\partial\mathcal{R}(D) = \left\{ R_{\mathcal{I}_L} \in \mathcal{R}(D) : R'_{\mathcal{I}_L} \leq R_{\mathcal{I}_L} \Rightarrow R'_{\mathcal{I}_L} = R_{\mathcal{I}_L}, \forall R'_{\mathcal{I}_L} \in \mathcal{R}(D) \right\}.$$

Clearly, any rate tuple inside $\mathcal{R}(D)$ is dominated by some rate tuple in $\partial\mathcal{R}(D)$. Therefore is no loss of generality to focus on $\partial\mathcal{R}(D)$.

Now we proceed to compute $\partial\mathcal{R}(D)$ for the quadratic Gaussian CEO problem. The closed-form expression of $\partial\mathcal{R}(D)$ is hard to get. Instead, we shall characterize the supporting hyperplanes of $\mathcal{R}(D)$, since the upper envelope of their union is exactly $\partial\mathcal{R}(D)$. The supporting hyperplanes of $\mathcal{R}(D)$ have the following parametric form:

$$\sum_{i=1}^{L}\alpha_i R_i = \varphi(\alpha_{\mathcal{I}_L})$$

where $\alpha_{\mathcal{I}_L}$ is a unit (l_1-norm) vector in $\mathbb{R}_+^L$ and

$$\varphi(\alpha_{\mathcal{I}_L}) = \min_{R_{\mathcal{I}_L}\in\mathcal{R}(D)} \sum_{i=1}^{L}\alpha_i R_i.$$

Since $\mathcal{R}(r_{\mathcal{I}_L})$ is a contra-polymatroid, by [31, Lemma 3.3], a solution to the optimization problem

$$\min \sum_{i=1}^{L}\alpha_i R_i \quad \text{subject to} \quad R_{\mathcal{I}_L} \in \mathcal{R}(r_{\mathcal{I}_L})$$

is attained at a vertex $R_{\mathcal{I}_L}(\pi^*)$ where π^* is any permutation such that $\alpha_{\pi^*(1)} \geq \cdots \geq \alpha_{\pi^*(L)}$. That is

$$\begin{aligned}&\min_{R_{\mathcal{I}_L}\in\mathcal{R}(r_{\mathcal{I}_L})} \sum_{i=1}^{L}\alpha_i R_i \\ &= \sum_{i=1}^{L}\alpha_i R_i(\pi^*) \\ &= \sum_{i=1}^{L-1}\left((\alpha_{\pi^*(i)} - \alpha_{\pi^*(i+1)}) \sum_{j=1}^{i} R_{\pi^*(j)}(\pi^*) \right) \\ &\quad + \alpha_{\pi^*(L)}\sum_{i=1}^{L} R_{\pi^*(i)}(\pi^*) \\ &= \sum_{i=1}^{L-1}(\alpha_{\pi^*(i)} - \alpha_{\pi^*(i+1)}) \\ &\quad \times \left(\frac{1}{2}\log\left(\frac{\frac{1}{\sigma_X^2} + \sum_{j=1}^{L}\frac{1-\exp(-2r_j)}{\sigma_{N_j}^2}}{\frac{1}{\sigma_X^2} + \sum_{j=i+1}^{L}\frac{1-\exp(-2r_{\pi^*(j)})}{\sigma_{N_{\pi^*(j)}}^2}} \right) + \sum_{j=1}^{i} r_{\pi^*(j)} \right) \\ &\quad + \alpha_{\pi^*(L)}\left(\frac{1}{2}\log\left(1 + \sigma_X^2\sum_{j=1}^{L}\frac{1-\exp(-2r_j)}{\sigma_{N_j}^2}\right) + \sum_{j=1}^{L} r_j \right).\end{aligned}$$

Hence we have

$$\begin{aligned}&\varphi(\alpha_{\mathcal{I}_L}) \\ &= \min_{r_{\mathcal{I}_L}\in\mathbb{R}_+^L} \sum_{i=1}^{L-1}(\alpha_{\pi^*(i)} - \alpha_{\pi^*(i+1)})\end{aligned}$$

$$\times\left(\frac{1}{2}\log\left(\frac{\frac{1}{\sigma_X^2}+\sum_{j=1}^{L}\frac{1-\exp(-2r_j)}{\sigma_{N_j}^2}}{\frac{1}{\sigma_X^2}+\sum_{j=i+1}^{L}\frac{1-\exp(-2r_{\pi^*(j)})}{\sigma_{N_{\pi^*(j)}}^2}}\right)+\sum_{j=1}^{i}r_{\pi^*(j)}\right)$$
$$+\alpha_{\pi^*(L)}\left(\frac{1}{2}\log\left(1+\sigma_X^2\sum_{j=1}^{L}\frac{1-\exp(-2r_j)}{\sigma_{N_j}^2}\right)+\sum_{j=1}^{L}r_j\right) \tag{9}$$

subject to

$$\frac{1}{\sigma_X^2}+\sum_{i=1}^{L}\frac{1-\exp(-2r_i)}{\sigma_{N_i}^2}\geq\frac{1}{D}. \tag{10}$$

Let $r_{\mathcal{I}_L}^*$ be the minimizer of the above optimization problem. Since we can decrease $r_{\pi^*(1)}$ to make the constraint in (10) tight and keep the sum in (9) decreasing at the same time,[5] we must have

$$\frac{1}{\sigma_X^2}+\sum_{i=1}^{L}\frac{1-\exp(-2r_i^*)}{\sigma_{N_i}^2}=\frac{1}{D}. \tag{11}$$

Introduce Lagrange multipliers $\lambda_{\mathcal{I}_L}\in\mathbb{R}_+^L$ for the inequality constraints $r_{\mathcal{I}_L}\in\mathbb{R}_+^L$ and a multiplier $\nu\in\mathbb{R}$ for the equality constraint (11). Define

$$\begin{aligned}
&G\left(r_{\mathcal{I}_L},\lambda_{\mathcal{I}_L},\nu\right)\\
&\quad=\sum_{i=1}^{L-1}\left(\alpha_{\pi^*(i)}-\alpha_{\pi^*(i+1)}\right)\\
&\qquad\times\left(\sum_{j=1}^{i}r_{\pi^*(j)}-\frac{1}{2}\log\left(\frac{D}{\sigma_X^2}+\sum_{j=i+1}^{L}\frac{D-D\exp(-2r_{\pi^*(j)})}{\sigma_{N_{\pi^*(j)}}^2}\right)\right)\\
&\qquad+\alpha_{\pi^*(L)}\left(\frac{1}{2}\log\frac{\sigma_X^2}{D}+\sum_{j=1}^{L}r_j\right)-\sum_{i=1}^{L}\lambda_i r_i\\
&\qquad-\nu\left(\frac{1}{\sigma_X^2}+\sum_{i=1}^{L}\frac{1-\exp(-2r_i)}{\sigma_{N_i}^2}\right).
\end{aligned}$$

[5]If $r_{\pi^*(1)}$ attains 0 but the constraint in (10) is still not tight, then apply the same procedure to $r_{\pi^*(2)}$ and so on.

The *Karush–Kuhn–Tucker* conditions [38] yield

$$\begin{aligned}
&\left.\frac{\partial G}{\partial r_{\pi^*(1)}}\right|_{r_{\pi^*(1)}=r_{\pi^*(1)}^*}\\
&\quad=-\frac{2\nu\exp\left(-2r_{\pi^*(1)}^*\right)}{\sigma_{N_{\pi^*(1)}}^2}+\alpha_{\pi^*(1)}-\lambda_{\pi^*(i)}=0,\\
&\left.\frac{\partial G}{\partial r_{\pi^*(k)}}\right|_{r_{\pi^*(k)}=r_{\pi^*(k)}^*}\\
&\quad=-\frac{\exp\left(-2r_{\pi^*(k)}^*\right)}{\sigma_{N_{\pi^*(k)}}^2}\sum_{i=1}^{k-1}\left(\alpha_{\pi^*(i)}-\alpha_{\pi^*(i+1)}\right)\\
&\qquad\times\left(\frac{1}{\sigma_X^2}+\sum_{j=i+1}^{L}\frac{1-\exp(-2r_{\pi^*(j)}^*)}{\sigma_{N_{\pi^*(j)}}^2}\right)^{-1}\\
&\qquad+\alpha_{\pi^*(k)}-\lambda_{\pi^*(k)}-\frac{2\nu\exp\left(-2r_{\pi^*(k)}^*\right)}{\sigma_{N_{\pi^*(k)}}^2}\\
&\quad=-\frac{\exp(-2r_{\pi^*(k)}^*)}{\sigma_{N_{\pi^*(k)}}^2}\sum_{i=1}^{k-1}\left(\alpha_{\pi^*(i)}-\alpha_{\pi^*(i+1)}\right)\\
&\qquad\times\left(\frac{1}{D}-\sum_{j=1}^{i}\frac{1-\exp(-2r_{\pi^*(j)}^*)}{\sigma_{N_{\pi^*(j)}}^2}\right)^{-1}\\
&\qquad+\alpha_{\pi^*(k)}-\lambda_{\pi^*(k)}-\frac{2\nu\exp(-2r_{\pi^*(k)}^*)}{\sigma_{N_{\pi^*(k)}}^2}\\
&\quad=0,\qquad k=2,3,\ldots,L.
\end{aligned}$$

Solving these equations, we get the expressions of $r_{\pi^*(1)}^*,\ldots,r_{\pi^*(L)}^*$ shown at the bottom of the page, where $\lambda_{\mathcal{I}_L}\in\mathbb{R}_+^L$ and $\nu\in\mathbb{R}$ should be chosen so that the distortion constraint (11) is satisfied. Suppose they are given by $\lambda_{\mathcal{I}_L}^*$ and ν^* respectively. If $\nu^*\geq 0$, by the complementary slackness condition (i.e., $\lambda_k^*>0\Rightarrow r_k^*=0$), one can readily show (12) and (13) at the bottom of the next page, where $\log^+(t)=\max(\log t,0)$. Leveraging (12) and (13), we can compute $r_{\mathcal{I}_L}^*$ successively from $r_{\pi^*(1)}^*$, $r_{\pi^*(2)}^*,\ldots,$ to $r_{\pi^*(L)}^*$ for any given ν. Note that if $\nu^*\geq 0$, then $r_1^*,\ldots,r_L^*$ given in (12) and (13) are monotone increasing functions of ν for $\nu\in[0,\nu^*]$; therefore, ν can be uniquely determined by

---

$$r_{\pi^*(1)}^*=\frac{1}{2}\log\left(\frac{2\nu}{(\alpha_{\pi^*(1)}-\lambda_{\pi^*(1)})\sigma_{N_{\pi^*(1)}}^2}\right)$$
$$r_{\pi^*(k)}^*=\frac{1}{2}\log\left(\frac{2\nu+\sum_{i=1}^{k-1}(\alpha_{\pi^*(i)}-\alpha_{\pi^*(i+1)})\left(\frac{1}{D}-\sum_{j=1}^{i}\frac{1-\exp(-2r_{\pi^*(j)}^*)}{\sigma_{N_{\pi^*(j)}}^2}\right)^{-1}}{(\alpha_{\pi^*(k)}-\lambda_{\pi^*(k)})\sigma_{N_{\pi^*(k)}}^2}\right),\qquad k=2,3,\ldots,L.$$

substituting (12) and (13) into the distortion constraint (11); moreover, the inequality

$$\frac{1}{\sigma_X^2}+\sum_{i=1}^{L}\frac{1-\exp\left(-2r_i^*\right)}{\sigma_{N_i}^2}\bigg|_{\nu=0}\leq\frac{1}{D}$$

must hold. If the above inequality is not satisfied, then we must have $\nu^*<0$, which implies $\lambda_{\pi(1)}^*>0$ and further implies $r_{\pi^*(1)}^*=0$. Now apply the same argument to $(r_{\pi^*(2)}^*,\ldots,r_{\pi^*(\tilde{L})}^*)$. Continue this procedure until $r_{\mathcal{I}_L}^*$ is determined.

In the above we have assumed that $\alpha_i>0$ for all $i\in\mathcal{I}_L$. Now suppose $\alpha_{\pi^*(1)}\geq\cdots\geq\alpha_{\pi^*(\tilde{L})}>0=\alpha_{\pi^*(\tilde{L}+1)}=\cdots=\alpha_{\pi^*(L)}$. We can let $r_{\pi^*(\tilde{L}+1)}^*=\cdots=r_{\pi^*(L)}^*=\infty$. If

$$\frac{1}{\sigma_X^2}+\sum_{i=\tilde{L}+1}^{L}\frac{1}{\sigma_{N_{\pi^*(i)}}^2}>\frac{1}{D}\tag{14}$$

then we have $r_{\pi^*(1)}^*=\cdots=r_{\pi^*(\tilde{L})}^*=0$ and correspondingly $\varphi(\alpha_{\mathcal{I}_L})=0$. Otherwise, use the method in the previous paragraph to compute $r_{\pi^*(1)}^*,\ldots,r_{\pi^*(\tilde{L})}^*$ with the distortion constraint (11) replaced by

$$\frac{1}{\sigma_X^2}+\sum_{i=1}^{\tilde{L}}\frac{1-\exp\left(-2r_{\pi^*(i)}^*\right)}{\sigma_{N_{\pi^*(i)}}^2}+\sum_{i=\tilde{L}+1}^{L}\frac{1}{\sigma_{N_{\pi^*(i)}}^2}=\frac{1}{D}.\tag{15}$$

Let $\mathcal{T}(\alpha_{\mathcal{I}_L},D)$ with $\alpha_i>0$ ($\forall i\in\mathcal{I}_L$) be a supporting hyperplane of $\mathcal{R}(D)$. By (7), we have

$$\begin{aligned}\mathcal{T}(\alpha_{\mathcal{I}_L},D)\cap\partial\mathcal{R}(D)&=\mathcal{T}(\alpha_{\mathcal{I}_L},D)\cap\mathcal{R}(D)\\&=\mathcal{T}\left(\alpha_{\mathcal{I}_L},D\right)\cap\left(\bigcup_{r_{\mathcal{I}_L}\in\mathcal{F}(D)}\mathcal{R}(r_{\mathcal{I}_L})\right).\end{aligned}$$

If $\mathcal{T}(\alpha_{\mathcal{I}_L},D)\cap\mathcal{R}(r_{\mathcal{I}_L})\neq\emptyset$ for some $r_{\mathcal{I}_L}\in\mathcal{F}(D)$, then we must have $R_{\mathcal{I}_L}(\pi^*)\in\mathcal{T}(\alpha_{\mathcal{I}_L},D)\cap\mathcal{R}(r_{\mathcal{I}_L})$, where $R_{\mathcal{I}_L}(\pi^*)$ is the vertex of $\mathcal{R}(r_{\mathcal{I}_L})$ associated with permutation π^*. Now it follows from the aforederived Lagrangian optimization result that $\mathcal{R}(r_{\mathcal{I}_L})=\mathcal{R}(r_{\mathcal{I}_L}^*)$. Therefore, we have

$$\mathcal{T}(\alpha_{\mathcal{I}_L},D)\cap\partial\mathcal{R}(D)=\mathcal{T}(\alpha_{\mathcal{I}_L},D)\cap\mathcal{R}(r_{\mathcal{I}_L}^*).$$

Clearly, $\mathcal{T}(\alpha_{\mathcal{I}_L},D)\cap\mathcal{R}(r_{\mathcal{I}_L}^*)$ is a subface of the dominant face of $\mathcal{R}(r_{\mathcal{I}_L}^*)$. Let $(\mathcal{B}_1',\ldots,\mathcal{B}_k')$ be a partition of $\mathcal{I}_L$ such that $\alpha_m=\alpha_n$ for any $\alpha_m,\alpha_n\in\mathcal{B}_i'$ $(i=1,2,\ldots,k)$ and $\alpha_m>\alpha_n$ for any $\alpha_m\in\mathcal{B}_i',\alpha_n\in\mathcal{B}_j'$ $(i<j)$. Let Ξ' be the set of permutations π on $\mathcal{I}_L$ such that

$$\left\{\pi\left(\sum_{j=1}^{i-1}|\mathcal{B}_j'|+1\right),\ldots,\pi\left(\sum_{j=1}^{i}|\mathcal{B}_j'|\right)\right\}=\mathcal{B}_i',\qquad i=1,2,\ldots,k.$$

$\mathcal{T}(\alpha_{\mathcal{I}_L},D)\cap\mathcal{R}(r_{\mathcal{I}_L}^*)$ has totally $|\Xi'|=\prod_{i=1}^{k}(|\mathcal{B}_i'|!)$ vertices, each of which is associated with a permutation $\pi\in\Xi'$. Furthermore, $\dim(\mathcal{T}(\alpha_{\mathcal{I}_L},D)\cap\mathcal{R}(r_{\mathcal{I}_L}^*))\leq L-k$, where the equality holds if these $|\Xi'|$ vertices are all distinct. It is worth noting that if $\alpha_1=\cdots=\alpha_L$, then $\mathcal{T}(\alpha_{\mathcal{I}_L},D)\cap\mathcal{R}(r_{\mathcal{I}_L}^*)$ is the minimum sum-rate region of $\mathcal{R}(D)$ [23].

Corollary 2.3: For the quadratic Gaussian CEO problem, every rate tuple $R_{\mathcal{I}_L}\in\partial\mathcal{R}(D)$ can be achieved via a K-successive Wyner–Ziv coding scheme for some $K\leq 2L-1$.

Proof: Since $\mathcal{R}(D)=\bigcup_{r_{\mathcal{I}_L}\in\mathcal{F}(D)}\mathcal{R}(r_{\mathcal{I}_L})$, for any rate tuple $R_{\mathcal{I}_L}\in\partial\mathcal{R}(D)$, there exists a vector $r_{\mathcal{I}_L}\in\mathcal{F}(D)$ such that $R_{\mathcal{I}_L}\in\mathcal{R}(r_{\mathcal{I}_L})$. Furthermore, by Definition 2.4, it is easy to see that $R_{\mathcal{I}_L}$ must be on the dominant face of $\mathcal{R}(r_{\mathcal{I}_L})$. The desired result now follows from Corollary 2.1. □

Remark: To get more detailed information about the coding complexity of a rate tuple $R_{\mathcal{I}_L}\in\partial\mathcal{R}(D)$, we can proceed as follows. Let $\mathcal{T}(\alpha_{\mathcal{I}_L},D)$ be the supporting hyperplane of $\partial\mathcal{R}(D)$ such that $R_{\mathcal{I}_L}\in\mathcal{T}(\alpha_{\mathcal{I}_L},D)\cap\partial\mathcal{R}(D)$. Use the Lagrangian optimization method to find $\mathcal{R}(r_{\mathcal{I}_L}^*)$ with $r_{\mathcal{I}_L}^*\in\mathcal{F}(D)$ such that $\mathcal{T}(\alpha_{\mathcal{I}_L},D)\cap\partial\mathcal{R}(D)=\mathcal{T}(\alpha_{\mathcal{I}_L},D)\cap\mathcal{R}(r_{\mathcal{I}_L}^*)$. Let $\mathcal{F}\subseteq\mathcal{T}(\alpha_{\mathcal{I}_L},D)\cap\mathcal{R}(r_{\mathcal{I}_L}^*)$ be the lowest dimensional face of $\mathcal{R}(r_{\mathcal{I}_L}^*)$ that contains $R_{\mathcal{I}_L}$. We can conclude that $R_{\mathcal{I}_L}$ is achievable via an $(L+\dim(\mathcal{F}))$-successive Wyner–Ziv coding scheme.

III. Distributed Successive Refinement

In the previous section, we have shown that the successive Wyner–Ziv coding scheme suffices to achieve any rate tuple on the boundary of the rate region for the quadratic Gaussian CEO problem. We shall extend this result to the multistage source coding scenario.

Definition 3.1: For $R_{\mathcal{I}_L,1}\leq R_{\mathcal{I}_L,2}\leq\cdots\leq R_{\mathcal{I}_L,M}$ and $D_1\geq D_2\geq\cdots\geq D_M$, we say the M-stage source coding

$$(R_{\mathcal{I}_L,1},D_1)\nearrow(R_{\mathcal{I}_L,2},D_2)\nearrow\cdots\nearrow(R_{\mathcal{I}_L,M},D_M)$$

$$r_{\pi^*(1)}^*=\frac{1}{2}\log^+\left(\frac{2\nu}{\alpha_{\pi^*(1)}\sigma_{N_{\pi^*(1)}}^2}\right)\tag{12}$$

$$r_{\pi^*(k)}^*=\frac{1}{2}\log^+\left(\frac{2\nu+\sum_{i=1}^{k-1}(\alpha_{\pi^*(i)}-\alpha_{\pi^*(i+1)})\left(\frac{1}{D}-\sum_{j=1}^{i}\frac{1-\exp(-2r_{\pi^*(j)}^*)}{\sigma_{N_{\pi^*(j)}}^2}\right)^{-1}}{\alpha_{\pi^*(k)}\sigma_{N_{\pi^*(k)}}^2}\right),\qquad k=2,3,\ldots,L\tag{13}$$

is feasible if for each $\epsilon > 0$, there exists an n_0 such that for $n > n_0$ there exist encoders:

$$f_{i,j}^{(n)} : \mathcal{Y}_i^n \to \mathcal{C}_{i,j}^{(n)}, \qquad i = 1,2,\ldots,L, \qquad j = 1,2,\ldots,M$$

and decoders:

$$g_j^{(n)} : \prod_{k=1}^{j} \prod_{i=1}^{L} \mathcal{C}_{i,k}^{(n)} \to \mathcal{X}^n, \qquad j = 1,2,\ldots,M$$

such that

$$\begin{aligned} \frac{1}{n}\log\left|\mathcal{C}_{i,j}^n\right| &\le R_{i,j} - R_{i,j-1} + \epsilon, \\ &\qquad i = 1,2,\ldots,L, \qquad j = 1,2,\ldots,M, \\ \frac{1}{n}\mathbb{E}\left[\sum_{t=1}^{n} d(X(t), \hat{X}_j(t))\right] &\le D_j + \epsilon, \\ &\qquad\qquad j = 1,2,\ldots,M \end{aligned}$$

where

$$\begin{aligned} \hat{\mathbf{X}}_j = g_j^{(n)} \Big(& f_{1,1}^{(n)}(\mathbf{Y}_1), \ldots, f_{L,1}^{(n)}(\mathbf{Y}_L), \ldots, \\ & f_{1,j}^{(n)}(\mathbf{Y}_1), \ldots, f_{L,j}^{(n)}(\mathbf{Y}_L) \Big), \qquad j = 1,2,\ldots,M. \end{aligned}$$

Here we assume $R_{\mathcal{I}_L,0} = (0,\ldots,0)$.

The following definition can be viewed as a natural generalization of successive refinement in single source coding [39]–[42] to the distributed source coding scenario.

Definition 3.2 (Distributed Successive Refinement): Let $D^*(R_{\mathcal{I}_L}) = \min\{D : R_{\mathcal{I}_L} \in \mathcal{R}(D)\}$. For $R_{\mathcal{I}_L,1} \le R_{\mathcal{I}_L,2} \le \cdots \le R_{\mathcal{I}_L,M}$, we say there exists an M-stage distributed successive refinement scheme from $R_{\mathcal{I}_L,1}$ to $R_{\mathcal{I}_L,2}$, to $\cdots\cdots$, to $R_{\mathcal{I}_L,M}$ if the M-stage source coding

$$\begin{aligned} (R_{\mathcal{I}_L,1}, D^*(R_{\mathcal{I}_L,1})) \nearrow (R_{\mathcal{I}_L,2}, D^*(R_{\mathcal{I}_L,2})) \nearrow \cdots \\ \nearrow (R_{\mathcal{I}_L,M}, D^*(R_{\mathcal{I}_L,M})) \end{aligned}$$

is feasible.

Theorem 3.1: For $R_{\mathcal{I}_L,1} \le R_{\mathcal{I}_L,2} \le \cdots \le R_{\mathcal{I}_L,M}$ and $D_1 \ge D_2 \ge \cdots \ge D_M$, the M-stage source coding

$$(R_{\mathcal{I}_L,1}, D_1) \nearrow (R_{\mathcal{I}_L,2}, D_2) \nearrow \cdots \nearrow (R_{\mathcal{I}_L,M}, D_M)$$

is feasible if there exist random variables $W_{\mathcal{I}_L,\mathcal{I}_M}$ jointly distributed with the generic source variables $(X, Y_{\mathcal{I}_L})$ such that

$$(R_{\mathcal{I}_L,j} - R_{\mathcal{I}_L,j-1}) \in \mathcal{R}(W_{\mathcal{I}_L,j} | W_{\mathcal{I}_L,j-1}), \qquad j = 1,2,\ldots,M$$

where $W_{\mathcal{I}_L,\mathcal{I}_M}$ satisfy the following properties:

1) $W_{i,1} \to W_{i,2} \to \cdots \to W_{i,M} \to Y_i \to (X, Y_{\mathcal{I}_L \setminus \{i\}}, W_{\mathcal{I}_L \setminus \{i\},\mathcal{I}_M})$ form a Markov chain for all $i \in \mathcal{I}_L$;
2) for each $j \in \mathcal{I}_M$, there exists a function $\hat{X}_j : \prod_{i=1}^{L} \mathcal{W}_{i,j} \to \mathcal{X}$ such that $\mathbb{E}d(X, \hat{X}_j(W_{\mathcal{I}_L,j})) \le D_j$.

Here $W_{\mathcal{I}_L,0}$ is assumed to be a constant vector.

Proof: By Theorem 2.1, we can see that each stage can be realized via a $(2L-1)$-successive Wyner–Ziv scheme. □

The M-stage source coding, if realized by concatenating M versions of $(2L-1)$-successive Wyner–Ziv coding schemes, is essentially a $(2ML-M)$-successive Wyner–Ziv coding scheme. But it is subject to more restricted conditions since a general $(2ML-M)$-successive Wyner–Ziv scheme (satisfying the rate constraints $R_{\mathcal{I}_L,M}$ and the distortion constraint D_M) may not be decomposable into M versions of $2L-1$ successive Wyner–Ziv scheme with the rate and distortion constraints satisfied at each stage.

In the remaining part of this section, we shall focus on the quadratic Gaussian CEO problem.

Lemma 3.1: For $R_{\mathcal{I}_L,1} \le R_{\mathcal{I}_L,2} \le \cdots \le R_{\mathcal{I}_L,M}$ and $D_1 \ge D_2 \ge \cdots \ge D_M$, the M-stage source coding

$$(R_{\mathcal{I}_L,1}, D_1) \nearrow (R_{\mathcal{I}_L,2}, D_2) \nearrow \cdots \nearrow (R_{\mathcal{I}_L,M}, D_M)$$

is feasible if there exist $r_{\mathcal{I}_L,j} \in \mathbb{R}_+^L$, $j = 1,2,\ldots,M$, satisfying

1) $r_{\mathcal{I}_L,j-1} \le r_{\mathcal{I}_L,j}$ for all $j \in \mathcal{I}_M$;
2) $1/\sigma_X^2 + \sum_{i=1}^{L}(1 - \exp(-2r_{i,j}))/\sigma_{N_i}^2 = 1/D_j$ for all $j \in \mathcal{I}_M$;

such that

$$\begin{aligned} &\sum_{i\in\mathcal{A}}(R_{i,j} - R_{i,j-1}) \\ &\quad \ge \frac{1}{2}\log\frac{1}{D_j} - \frac{1}{2}\log\Bigg(\frac{1}{\sigma_X^2} + \sum_{i\in\mathcal{A}}\frac{1-\exp(-2r_{i,j-1})}{\sigma_{N_i}^2} \\ &\qquad\qquad + \sum_{i\in\mathcal{I}_L\setminus\mathcal{A}}\frac{1-\exp(-2r_{i,j})}{\sigma_{N_i}^2}\Bigg) \\ &\qquad + \sum_{i\in\mathcal{A}}(r_{i,j} - r_{i,j-1}) \\ &\qquad\qquad \forall\, j \in \mathcal{I}_M \qquad \forall \text{ nonempty set } \mathcal{A} \subseteq \mathcal{I}_L. \end{aligned}$$

Here we assume $r_{\mathcal{I}_L,0} = (0,\ldots,0)$.

Proof: Let $W_{i,M} = Y_i + T_{i,M}$ and $W_{i,j} = W_{i,j+1} + T_{i,j}$ $(j \in \mathcal{I}_{M-1})$, where $T_{i,j} \sim \mathcal{N}(0, \sigma_{T_{i,j}}^2)$, $i \in \mathcal{I}_L$, $j \in \mathcal{I}_M$, are mutually independent and also independent of $(X, Y_{\mathcal{I}_L})$. Let

$$r_{i,j} = I(Y_i; W_{i,j} \,|\, X) = \frac{1}{2}\log\frac{\sigma_{N_i}^2 + \sum_{k=j}^{M}\sigma_{T_{i,k}}^2}{\sum_{k=j}^{M}\sigma_{T_{i,k}}^2} \tag{16}$$

and $\mathbb{E}(X - \mathbb{E}(X|W_{\mathcal{I}_L,j}))^2 = D_j$ for all $j \in \mathcal{I}_M$, i.e.,

$$\frac{1}{\sigma_X^2} + \sum_{i=1}^{L}\frac{1-\exp(-2r_{i,j})}{\sigma_{N_i}^2} = \frac{1}{D_j}.$$

By Theorem 3.1, for any $R_{\mathcal{I}_L,1} \le R_{\mathcal{I}_L,2} \le \cdots \le R_{\mathcal{I}_L,M}$ with

$$(R_{\mathcal{I}_L,j} - R_{\mathcal{I}_L,j-1}) \in \mathcal{R}(W_{\mathcal{I}_L,j} \,|\, W_{\mathcal{I}_L,j-1}) \qquad \forall\, j \in \mathcal{I}_M \tag{17}$$

the M-stage source coding

$$(R_{\mathcal{I}_L,1}, D_1) \nearrow (R_{\mathcal{I}_L,2}, D_2) \nearrow \cdots \nearrow (R_{\mathcal{I}_L,M}, D_M)$$

is feasible. We can compute (17) explicitly as follows:

$$\begin{aligned}
&\sum_{i\in\mathcal{A}}(R_{i,j} - R_{i,j-1})\\
&\quad\geq I\left(Y_A; W_{\mathcal{A},j} | W_{\mathcal{I}_L\setminus\mathcal{A},j}, W_{\mathcal{A},j-1}\right)\\
&\quad= I\left(X, Y_A; W_{\mathcal{A},j} | W_{\mathcal{I}_L\setminus\mathcal{A},j}, W_{\mathcal{A},j-1}\right)\\
&\quad= I\left(X; W_{\mathcal{A},j} | W_{\mathcal{I}_L\setminus\mathcal{A},j}, W_{\mathcal{A},j-1}\right)\\
&\qquad + I(Y_A; W_{\mathcal{A},j} \,|\, X) - I(Y_A; W_{\mathcal{A},j-1}|X)\\
&\quad= h\left(X | W_{\mathcal{I}_L\setminus\mathcal{A},j}, W_{\mathcal{A},j-1}\right)\\
&\qquad - h\left(X \,|\, W_{\mathcal{I}_L,j}\right) + \sum_{i\in\mathcal{A}}(r_{i,j} - r_{i,j-1})\\
&\quad= \frac{1}{2}\log\frac{1}{D_j} - \frac{1}{2}\log\left(\frac{1}{\sigma_X^2} + \sum_{i\in\mathcal{A}}\frac{1-\exp(-2r_{i,j-1})}{\sigma_{N_i}^2}\right.\\
&\qquad\qquad\left. + \sum_{i\in\mathcal{I}_L\setminus\mathcal{A}}\frac{1-\exp(-2r_{i,j})}{\sigma_{N_i}^2}\right)\\
&\qquad + \sum_{i\in\mathcal{A}}(r_{i,j} - r_{i,j-1}) \qquad \forall \text{ nonempty set } \mathcal{A} \subseteq \mathcal{I}_L.
\end{aligned}$$

The proof is complete. □

Lemma 3.2 ([34, Lemma 1]): Let $f^{(n)}_{\mathcal{I}_L,\mathcal{I}_j} = (f^{(n)}_{i,\mathcal{I}_j})_{i\in\mathcal{I}_L}$, $j = 1, 2, \ldots, M$, where $f^{(n)}_{i,\mathcal{I}_j}$ is the abbreviation of $(f^{(n)}_{i,1}(\mathbf{Y}_i), \ldots, f^{(n)}_{i,j}(\mathbf{Y}_i))$. If there exist functions $g^{(n)}_j(\cdot)$, $j = 1, 2, \ldots, M$, such that

$$\frac{1}{n}\mathbb{E}\left[\sum_{t=1}^{n}\left(X(t) - \hat{X}_j(t)\right)^2\right] \leq D_j$$

with $\hat{\mathbf{X}}_j = g^{(n)}_j(f^{(n)}_{\mathcal{I}_L,\mathcal{I}_j})$, then

$$\frac{1}{n}I\left(\mathbf{X}; f^{(n)}_{\mathcal{I}_L,\mathcal{I}_j}\right) \geq \frac{1}{2}\log\frac{\sigma_X^2}{D_j}.$$

The next lemma is a direct application of [37, Lemma 3.1] (see also [36, Lemma 3]) with $C_i = f^{(n)}_{i,\mathcal{I}_k} (\forall i \in \mathcal{A})$ and $C_i = f^{(n)}_{i,\mathcal{I}_j}$ $(\forall i \in \mathcal{I}_L\setminus\mathcal{A})$.

Lemma 3.3: Let $r_{i,j} = \frac{1}{n}I(\mathbf{Y}; f^{(n)}_{i,\mathcal{I}_j}|\mathbf{X})$, $i \in \mathcal{I}_L, j \in \mathcal{I}_M$. We have, for all $0 \leq j < k \leq M$

$$\begin{aligned}
&\frac{1}{\sigma_X^2}\exp\left(\frac{2}{n}I(\mathbf{X}; f^{(n)}_{\mathcal{A},\mathcal{I}_k}, f^{(n)}_{\mathcal{I}_L\setminus\mathcal{A},\mathcal{I}_j})\right)\\
&\quad\leq \frac{1}{\sigma_X^2} + \sum_{i\in\mathcal{A}}\frac{1-\exp(-2r_{i,k})}{\sigma_{N_i}^2} + \sum_{i\in\mathcal{I}_L\setminus\mathcal{A}}\frac{1-\exp(-2r_{i,j})}{\sigma_{N_i}^2}
\end{aligned} \tag{18}$$

where $f^{(n)}_{i,0}$ $(i \in \mathcal{I}_L)$ are constant functions and $r_{\mathcal{I}_L,0} = (0, \ldots, 0)$.

Lemma 3.4: For $R_{\mathcal{I}_L,1} \leq R_{\mathcal{I}_L,2} \leq \cdots \leq R_{\mathcal{I}_L,M}$ and $D_1 \geq D_2 \geq \cdots \geq D_M$, if the M-stage source coding

$$(R_{\mathcal{I}_L,1}, D_1) \nearrow (R_{\mathcal{I}_L,2}, D_2) \nearrow \cdots \nearrow (R_{\mathcal{I}_L,M}, D_M)$$

is feasible, then there exist $r_{\mathcal{I}_L,j} \in \mathbb{R}^L_+$, $j = 1, 2, \ldots, M$, satisfying

1) $r_{\mathcal{I}_L,j-1} \leq r_{\mathcal{I}_L,j}$ for all $j \in \mathcal{I}_M$;
2) $1/\sigma_X^2 + \sum_{i=1}^{L}(1 - \exp(-2r_{i,j}))/\sigma_{N_i}^2 \geq 1/D_j$ for all $j \in \mathcal{I}_M$;

such that

$$\begin{aligned}
&\sum_{i\in\mathcal{A}}(R_{i,k} - R_{i,j})\\
&\quad\geq \frac{1}{2}\log\frac{1}{D_k} - \frac{1}{2}\log\left(\frac{1}{\sigma_X^2} + \sum_{i\in\mathcal{A}}\frac{1-\exp(-2r_{i,j})}{\sigma_{N_i}^2}\right.\\
&\qquad\qquad\left. + \sum_{i\in\mathcal{I}_L\setminus\mathcal{A}}\frac{1-\exp(-2r_{i,k})}{\sigma_{N_i}^2}\right)\\
&\qquad + \sum_{i\in\mathcal{A}}(r_{i,k} - r_{i,j})\\
&\qquad\qquad \forall\, 0 \leq j < k \leq M \qquad \forall \text{ nonempty set } \mathcal{A} \subseteq \mathcal{I}_L.
\end{aligned}$$

Here $r_{\mathcal{I}_L,0} = (0, \ldots, 0)$.

Proof: Let $r_{i,j} = I(\mathbf{Y}_i; f^{(n)}_{i,\mathcal{I}_j}|\mathbf{X})/n$, $i \in \mathcal{I}_L, j \in \mathcal{I}_M$. It is clear that $r_{\mathcal{I}_L,j-1} \leq r_{\mathcal{I}_L,j}$ for all $j \in \mathcal{I}_M$. Substituting $\mathcal{A} = \mathcal{I}_L$ into (18), we get

$$\begin{aligned}
&\frac{1}{\sigma_X^2} + \sum_{i=1}^{L}\frac{1-\exp(-2r_{i,k})}{\sigma_{N_i}^2}\\
&\quad\geq \frac{1}{\sigma_X^2}\exp\left(\frac{2}{n}I\left(\mathbf{X}; f^{(n)}_{\mathcal{I}_L,\mathcal{I}_k}\right)\right) \geq \frac{1}{D_k} \qquad \forall\, k \in \mathcal{I}_M
\end{aligned}$$

where the last inequality follows from Lemma 3.2.

Furthermore, we have

$$\begin{aligned}
&\sum_{i\in\mathcal{A}}(R_{i,k} - R_{i,j})\\
&\quad\geq \frac{1}{n}\sum_{i\in\mathcal{A}}\sum_{s=j+1}^{k}H\left(f^{(n)}_{i,s}\right) \geq \frac{1}{n}H\left(f^{(n)}_{\mathcal{A},s}, s = j+1, \ldots, k\right)\\
&\quad\geq \frac{1}{n}I\left(\mathbf{Y}_{\mathcal{A}}; f^{(n)}_{\mathcal{A},s}, s = j+1, \ldots, k \,\middle|\, f^{(n)}_{\mathcal{I}_L\setminus\mathcal{A},\mathcal{I}_k}, f^{(n)}_{\mathcal{A},\mathcal{I}_j}\right)\\
&\quad= \frac{1}{n}I\left(\mathbf{X}, \mathbf{Y}_{\mathcal{A}}; f^{(n)}_{\mathcal{A},s}, s = j+1, \ldots, k \,\middle|\, f^{(n)}_{\mathcal{I}_L\setminus\mathcal{A},\mathcal{I}_k}, f^{(n)}_{\mathcal{A},\mathcal{I}_j}\right)\\
&\quad= \frac{1}{n}I\left(\mathbf{X}; f^{(n)}_{\mathcal{A},s}, s = j+1, \ldots, k \,\middle|\, f^{(n)}_{\mathcal{I}_L\setminus\mathcal{A},\mathcal{I}_k}, f^{(n)}_{\mathcal{A},\mathcal{I}_j}\right)\\
&\qquad + \frac{1}{n}\sum_{i\in\mathcal{A}}I\left(\mathbf{Y}_i; f^{(n)}_{i,s}, s = j+1, \ldots, k \,\middle|\, \mathbf{X}, f^{(n)}_{i,\mathcal{I}_j}\right)\\
&\quad= \frac{1}{n}I\left(\mathbf{X}; f^{(n)}_{\mathcal{I}_L,\mathcal{I}_k}\right) - \frac{1}{n}I\left(\mathbf{X}; f^{(n)}_{\mathcal{I}_L\setminus\mathcal{A},\mathcal{I}_k}, f^{(n)}_{\mathcal{A},\mathcal{I}_j}\right)\\
&\qquad + \sum_{i\in\mathcal{A}}(r_{i,k} - r_{i,j})\\
&\quad\geq \frac{1}{2}\log\frac{1}{D_k} - \frac{1}{2}\log\left(\frac{1}{\sigma_X^2} + \sum_{i\in\mathcal{A}}\frac{1-\exp(-2r_{i,j})}{\sigma_{N_i}^2}\right.\\
&\qquad\qquad\left. + \sum_{i\in\mathcal{I}_L\setminus\mathcal{A}}\frac{1-\exp(-2r_{i,k})}{\sigma_{N_i}^2}\right)\\
&\qquad + \sum_{i\in\mathcal{A}}(r_{i,k} - r_{i,j})
\end{aligned} \tag{19}$$

where (19) follows from Lemma 3.2 and Lemma 3.3. Now the proof is complete. □

Lemma 3.5: For any $R_{\mathcal{I}_L} \in \mathbb{R}_+^L$, there exists a unique $r_{\mathcal{I}_L} \in \mathbb{R}_+^L$ satisfying

1) Constraint 1:

$$\frac{1}{\sigma_X^2} + \sum_{i=1}^{L} \frac{1-\exp(-2r_i)}{\sigma_{N_i}^2} \geq \frac{1}{D^*\left(R_{\mathcal{I}_L}\right)} \tag{20}$$

2) Constraint 2: for any nonempty set $\mathcal{A} \subseteq \mathcal{I}_L$

$$\sum_{i\in\mathcal{A}} R_i \geq \frac{1}{2}\log\frac{1}{D^*\left(R_{\mathcal{I}_L}\right)} - \frac{1}{2}\log\left(\frac{1}{\sigma_X^2} + \sum_{i\in\mathcal{I}_L\setminus\mathcal{A}} \frac{1-\exp(-2r_i)}{\sigma_{N_i}^2}\right) + \sum_{i\in\mathcal{A}} r_i. \tag{21}$$

Denote this $r_{\mathcal{I}_L}$ by $r^*_{\mathcal{I}_L}(R_{\mathcal{I}_L})$. We have

$$\frac{1}{\sigma_X^2} + \sum_{i=1}^{L} \frac{1-\exp\left(-2r_i^*\left(R_{\mathcal{I}_L}\right)\right)}{\sigma_{N_i}^2} = \frac{1}{D^*\left(R_{\mathcal{I}_L}\right)} \tag{22}$$

and

$$\sum_{i=1}^{L} R_i = \frac{1}{2}\log\frac{\sigma_X^2}{D^*\left(R_{\mathcal{I}_L}\right)} + \sum_{i=1}^{L} r_i^*\left(R_{\mathcal{I}_L}\right). \tag{23}$$

Proof: See the Appendix. □

Now we are ready to prove the main result of this section.

Theorem 3.2: For $R_{\mathcal{I}_L,1} \leq R_{\mathcal{I}_L,2} \leq \cdots \leq R_{\mathcal{I}_L,M}$, there exists an M-stage distributed successive refinement scheme from $R_{\mathcal{I}_L,1}$ to $R_{\mathcal{I}_L,2}$, to $\cdots\cdots$, to $R_{\mathcal{I}_L,M}$ if and only if

$$r^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j-1}) \leq r^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j}) \qquad \forall j \in \mathcal{I}_M \tag{24}$$

and

$$\begin{aligned}\sum_{i\in\mathcal{A}}(R_{i,j} - R_{i,j-1}) &\geq \frac{1}{2}\log\frac{1}{D_j^*\left(R_{\mathcal{I}_L,j}\right)} \\ &\quad - \frac{1}{2}\log\left(\frac{1}{\sigma_X^2} + \sum_{i\in\mathcal{A}} \frac{1-\exp\left(-2r_i^*\left(R_{\mathcal{I}_L,j-1}\right)\right)}{\sigma_{N_i}^2} + \sum_{i\in\mathcal{I}_L\setminus\mathcal{A}} \frac{1-\exp\left(-2r_i^*\left(R_{\mathcal{I}_L,j}\right)\right)}{\sigma_{N_i}^2}\right) \\ &\quad + \sum_{i\in\mathcal{A}}\left(r_i^*(R_{\mathcal{I}_L,j}) - r_i^*\left(R_{\mathcal{I}_L,j-1}\right)\right) \\ &\qquad \forall j \in \mathcal{I}_M, \forall \text{ nonempty set } \mathcal{A} \subseteq \mathcal{I}_L. \end{aligned} \tag{25}$$

Here $R_{\mathcal{I}_L,0} = r^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,0}) = (0,\ldots,0)$.

Proof: For every $j \in \mathcal{I}_M$, let $D_j = D^*(R_{\mathcal{I}_L,j})$ in Lemma 3.4. Suppose the vector sequence $r_{\mathcal{I}_L,j}$ $(j = 1,2,\ldots,M)$ satisfies all the constraints in Lemma 3.4. By Lemma 3.5, we must have

$$\frac{1}{\sigma_X^2} + \sum_{i=1}^{L} \frac{1-\exp(-2r_{i,j})}{\sigma_{N_i}^2} = \frac{1}{D^*\left(R_{\mathcal{I}_L,j}\right)}.$$

So the constraints in Lemma 3.4 imply the conditions in Lemma 3.1. Therefore, the conditions in Lemma 3.1 are necessary and sufficient. Furthermore, by Lemma 3.5, $r_{\mathcal{I}_L,j}$, if it exists, must be equal to $r^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j})$. The proof is complete. □

Remark: Applying (23) and then (22), we get

$$\begin{aligned}\sum_{i=1}^{L}(R_{i,j} - R_{i,j-1}) &= \frac{1}{2}\log\frac{\sigma_X^2}{D^*\left(R_{\mathcal{I}_L,j}\right)} + \sum_{i=1}^{L} r_i^*\left(R_{\mathcal{I}_L,j}\right) \\ &\quad - \frac{1}{2}\log\frac{\sigma_X^2}{D^*\left(R_{\mathcal{I}_L,j-1}\right)} - \sum_{i=1}^{L} r_i^*\left(R_{\mathcal{I}_L,j-1}\right) \\ &= \frac{1}{2}\log\frac{1}{D^*\left(R_{\mathcal{I}_L,j}\right)} \\ &\quad - \frac{1}{2}\log\left(\frac{1}{\sigma_X^2} + \sum_{i=1}^{L} \frac{1-\exp\left(-2r_i^*\left(R_{\mathcal{I}_L,j-1}\right)\right)}{\sigma_{N_i}^2}\right) \\ &\quad + \sum_{i=1}^{L}\left(r_i^*\left(R_{\mathcal{I}_L,j}\right) - r_i^*\left(R_{\mathcal{I}_L,j-1}\right)\right) \qquad \forall j \in \mathcal{I}_M. \end{aligned} \tag{26}$$

Hence in (25) the constraints on $\sum_{i=1}^{L}(R_{i,j} - R_{i,j-1}), j = 1,2,\ldots,M$, are tight.

The sequential structure of (24) and (25) leads straightforwardly to the following result.

Corollary 3.1: For $R_{\mathcal{I}_L,1} \leq R_{\mathcal{I}_L,2} \leq \cdots \leq R_{\mathcal{I}_L,M}$, there exists an M-stage distributed successive refinement scheme from $R_{\mathcal{I}_L,1}$ to $R_{\mathcal{I}_L,2}$, to $\cdots\cdots$, to $R_{\mathcal{I}_L,M}$ if and only if there exist a sequence of 2-stage distributed successive refinement schemes from $R_{\mathcal{I}_L,j-1}$ to $R_{\mathcal{I}_L,j}$, $j = 2,\ldots,M$.

Corollary 3.1 shows that for the quadratic Gaussian CEO problem, we only need to focus on two-stage distributed successive refinement.

By (16), each monotone increasing vector sequence $r_{\mathcal{I}_L,j}$ $(j = 1,2,\ldots,M)$ is associated with a unique $\sigma^2_{T_{\mathcal{I}_L,j}}$ $(j = 1,2,\ldots,M)$ and thus a unique $W_{\mathcal{I}_L,j}$ $(j = 1,2,\ldots,M)$. We shall let $W^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j})$ denote the $W_{\mathcal{I}_L,j}$ that is associated with $r^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j})$ $(j = 1,2,\ldots,M)$, and let $W^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,0})$ be a deterministic vector. Now we state Theorem 3.2 in the following equivalent form, which highlights the underlying geometric structure.

Corollary 3.2: For $R_{\mathcal{I}_L,1} \leq R_{\mathcal{I}_L,2} \leq \cdots \leq R_{\mathcal{I}_L,M}$, there exists an M-stage distributed successive refinement scheme from $R_{\mathcal{I}_L,1}$ to $R_{\mathcal{I}_L,2}$, to $\cdots\cdots$, to $R_{\mathcal{I}_L,M}$ if and only if $r^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j-1}) \leq r^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j}), \forall\ j \in \mathcal{I}_M$, and $(R_{\mathcal{I}_L,j} - R_{\mathcal{I}_L,j-1}) \in \mathcal{D}(W^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j})|W^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j-1}))$, $\forall j \in \mathcal{I}_M$, where $\mathcal{D}(W^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j})|W^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j-1}))$ is the dominant face of $\mathcal{R}(W^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j})|W^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j-1}))$.

Proof: It is easy to verify that (25) is equivalent to

$$\sum_{i\in\mathcal{A}}(R_{i,j}-R_{i,j-1}) \geq I\left(Y_{\mathcal{A}};W^*_{\mathcal{A}}(R_{\mathcal{I}_L,j})\,|\,W^*_{\mathcal{I}_L\setminus\mathcal{A}}(R_{\mathcal{I}_L,j}),W^*_{\mathcal{A}}(R_{\mathcal{I}_L,j-1})\right) \quad \forall\, j\in\mathcal{I}_M, \forall \text{ nonempty set } \mathcal{A}\subseteq\mathcal{I}_L$$

which, by Definition 2.3, is equivalent to

$$(R_{\mathcal{I}_L,j}-R_{\mathcal{I}_L,j-1}) \in \mathcal{R}\left(W^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j})\,|\,W^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j-1})\right) \qquad \forall j\in\mathcal{I}_M.$$

Furthermore, (26) is equivalent to

$$\sum_{i=1}^{L}(R_{i,j}-R_{i,j-1}) = I\left(Y_{\mathcal{I}_L};W^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j})\,|\,W^*_{\mathcal{I}_L,j-1}(R_{\mathcal{I}_L,j})\right) \qquad \forall j\in\mathcal{I}_M$$

which means $R_{\mathcal{I}_L,j}-R_{\mathcal{I}_L,j-1}$ is on the dominant face of $\mathcal{R}(W^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j})|W^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j-1})), \forall j\in\mathcal{I}_M$. □

Remark: Let $\mathcal{F}_j$ be the lowest dimensional face of $\mathcal{D}(W^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j})\,|\,W^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,j-1}))$ that contains $R_{\mathcal{I}_L,j}-R_{\mathcal{I}_L,j-1}$. By the discussion in the preceding section, we can see that if an M-stage distributed successive refinement scheme exists, then it can be realized via an $(ML+\sum_{j=1}^{M}dim(\mathcal{F}_j))$-successive Wyner–Ziv coding scheme.

Now we proceed to compute $r^*_{\mathcal{I}_L}(R_{\mathcal{I}_L})$. In view of (7) and Lemma 3.5, it is easy to show that $r^*_{\mathcal{I}_L}(R_{\mathcal{I}_L})$ is the maximizer to the following optimization problem:

$$\max_{r_{\mathcal{I}_L}\in\mathbb{R}^L_+} \frac{1}{\sigma_X^2}+\sum_{i=1}^{L}\frac{1-\exp(-2r_i)}{\sigma^2_{N_i}} \tag{27}$$

subject to

$$\frac{1}{2}\log\left(\frac{\frac{1}{\sigma_X^2}+\sum_{i=1}^{L}\frac{1-\exp(-2r_i)}{\sigma^2_{N_i}}}{\frac{1}{\sigma_X^2}+\sum_{i\in\mathcal{I}_\mathcal{L}\setminus\mathcal{A}}\frac{1-\exp(-2r_i)}{\sigma^2_{N_i}}}\right)+\sum_{i\in\mathcal{A}}r_i\leq\sum_{i\in\mathcal{A}}R_i \quad \forall \text{ nonempty set } \mathcal{A}\subset\mathcal{I}_L \tag{28}$$

and

$$\frac{1}{2}\log\left(\frac{1}{\sigma_X^2}+\sum_{i=1}^{L}\frac{1-\exp(-2r_i)}{\sigma^2_{N_i}}\right)+\frac{1}{2}\log\sigma_X^2+\sum_{i=1}^{L}r_i=\sum_{i=1}^{L}R_i. \tag{29}$$

This problem essentially amounts to finding the contra-polymatroid $\mathcal{R}(r_{\mathcal{I}_L})$ that contains $R_{\mathcal{I}_L}$ and has the minimum achievable distortion $D(r_{\mathcal{I}_L}) = (1/\sigma_X^2+\sum_{i=1}^{L}(1-\exp(-2r_i))/\sigma^2_{N_i})^{-1}$.

We shall first consider the case where $R_1=R_2=\cdots=R_L=R$ and $\sigma^2_{N_1}=\sigma^2_{N_2}=\cdots=\sigma^2_{N_L}=\sigma^2_N$. For this symmetric case, it is easy to show that

$$r^*_1(R_{\mathcal{I}_L})=r^*_2(R_{\mathcal{I}_L})=\cdots=r^*_L(R_{\mathcal{I}_L})=r^*(R_{\mathcal{I}_L})$$

where $r^*(R_{\mathcal{I}_L})$ is the unique solution to the equation

$$\frac{1}{2}\log\left(\frac{1}{\sigma_X^2}+\frac{L-L\exp(-2r^*(R_{\mathcal{I}_L}))}{\sigma_N^2}\right)+\frac{1}{2}\log\sigma_X^2+Lr^*(R_{\mathcal{I}_L})=LR. \tag{30}$$

Moreover, it follows from Lemma 3.5 that

$$\frac{1}{D^*(R_{\mathcal{I}_L})}=\frac{1}{\sigma_X^2}+\frac{L-L\exp(-2r^*(R_{\mathcal{I}_L}))}{\sigma_N^2}.$$

Therefore, we have the equation at the bottom of the page. Let $R_{1,1}=R_{2,1}=\cdots=R_{L,1}=R_1$ and $R_{1,2}=R_{2,2}=\cdots=R_{L,2}=R_2$ with $R_2\geq R_1$. In view of (30), one can readily show that $r^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,2})\geq r^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,1})$. It is also easy to verify that $(R_{\mathcal{I}_L,2}-R_{\mathcal{I}_L,1})\in\mathcal{D}(W^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,2})|W^*_{\mathcal{I}_L}(R_{\mathcal{I}_L,1}))$. Therefore, by Corollary 3.2, there exists a distributed successive refinement scheme from $R_{\mathcal{I}_L,1}$ to $R_{\mathcal{I}_L,2}$.

For small L, it is relatively easy to get a parametric expression of $r^*_{\mathcal{I}_L}(R_{\mathcal{I}_L})$ via the following approach: first characterize $r^*_{\mathcal{I}_L}(R_{\mathcal{I}_L})$ for $R_{\mathcal{I}_L}\in\partial\mathcal{R}(D)$ by studying the supporting hyperplanes of $\partial\mathcal{R}(D)$ for fixed D, and then vary D to get $r^*_{\mathcal{I}_L}(R_{\mathcal{I}_L})$ for all $R_{\mathcal{I}_L}$. To obtain a concrete understanding, we shall study the special case $L=2$. It is easy to see that $R_{\mathcal{I}_2}$ is either a vertex of $\mathcal{R}(r^*_{\mathcal{I}_2}(R_{\mathcal{I}_2}))$ or an interior point of the dominant face (which is a line segment) of $\mathcal{R}(r^*_{\mathcal{I}_2}(R_{\mathcal{I}_2}))$. For the first case, $r^*_{\mathcal{I}_2}(R_{\mathcal{I}_2})$ is completely determined. For the second case, $R_{\mathcal{I}_2}$ must be on the minimum sum-rate line of $\partial\mathcal{R}(D^*(R_{\mathcal{I}_2}))$. Hence we need to study only one supporting line of $\partial\mathcal{R}(D)$, namely, $R_1+R_2=\min_{R_{\mathcal{I}_2}\in\partial\mathcal{R}(D)}(R_1+R_2)$, which has been characterized for all D in [23].

Without loss of generality, we assume $\sigma^2_{N_1}\leq\sigma^2_{N_2}$. Let

$$L(D)=\max\left\{k\in\mathcal{I}_2:\frac{k}{\sigma^2_{N_k}}+\frac{1}{D}-\frac{1}{D_{\min}(k)}\geq 0\right\}$$

where

$$D_{\min}(k)=\left(\frac{1}{\sigma_X^2}+\sum_{i=1}^{k}\frac{1}{\sigma^2_{N_i}}\right)^{-1}.$$

Let $\tilde{D}$ be the unique solution to the following equation:

$$\frac{1}{2}\log\left(\frac{\sigma_X^2}{D}\prod_{i=1}^{L(D)}\left(\frac{L(D)}{\sigma^2_{N_i}\left(\frac{1}{D_{\min}(L(D))}-\frac{1}{D}\right)}\right)\right)=R_1+R_2.$$

$$R=\frac{1}{2L}\log\left\{\frac{\sigma_X^2}{D^*(R_{\mathcal{I}_L})}\left(\frac{D^*(R_{\mathcal{I}_L})\sigma_X^2L}{D^*(R_{\mathcal{I}_L})\sigma_X^2L-\sigma_X^2\sigma_N^2+D^*(R_{\mathcal{I}_L})\sigma_N^2}\right)^L\right\}.$$

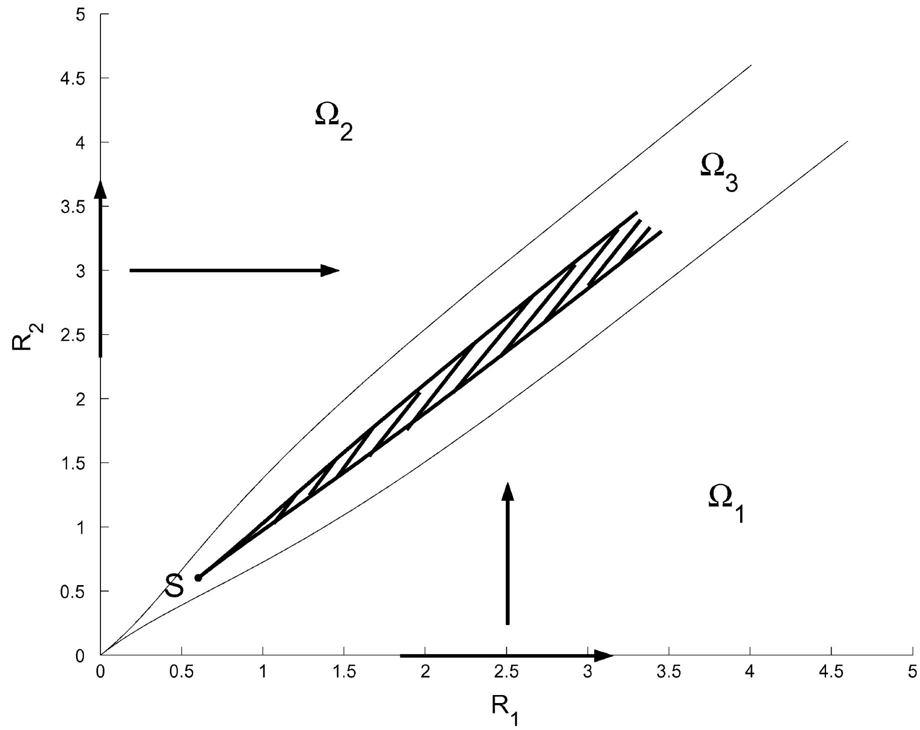

Fig. 4. Distributed successive refinement for the quadratic Gaussian CEO problem.

Let

$$\tilde{r}_1 = \frac{1}{2}\log\left(\frac{L(\tilde{D})}{\sigma_{N_1}^2\left(\frac{1}{D_{\min}(L(\tilde{D}))} - \frac{1}{\tilde{D}}\right)}\right)$$

$$\tilde{r}_2 = \begin{cases} 0, & L_{\tilde{D}} = 1 \\ \frac{1}{2}\log\left(\frac{2}{\sigma_{N_2}^2}\left(\frac{1}{\sigma_X^2} + \frac{1}{\sigma_{N_1}^2} + \frac{1}{\sigma_{N_2}^2} - \frac{1}{\tilde{D}}\right)^{-1}\right), & L_{\tilde{D}} = 2. \end{cases}$$

We have the following:

1) if

$$R_1 \geq \frac{1}{2}\log\left(\frac{1}{\sigma_X^2} + \frac{1 - \exp(-2\tilde{r}_1)}{\sigma_{N_1}^2}\right) + \frac{1}{2}\log\sigma_X^2 + \tilde{r}_1 \quad (31)$$

then

$$\frac{1}{2}\log\left(\frac{1}{\sigma_X^2} + \frac{1 - \exp\left(-2r_1^*\left(R_{\mathcal{I}_2}\right)\right)}{\sigma_{N_1}^2}\right) + \frac{1}{2}\log\sigma_X^2 + r_1^*\left(R_{\mathcal{I}_2}\right) = R_1 \quad (32)$$

$$\frac{1}{2}\log\left(1 + \sigma_X^2\sum_{i=1}^{2}\frac{1 - \exp\left(-2r_i^*\left(R_{\mathcal{I}_2}\right)\right)}{\sigma_{N_i}^2}\right) + r_1^*\left(R_{\mathcal{I}_2}\right) + r_2^*\left(R_{\mathcal{I}_2}\right) = R_1 + R_2 \quad (33)$$

2) if

$$R_2 \geq \frac{1}{2}\log\left(\frac{1}{\sigma_X^2} + \frac{1 - \exp(-2\tilde{r}_2)}{\sigma_{N_2}^2}\right) + \frac{1}{2}\log\sigma_X^2 + \tilde{r}_2 \quad (34)$$

then

$$\frac{1}{2}\log\left(\frac{1}{\sigma_X^2} + \frac{1 - \exp\left(-2r_2^*\left(R_{\mathcal{I}_2}\right)\right)}{\sigma_{N_2}^2}\right) + \frac{1}{2}\log\sigma_X^2 + r_2^*\left(R_{\mathcal{I}_2}\right) = R_2 \quad (35)$$

$$\frac{1}{2}\log\left(1 + \sigma_X^2\sum_{i=1}^{2}\frac{1 - \exp\left(-2r_i^*\left(R_{\mathcal{I}_2}\right)\right)}{\sigma_{N_i}^2}\right) + r_1^*\left(R_{\mathcal{I}_2}\right) + r_2^*\left(R_{\mathcal{I}_2}\right) = R_1 + R_2 \quad (36)$$

3) otherwise $r_i^*(R_{\mathcal{I}_2}) = \tilde{r}_i$, $i = 1, 2$.

The above three conditions essentially divide $\mathbb{R}_+^2$ into three regions. Define

$$\Omega_1 = \left\{R_{\mathcal{I}_2} \in \mathbb{R}_+^2 : R_1 \geq \frac{1}{2}\log\left(\frac{1}{\sigma_X^2} + \frac{1 - \exp(-2\tilde{r}_1)}{\sigma_{N_1}^2}\right) + \frac{1}{2}\log\sigma_X^2 + \tilde{r}_1\right\}$$

$$\Omega_2 = \left\{R_{\mathcal{I}_2} \in \mathbb{R}_+^2 : R_2 \geq \frac{1}{2}\log\left(\frac{1}{\sigma_X^2} + \frac{1 - \exp(-2\tilde{r}_2)}{\sigma_{N_2}^2}\right) + \frac{1}{2}\log\sigma_X^2 + \tilde{r}_2\right\}$$

$$\Omega_3 = \left\{R_{\mathcal{I}_2} \in \mathbb{R}_+^2 : R_i \leq \frac{1}{2}\log\left(\frac{1}{\sigma_X^2} + \frac{1 - \exp(-2\tilde{r}_i)}{\sigma_{N_i}^2}\right) + \frac{1}{2}\log\sigma_X^2 + \tilde{r}_i, i = 1, 2\right\}.$$

Typical shapes of Ω_1, Ω_2, and Ω_3 are plotted in Fig. 4. Every rate pair $R_{\mathcal{I}_2} \in \Omega_1 \cup \Omega_2$ is a vertex of $\mathcal{R}(r_{\mathcal{I}_2}^*(R_{\mathcal{I}_2}))$ and thus is associated with a 2-successive Wyner–Ziv coding scheme.

Every rate pair $R_{\mathcal{I}_2}$ strictly inside Ω_3 is an interior point of the dominant face of $\mathcal{R}(r^*_{\mathcal{I}_2}(R_{\mathcal{I}_2}))$ and thus is associated with a 3-successive Wyner–Ziv coding scheme. Hence, there is a clear distinction between (Ω_1, Ω_2) and Ω_3. It will be seen that this difference has interesting implications on distributed successive refinement.

Henceforth we shall assume $R_{\mathcal{I}_2,1} \le R_{\mathcal{I}_2,2}$.

Claim 3.1: $r^*_{\mathcal{I}_2}(R_{\mathcal{I}_2,1})) \le r^*_{\mathcal{I}_2}(R_{\mathcal{I}_2,2})$.

Proof: If both $R_{\mathcal{I}_2,1}$ and $R_{\mathcal{I}_2,2}$ are in Ω_1 or both $R_{\mathcal{I}_2,1}$ and $R_{\mathcal{I}_2,2}$ are in Ω_2, the claim can be easily verified by checking the (32), (33), (35) and (36). Since $\tilde{r}_1$ and $\tilde{r}_2$ are monotone increasing functions of $R_1 + R_2$, the claim is also true when both $R_{\mathcal{I}_2,1}$ and $R_{\mathcal{I}_2,2}$ are in Ω_3.

Now consider the general case when $R_{\mathcal{I}_2,1}$ and $R_{\mathcal{I}_2,2}$ are in different regions, say, $R_{\mathcal{I}_2,1} \in \Omega_1$ and $R_{\mathcal{I}_2,2} \in \Omega_3$. Suppose the line segment connecting $R_{\mathcal{I}_2,1}$ and $R_{\mathcal{I}_2,2}$ intersects the boundary of Ω_1 and Ω_3 at point $R'_{\mathcal{I}_2}$. We have $r^*_{\mathcal{I}_2}(R_{\mathcal{I}_2,1}) \le r^*_{\mathcal{I}_2}(R'_{\mathcal{I}_2})$ since both $R_{\mathcal{I}_2,1}$ and $R'_{\mathcal{I}_2}$ are in Ω_1, and $r^*_{\mathcal{I}_2}(R'_{\mathcal{I}_2}) \le r^*_{\mathcal{I}_2}(R_{\mathcal{I}_2,2})$ since both $R'_{\mathcal{I}_2}$ and $R_{\mathcal{I}_2,2}$ are in Ω_3. Hence $r^*_{\mathcal{I}_2}(R_{\mathcal{I}_2,1}) \le r^*_{\mathcal{I}_2}(R_{\mathcal{I}_2,2})$. The other cases can be verified in a similar way. □

Remark: Note that $r^*_{\mathcal{I}_2}(R_{\mathcal{I}_2,1}) \le r^*_{\mathcal{I}_2}(R_{\mathcal{I}_2,2})$ is given as a condition in Theorem 3.2 and Corollary 3.2. However, we see that for the case $L = 2$, this condition is redundant since it is implied by the fact that $R_{\mathcal{I}_2,1} \le R_{\mathcal{I}_2,2}$. We conjecture that this condition is also redundant for general L.

Claim 3.2: If both $R_{\mathcal{I}_2,1}$ and $R_{\mathcal{I}_2,2}$ are in Ω_1, then there exists a distributed successive refinement scheme from $R_{\mathcal{I}_2,1}$ to $R_{\mathcal{I}_2,2}$ if and only if $R_{1,2} = R_{1,1}$ or $R_{2,1} = 0$.

Proof: If $R_{1,2} = R_{1,1}$, by (32) we have $r^*_1(R_{\mathcal{I}_2,2}) = r^*_1(R_{\mathcal{I}_2,1})$. It is easy to verify that the conditions in Theorem 3.2 are all satisfied. If $R_{2,1} = 0$, by (32) and (33), we have $r^*_2(R_{\mathcal{I}_2,1}) = 0$. Again, it is easy to verify that the conditions in Theorem 3.2 are all satisfied.

Now suppose there exists a distributed successive refinement scheme from $R_{\mathcal{I}_2,1}$ to $R_{\mathcal{I}_2,2}$. Since both $R_{\mathcal{I}_2,1}$ and $R_{\mathcal{I}_2,2}$ are in Ω_1, by (32) and (33)

$$
\begin{aligned}
&R_{2,2} - R_{2,1}\\
&= \frac{1}{2}\log\left(\frac{1}{\sigma_X^2}+\sum_{i=1}^{2}\frac{1-\exp\left(-2r_i^*\left(R_{\mathcal{I}_2,2}\right)\right)}{\sigma_{N_i}^2}\right)+r_2^*\left(R_{\mathcal{I}_2,2}\right)\\
&\quad -\frac{1}{2}\log\left(\frac{1}{\sigma_X^2}+\frac{1-\exp\left(-2r_1^*\left(R_{\mathcal{I}_2,2}\right)\right)}{\sigma_{N_1}^2}\right)\\
&\quad -\frac{1}{2}\log\left(\frac{1}{\sigma_X^2}+\sum_{i=1}^{2}\frac{1-\exp\left(-2r_i^*\left(R_{\mathcal{I}_2,1}\right)\right)}{\sigma_{N_i}^2}\right)-r_2^*\left(R_{\mathcal{I}_2,1}\right)\\
&\quad +\frac{1}{2}\log\left(\frac{1}{\sigma_X^2}+\frac{1-\exp\left(-2r_1^*\left(R_{\mathcal{I}_2,1}\right)\right)}{\sigma_{N_1}^2}\right)\\
&= \frac{1}{2}\log\frac{1}{D^*\left(R_{\mathcal{I}_2,2}\right)}\\
&\quad -\frac{1}{2}\log\left(\frac{1}{\sigma_X^2}+\frac{1-\exp\left(-2r_1^*\left(R_{\mathcal{I}_2,2}\right)\right)}{\sigma_{N_1}^2}\right)\\
&\quad + r_2^*\left(R_{\mathcal{I}_2,2}\right) - r_2^*\left(R_{\mathcal{I}_2,1}\right)\\
&\quad -\frac{1}{2}\log\left(\frac{1}{\sigma_X^2}+\sum_{i=1}^{2}\frac{1-\exp\left(-2r_i^*\left(R_{\mathcal{I}_2,1}\right)\right)}{\sigma_{N_i}^2}\right)\\
&\quad +\frac{1}{2}\log\left(\frac{1}{\sigma_X^2}+\frac{1-\exp\left(-2r_1^*\left(R_{\mathcal{I}_2,1}\right)\right)}{\sigma_{N_1}^2}\right).
\end{aligned}
$$

By Theorem 3.2, we must have

$$
\begin{aligned}
&\frac{1}{2}\log\frac{1}{D^*\left(R_{\mathcal{I}_2,2}\right)}\\
&\quad -\frac{1}{2}\log\left(\frac{1}{\sigma_X^2}+\frac{1-\exp\left(-2r_1^*\left(R_{\mathcal{I}_2,2}\right)\right)}{\sigma_{N_1}^2}\right)\\
&\quad + r_2^*\left(R_{\mathcal{I}_2,2}\right) - r_2^*\left(R_{\mathcal{I}_2,1}\right)\\
&\quad -\frac{1}{2}\log\left(\frac{1}{\sigma_X^2}+\sum_{i=1}^{2}\frac{1-\exp\left(-2r_i^*\left(R_{\mathcal{I}_2,1}\right)\right)}{\sigma_{N_i}^2}\right)\\
&\quad +\frac{1}{2}\log\left(\frac{1}{\sigma_X^2}+\frac{1-\exp\left(-2r_1^*\left(R_{\mathcal{I}_2,1}\right)\right)}{\sigma_{N_1}^2}\right)\\
&\qquad \ge \frac{1}{2}\log\frac{1}{D^*\left(R_{\mathcal{I}_2,2}\right)}-\frac{1}{2}\log\left(\frac{1}{\sigma_X^2}+\frac{1-\exp\left(-2r_1^*\left(R_{\mathcal{I}_2,2}\right)\right)}{\sigma_{N_1}^2}\right.\\
&\qquad\qquad \left.+\frac{1-\exp(-2r_2^*\left(R_{\mathcal{I}_2,1}\right))}{\sigma_{N_2}^2}\right)\\
&\qquad + r_2^*\left(R_{\mathcal{I}_2,2}\right) - r_2^*\left(R_{\mathcal{I}_2,1}\right)
\end{aligned}
$$

which, after some algebraic manipulation, is equivalent to $r_1^*(R_{\mathcal{I}_2,2})r_2^*(R_{\mathcal{I}_2,1}) \le r_1^*(R_{\mathcal{I}_2,1})r_2^*(R_{\mathcal{I}_2,1})$. Then we have either $r_2^*(R_{\mathcal{I}_2,1}) = 0$ or $r_1^*(R_{\mathcal{I}_2,2}) \le r_1^*(R_{\mathcal{I}_2,1})$, which further implies $r_1^*(R_{\mathcal{I}_2,2}) = r_1^*(R_{\mathcal{I}_2,1})$. Hence, by (32) and (33), we have $R_{1,2} = R_{1,1}$ or $R_{2,1} = 0$. □

The following claim follows by symmetry.

Claim 3.3: If both $R_{\mathcal{I}_2,1}$ and $R_{\mathcal{I}_2,2}$ are in Ω_2, then there exists a distributed successive refinement scheme from $R_{\mathcal{I}_2,1}$ to $R_{\mathcal{I}_2,2}$ if and only if $R_{2,2} = R_{2,1}$ or $R_{1,1} = 0$.

Remark: Claims 3.2 and 3.3 imply that there exists a distributed successive refinement scheme from $R_{\mathcal{I}_2,1}$ to $R_{\mathcal{I}_2,2}$ if $R_{\mathcal{I}_2,1}$ and $R_{\mathcal{I}_2,2}$ are on the R_1-axis or $R_{\mathcal{I}_2,1}$ and $R_{\mathcal{I}_2,2}$ are on the R_2-axis. Actually in this case, distributed successive refinement reduces to conventional successive refinement in single source coding[6] [41]. Furthermore, it is clear from Fig. 4 that if $R_{11} = R_{12} = R_1$ and $R_{21} \le R_{22} < \infty$, a distributed successive refinement scheme from $R_{\mathcal{I}_2,1}$ to $R_{\mathcal{I}_2,2}$ always exists when R_1 is sufficiently large since both $R_{\mathcal{I}_2,1}$ and $R_{\mathcal{I}_2,2}$ must be in Ω_1; likewise, if $R_{21} = R_{22} = R_2$ and $R_{11} \le R_{12} < \infty$, a distributed successive refinement scheme from $R_{\mathcal{I}_2,1}$ to $R_{\mathcal{I}_2,2}$ always exists when R_2 is sufficiently large since both $R_{\mathcal{I}_2,1}$ and $R_{\mathcal{I}_2,2}$ must be in Ω_2. Note that for the extreme case where $R_{11} = R_{12} = \infty$, $\sigma^2_{N_2} = 0$ (or $R_{21} = R_{22} = \infty$, $\sigma^2_{N_1} = 0$), the CEO problem reduces to the Wyner–Ziv problem, and this result has been derived in [43].

Claim 3.4: Suppose $R_{1,1} > 0, R_{2,1} > 0$. Then there is no distributed successive refinement scheme from $R_{\mathcal{I}_2,1}$ to $R_{\mathcal{I}_2,2}$

[6]There is a slight difference since in the single encoder case, the CEO problem becomes the noisy (single) source coding problem. But the generalization of successive refinement in single source coding to noisy (single) source coding is straightforward.

if either $R_{\mathcal{I}_2,1} \in \Omega_1$ and $R_{\mathcal{I}_2,2} \in \Omega_2$ or $R_{\mathcal{I}_2,1} \in \Omega_2$ and $R_{\mathcal{I}_2,2} \in \Omega_1$.

Proof: We shall prove only the case in which $R_{\mathcal{I}_2,1} \in \Omega_1$ and $R_{\mathcal{I}_2,2} \in \Omega_2$. The other one follows by symmetry.

By (32) and (33), $R_{1,1} > 0$ and $R_{2,1} > 0$ implies $r_1^*(R_{\mathcal{I}_2,1}) > 0$ and $r_2^*(R_{\mathcal{I}_2,1}) > 0$, which further implies $r_2^*(R_{\mathcal{I}_2,2}) > 0$ by Claim 3.1. Now it follows from (32), (33), (35), and (36) that

$$\begin{aligned}
&R_{1,2} - R_{1,1}\\
&\quad= \frac{1}{2}\log\left(\frac{1}{\sigma_X^2}+\sum_{i=1}^{2}\frac{1-\exp\left(-2r_i^*\left(R_{\mathcal{I}_2,2}\right)\right)}{\sigma_{N_i}^2}\right)+r_1^*\left(R_{\mathcal{I}_2,2}\right)\\
&\qquad-\frac{1}{2}\log\left(\frac{1}{\sigma_X^2}+\frac{1-\exp\left(-2r_2^*\left(R_{\mathcal{I}_2,2}\right)\right)}{\sigma_{N_2}^2}\right)\\
&\qquad-\frac{1}{2}\log\left(\frac{1}{\sigma_X^2}+\frac{1-\exp\left(-2r_1^*\left(R_{\mathcal{I}_2,1}\right)\right)}{\sigma_{N_1}^2}\right)\\
&\qquad-\frac{1}{2}\log\sigma_X^2 - r_1^*\left(R_{\mathcal{I}_2,1}\right)\\
&\quad= \frac{1}{2}\log\frac{1}{D^*\left(R_{\mathcal{I}_2,2}\right)}\\
&\qquad-\frac{1}{2}\log\left(\frac{1}{\sigma_X^2}+\frac{1-\exp\left(-2r_2^*\left(R_{\mathcal{I}_2,2}\right)\right)}{\sigma_{N_2}^2}\right)-\frac{1}{2}\log\sigma_X^2\\
&\qquad-\frac{1}{2}\log\left(\frac{1}{\sigma_X^2}+\frac{1-\exp\left(-2r_1^*\left(R_{\mathcal{I}_2,1}\right)\right)}{\sigma_{N_1}^2}\right)\\
&\qquad+r_1^*\left(R_{\mathcal{I}_2,2}\right)-r_1^*\left(R_{\mathcal{I}_2,1}\right)
\end{aligned}$$

which is strictly less than

$$\begin{aligned}
&\frac{1}{2}\log\frac{1}{D^*\left(R_{\mathcal{I}_2,2}\right)}-\frac{1}{2}\log\Bigg(\frac{1}{\sigma_X^2}+\frac{1-\exp\left(-2r_2^*\left(R_{\mathcal{I}_2,2}\right)\right)}{\sigma_{N_2}^2}\\
&\qquad\qquad\qquad\qquad+\frac{1-\exp\left(-2r_1^*\left(R_{\mathcal{I}_2,1}\right)\right)}{\sigma_{N_1}^2}\Bigg)\\
&\quad+r_1^*\left(R_{\mathcal{I}_2,2}\right)-r_1^*\left(R_{\mathcal{I}_2,1}\right)
\end{aligned}$$

if $r_1^*(R_{\mathcal{I}_2,1}) > 0$ and $r_2^*(R_{\mathcal{I}_2,2}) > 0$. Thus by Theorem 3.2, a distributed successive refinement scheme cannot exist. $\square$

In Fig. 4, the arrows denote the possible directions for distributed successive refinement in Ω_1 and Ω_2. For illustration, we pick a point s in Ω_3. The dark region is the set of points to which there exists a distributed successive refinement scheme from s. It can be seen that the possible directions for distributed successive refinement are different in these three regions.

IV. Conclusion

We have discussed two closely related problems in distributed source coding: the first one is how to decompose a high complexity distributed source code into low complexity codes; the second one is how to construct a high rate distributed source code using low rate codes via distributed successive refinement. It turns out that, at least for the quadratic Gaussian CEO problem, the successive Wyner–Ziv coding scheme gives the answer to both problems. Successive Wyner–Ziv coding has several desirable features such as low complexity and robustness. Moreover, its concatenable chain structure seems especially attractive in wireless sensor networks, where channels are subject to fluctuation. Indeed, by properly converting a high-rate distributed source code to a multistage code via successive Wyner–Ziv coding, one can adaptively match source rates to fluctuating channel rates.

Appendix
Proof of Lemma 3.5

For any $r_{\mathcal{I}_L} \in \mathbb{R}_+^L$ and $D > 0$, define two set functions $f(\cdot, r_{\mathcal{I}_L})$ and $f(\cdot, r_{\mathcal{I}_L}, D)$:

$$\begin{aligned}
&f(\mathcal{A}, r_{\mathcal{I}_L})\\
&\quad=\frac{1}{2}\log\left(\frac{\dfrac{1}{\sigma_X^2}+\displaystyle\sum_{i=1}^{L}\frac{1-\exp(-2r_i)}{\sigma_{N_i}^2}}{\dfrac{1}{\sigma_X^2}+\displaystyle\sum_{i\in\mathcal{I}_L\setminus\mathcal{A}}\frac{1-\exp(-2r_i)}{\sigma_{N_i}^2}}\right)\\
&\qquad+\sum_{i\in\mathcal{A}}r_i,\qquad \mathcal{A}\subseteq\mathcal{I}_L\\
&f(\mathcal{A}, r_{\mathcal{I}_L}, D)\\
&\quad=\frac{1}{2}\log\frac{1}{D}-\frac{1}{2}\log\left(\frac{1}{\sigma_X^2}+\sum_{i\in\mathcal{I}_L\setminus\mathcal{A}}\frac{1-\exp(-2r_i)}{\sigma_{N_i}^2}\right)\\
&\qquad+\sum_{i\in\mathcal{A}}r_i,\qquad \mathcal{A}\subseteq\mathcal{I}_L.
\end{aligned}$$

Note that $f(\cdot, r_{\mathcal{I}_L})$ is a rank function and induces the contra-polymatroid $\mathcal{R}(r_{\mathcal{I}_L})$ defined in (6). It can be verified that for any $r_{\mathcal{I}_L}$ satisfying $r_i > 0$ ($\forall i \in \mathcal{I}_L$) and nonempty sets $\mathcal{S}, \mathcal{T} \subseteq \mathcal{I}_L$, if $\mathcal{S} \nsubseteq \mathcal{T}$ and $\mathcal{T} \nsubseteq \mathcal{S}$, then

$$f\left(\mathcal{S}, r_{\mathcal{I}_L}\right)+f\left(\mathcal{T}, r_{\mathcal{I}_L}\right)<f\left(\mathcal{S}\cup\mathcal{T}, r_{\mathcal{I}_L}\right)+f\left(\mathcal{S}\cap\mathcal{T}, r_{\mathcal{I}_L}\right) \tag{37}$$

$$\begin{aligned}
&f\left(\mathcal{S}, r_{\mathcal{I}_L}, D\right)+f\left(\mathcal{T}, r_{\mathcal{I}_L}, D\right)\\
&\quad<f\left(\mathcal{S}\cup\mathcal{T}, r_{\mathcal{I}_L}, D\right)+f\left(\mathcal{S}\cap\mathcal{T}, r_{\mathcal{I}_L}, D\right).
\end{aligned}\tag{38}$$

It was shown in [37] that

$$\begin{aligned}
\mathcal{R}(D) = \bigcup_{r_{\mathcal{I}_L}\in\mathcal{F}(D)}\Bigg\{R_{\mathcal{I}_L} : \sum_{i\in\mathcal{A}}R_i \geq f\left(\mathcal{A}, r_{\mathcal{I}_L}, D\right)\\
\forall \text{ nonempty set } \mathcal{A}\subseteq\mathcal{I}_L\Bigg\}
\end{aligned}\tag{39}$$

where $\mathcal{F}(D)$ is defined in (8). Hence there must exist a vector $r_{\mathcal{I}_L} \in \mathbb{R}_+^L$ satisfying the constraints (20) and (21) in Lemma 3.5, i.e.

$$\sum_{i\in\mathcal{A}}R_i \geq f(\mathcal{A}, r_{\mathcal{I}_L}, D^*\left(R_{\mathcal{I}_L}\right)) \qquad \forall \text{ nonempty set } \mathcal{A}\subseteq\mathcal{I}_L \tag{40}$$

and

$$\frac{1}{\sigma_X^2}+\sum_{i=1}^{L}\frac{1-\exp(-2r_i)}{\sigma_{N_i}^2}\geq\frac{1}{D^*\left(R_{\mathcal{I}_L}\right)}. \tag{41}$$

Let $\mathcal{G} = \{i \in \mathcal{I}_L : r_i > 0\}$. Note that (40) and (41) can be reduced to the following inequalities:

$$\sum_{i \in \mathcal{A}} R_i \geq f\left(\mathcal{A}, r_{\mathcal{I}_L}, D^*\left(R_{\mathcal{I}_L}\right)\right) \qquad \forall \text{ nonempty set } \mathcal{A} \subseteq \mathcal{G}$$

and

$$\frac{1}{\sigma_X^2} + \sum_{i \in \mathcal{G}} \frac{1 - \exp(-2r_i)}{\sigma_{N_i}^2} \geq \frac{1}{D^*\left(R_{\mathcal{I}_L}\right)}.$$

Without loss of generality, we shall assume $\mathcal{G} = \mathcal{I}_L$; otherwise, by restricting to the set $\mathcal{G}$, the following argument can still be applied. Note that for any nonempty sets $\mathcal{S}, \mathcal{T} \subseteq \mathcal{I}_L$ such that the constraints on $\sum_{i \in \mathcal{S}} R_i$ and $\sum_{i \in \mathcal{T}} R_i$ are tight in (40), we have

$$\begin{aligned} f\left(\mathcal{S}, r_{\mathcal{I}_L}, D^*\left(R_{\mathcal{I}_L}\right)\right) &+ f\left(\mathcal{T}, r_{\mathcal{I}_L}, D^*\left(R_{\mathcal{I}_L}\right)\right) \\ &= \sum_{i \in \mathcal{S}} R_i + \sum_{i \in \mathcal{T}} R_i \\ &= \sum_{i \in \mathcal{S} \cup \mathcal{T}} R_i + \sum_{i \in \mathcal{S} \cap \mathcal{T}} R_i \\ &\geq f\left(\mathcal{S} \cup \mathcal{T}, r_{\mathcal{I}_L}, D^*\left(R_{\mathcal{I}_L}\right)\right) + f\left(\mathcal{S} \cap \mathcal{T}, r_{\mathcal{I}_L}, D^*\left(R_{\mathcal{I}_L}\right)\right). \end{aligned}$$

Therefore, it follows from (38) that either $\mathcal{S} \subseteq \mathcal{T}$ or $\mathcal{T} \subseteq \mathcal{S}$. Let $\tilde{\mathcal{A}} = \bigcap_{k \in \mathcal{I}_K} \mathcal{A}_k$, where $\mathcal{A}_k$ $(k \in \mathcal{I}_K)$ are the sets for which the constraints on $\sum_{i \in \mathcal{A}_k} R_i$ are tight in (40). If there is no such $\mathcal{A}_k$, let $\tilde{\mathcal{A}} = \mathcal{I}_L$. Thus $\tilde{\mathcal{A}}$ is always nonempty. Now suppose

$$\frac{1}{\sigma_X^2} + \sum_{i=1}^{L} \frac{1 - \exp(-2r_i)}{\sigma_{N_i}^2} > \frac{1}{D^*\left(R_{\mathcal{I}_L}\right)}. \tag{42}$$

Picking any $i^* \in \tilde{\mathcal{A}}$, we can decrease r_{i^*} to $r_{i^*} - \delta$ for some $\delta > 0$ so that all the constraints in (40) and (42) become nontight. Then we can decrease $D^*(R_{\mathcal{I}_L})$ to $D^*(R_{\mathcal{I}_L}) - \epsilon$ for some $\epsilon > 0$ without violating any constraints in (40) and (42). It follows from (39) that $R_{\mathcal{I}_L} \in \mathcal{R}(D^*(R_{\mathcal{I}_L}) - \epsilon)$, which is contradictory to the definition of $D^*(R_{\mathcal{I}_L})$. Hence we must have

$$\frac{1}{\sigma_X^2} + \sum_{i=1}^{L} \frac{1 - \exp(-2r_i)}{\sigma_{N_i}^2} = \frac{1}{D^*\left(R_{\mathcal{I}_L}\right)}. \tag{43}$$

Now we proceed to show that $r_{\mathcal{I}_L}$ must be unique. It is easy to verify the following facts: 1) $1/\sigma_X^2 + \sum_{i=1}^{L}(1 - \exp(-2r_i))/\sigma_{N_i}^2$ is a strictly concave function of $r_{\mathcal{I}_L}$; 2) for any nonempty set $\mathcal{A} \subseteq \mathcal{I}_L$, $f(\mathcal{A}, r_{\mathcal{I}_L}, D)$ is convex in $r_{\mathcal{I}_L}$. Suppose both $r'_{\mathcal{I}_L}$ and $r''_{\mathcal{I}_L} \in \mathbb{R}_+^L$ satisfy the constraints (40) and (41), and there exists some i^* such that $r'_{i^*} \neq r''_{i^*}$. We shall first show that r'_{i^*}, r''_{i^*} are both finite. If not, without loss of generality suppose $r'_{i^*} = \infty$, which implies that $R_{i^*} = \infty$. Now construct a new vector $r'''_{\mathcal{I}_L}$ such that $r'''_i = r'_i = \infty$ if $i = i^*$, and $r'''_i = r''_i$ otherwise. It is easy to check that $r'''_{\mathcal{I}_L}$ satisfies the constraints (40) and (41). But we have

$$\begin{aligned} \frac{1}{\sigma_X^2} &+ \sum_{i=1}^{L} \frac{1 - \exp(-2r'''_i)}{\sigma_{N_i}^2} \\ &> \frac{1}{\sigma_X^2} + \sum_{i=1}^{L} \frac{1 - \exp(-2r''_i)}{\sigma_{N_i}^2} = \frac{1}{D^*\left(R_{\mathcal{I}_L}\right)} \end{aligned}$$

which is contradictory to (43). Now let $\bar{r}_i = (r'_i + r''_i)/2$ for all $i \in \mathcal{I}_L$. Note that $\bar{r}_{i^*}$ is equal to neither r'_{i^*} nor r''_{i^*} since $r'_{i^*} \neq r''_{i^*}$ and both are finite. It is obvious that $\bar{r}_{\mathcal{I}_L} \in \mathbb{R}_+^L$. Furthermore, we have

$$\begin{aligned} \frac{1}{\sigma_X^2} &+ \sum_{i=1}^{L} \frac{1 - \exp(-2\bar{r}_i)}{\sigma_{N_i}^2} \\ &\geq \frac{1}{\sigma_X^2} + \frac{1}{2}\sum_{i=1}^{L} \frac{1 - \exp(-2r'_i)}{\sigma_{N_i}^2} + \frac{1}{2}\sum_{i=1}^{L} \frac{1 - \exp(-2r''_i)}{\sigma_N^2} \\ &\geq \frac{1}{D^*\left(R_{\mathcal{I}_L}\right)} \end{aligned} \tag{44}$$

and

$$\begin{aligned} \sum_{i \in \mathcal{A}} R_i &\geq \frac{1}{2} f\left(\mathcal{A}, r'_{\mathcal{I}_L}, D^*\left(R_{\mathcal{I}_L}\right)\right) + \frac{1}{2} f\left(\mathcal{A}, r''_{\mathcal{I}_L}, D^*\left(R_{\mathcal{I}_L}\right)\right) \\ &\geq f(\mathcal{A}, \bar{r}_{\mathcal{I}_L}, D^*\left(R_{\mathcal{I}_L}\right)) \qquad \forall \text{ nonempty set } \mathcal{A} \subseteq \mathcal{I}_L. \end{aligned}$$

Hence $\bar{r}_{\mathcal{I}_L}$ satisfies the constraints (40) and (41). Since $1/\sigma_X^2 + \sum_{i=1}^{L}(1 - \exp(-2r_i))/\sigma_{N_i}^2$ is a strictly concave function of $r_{\mathcal{I}_L}$, the first inequality in (44) is strict, which results in a contradiction with (43).

Note that (22) follows from (43). So only (23) remains to be proved. We shall first show that $r_i^*(R_{\mathcal{I}_L}) = 0$ implies $R_i = 0$. Without loss of generality, suppose $r_L^*(R_{\mathcal{I}_L}) = 0$. Then it is easy to see that (40) still holds if we set $R_L = 0$ on its left hand side. So if $R_L > 0$, we can increase $r_L^*(R_{\mathcal{I}_L})$ by a small amount without violating (40) and (41), which is contradictory to the fact that $r_L^*(R_{\mathcal{I}_L})$ is unique. Hence without loss of generality, we can assume $r_i^*(R_{\mathcal{I}_L}) > 0$ for all $i \in \mathcal{I}_L$; otherwise, by restricting to the set $\mathcal{G} = \{i \in \mathcal{I}_L : r_i^*(R_{\mathcal{I}_L}) > 0\}$, the following argument can still be applied. Since (22) holds, the right-hand side of (40) becomes $f(\mathcal{A}, r_{\mathcal{I}_L}^*(R_{\mathcal{I}_L}))$. By (37), it can be shown that if in (40) the constraints on $\sum_{i \in \mathcal{S}} R_i$ and $\sum_{i \in \mathcal{T}} R_i$ are tight, then either $\mathcal{S} \subseteq \mathcal{T}$ or $\mathcal{T} \subseteq \mathcal{S}$. Let $\tilde{\mathcal{A}}' = \bigcup_{k \in \mathcal{I}_K} \mathcal{A}_k$, where $\mathcal{A}_k$ $(k \in \mathcal{I}_K)$ are the sets for which the constraints on $\sum_{i \in \mathcal{A}_k} R_i$ are tight in (40). If there is no such $\mathcal{A}_k$, let $\tilde{\mathcal{A}}' = \emptyset$. If $\tilde{\mathcal{A}}' = \mathcal{I}_L$, the proof is complete. Otherwise, pick any $i^* \in \mathcal{I}_L \backslash \tilde{\mathcal{A}}'$; we can increase $r_{i^*}^*(R_{\mathcal{I}_L})$ to $r_{i^*}^*(R_{\mathcal{I}_L}) + \delta$ for some $\delta > 0$ without violating any constraints in (20) and (21), which is contradictory to the uniqueness of $r_{i^*}^*(R_{\mathcal{I}_L})$.

REFERENCES

[1] D. Schonberg, K. Ramchandran, and S. S. Pradhan, "Distributed code constructions for the entire Slepian-Wolf rate region for arbitrarily correlated sources," in *Proc. DCC04*, Snowbird, UT, Mar. 2004.

[2] T. P. Coleman, A. H. Lee, M. Médard, and M. Effros, "Low-complexity approaches to Slepian-Wolf near-lossless distributed data compression," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3546–3561, Aug. 2006.

[3] S. Cheng and Z. Xiong, "Successive refinement for the Wyner-Ziv problem and layered code design," in *Proc. DCC04*, Snowbird, UT, Mar. 2004.

[4] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, July 1973.

[5] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[6] J. Edmonds, "Submodular functions, matroids and certain polyhedra," in *Combinatorial Structures and Their Applications*, R. Guy, H. Hanani, N. Sauer, and J. Schonheim, Eds. New York: Gordon and Breach, 1970, pp. 69–87.

[7] T. S. Han, "Source coding with cross observations at the encoders," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 3, pp. 360–361, May 1979.
[8] T. S. Han, "Slepian-Wolf-Cover theorem for networks of channels," *Inf. Contr.*, vol. 47, pp. 67–83, Oct. 1980.
[9] B. Rimoldi and R. Urbanke, "Asynchronous Slepian-Wolf coding via source-splitting," in *IEEE Int. Symp. Inf. Theory*, Ulm, Germany, Jun. 29–Jul. 4, 1997, p. 271.
[10] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1250–1276, Jun. 2002.
[11] A. B. Carleial, "On the capacity of multiple-terminal communication networks," Ph.D. dissertation, Stanford Univ., Stanford, CA, Aug. 1975.
[12] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple- access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 364–375, Mar. 1996.
[13] A. J. Grant, B. Rimoldi, R. L. Urbanke, and P. A. Whiting, "Rate-splitting multiple access for discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 873–890, Mar. 2001.
[14] B. Rimoldi, "Generalized time sharing: A low-complexity capacity-achieving multiple- access technique," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2432–2442, Sep. 2001.
[15] S. I. Gel'fand and M. S. Pinsker, "Coding of sources on the basis of observations with incomplete information," *Problems of Information Transmission*, vol. 15, no. 2, pp. 115–125, 1979.
[16] T. J. Flynn and R. M. Gray, "Encoding of correlated observations," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 6, pp. 773–787, Nov. 1987.
[17] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.
[18] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, ser. CISM Courses and Lectures, G. Longo, Ed. Vienna/New York: Springer-Verlag, 1978, pp. 171–231.
[19] S. Y. Tung, "Multiterminal source coding," Ph.D. dissertation, School of Elect. Eng., Cornell Univ., Ithaca, NY, May 1978.
[20] 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. Inf. Theory*, vol. IT-25, no. 6, pp. 664–666, Nov. 1979.
[21] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 2, pp. 219–221, Mar. 1979.
[22] P. Viswanath, "Sum rate of multiterminal gaussian source coding," in *Advances in Network Information Theory, DIMACS: Series in Discrete Mathematics and Theoretical Computer Science*, P. Gupta, G. Kramer, and A. Wijngaarden, Eds. Providence, RI: American Mathematical Society, 2004, pp. 43–60.
[23] J. Chen, X. Zhang, T. Berger, and S. B. Wicker, "An upper bound on the sum- Rate distortion function and its corresponding rate allocation schemes for the CEO problem," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 977–987, Aug. 2004.
[24] H. G. Eggleston, *Convexity*. Cambridge, U.K.: Cambridge Univ. Press, 1958.
[25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[26] S. C. Draper and G. W. Wornell, "Side information aware coding strategies for sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 966–976, Aug. 2004.
[27] B. Rimoldi and R. Urbanke, "On the structure of the dominant face of multiple access channels," in *Proc. Inf. Theory Commun. Workshop*, Kruger National Park, South Africa, Jun. 20–25, 1999, pp. 12–14.
[28] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Inf. Theory*, Budapest, Hungary, 1973, pp. 23–52, Hungarian Acad. Sci..
[29] H. Liao, "Multiple access channels," Ph.D. dissertation, Dep. Elec. Eng., Univ. Hawaii, Honolulu, HI, 1972.
[30] T. S. Han, "The capacity region of general multiple-access channel with certain correlated sources," *Inf. Contr.*, vol. 40, pp. 37–60, Jan. 1979.
[31] D. N. C. Tse and S. V. Hanly, "Multiaccess fading channels—Part I: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2796–2815, Nov. 1998.
[32] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: General sources," *Inf. Contr.*, vol. 38, pp. 60–80, Jul. 1978.
[33] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1912–1923, Nov. 1997.
[34] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.
[35] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1549–1559, Sept. 1997.
[36] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2577–2593, Jul. 2005.
[37] V. Prabhakaran, D. Tse, and K. Ramchandran, "Rate region of the quadratic Gaussian CEO problem," in *Proc. Int. Symp. Inf. Theory*, Chicago, IL, Jun. 27–Jul. 2, 2004, p. 119 [Online]. Available: http://www.eecs.berkeley.edu/~vinodmp/publications/ISIT04a.pdf
[38] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[39] V. Koshelev, "Hierarchical coding of discrete sources," *Probl. Pered. Inf.*, vol. 16, no. 3, pp. 31–49, 1980.
[40] V. Koshelev, "Estimation of mean error for a discrete successive-approximation scheme," *Probl. Pered. Inf.*, vol. 17, no. 3, pp. 20–33, 1981.
[41] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 269–274, Mar. 1991.
[42] B. Rimoldi, "Successive refinement of information: Characterization of the achievable rates," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 253–259, Jan. 1994.
[43] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1636–1654, Aug. 2004.