Analysis and Code Design for the Binary CEO Problem Under Logarithmic Loss

Mahdi Nangir, Reza Asvadi, *Member, IEEE*, Mahmoud Ahmadian-Attari, *Member, IEEE*, and Jun Chen, *Senior Member, IEEE*

***Abstract*— In this paper, we propose an efficient coding scheme for the binary Chief Executive Officer (CEO) problem under logarithmic loss criterion. Courtade and Weissman obtained the exact rate-distortion bound for a two-link binary CEO problem under this criterion. We find optimal parameters of the binary symmetric test-channel model for the encoder of each link by using the given bound. Furthermore, an efficient coding scheme based on compound low-density generator matrix (LDGM)–low-density parity-check (LDPC) codes is presented to achieve the theoretical rates. In the proposed encoding scheme, a binary quantizer using LDGM codes and a syndrome generator using LDPC codes are applied. The proposed decoder employs a sum-product algorithm and a soft estimator to produce an approximate *a posteriori* distribution of the source bits given the data received through both links. Our numerical examples verify a close performance of the proposed coding scheme to the theoretical bound in several cases.**

***Index Terms*— Binary CEO problem, logarithmic loss (log-loss), test channel model, compound LDGM-LDPC codes, soft CEO decoder.**

I. Introduction

THE CHIEF EXECUTIVE OFFICER (CEO) problem is defined by Berger *et al.* for distributed source coding of multi-observation of a source corrupted by independent noises [2]. By using the compressed observations, a fusion center makes an estimation of the source at the receiver with an acceptable distortion between original and estimated symbols. In the last two decades, there has been an explosion of studies on the theoretical bounds of the transmission rate in the CEO problem in the case of noisy observations of a Gaussian source corrupted by independent additive Gaussian noises [3]–[6]. This case is usually known as the quadratic Gaussian CEO problem. The CEO problem empirically emerges in wireless sensor networks, where a particular phenomenon is measured by some separate and independent sensors in a noisy environment.

A tight upper bound on the sum-rate-distortion function of the quadratic Gaussian CEO problem and the optimal rate allocation scheme are provided in [7]. Alternatively, studies like [8]–[10] present various coding schemes to achieve any point of the rate-distortion region of the quadratic Gaussian CEO problem. Moreover, an optimal coding scheme based on the successive Wyner-Ziv coding structure is applied to achieve the bounds of the quadratic Gaussian CEO in [8].

The case of a binary source with observations corrupted by binary noises, called the binary CEO problem, has been paid less attention during these years. In general, the exact rate-distortion bound of this case and its associated multi-terminal source coding problem are open problems in information theory. The most common criterion for measuring distortion in the binary case is the Hamming distortion measure [11]. The binary CEO problem appears in cooperative digital communication networks where some correlated remote sources are being sent to a central receiver via paralleled channels with independent noises.

A lower bound for the rate-distortion region of a two-link binary CEO problem is established in [11] using the Hamming distortion benchmark. The Berger-Tung inner and outer bounds [12] are exploited for this case which are not tight under the Hamming distortion criterion. Some useful bounds on the rate-distortion performance of the binary CEO problem under the Hamming distortion measure are given in [13] and [14]. The prior studies on the binary CEO problem in [11], [13], [15], and [16] consider that the correlated observations are transmitted through AWGN channels, and hence their encoders apply a channel coding to protect the transmitted data. Thus, the problem definition in those papers differs from the standard CEO problem, defined in [2], for which the transmission links are assumed to be noiseless and the encoders employ source coding schemes, alternatively. In contrast, we follow the lossy distributed source coding framework in the binary CEO problem. Thus, our goal is to achieve the maximum compression of the correlated noisy observations for sending them through noiseless channels with minimum distortion.

In recent years, the logarithmic loss (or simply log-loss) has gained popularity as a distortion measure in lossy source coding where the reconstructions are given in the form of

Manuscript received January 1, 2018; revised May 19, 2018 and July 27, 2018; accepted July 29, 2018. Date of publication August 6, 2018; date of current version December 14, 2018. This paper was presented at the 29th Biennial Symposium on Communications, Toronto, Canada, 2018 [1]. The associate editor coordinating the review of this paper and approving it for publication was R. F. Schaefer. *(Corresponding author: Reza Asvadi.)*

M. Nangir and M. Ahmadian-Attari are with the Faculty of Electrical Engineering, K. N. Toosi University of Technology, Tehran 1631714191, Iran (e-mail: mahdinangir@ee.kntu.ac.ir; mahmoud@eetd.kntu.ac.ir).

R. Asvadi is with the Faculty of Electrical Engineering, Shahid Beheshti University, Tehran 1983963113, Iran (e-mail: r_asvadi@sbu.ac.ir).

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

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2018.2863377

probability distributions [17], [18]. Such soft reconstructions are more suitable than hard reconstructions for applications like classification, learning, data mining, and prediction. Likewise, deep neural networks used for image classification commonly output a distribution which assigns different probabilities to different classes. In this paper, we focus on the binary CEO problem under the log-loss criterion. One of the advantages of considering the CEO problem under the log-loss is that the corresponding rate-distortion region is known [19]. Consequently, the rate-distortion performance of a designed coding scheme relative to the fundamental limit can be measured with more accuracy.

Although the CEO problem under the log-loss is a distributed source coding problem, it is closely related to the channel coding problem (especially the compress-and-forward scheme) for Cloud-Radio Access Networks (C-RANs) [20]. Indeed, the results obtained in this work, when properly interpreted, are applicable to C-RANs as well. Specifically, consider the following scenario: in a C-RAN, a source sends an encoded message using binary signaling to two base stations (BSs) via two binary symmetric channels; the BSs compress their respective observations and forward the compressed data to the cloud center over fiber optic links which can be modeled as bit pipes due to their low bit error rate performance; finally, the cloud center decodes the transmitted message based on the received data. It can be shown that, in a certain sense, maximizing the end-to-end throughput of this C-RAN is equivalent to minimizing the log-loss of the binary CEO problem.

Our main contributions in this paper can be considered in the contexts of both information theory and coding theory. First, an exact rate-distortion bound is presented for a two-link binary CEO problem under the log-loss distortion. Next, we adopt a binary symmetric test channel model for each encoder in the binary CEO problem and compute the corresponding optimal parameters. Finally, efficient encoding and decoding schemes are proposed by utilizing the compound LDGM-LDPC codes and iterative message-passing algorithms. We show that the rate-distortion performance of the proposed coding scheme is close to the theoretical bounds.

The organization of this paper is as follows. In Section II, the problem definition, preliminaries, and notations are provided. Information theoretic aspects of the binary CEO problem under the log-loss are described in Section III. Optimal parameters of the binary symmetric test channel model are also presented in this section. Next in Section IV, we provide the designed encoding and decoding schemes in details. Numerical results and discussions are presented in Section V. Finally, Section VI draws the conclusion.

II. Preliminaries

In this paper, we use uppercase letters for denoting a random variable like one used in [19]. The realization of random variables are denoted by lowercase letters and the alphabet sets of random variables are denoted by calligraphic letters. Throughout this paper, the logarithm is to base 2. In the Tanner graph representation of codes, first subscript shows the index

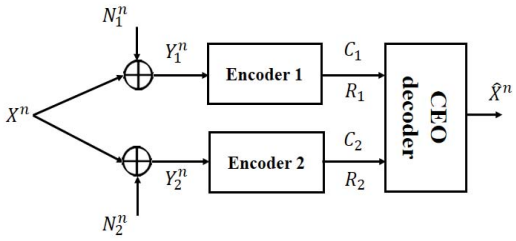


Fig. 1. Block diagram of the two-link binary CEO problem.

of each associated link for any length, rate, distortion, and etc. Some other used notations are as follows: $p * d = p(1-d) + d(1-p)$ is binary convolution of d and p, for $0 \leq p, q \leq 1$ and $[x]^+ = \max\{0, x\}$. Let $h_b(x) = -x\log x - (1-x)\log(1-x)$ be the binary entropy function where its first and second derivatives are, respectively, $h_b'(x) = \log\left(\frac{1-x}{x}\right)$ and $h_b''(x) = -\frac{\log e}{x(1-x)}$, where $e \approx 2.7182$. The functions $h_b(x)$, $h_b'(x)$, and $h_b''(x)$ are, respectively, increasing, decreasing, and increasing functions in $x \in (0, 0.5]$.

A. System Model and Definitions

Consider a communication system consisting of an independent and identically distributed (i.i.d.) binary symmetric source (BSS) and its two noisy observations being transmitted via two parallel links as depicted in Fig. 1. Let X^n, ${Y_1}^n$, and ${Y_2}^n$ denote a sequence of the BSS and two noisy observations of it, on the first and the second links with length n, respectively. Observation noises N_1^n and N_2^n are independent from each other and are i.i.d. binary sequences generated by Bernoulli distributions with crossover parameters p_1 and p_2 associated to the first and the second links, respectively. Consider ${Y_1}^n$ and ${Y_2}^n$ are encoded to C_1 and C_2, and then they are sent to the CEO joint decoder. Note that $C_1 \leftrightarrow Y_1^n \leftrightarrow X^n \leftrightarrow Y_2^n \leftrightarrow C_2$ form a Markov chain. At the decoder, the binary sequence $\hat{X}^n$ is reconstructed in the joint decoder of CEO by using (C_1, C_2).

Each encoder consists of a function f_i, $i = 1, 2$, which compresses the observation as follows:

$$f_i(Y_i^n) = C_i, \quad \text{where } Y_i^n \in \mathcal{Y}_i^n = \{0,1\}^n \text{ and } C_i \in \mathcal{C}_i, \quad \text{for } i = 1, 2. \tag{1}$$

The CEO decoder is a function g which maps the ordered pair (C_1, C_2) to the reconstruction $\hat{X}^n$,

$$g(C_1, C_2) = \hat{X}^n, \quad \text{where } (C_1, C_2) \in \mathcal{C}_1 \times \mathcal{C}_2. \tag{2}$$

In the lossy source coding theory, Hamming distance is a prevalent and a classic criterion for measuring the average number of flipped bits between the estimated binary sequence $\hat{X}^n$ compared to the original binary sequence X^n, and is denoted by $d_{\mathrm{H}}(X^n, \hat{X}^n) \triangleq \frac{1}{n}\sum_{j=1}^{n}[x_j \oplus \hat{x}_j]$, where $\oplus$ means the binary sum operation. If the estimated sequence may not necessarily be binary, another criterion is needed to measure the distance between these two sequences with different alphabets. In this case, an efficient conditional entropy-based distortion measure is used where probability distributions of the original source alphabet, binary in this paper, is the same as

the one that will be used in the reconstructed source alphabet. This distortion measure is called log-loss.

Definition 1: Symbol-wise log-loss between a source symbol x_j and its reconstruction $\hat{x}_j$ is defined as follows:

$$d(x_j, \hat{x}_j) = \log\big(\frac{1}{\hat{x}_j(x_j)}\big), \quad j = 1, 2, \cdots, n. \tag{3}$$

where $\hat{x}_j(x_j)$ generally depends on (c_1, c_2). The total value of log-loss distortion between x^n and $\hat{x}^n$ is obtained by averaging over all the symbols, i.e.,

$$d(x^n, \hat{x}^n) = \frac{1}{n}\sum_{j=1}^{n} \log\big(\frac{1}{\hat{x}_j(x_j)}\big). \tag{4}$$

Definition 2: A rate distortion vector (R_1, R_2, D) is called strict-sense achievable for a distortion measure $d(\cdot,\cdot)$, if there exist functions f_1, f_2, and g according to (1) and (2) such that for length n,

$$\begin{aligned} R_i &\geq \frac{1}{n}\log\big|\mathcal{C}_i\big|, \quad \text{for } i = 1, 2; \\ D &\geq \mathbb{E}d(X^n, \hat{X}^n), \end{aligned} \tag{5}$$

where $\mathbb{E}(\cdot)$ denotes expectation function.

Definition 3: The closure of the set of all strict-sense achievable vectors (R_1, R_2, D) is called achievable rate-distortion region of the binary CEO problem and is denoted by $\overline{\mathcal{RD}}^{\star}_{\text{CEO}}$. Furthermore, $\mathcal{RD}^{i}_{\text{CEO}}$ and $\mathcal{RD}^{o}_{\text{CEO}}$ denote inner and outer bounds of the rate-distortion region, respectively.

B. Message-Passing Algorithms

In our proposed coding scheme, we apply different types of message-passing algorithms depending on their applications. The Bias-Propagation (BiP) algorithm [21] is applied for lossy compression of a given binary source. It maps each output sequence of the source to a codeword of a given low-density generator matrix (LDGM) code which has the nearest Hamming distance to the source output. It achieves the rate-distortion bound of the BSS, and hence it is usually known as a binary quantizer in the context of source coding. Specifically, we can approach a target binary rate-distortion pair $(R, D) = (1 - h_b(d), d)$ by employing LDGM codes. In each round of this algorithm, a bias value for each variable node is calculated and then it is compared with a threshold. Regarding this comparison, the values of at least one of the variable nodes is determined in the quantized sequence. This process continues until values of all the variable nodes are fixed. Details of the BiP algorithm including update equations and damping process are presented in [21] and [22].

Another useful message passing algorithm is the Sum-Product (SP) algorithm [23] that is basically a decoding technique for low-density parity-check (LDPC) codes with an specified code rate and a degree distribution. In distributed lossless source coding, this algorithm is widely used as a syndrome decoder for finding the nearest sequence to a particular sequence called side information using the given syndrome (see [24] and [25]). The iterative routine for executing this algorithm is given in [26]. This algorithm can asymptotically achieve zero bit-error-rate (BER) for a target code rate equal to the capacity of a virtual channel between the original source and the side information.

III. The Information-Theoretic Aspect

In this section, we investigate the information-theoretic aspect of the binary CEO problem. We review existent bounds on the rate-distortion performance of this problem and then find optimal parameters of the binary symmetric test channel model. The Berger-Tung inner and outer bounds [12] are not generally tight, especially in the binary CEO problem case with Hamming distortion measure. If there exists a gap between the inner and the outer bounds, then measuring and comparing the rate-distortion performance of designed codes are inaccurate. Therefore, the existence of a tight bound seems crucial for the performance analysis of a code design. Because of this, the Berger-Tung coding scheme is not optimal for the binary CEO problem under the Hamming loss in the sense of achieving the exact rate-distortion bound [12]. In our proposed coding scheme, the total distortion is measured by using the log-loss definition (4).

A. Binary CEO Problem Bound Under the Log-Loss

Theoretical rate-distortion bound of the binary CEO problem is unknown when distortion measure is the Hamming distance, however, the inner and the outer bounds are only available for this case. Alternatively, if the log-loss criterion being used to measure distortion, the rate-distortion region is exactly established. Specifically, the classical Berger-Tung scheme yields the following inner bound of $\overline{\mathcal{RD}}^{\star}_{\text{CEO}}$ according to [19, Definition 3 and Th. 1]. Let $(R_1, R_2, D) \in \mathcal{RD}^{i}_{\text{CEO}}$, if and only if, there exists a joint distribution of the form

$$p(x)p(y_1|x)p(y_2|x)p(u_1|y_1, q)p(u_2|y_2, q)p(q), \tag{6}$$

where $|\mathcal{U}_i| \leq |\mathcal{Y}_i|$ for $i = 1, 2$, and $|\mathcal{Q}| \leq 4$, which satisfies

$$\begin{aligned} R_1 &\geq I(Y_1; U_1|U_2, Q), \\ R_2 &\geq I(Y_2; U_2|U_1, Q), \\ R_1 + R_2 &\geq I(Y_1, Y_2; U_1, U_2|Q), \\ D &\geq H(X|U_1, U_2, Q). \end{aligned} \tag{7}$$

Furthermore, an outer bound is provided for the binary CEO problem under the log-loss according to [19, Definition 4 and Th. 2]. Let $(R_1, R_2, D) \in \mathcal{RD}^{o}_{\text{CEO}}$, if and only if, there exists a joint distribution of the form (6) satisfies the following inequalities,

$$\begin{aligned} R_1 &\geq [I(Y_1; U_1|X, Q) + H(X|U_2, Q) - D]^{+}, \\ R_2 &\geq [I(Y_2; U_2|X, Q) + H(X|U_1, Q) - D]^{+}, \\ R_1 + R_2 &\geq [I(Y_1; U_1|X, Q) + I(Y_2; U_2|X, Q) \\ &\qquad + H(X) - D]^{+}, \\ D &\geq H(X|U_1, U_2, Q). \end{aligned} \tag{8}$$

The most important result in [19], related to our work, is *Theorem 3* and its extension. It states that the bounds in (7) and (8) are the same, yielding a computable characterization of the rate-distortion region under the log-loss. We shall focus

on the two-link binary CEO problem,[1] for which it suffices to have $|\mathcal{Q}| \leq 4$ and $|\mathcal{U}_i| \leq |\mathcal{Y}_i| = 2$, $i = 1, 2$. To find a complete characterization of the sum-rate-distortion function for this problem, we should solve the following optimization problem:

$$\min_{p(u_1|y_1,q)p(u_2|y_2,q)p(q)} I(U_1, U_2; Y_1, Y_2|Q), \quad \text{s.t. } H(X|U_1, U_2, Q) = D_0, \tag{9}$$

where $H(X|Y_1, Y_2) \leq D_0 \leq 1$. This optimization problem can be written in the following unconstrained form:

$$\min_{p(u_1|y_1,q)p(u_2|y_2,q)p(q)} H(X|U_1, U_2, Q) + \mu I(U_1, U_2; Y_1, Y_2|Q), \tag{10}$$

where μ is the Lagrangian multiplier. Note that

$$\begin{aligned} &H(X|U_1, U_2, Q) + \mu I(U_1, U_2; Y_1, Y_2|Q) \\ &\quad = \sum_{q\in\mathcal{Q}} p(q)[H(X|U_1, U_2, Q = q) \\ &\qquad + \mu I(U_1, U_2; Y_1, Y_2|Q = q)] \\ &\quad \geq \min_{q\in\mathcal{Q}} H(X|U_1, U_2, Q = q) \\ &\qquad + \mu I(U_1, U_2; Y_1, Y_2|Q = q). \end{aligned} \tag{11}$$

Assume that the minimum in the above expression is achieved at q^*. Now, we define U_1^*, U_2^*, and Q^* joint distributed with X, Y_1, and Y_2 according to

$$\begin{aligned} &p_{X,Y_1,Y_2}(x, y_1, y_2) \\ &\quad \times p_{U_1^*|Y_1,Q^*}(u_1^*|y_1, q^*) p_{U_2^*|Y_2,Q^*}(u_2^*|y_2, q^*) p_{Q^*}(q^*). \end{aligned} \tag{12}$$

If we set $p_{U_1^*|Y_1,Q^*} = p_{U_1|Y_1,Q}$, $p_{U_2^*|Y_2,Q^*} = p_{U_2|Y_2,Q}$, $p_{Q^*}(q^*) = 1$, and $p_{Q^*}(q) = 0$ for $q \neq q^*$, then

$$\begin{aligned} &H(X|U_1, U_2, Q) + \mu I(U_1, U_2; Y_1, Y_2|Q) \\ &\quad \geq H(X|U_1, U_2, Q = q^*) + \mu I(U_1, U_2; Y_1, Y_2|Q = q^*) \\ &\quad = H(X|U_1^*, U_2^*, Q^*) + \mu I(U_1^*, U_2^*; Y_1, Y_2|Q^*). \end{aligned} \tag{13}$$

Therefore, this newly constructed (U_1^*, U_2^*, Q^*) achieves the lower bound. In summary, we should assign all weights to a particular realization of Q that achieves the minimum, and consequently for the purpose of characterizing the sum-rate-distortion function, there is no loss of generality in assuming that Q is a constant, which leads to the following simplified optimization problem:

$$\min_{p(u_1|y_1)p(u_2|y_2)} H(X|U_1, U_2) + \mu I(U_1, U_2; Y_1, Y_2) \triangleq F. \tag{14}$$

B. Binary Symmetric Test Channel Model for the Encoders

We shall assume[2] that $p(u_i|y_i)$ is a binary symmetric channel with crossover probability d_i, $i = 1, 2$. Consequently, after some calculus manipulations in (7), the rate-distortion bounds are expressed by:

$$\begin{aligned} R_1 &\geq h_b(p * d) - h_b(d_1), \\ R_2 &\geq h_b(p * d) - h_b(d_2), \\ R &\triangleq R_1 + R_2 \geq 1 + h_b(p * d) - h_b(d_1) - h_b(d_2), \\ D &\geq h_b(p_1 * d_1) + h_b(p_2 * d_2) - h_b(p * d), \end{aligned} \tag{15}$$

where $p \triangleq p_1 * p_2$ and $d \triangleq d_1 * d_2$. In this case, the optimization problem (14) is equivalent to:

$$\begin{aligned} \min_{0\leq d_1,d_2\leq 0.5} \Big[&h_b(p_1 * d_1) + h_b(p_2 * d_2) - h_b(p * d) \\ &+ \mu\big(1 + h_b(p * d) - h_b(d_1) - h_b(d_2)\big)\Big], \end{aligned} \tag{16}$$

for any μ. The following example indicates that setting $d_1 = d_2$ is not necessarily an optimal choice even if $p_1 = p_2$.

Example 1: Let consider $p_1 = p_2 = 0.1$ and also assume that the minimum achievable sum-rate R is fixed to 0.6, i.e., $1 + h_b(p * d) - h_b(d_1) - h_b(d_2) = 0.6$. First, let $d_1 = d_2$. By a simple calculation, $d_1 = d_2 = 0.177$ is obtained, and then the minimum achievable distortion D will be equal to 0.6474. Alternatively, let presume that the total information is only sent over the first link, i.e., $d_2 = 0.5$. Thus, d_1 and D are calculated as 0.0795 and 0.6428, respectively. Consequently, the distortion value by using only one of the links is unexpectedly smaller than the case that both of them are used, and hence finding optimum values of d_1 and d_2 is interesting. Optimality in this case means achieving the minimum achievable log-loss distortion subject to a given minimum achievable sum-rate. First of all, we show that the optimization problem (9) is not convex even with the restriction to the binary symmetric test channel model.

Theorem 1: The lower bounds of the sum-rate R and distortion D in (15) are neither convex nor concave in terms of variables (d_1, d_2).

Proof: For ease of notation, we denote these lower bounds by R and D. From (15) we have:

$$\begin{aligned} \frac{\partial^2 R}{\partial d_i^2} &= (1 - 2p * d_{3-i})^2 h_b''(p * d) - h_b''(d_i), \\ \frac{\partial^2 R}{\partial d_1 \partial d_2} &= (1 - 2p * d_1)(1 - 2p * d_2) h_b''(p * d) \\ &\quad - 2(1 - 2p) h_b'(p * d), \\ \frac{\partial^2 D}{\partial d_i^2} &= (1 - 2p_i)^2 h_b''(p_i * d_i) \\ &\quad - (1 - 2p * d_{3-i})^2 h_b''(p * d), \\ \frac{\partial^2 D}{\partial d_1 \partial d_2} &= -(1 - 2p * d_1)(1 - 2p * d_2) h_b''(p * d) \\ &\quad + 2(1 - 2p) h_b'(p * d). \end{aligned} \tag{17}$$

The Hessian matrices of the rate and distortion are, respectively, $H_R = [\frac{\partial^2 R}{\partial d_i \partial d_j}]$ and $H_D = [\frac{\partial^2 D}{\partial d_i \partial d_j}]$, for $i, j = 1, 2$. By defining $q_i \triangleq p * d_{3-i}$, obviously $q_i \geq p_i$. After some

[1] The extension to the m-link case is straightforward.

[2] Although we have not been able to find an analytical proof, extensive numerical results suggest that this assumption incurs no loss of optimality for the purpose of solving (14).

calculations, we have:

$$\frac{\partial^2 R}{\partial d_i^2} = \frac{\log e\ q_i(1-q_i)}{q_i * d_i(1-q_i * d_i)d_i(1-d_i)} \geq 0,$$
$$\frac{\partial^2 D}{\partial d_i^2} = \frac{\log e\ [p_i(1-p_i)-q_i(1-q_i)]}{p_i * d_i(1-p_i * d_i)q_i * d_i(1-q_i * d_i)} \leq 0. \quad (18)$$

For fixed values of d_i, sum-rate R and distortion D are, respectively, a convex and a concave single-variable functions in terms of d_{3-i}, for $i = 1, 2$ according to (18). We show that the determinant of the Hessian matrices for R and D are not positive with a counter-example.

$$\begin{aligned}\det\big[H_R\big] &= \frac{\partial^2 R}{\partial d_1^2}\frac{\partial^2 R}{\partial d_2^2} - \big(\frac{\partial^2 R}{\partial d_1 \partial d_2}\big)^2 \\ &= \big(\frac{\log e\ q_1(1-q_1)}{p * d(1-p * d)d_1(1-d_1)}\big) \\ &\quad \times \big(\frac{\log e\ q_2(1-q_2)}{p * d(1-p * d)d_2(1-d_2)}\big) \\ &\quad - \big(2(1-2p)\log\big[\frac{1-p * d}{p * d}\big] \\ &\quad + \frac{\log e\ (1-2q_1)(1-2q_2)}{p * d(1-p * d)}\big)^2. \end{aligned} \quad (19)$$

Let calculate $\det\big[H_R\big]$ for $d_1 = d_2 = 0.1$ where $p_1 \to 0$ and $p_2 \to 0$. In this case, $d = 0.18$, $p \to 0$, $q_1 \to 0.1$, and $q_2 \to 0.1$. Hence,

$$\begin{aligned}\det\big[H_R\big] &= \big(\frac{0.1298}{0.0133}\big) \times \big(\frac{0.1298}{0.0133}\big) - \big(4.3752 + 6.2555\big)^2 \\ &= -17.7659 < 0. \end{aligned} \quad (20)$$

This counter-example shows that R is neither convex nor concave in general. Similarly, for distortion D, we have:

$$\begin{aligned}\det\big[H_D\big] &= \frac{\partial^2 D}{\partial d_1^2}\frac{\partial^2 D}{\partial d_2^2} - \big(\frac{\partial^2 D}{\partial d_1 \partial d_2}\big)^2 \\ &= \big(\frac{\log e\ [q_1(1-q_1)-p_1(1-p_1)]}{p * d(1-p * d)p_1 * d_1(1-p_1 * d_1)}\big) \\ &\quad \times \big(\frac{\log e\ [q_2(1-q_2)-p_2(1-p_2)]}{p * d(1-p * d)p_2 * d_2(1-p_2 * d_2)}\big) \\ &\quad - \big(2(1-2p)\log\big[\frac{1-p * d}{p * d}\big] \\ &\quad + \frac{\log e\ (1-2q_1)(1-2q_2)}{p * d(1-p * d)}\big)^2. \end{aligned} \quad (21)$$

Now we calculate $\det\big[H_D\big]$ for $p_1 = p_2 = 0.1$ when $d_1 \to 0$ and $d_2 \to 0$. In this case, $p = 0.18$, $d \to 0$, $q_1 \to 0.18$, and $q_2 \to 0.18$. Hence,

$$\begin{aligned}\det\big[H_D\big] &= \big(\frac{0.083}{0.0133}\big) \times \big(\frac{0.083}{0.0133}\big) - \big(2.8002 + 4.0036\big)^2 \\ &= -7.3465 < 0. \end{aligned} \quad (22)$$

This counter-example also shows that D is neither convex nor concave in general. □

Regarding the above theorem, the objective function $F(\mu) \triangleq D + \mu R$ in (16), which is a two-dimensional function of (d_1, d_2), is not convex in general. For a fixed value of μ, solution of (16) is an ordered pair denoted by (d_1^*, d_2^*) which achieves minimum value of F. If $p_1 = p_2$, only pairs that satisfy $d_1^* \leq d_2^*$ will be considered as an acceptable solution due to the symmetry. Now, the solution of the problem (16) is found via a brute-force search on the plane (d_1, d_2) with sufficiently small step-sizes.

There exist two definite boundary points in $F(\mu)$. First, it arises in $\mu = 0$ where the objective function equals distortion, and hence the minimum value $D_{\min}$ equals $H(X|Y_1, Y_2)$ for $(d_1^*, d_2^*) = (0, 0)$. Second, it occurs in $\mu_{\max}$ by which the minimum value of the objective function equals to $H(X) = 1$ for all μ values equal or greater than $\mu_{\max}$, i.e., $F_{\min}(\mu) = 1$ for $\forall \mu \geq \mu_{\max}$. In the latter boundary point, the solution is located in $(d_1^*, d_2^*) = (0.5, 0.5)$ and the sum-rate $R = 0$. Thus, it is sufficient to study behavior of the objective function between these two boundary points, i.e., $0 \leq \mu \leq \mu_{\max}$. Our results show that location of the solutions depends on the value of noise parameters p_1 and p_2 whether they are equal or not. In Fig. 2, location of the solution points (d_1^*, d_2^*) are depicted for several cases. When $p_1 = p_2$, as it is seen in these curves, there exist two threshold values for parameter μ, denoted by μ_{t_1} and μ_{t_2}, related to non-smooth critical points of the curves. These critical points are used to categorize location of the optimum solutions. In *Region 1*, $0 \leq \mu \leq \mu_{t_1}$, the optimum points (d_1^*, d_2^*) are located on the line $d_1^* = d_2^*$. In *Region 2*, $\mu_{t_1} \leq \mu \leq \mu_{t_2}$, the optimum points are located on a curve such that $d_1^* < d_2^* < 0.5$. In *Region 3*, $\mu_{t_2} \leq \mu \leq \mu_{\max}$, the solutions are located on the boundary points. However, when $p_1 \neq p_2$, there is only one threshold value for μ denoted by μ_t corresponding to the critical point of the curve. In *Region 1*, $0 \leq \mu \leq \mu_t$, the solutions are located on a curve such that $d_1^* < d_2^* < 0.5$, and in *Region 2*, $\mu_t \leq \mu \leq \mu_{\max}$, the optimum points are located on the boundary points. As it is seen in Fig. 2, one of the links becomes useless for sending encoded observations to the decoder when the difference $p_2 - p_1$ slightly increases. To confirm these solutions, all roots of the gradient equation $\nabla F = [\frac{\partial F}{\partial d_1}, \frac{\partial F}{\partial d_2}] = [0, 0]$ are calculated. These roots give all possible optimum points except the boundary points, hence:

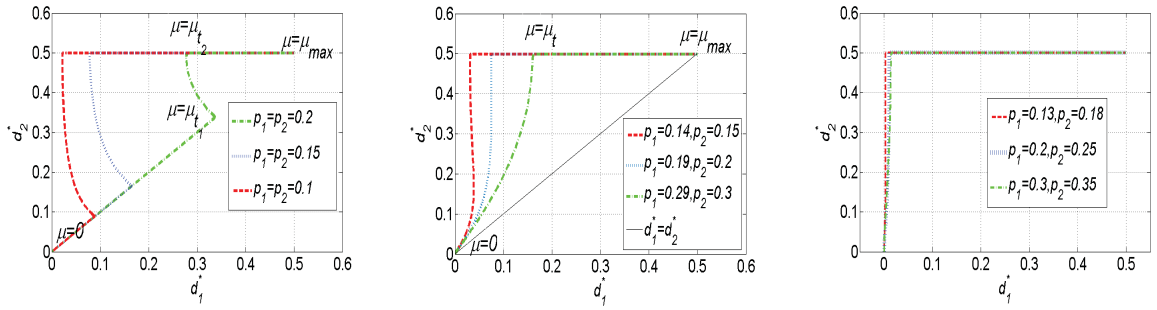
$$(1-2p_i)h_b'(d_i * p_i) - \mu h_b'(d_i) + (\mu - 1)(1-2q_i)h_b'(d * p) = 0, \quad (23)$$

for $i = 1, 2$. This non-linear system of equations does not have any closed-form solution and it is generally solved by using numerical methods such as Newton's method [27]. Next, the Hessian matrix $H_F = H_D + \mu H_R$ is calculated in these roots to check whether a possible point is exactly an optimum point or not.

$$H_F = \big[\frac{\partial^2 F}{\partial d_i \partial d_j}\big]; \quad i, j \in \{1, 2\}. \quad (24)$$

If H_F is a positive definite matrix in a root of $\nabla F = [0, 0]$, then it is a solution. Otherwise, since there is not such a point, the solution is only located on the boundary points, i.e., at least one of the d_i^*s equals 0 or 0.5. Due to the non-convexity of the optimization problem and the non-linearity of the rate and distortion expressions, we give an asymptotic analysis of the problem (16). In this regard, we consider a high-resolution regime.

Lemma 1: Assume $K = (1-2p)\log\big(\frac{1-p}{p}\big)$ and $K_i = (1-2p_i)\log\big(\frac{1-p_i}{p_i}\big)$ for $i = 1, 2$. For $x \to 0$, through applying



(a) For $p_2 = p_1$. (b) For $p_2 - p_1 = 0.01$. (c) For $p_2 - p_1 = 0.05$.

Fig. 2. Location of the optimum points (d_1^*, d_2^*).

the Taylor expansion we have

$$\begin{aligned} h_b(x) &= -x\log(x) + x + O(x^2), \\ h_b(x*p) - h_b(p) &= Kx + O(x^2). \end{aligned} \tag{25}$$

Theorem 2: Location of the solution points of (16), when $d_1 \to 0$ and $d_2 \to 0$, is as follows:

$$d_2 \approx 2^{\frac{K(K_2-K_1)}{(K_1-K)}} d_1^{\frac{K_2-K}{K_1-K}}. \tag{26}$$

Proof: Let $R_0 = 1 + h_b(p)$ and $D_0 = h_b(p_1) + h_b(p_2) - h_b(p)$ be sum-rate and distortion at $(d_1, d_2) = (0, 0)$. According to (25), if $(d_1, d_2) \to (0, 0)$, then we have:

$$\begin{aligned} R - R_0 &= h_b(p*d) - h_b(p) - h_b(d_1) - h_b(d_2), \\ &= Kd + d_1\log d_1 - d_1 + d_2\log d_2 - d_2 \\ &\quad + O(\max\{d^2, d_1^2, d_2^2\}) \\ &= (K-1)(d_1+d_2) + d_1\log d_1 + d_2\log d_2 \\ &\quad + O(\max\{d_1^2, d_2^2\}), \end{aligned} \tag{27}$$

$$\begin{aligned} D - D_0 &= h_b(p_1*d_1) - h_b(p_1) + h_b(p_2*d_2) \\ &\quad - h_b(p_2) - h_b(p*d) + h_b(p) \\ &= K_1d_1 + K_2d_2 - Kd + O(\max\{d^2, d_1^2, d_2^2\} \\ &= (K_1-K)d_1 + (K_2-K)d_2 + O(\max\{d_1^2, d_2^2\}). \end{aligned} \tag{28}$$

By ignoring the high-order terms, the following convex optimization problem is obtained:

$$\begin{aligned} \min_{0\le d_1,d_2\le 0.5} \ & (K-1)(d_1+d_2) + d_1\log d_1 + d_2\log d_2 \\ \text{s.t. } & (K_1-K)d_1 + (K_2-K)d_2 = D - D_0. \end{aligned} \tag{29}$$

The Lagrangian L of the above minimization problem for the Lagrangian multiplier λ is given by:

$$\begin{aligned} L = (K-1)(d_1+d_2) + d_1\log d_1 + d_2\log d_2 \\ + \lambda[(K_1-K)d_1 + (K_2-K)d_2 - D + D_0]. \end{aligned} \tag{30}$$

The gradient equation implies:

$$\frac{\partial L}{\partial d_i} = K + \log d_i + \lambda(K_i - K) = 0, \quad \text{for} \quad i = 1, 2. \tag{31}$$

Finally, by canceling λ in the above two equations, it is concluded that:

$$\begin{aligned} K(K_2-K_1) + (K_2-K)\log d_1 - (K_1-K)\log d_2 = 0 \\ \Rightarrow d_2 = 2^{\frac{K(K_2-K_1)}{(K_1-K)}} d_1^{\frac{K_2-K}{K_1-K}}. \end{aligned} \tag{32}$$

□

Corollary 1: In general, without assuming the test channels to be symmetric, one can still show that $R - R_0$ can be approximated by a convex function while $D - D_0$ can be approximated by a linear function, in the high-resolution regime. As a consequence, computation of the rate-distortion function can be approximately formulated as a convex optimization problem. A direct implication of this convex optimization formulation is that the binary symmetric test channel is asymptotically optimal in the high-resolution regime.

Corollary 2: The slope of the tangent lines to the curve of location of optimum points in $(d_1^*, d_2^*) = (0, 0)$ are, respectively, 1, ∞, or 0 when $p_1 = p_2$, $p_1 < p_2$, or $p_1 > p_2$.

In the following figure, location of the optimum points in a high-resolution regime[3] and the curve (26) for Fig. 2(b) are depicted. As it is obvious, these curves are approximately the same. By the following lemma, the asymptotic analysis of the problem is investigated around $(d_1^*, d_2^*) = (0.5, 0.5)$.

Lemma 2: The maximum value of the parameter μ occurs in $(R, D) = (0, 1)$ when $(d_1^*, d_2^*) = (0.5, 0.5)$ and it equals:

$$\mu_{\max} = \max\{(1-2p_1)^2, (1-2p_2)^2\}. \tag{33}$$

Proof: Consider the rate and distortion of (15) are denoted by $R_e = 0$ and $D_e = 1$ if $(d_1^*, d_2^*) = (0.5, 0.5)$. Calculation of the following fraction limit, which equals the slope of the tangent line to the sum-rate-distortion curve, is desired;

$$\begin{aligned} &\lim_{(d_1,d_2)\to(0.5,0.5)} \frac{D_e - D}{R - R_e} \\ &= \lim_{(d_1,d_2)\to(0.5,0.5)} \frac{1 + h_b(p*d) - h_b(p_1*d_1) - h_b(p_2*d_2)}{1 + h_b(p*d) - h_b(d_1) - h_b(d_2)} \\ &= \frac{0}{0}! \end{aligned} \tag{34}$$

By applying L'Hopital's rule and differentiation with respect to d_1, the limit (34) equals (35), as shown at the top of

[3] d_1 and d_2 are $O(10^{-4})$.

$$\lim_{(d_1,d_2)\to(0.5,0.5)} \frac{(1+\ell-2d_2-2\ell d_1)(1-2p)h_b'(p*d)-(1-2p_1)h_b'(p_1*d_1)-\ell(1-2p_2)h_b'(p_2*d_2)}{(1+\ell-2d_2-2\ell d_1)(1-2p)h_b'(p*d)-h_b'(d_1)-\ell h_b'(d_2)}, \tag{35}$$

the next page, where $\ell = \frac{\partial d_2}{\partial d_1}$. The above fraction is again ambiguous, therefore we need another differentiation with respect to d_1 from both the numerator and the denominator. After differentiation and letting $(d_1, d_2) \to (0.5, 0.5)$,

$$\begin{aligned}\lim_{(d_1,d_2)\to(0.5,0.5)} \frac{D_e - D}{R - R_e} &= \lim_{(d_1,d_2)\to(0.5,0.5)} \\ &\frac{0-(1-2p_1)^2 h_b''(p_1*d_1)-\ell^2(1-2p_2)^2 h_b''(p_2*d_2)}{0-h_b''(d_1)-\ell^2 h_b''(d_2)} \\ &= \frac{-(1-2p_1)^2 h_b''(0.5)-\ell^2(1-2p_2)^2 h_b''(0.5)}{-h_b''(0.5)-\ell^2 h_b''(0.5)} \\ &= \frac{(1-2p_1)^2+\ell^2(1-2p_2)^2}{1+\ell^2} \triangleq g(\ell). \end{aligned} \tag{36}$$

In the curve of the sum-rate-distortion bound, the value of (34) is maximized. Hence, the maximum of $g(\ell)$ is desirable.

$$\begin{aligned} g'(\ell) &= \frac{\partial g(\ell)}{\partial \ell} \\ &= \frac{\big(2\ell(1-2p_2)^2\big)(1+\ell^2)-2\ell\big((1-2p_1)^2+\ell^2(1-2p_2)^2\big)}{(1+\ell^2)^2} \\ &= 0 \Leftrightarrow \ell = 0 \quad \text{or } \ell = \infty. \end{aligned} \tag{37}$$

Therefore, the maximum value of the parameter μ is obtained from (37) as in (33). □

Corollary 3: According to (37), the slope of the tangent line to the curve of location of the optimum points is 0 if $p_1 < p_2$, and it is ∞ if $p_1 > p_2$. For the case $p_1 = p_2$, both 0 and ∞ are acceptable as the slope of the tangent line to the curve of location of the optimum points due to the continuity of the rate and the distortion functions.

In practical applications, parameters of the observation noises are small values. Hence, we investigate our problem with more details when at least one of the noise parameters is a very small value. Thus, we may assume without loss of generality that $p_1 \to 0$, then from continuity:

$$\begin{aligned} R &\approx 1 + h_b(p_2 * d) - h_b(d_1) - h_b(d_2), \\ D &\approx h_b(d_1) + h_b(p_2 * d_2) - h_b(p_2 * d) \\ &\approx 1 - R + h_b(p_2 * d_2) - h_b(d_2). \end{aligned} \tag{38}$$

The behavior of low noise case is expressed by the following theorem.

Theorem 3: If $p_1 \to 0$, then only two cases can occur: either (i) $d_2^* = 0.5$, or (ii) $d_1^* \to 0$. *Proof:* [Proof] It is sufficient to prove that if $0 \le d_2^* < 0.5$, then $d_1^* \to 0$. First, we shall declare that $h_b(q * x) - h_b(x)$ is a positive and a decreasing function in terms of x, where $0 \le x \le 0.5$ and $0 < q \le 0.5$. Therefore, it takes its maximum value in $x = 0$ for any q. Now compute the objective function in the case $p_1 \to 0$. Due to (38), we have:

$$F(\mu) = D + \mu R \approx 1 + (\mu - 1)R + h_b(p_2 * d_2) - h_b(d_2). \tag{39}$$

Assume that $0 \le d_2^* = c < 0.5$, where c is a constant value. The latter optimization problem becomes as follows:

$$\begin{aligned} \min_{0\le d_1\le 0.5,\ d_2=c} &\Big[(\mu-1)\big(1+h_b(p_2*d)-h_b(d_1)-h_b(d_2)\big) \\ &+ h_b(p_2*d_2)-h_b(d_2)\Big] \equiv \min_{0\le d_1\le 0.5}(\mu-1)\big(h_b(q*d_1) \\ &-h_b(d_1)\big) \equiv \max_{0\le d_1\le 0.5}\big(h_b(q*d_1)-h_b(d_1)\big), \end{aligned} \tag{40}$$

where $0 < q = p_2 * c \le 0.5$. Obviously, solution of the above problem is $d_1^* = 0$ and due to the approximations in our calculations, we shall have $d_1^* \to 0$. □

Corollary 4: If p_2 is sufficiently larger than p_1, then $p = p_1 * p_2 \approx p_2$ and the same situation of Theorem 3 occurs. Similarly, if $\frac{p_2-p_1}{p_2} = \alpha \to 1$, then $p = p_1 * p_2 = (1-\alpha)p_2 + p_2 - 2(1-\alpha)p_2^2 \approx p_2$, and hence Theorem 3 is used again, as depicted in the case of Fig. 2(c). In order to have more intuition to the result of Theorem 3, some other cases are provided in the Fig. 4.

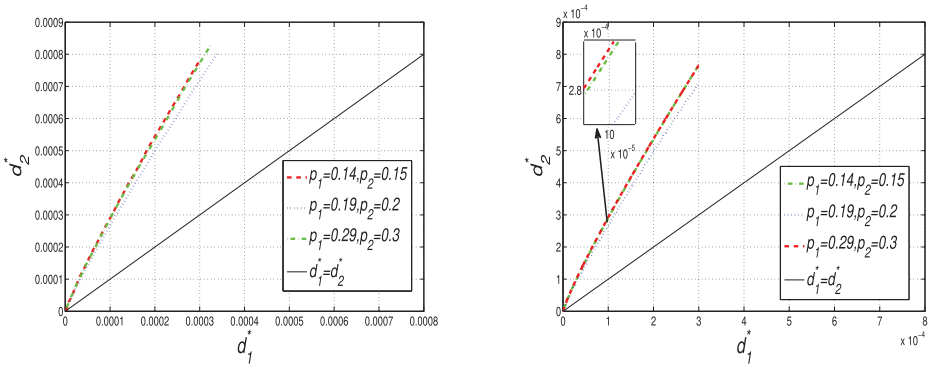
IV. The Proposed Coding Scheme

In this section, a practical coding scheme is introduced to achieve the calculated rate-distortion bounds. In Fig. 5, values of the theoretical bound of sum-rate versus distortion are displayed for several noise parameters to evaluate performance of the designed codes. The value of gap between the achieved point and the theoretical bound is employed as a performance criterion. The structure of the designed code significantly differs whether only one of the links or both of them are engaged in information sending. Obviously, when only one link sends information, our problem reduces to a point-to-point lossy source coding problem. Furthermore, for any (d_1, d_2), there exists a particular achievable rate region which is characterized by the corner and intermediate points in its boundary. In the following, the proposed encoding and decoding schemes are separately illustrated for achieving the corner points and the intermediate points of the bound in the achievable rate region.

A. Coding Scheme for the Corner Points

According to the exact rate-distortion bound (7), we have the following bound for the rate of i-th link and the sum-rate:

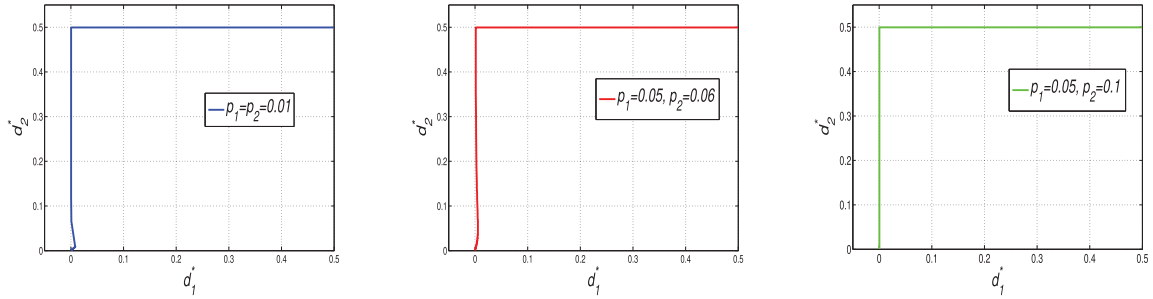
$$\begin{aligned} R_i &\ge I(Y_i; U_i|U_{3-i}) = H(U_i|U_{3-i}) - H(U_i|Y_i, U_{3-i}) \\ &\overset{(a)}{=} H(U_i|U_{3-i}) - H(U_i|Y_i) \\ &= h_b(d*p) - h_b(d_i), \quad \text{for } i = 1, 2, \\ R_1 + R_2 &\ge I(Y_1, Y_2; U_1, U_2) = H(U_1, U_2) \\ &\quad - H(U_1, U_2|Y_1, Y_2) \\ &\overset{(a)}{=} 1 + H(U_1|U_2) - H(U_1|Y_1) - H(U_2|U_2) \\ &= 1 + h_b(p*d) - h_b(d_1) - h_b(d_2), \end{aligned} \tag{41}$$



(a) Depiction of Fig. 2(b) in a high-resolution regime.

(b) Direct plotting of (26).

Fig. 3. Comparison of the location of optimum points and curves of (26).



(a) If $p_2 = p_1$.

(b) If $p_2 - p_1 = 0.01$.

(c) If $p_2 - p_1 = 0.05$.

Fig. 4. Location of the optimum points (d_1^*, d_2^*).

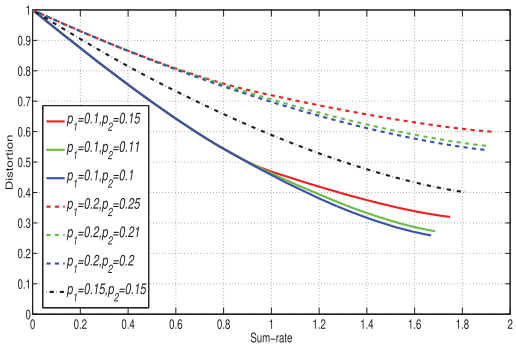


Fig. 5. The sum-rate-distortion function of the binary CEO problem for some noise parameters.

where (a) follows since $U_1 \overset{d_1}{\leftrightarrow} Y_1 \overset{p_1}{\leftrightarrow} X \overset{p_2}{\leftrightarrow} Y_2 \overset{d_2}{\leftrightarrow} U_2$ from a Markov chain. Since a conventional rate-distortion quantizer can asymptotically achieve the compression rate $1 - h_b(d_i)$ for distortion being assumed d_i, it is impossible to get close to (41) by only using the rate-distortion quantizer. Hence, another lossless source encoder should be utilized after the conventional rate-distortion quantizer for achieving the rate (41) in the i-th link. We use an LDGM quantizer concatenated with a Syndrome-Generator (SG), inspired by the "quantize-and-bin" idea in the context of information theory. The dominant face of the rate region is a line segment connecting two end points (R_1', R_2') and (R_1'', R_2''), where

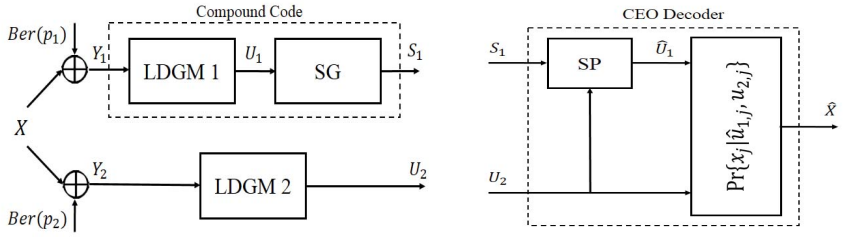
$$(R_1', R_2') = \big(h_b(p * d) - h_b(d_1), 1 - h_b(d_2)\big), \qquad (42)$$

and

$$(R_1'', R_2'') = \big(1 - h_b(d_1), h_b(p * d) - h_b(d_2)\big). \qquad (43)$$

We consider a coding scheme for achieving (R_1', R_2'). A similar method can be applied for achieving (R_1'', R_2''). Encoder 1 quantizes y_1^n to u_1^n using an LDGM code of rate $R_{1,1} = \frac{m_1}{n}$, then it computes the syndrome $s_1^{k_1} = u_1^n H_1^T$, where H_1 is the parity-check matrix of an LDPC code of rate $R_{1,2} = \frac{m_1 - k_1}{n}$. We do this process of quantize and bin by employing a compound LDGM-LDPC code. It is notable that the total length of the obtained syndrome equals $n - m_1 + k_1$, where its first $n - m_1$ bits are zero because the LDPC code is nested with the LDGM code [22]. Hence, only k_1 non-zero bits are sent to the decoder and the total rate is $R_1 = R_{1,1} - R_{1,2} = \frac{k_1}{n}$. Encoder 2 quantizes y_2^n to u_2^n using an LDGM code of rate $R_{2,1} = \frac{m_2}{n}$, then u_2^n is sent to the decoder. In the second link, the total rate is $R_2 = \frac{m_2}{n}$. The block diagram of the proposed scheme is shown in Fig. 6.

At the decoder side, the syndrome $s_1^{k_1}$ with u_2^n as a side information are used to decode u_1^n, denoted by $\hat{u}_1^n$, by applying a SP algorithm. Finally, at the decoder, calculation of the soft estimation $\hat{x}_j = \Pr\{x_j | \hat{u}_{1,j}, u_{2,j}\}$ completes the decoding process. Here $\Pr\{x_j | \hat{u}_{1,j}, u_{2,j}\} \triangleq p_{X|U_1,U_2}(x_j | \hat{u}_{1,j}, u_{2,j})$,



(a) The proposed encoding scheme based on compound structure.

(b) Two-link joint decoding structure.

Fig. 6. The proposed coding scheme for achieving a corner point.

and the conditional distribution $p_{X|U_1,U_2}$ can be deduced from the joint distribution in (6) once d_1 and d_2 are given (assuming Q is a constant).

A compound LDGM-LDPC code includes nested LDGM and LDPC codes with the following parity-check matrices:

$$H_{\text{LDPC}} = \begin{bmatrix} H_{\text{LDGM}} \\ \Delta H \end{bmatrix}, \tag{44}$$

where H_{LDPC} and H_{LDGM} are, respectively, parity-check matrices of the LDPC and LDGM codes. Let assume their sizes are $(n-m+k)\times n$ and $(n-m)\times n$, respectively. We have used the compound LDGM-LDPC structure to achieve theoretical bound of the Wyner-Ziv problem in [22]. We denote the mentioned compound code by $\mathcal{C}_{H_{\text{LDPC}}}(n,m,k)$.

For achieving a corner point, we employ a compound code $\mathcal{C}_{H^{(1)}_{\text{LDPC}}}(n,m_1,k_1)$ in the first link, and a single LDGM code with the generator matrix $G^{(2)}$ of size $m_2\times n$ in the second link. The observation y_1^n is quantized to an LDGM codeword u_1^n by applying the BiP algorithm with the generator matrix $G^{(1)}$. Hence, $u_1^n H^{(1)}_{\text{LDGM}} = \underbrace{[0\ \cdots\ 0]}_{n-m_1}$. Next, the syndrome $u_1^n H^{(1)}_{\text{LDPC}} = [\underbrace{0\ \cdots\ 0}_{n-m_1},\ \underbrace{u_1^n \Delta H^T}_{s_1^{k_1}}]$ is calculated and only $s_1^{k_1}$ is sent to the CEO decoder.

B. Coding Scheme for the Intermediate Points

Consider the following intermediate point located in the dominant face of the achievable rate region,

$$(R_1^*, R_2^*) = \big(h_b(p*d) - h_b(d_1) + \delta, 1 - h_b(d_2) - \delta\big), \tag{45}$$

where $0<\delta<1-h_b(p*d)$. Obviously, $R_1^*\le 1-h_b(d_1)$ and $R_2^*\le 1-h_b(d_2)$. Therefore, a lossless compression is needed in each link. In the i-th link, we use a compound code $\mathcal{C}_{H^{(i)}_{\text{LDPC}}}(n,m_i,k_i)$, for $i=1,2$. First step of encoding includes quantizing the observations y_i^n to u_i^n by using the BiP algorithm on the LDGM codes associated with the parity-check matrices $H^{(i)}_{\text{LDGM}}$. In the second step, the syndromes $u_i^n H^{(i)}_{\text{LDPC}} = [\underbrace{0\ \cdots\ 0}_{n-m_i},\ \underbrace{u_i^n \Delta {H^{(i)}}^T}_{s_i^{k_i}}]$ are calculated, then only $s_1^{k_1}$ and $s_2^{k_2}$ are sent to the CEO decoder.

In the decoder, we propose a Joint Sum-Product (JSP) algorithm which is a modified version of the SP algorithm. In this algorithm, the received syndromes $s_1^{k_1}$ and $s_2^{k_2}$ are respectively located in the check nodes of the LDPC codes with parity-check matrices $H^{(1)}_{\text{LDPC}}$ and $H^{(2)}_{\text{LDPC}}$. The JSP includes r rounds and each round includes l iterations. At the starting point of the JSP, initial LLRs in the variable nodes are set based on a random side information in each SP. At the end of each round, which includes update equations in the check and variable nodes, the bit values of the variable nodes are calculated according to the decision rule of the SP algorithm, where it maps the non-negative LLRs to bit 0 and the negative LLRs to bit 1. In the next round, these updated bit values in the variable nodes are used as a new side information for calculating new initial LLR values. Finally, after r rounds, $\hat{u}_1^n$ and $\hat{u}_2^n$ are decoded based on the decision rule of the SP algorithm in the variable nodes of the LDPC codes. An EXIT chart analysis is presented in [28] for a similar JSP decoder which shows the capacity approaching property with two parallel and collaborative SP decoders. Similar to the decoding scheme of the corner points, the soft estimation $\hat{x}_j = \Pr\{x_j|\hat{u}_{1,j},\hat{u}_{2,j}\}$ accomplishes the decoding process.

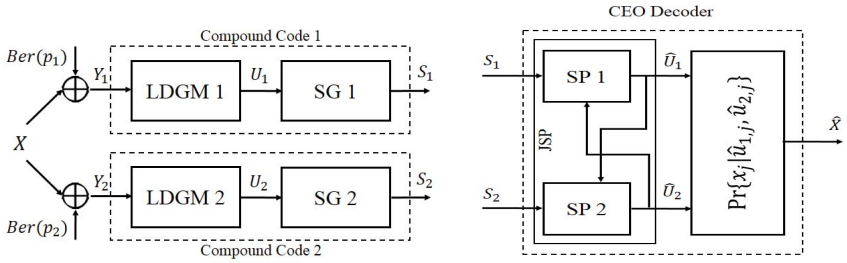
In the joint decoding scheme, the received sequences s_1^n and s_2^n are simultaneously decoded. If we look at this situation like two point-to-point lossy source coding problems, then we have to recover the noisy observations y_1^n and y_2^n, each of which is received with compression rates R_1 and R_2 and acceptable distortions, respectively. As a case, let these distortions be d_1 and d_2, respectively. A major part of distortions d_1 and d_2 arises from the LDGM quantization and a negligible part is from the syndrome-decoding. Let assume the LDPC code rate in the i-th link is denoted by $R_{i,2}$, for $i=1,2$. Furthermore, consider the associated distortion of each link, i.e., BER of the syndrome-decoding part, is denoted by $d_{i,2}$, for $i=1,2$. Using the compound LDGM-LDPC structure, the total rate and the distortion in each link are as follows:

$$d_i = d_{i,1} * d_{i,2} \approx d_{i,1},\ R_i = R_{i,1} - R_{i,2},\quad \text{for } i=1,2. \tag{46}$$

After reconstruction of the observations with distortions d_1 and d_2, denoted by $\hat{u}_1^n$ and $\hat{u}_2^n$, the soft reconstruction of the original binary source $\hat{x}^n$ is estimated.

C. A Practical Analysis for the Proposed Coding Scheme

Some coding parameters are affected by the information theoretical limits, that should be considered in the code design procedure. In the following notations, any ϵ denotes



(a) The proposed encoding scheme based on compound structures.

(b) Two-link joint decoding structure.

Fig. 7. The proposed coding scheme for achieving intermediate points.

a sufficiently small positive value. In the coding scheme for a corner point, the relation between the rate-distortion and the block lengths of employed LDGM and LDPC codes are as follows:

$$\begin{aligned} R_{1,1} &= \frac{m_1}{n} = 1 - h_b(d_{1,1}) + \epsilon_{1,1}, \\ R_{1,2} &= \frac{m_1 - k_1}{n} = 1 - h_b(d_{1,1} * d_{2,1} * p_1 * p_2) - \epsilon_{1,2}, \\ R_{2,1} &= \frac{m_2}{n} = 1 - h_b(d_{2,1}) + \epsilon_{2,1}, \quad R_{2,2} = 0. \end{aligned} \tag{47}$$

From (46) and (47), it is simply concluded that:

$$\begin{aligned} R_1 &= R_{1,1} - R_{1,2} = \frac{k_1}{n} \\ &= h_b(d_{1,1} * d_{2,1} * p_1 * p_2) - h_b(d_{1,1}) + \underbrace{\epsilon_{1,1} + \epsilon_{1,2}}_{\epsilon_1} \\ &\overset{\text{(b)}}{\approx} h_b(d * p) - h_b(d_1) + \epsilon_1, \\ R_2 &= R_{2,1} - R_{2,2} = \frac{k_2}{n} = 1 - h_b(d_{2,1}) + \epsilon_{2,1} \\ &\overset{\text{(b)}}{\approx} 1 - h_b(d_2) + \epsilon_{2,1}, \end{aligned} \tag{48}$$

where (b) follows from the continuity of the function $h_b(x)$. The above approximation expresses that achieving the rate bound (41) is possible by using our proposed method in each link. In the decoding side, the SP algorithm of Fig. 6(b) uses the LDPC code of rate $R_{1,2}$, that is smaller than the capacity of the virtual channel between the side information U_2 and the target sequence U_1.[4] Therefore, U_1 is decoded with a low BER, i.e., $d_{1,2} \approx 0$, by using a good channel decoder. Clearly, in this case $d_{2,2} = 0$ and $d_2 = d_{2,1}$.

In the coding scheme for an intermediate point (45), the relation between the rate-distortion and the block lengths of each employed LDGM and LDPC codes are as follows:

$$\begin{aligned} R_{1,1} &= \frac{m_1}{n} = 1 - h_b(d_{1,1}) + \epsilon_{1,1}, \\ R_{1,2} &= \frac{m_1 - k_1}{n} = 1 - h_b(d_{1,1} * d_{2,1} * p_1 * p_2) - \delta - \epsilon_{1,2}, \\ R_{2,1} &= \frac{m_2}{n} = 1 - h_b(d_{2,1}) + \epsilon_{2,1}, \\ R_{2,2} &= \frac{m_2 - k_2}{n} = \delta - \epsilon_{2,2}. \end{aligned} \tag{49}$$

[4]This capacity equals $1 - h_b(p * d)$.

From (46) and (49), it is simply concluded that:

$$\begin{aligned} R_1 &= R_{1,1} - R_{1,2} = \frac{k_1}{n} \\ &= h_b(d_{1,1} * d_{2,1} * p_1 * p_2) - h_b(d_{1,1}) + \delta + \underbrace{\epsilon_{1,1} + \epsilon_{1,2}}_{\epsilon_1} \\ &\overset{\text{(b)}}{\approx} h_b(d * p) - h_b(d_1) + \delta + \epsilon_1, \\ R_2 &= R_{2,1} - R_{2,2} = \frac{k_2}{n} = 1 - h_b(d_{2,1}) - \delta + \underbrace{\epsilon_{2,1} + \epsilon_{2,2}}_{\epsilon_2} \\ &\overset{\text{(b)}}{\approx} 1 - h_b(d_2) - \delta + \epsilon_2, \end{aligned} \tag{50}$$

where (b) follows from the continuity of the function $h_b(x)$. The above approximation expresses that achieving the rate bound (41), for an intermediate point (45), is possible by utilizing the proposed method. In the decoding side, the JSP algorithm of Fig. 7(b) uses the LDPC codes of rates $R_{1,2}$ and $R_{2,2}$. From (49) and $0 < \delta < 1 - h_b(p * d)$,

$$\begin{aligned} R_{1,2} &\approx 1 - h_b(d * p) - \delta - \epsilon_{1,2} < 1 - h_b(d * p), \\ R_{2,2} &= \delta - \epsilon_{2,2} < 1 - h_b(d * p). \end{aligned} \tag{51}$$

Therefore, the rates of LDPC codes in the SP1 and the SP2 algorithms are smaller than the capacity of the virtual channel between the side information U_1 and U_2. This implies that the SP algorithms can decode U_1 and U_2 with low BERs, i.e., $d_{i,2} \approx 0$ for $i = 1, 2$, for sufficiently large n, r, and l.

For the empirical distortion D_{em} in (4) with $\hat{x}_j(x_j) = \Pr\{x_j | u_{1,j}, u_{2,j}\}$, we have

$$\begin{aligned} D_{\text{em}} &= \frac{1}{n} \sum_{j=1}^{n} \log[\frac{1}{\Pr\{x_j | u_{1,j}, u_{2,j}\}}] \\ &= \sum_{x, u_1, u_2} \Pr_{\text{em}}\{x, u_1, u_2\} \log[\frac{1}{\Pr\{x | u_1, u_2\}}], \end{aligned} \tag{52}$$

where $\Pr_{\text{em}}\{x, u_1, u_2\}$ is the empirical distribution induced by (x^n, u_1^n, u_2^n). The theoretical distortion bound is given by $D_{\text{th}} = H(X|U_1, U_2)$. Clearly, we have $D_{\text{em}} \approx D_{\text{th}}$ if $\Pr_{\text{em}}\{x, u_1, u_2\}$ is close to $\Pr\{x, u_1, u_2\}$.

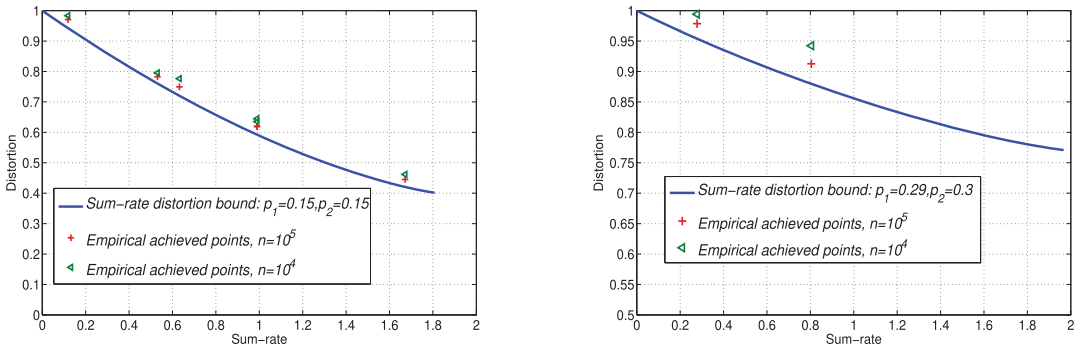
V. Results and Discussions

In this section, some numerical results are given for indicating the rate-distortion performance of the proposed coding scheme at different regions. For all of the LDPC codes,

TABLE I

NUMERICAL RESULTS OF THE PROPOSED ENCODING AND DECODING METHODS

(p_1, p_2)	Region-C/I	n	m_1, m_2	k_1, k_2	$d_{1,1}, d_1$	$d_{2,1}, d_2$	μ	R_{th}	D_{th}	D_{em}	Gap
(0.15, 0.15)	1-I	10^4	9200, 9200	8500, 8500	0.0144, 0.0175	0.0144, 0.0175	0.168	1.6722	0.4204	0.4617	0.0413
(0.15, 0.15)	1-I	10^4	5400, 5400	5100, 5100	0.1028, 0.1071	0.1028, 0.1071	0.326	0.9898	0.5925	0.645	0.0525
(0.15, 0.15)	1-C	10^4	5400, 5400	4700, 5400	0.1028, 0.1076	0.1028, 0.1028	0.326	0.9898	0.5925	0.6355	0.043
(0.15, 0.15)	2-I	10^4	5400, 1300	5300, 1200	0.1028, 0.1066	0.3055, 0.3078	0.3854	0.6319	0.7206	0.7766	0.056
(0.15, 0.15)	3-C	10^4	5400, –	5400, –	0.1028, 0.1028	0.5, 0.5	0.4043	0.531	0.7601	0.7955	0.0354
(0.15, 0.15)	3-C	10^4	1300, –	1300, –	0.3055, 0.3055	0.5, 0.5	0.4532	0.1187	0.9427	0.9835	0.0408
(0.15, 0.15)	1-I	10^5	92000, 92000	85000, 85000	0.012, 0.015	0.012, 0.015	0.168	1.6722	0.4204	0.4451	0.0247
(0.15, 0.15)	1-I	10^5	54000, 54000	51000, 51000	0.1003, 0.1043	0.1003, 0.1043	0.326	0.9898	0.5925	0.6203	0.0278
(0.15, 0.15)	1-C	10^5	54000, 54000	47000, 54000	0.1003, 0.105	0.1003, 0.1003	0.326	0.9898	0.5925	0.6184	0.0259
(0.15, 0.15)	2-I	10^5	54000, 13000	53000, 12000	0.1003, 0.1037	0.3018, 0.304	0.3854	0.6319	0.7206	0.7494	0.0288
(0.15, 0.15)	3-C	10^5	54000, –	54000, –	0.1003, 0.1003	0.5, 0.5	0.4043	0.531	0.7601	0.7826	0.0225
(0.15, 0.15)	3-C	10^5	13000, –	13000, –	0.3018, 0.3018	0.5, 0.5	0.4532	0.1187	0.9427	0.9707	0.028
(0.29, 0.3)	1-I	10^4	5420, 2920	5400, 2900	0.1025, 0.1066	0.2044, 0.208	0.1283	0.8044	0.8797	0.9423	0.0626
(0.29, 0.3)	2-C	10^4	2900, –	2900, –	0.2046, 0.2046	0.5, 0.5	0.157	0.278	0.9537	0.9942	0.0405
(0.29, 0.3)	1-I	10^5	54200, 29200	54000, 29000	0.1, 0.1038	0.202, 0.2052	0.1283	0.8044	0.8797	0.9127	0.033
(0.29, 0.3)	2-C	10^5	29000, –	29000, –	0.2025, 0.2025	0.5, 0.5	0.157	0.278	0.9537	0.9786	0.0249
(0.01, 0.01)	1-I	10^4	9900, 9900	6000, 6000	0.0032, 0.0053	0.0032, 0.0053	0.4	1.1161	0.0268	0.0372	0.0104
(0.01, 0.01)	1-I	10^5	99000, 99000	60000, 60000	0.0028, 0.0047	0.0028, 0.0047	0.4	1.1161	0.0268	0.0329	0.0061
(0.05, 0.1)	1-C	10^4	10000, 400	10000, 300	0, 0	0.405, 0.4075	0.253	1.014	0.2829	0.31	0.0271
(0.05, 0.1)	1-C	10^5	100000, 4000	100000, 3000	0, 0	0.4026, 0.4043	0.253	1.014	0.2829	0.3011	0.0182



(a) $p_1 = p_2 = 0.15$. (b) $(p_1, p_2) = (0.29, 0.3)$.

Fig. 8. The sum-rate-distortion performance of the proposed coding scheme.

the optimized degree distributions over the binary symmetric channel are employed.[5] However, the check-regular and variable-Poisson LDGM codes nested with the LDPC codes are designed similar to the code design method in [22]. In order to achieve some target optimum crossover probability pairs (d_1^*, d_2^*) by practical coding methods, we have applied our proposed coding scheme with the lengths of $n = 10^4,\ 10^5$ for various cases of the noise parameters including $(p_1, p_2) = (0.15, 0.15)$ and $(0.29, 0.3)$. We have also implemented our proposed coding scheme for two low-noise cases $(p_1, p_2) = (0.01, 0.01)$ and $(0.05, 0.1)$.

In the BiP algorithm, the parameters $t = 0.8$, $\gamma_i \approx 2\ R_{i,1} = 2\frac{m_i}{n}$ are selected for $i = 1, 2$. Maximum number of iterations in each round of this algorithm is set to be 25. In the SP algorithm, maximum number of iterations is set to be 100. Also, in the JSP algorithm, $r = 15$ and $l = 40$. All of the reported values for the empirical distortions are averaged over 50 runs. Parameters of the employed codes and their results are presented in Table I. The corner and the intermediate points are indicated by symbols C and I in the second column of the table, respectively. The gap value is equal to difference between the empirical distortion D_{em} and the theoretical distortion D_{th} and it evaluates performance of the designed codes.

[5]These degree distributions are available in [29] for some rates.

For the case of equal noises $(p_1, p_2) = (0.15, 0.15)$, the gap values is about 0.03 to 0.06 for the block length 10^4, as indicated in Table. I. It is obvious that by increasing the target distortions, the gap values increase. As the block length is set to 10^5, the gap value decreases in the range between 0.02 to 0.03. Performance of the sum-rate versus distortion is depicted in Fig. 8(a) for the proposed coding scheme. Simulation results confirm that performance of the sum-rate in terms of distortion is very close to the theoretical bounds for the empirically achieved points. The theoretical bounds are asymptotically achievable by employing the proposed coding scheme as well. For the case of unequal noise parameters $(p_1, p_2) = (0.29, 0.3)$, the achieved gaps are slightly more than that one for the case of equal noise parameters with approximately the same target distortions and block length. This observation expresses that increasing noise parameters p_1 and p_2 causes a higher gap values in distortion. Similarly, by increasing the block length to 10^5, the gap value is decreased by about less than half of the result for block length 10^4, as mentioned in Table. I. Performance of the sum-rate in terms of distortion for the proposed coding scheme is depicted in Fig. 8(b).

In the low-noise case, as it is seen in the Table. I, the achieved gap depends on the value of the target distortions, similar to the prior cases. The achieved gap with our coding scheme is about 0.006 to 0.011 for the

case $(p_1, p_2) = (0.01, 0.01)$, and it is about 0.018 to 0.028 for the case $(p_1, p_2) = (0.05, 0.1)$. Generally, the gap value for the corner points is smaller than that of the intermediate points. This gap can be reduced by increasing r and l.

VI. Conclusion

We have investigated the two-link binary CEO problem under the log-loss from both information-theoretic and coding-theoretic perspectives. A binary symmetric test channel model is adopted for each encoder, and its optimal parameters are computed numerically and analyzed in the asymptotic high-resolution regime. We have also proposed a practical coding scheme based on low-density graph codes and message-passing algorithms. Specifically, the compound LDGM-LDPC codes are utilized at each encoder for performing binary quantization and syndrome-generation. The decoder employs the SP algorithms based on the optimized LDPC codes for the binary symmetric channel and outputs the final reconstruction value in the form of a probability distribution. The effectiveness of the proposed scheme is confirmed by the simulation results.

References

[1] M. Nangir, R. Asvadi, M. Ahmadian-Attari, and J. Chen. (2018). "Binary CEO problem under log-loss with BSC test-channel model." [Online]. Available: https://arxiv.org/abs/1801.02976

[2] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[3] 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.

[4] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1549–1559, Sep. 1997.

[5] V. Prabhakaran, D. Tse, and K. Ramachandran, "Rate region of the quadratic Gaussian CEO problem," in *Proc. IEEE ISIT*, Chicago, IL, USA, Jun./Jul. 2004, p. 119.

[6] Y. Oohama, "Distributed source coding of correlated Gaussian remote sources," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5059–5085, Aug. 2012.

[7] 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.

[8] J. Chen and T. Berger, "Successive Wyner–Ziv coding scheme and its application to the quadratic Gaussian CEO Problem," *IEEE Trans. Inf. Theory.*, vol. 54, no. 4, pp. 1586–1603, Apr. 2008.

[9] H. Behroozi and M. R. Soleymani, "Optimal rate allocation in successively structured Gaussian CEO problem," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 627–632, Feb. 2009.

[10] H. Behroozi and M. R. Soleymani, "Performance of the successive coding strategy in the CEO problem," in *Proc. IEEE GLOBECOM*, St. Louis, MO, USA, Nov./Dec. 2005, p. 6.

[11] X. He, X. Zhou, P. Komulainen, M. Juntti, and T. Matsumoto, "A lower bound analysis of Hamming distortion for a binary ceo problem with joint source-channel coding," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 343–353, Jan. 2016.

[12] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[13] X. He, X. Zhou, M. Juntti, and T. Matsumoto, "A rate-distortion region analysis for a binary CEO problem," in *Proc. IEEE VTC*, Nanjing, China, May 2016, pp. 1–5.

[14] X. He, X. Zhou, M. Juntti, and T. Matsumoto, "Data and error rate bounds for binary data gathering wireless sensor networks," in *Proc. IEEE SPAWC*, Stockholm, Sweden, Jun./Jul. 2015, pp. 505–509.

[15] A. Razi and A. Abedi, "Convergence analysis of iterative decoding for binary CEO problem," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2944–2954, May 2014.

[16] J. Haghighat, H. Behroozi, and D. V. Plant, "Joint decoding and data fusion in wireless sensor networks using turbo codes," in *Proc. IEEE PIMRC*, Cannes, France, Sep. 2008, pp. 1–5.

[17] A. No and T. Weissman, "Universality of logarithmic loss in lossy compression," in *Proc. IEEE ISIT*, Hong Kong, Jun. 2015, pp. 2166–2170.

[18] Y. Shkel, M. Raginsky, and S. Verdú, "Universal lossy compression under logarithmic loss," in *Proc. IEEE ISIT*, Aachen, Germany, Jun. 2017, pp. 1157–1161.

[19] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 740–761, Jan. 2014.

[20] Y. Zhou, Y. Xu, W. Yu, and J. Chen, "On the optimal fronthaul compression and decoding strategies for uplink cloud radio access networks," *IEEE Trans. Inf. Theory.*, vol. 62, no. 12, pp. 7402–7418, Dec. 2016.

[21] T. Filler, "Minimizing embedding impact in steganography using low density codes," M.S. thesis, Dept. Elect. Comput. Eng., Binghamton Univ., Binghamton, NY, USA, 2007.

[22] M. Nangir, M. Ahmadian-Attari, and R. Asvadi, "Binary Wyner–Ziv code design based on compound LDGM-LDPC structures," *IET Commun.*, vol. 12, no. 4, pp. 375–383, 2018.

[23] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[24] G. Caire, S. Shamai (Shitz), and S. Verdú, "Lossless data compression with error correcting codes," in *Proc. IEEE ISIT*, Yokohama, Japan, Jun./Jul. 2003, p. 22.

[25] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 626–643, Mar. 2003.

[26] A. D. Liveris, Z. Xiong, and C. N. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Commun. Lett.*, vol. 6, no. 10, pp. 440–442, Oct. 2002.

[27] C. Remani, "Numerical methods for solving systems of nonlinear equations," Dept. Math. Sci., Lakehead Univ., Thunder Bay, ON, Canada, 2013. [Online]. Available: https://www.lakeheadu.ca/sites/default/files/uploads/77/docs/RemaniFinal.pdf

[28] M. Khas, H. Saeedi, and R. Asvadi, "LDPC code design for correlated sources using EXIT charts," in *Proc. IEEE ISIT*, Aachen, Germany, Jun. 2017, pp. 2945–2949.

[29] D. H. Schonberg, "Practical distributed source coding and its application to the compression of encrypted data," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Univ. California, Berkeley, CA, USA, 2007.