Title
A Lower Bound Analysis of Hamming Distortion for a Binary CEO Problem with Joint Source-Channel Coding

Author(s)
He, Xin; Zhou, Xiaobo; Komulainen, Petri; Juntti, Markku; Matsumoto, Tad

Citation
IEEE Transactions on Communications, 64(1): 343-353

Issue Date
2015-11-09

Type
Journal Article

Text version
author

URL
http://hdl.handle.net/10119/13000

Rights
This is the author's version of the work. Copyright (C) 2015 IEEE. IEEE Transactions on Communications, 64(1), 2015, 343-353. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Description
A Lower Bound Analysis of Hamming Distortion for a Binary CEO Problem with Joint Source-Channel Coding

Xin He, Xiaobo Zhou, Member, IEEE, Petri Komulainen, Member, IEEE, Markku Juntti, Senior Member, IEEE and Tad Matsumoto Fellow, IEEE

Abstract—A two-node binary chief executive officer (CEO) problem is investigated. Noise-corrupted versions of a binary sequence are forwarded by two nodes to a single destination node over orthogonal additive white Gaussian noise (AWGN) channels. We first reduce the binary CEO problem to a binary multiterminal source coding problem, of which an outer bound for the rate-distortion region is derived. The distortion function is then established by evaluating the relationship between the binary CEO and multiterminal source coding problems. A lower bound approximation on the Hamming distortion (HD) is obtained by minimizing a distortion function subject to constraints obtained based on the source-channel separation theorem. Encoding/decoding algorithms using concatenated convolutional codes and a joint decoding scheme are used to verify the lower bound on the HD. It is found that the theoretical lower bounds on the HD and the computer simulation based bit error rate performance curves have the same tendencies. The differences in the threshold signal-to-noise ratio between the theoretical lower bounds and those obtained by simulations are around 1.5 dB in AWGN channel. The theoretical lower bound on the HD in block Rayleigh fading channel is also evaluated by performing Monte Carlo simulation.

Index Terms—Hamming distortion lower bound, binary CEO problem, binary multiterminal source coding, rate-distortion outer bound.

I. INTRODUCTION

The chief executive officer (CEO) problem where a CEO aims at reproducing a common source which cannot be directly observed was first introduced by Berger et al. in [1]. The CEO problem has attracted a lot of attention, not only from pure information theoretic interest but also from practical application viewpoint such as wireless sensor networks (WSNs).

The quadratic Gaussian CEO problem, a particular case of the CEO problem, where the source and the multiple observations are assumed to be jointly Gaussian distributed, was studied in [2], [3], where an explicit form of the rate-distortion function was derived. Chen et al. derived an upper bound on the sum-rate for the CEO problem and proposed rate allocation schemes by exploiting the contra-polymatroid structure of the achievable rate region in [4]. Besides the theoretical work, the performance on minimum achievable distortion of a successive coding strategy [5] based on a generalization of Wyner-Ziv source coding was evaluated in [6], and the optimal rate allocation scheme to achieve the minimum distortion under a sum-rate constraint was further proposed in [7] for the successively structured Gaussian CEO problem.

In this work, we focus on a binary CEO problem, in which a binary source is estimated by the CEO through multiple deployed nodes. There are many practical applications of the binary CEO problem, for example, a binary data gathering sensor network, distributed detection using multiple sensors [8], power-distortion tradeoff in sensor networks [9], and wireless mesh network (WMN) with lossy forwarding (LF) [10].

An iterative joint decoding algorithm for the binary data gathering WSN, which is a direct application of the binary CEO problem, was proposed in [11], where a convolutional code is applied at the sensor node. A coding scheme based on the parallel concatenated convolutional codes was proposed in [12], where the extrinsic log-likelihood ratios (LLRs) are weighted by the observation error probabilities at the decoder. Moreover, the capacity of the equivalent parallel channel was derived to verify the bit error rate (BER) performance taking into account the error probability of the observed data sequence. In [13], an adaptive bi-modal decoder for a binary source estimation involving two sensors was proposed based on the modified extrinsic information transfer (EXIT) chart analysis. A convergence property analysis of the iterative decoding algorithm was presented based on the modified EXIT chart analysis for binary data gathering WSNs in [14]. It shows that iterative process is less useful if the channel quality is very good or the observation accuracy is very low. In [10], we proposed an encoding/decoding technique which can significantly improve the BER performance by exploiting the correlation...
knowledge through the LLR updating function [15], for both a binary independently and identically distributed (i.i.d.) source and a binary Markov source. In [16], we proposed a non-negative constrained iterative algorithm for estimating the observation error probabilities in a WSN having an arbitrary number of sensors.

For the binary CEO problem, most of the previous works focus on the design of practical encoding/decoding algorithms. This motivates us to analyze a problem that how small a distortion level the CEO can achieve from the rate-distortion perspective. Hence, the primary goals of this work are to theoretically provide a lower bound on the Hamming distortion (HD), corresponding to the bit error probability (BEP), and to verify the lower bound by making performance comparison with a practical encoding/decoding algorithm. The major contribution is the derivation of a HD approximation based on an information-theoretic outer bound resulting in a lower bound (approximation) on the HD (or BEP) for a binary CEO problem. Instead, we focus on the derivation of a lower bound that is formulated in the framework of convex optimization. It is derived by providing the converse proof. We separate the stages for JSC decoding and the final decision on the common source [14], [17], [18]. Hence, the primary goals of this work are to focus on the design of practical encoding/decoding algorithms.

The system model of estimating a single source through two nodes is depicted in Fig. 1. A common i.i.d. source $X$ produces a sequence $x^n = \{x(t)\}_{t=1}^n$ by taking values from a binary set $\mathcal{X} = \{0, 1\}$ with equal probability. Source $X$ is observed by two nodes and forwarded to a single destination. Due to the inaccuracy of the estimation and/or limited received signal power at nodes, such as in WSN and WMN, the sequences received by the nodes may contain errors$^1$, and the nodes still forward the erroneous sequences to the destination, which is referred as LF [20], [21]. The error probabilities $Pr(x_1(t) \neq x(t))$ and $Pr(x_2(t) \neq x(t))$ are denoted as $p_1$ and $p_2$, respectively, i.e., $Pr(x_1(t) = 1) = p_1$ for the binary noise sequence $z_1^n = \{z_1(t)\}_{t=1}^n$, $i = 1, 2$. At the nodes, the noisy versions $x_1^n = \{x_1(t)\}_{t=1}^n$ and $x_2^n = \{x_2(t)\}_{t=1}^n$ of $x^n$ are separately encoded by two JSC encoders to generate symbol sequences $s_1^{k_1} = \{s_1(t)\}_{t=1}^{k_1}$ and $s_2^{k_2} = \{s_2(t)\}_{t=1}^{k_2}$ with coding rates $r_1 = n/k_1, i = 1, 2$. The symbol sequences $s_1^{k_1}$ and $s_2^{k_2}$ are then transmitted over two orthogonal AWGN channels, as

$$y_i^{k_i} = h_i \cdot s_i^{k_i} + w_i^{k_i}, \quad i = 1, 2,$$

where $h_i$ and $w_i^{k_i} = \{w(t)\}_{t=1}^{k_i}$ represent the channel gain and the AWGN sequence at the destination, respectively. The orthogonality can be achieved by any scheduled multiple access scheme, like time division multiple access (TDMA), i.e., $s_1^{k_1}$ and $s_2^{k_2}$ can be transmitted at different time intervals. The destination performs JSC decoding to form estimates $\hat{x}_i^n$ of the sequences $x_i^n$, $i = 1, 2$. We define the expected Hamming distortion measures $E[\frac{1}{n} \sum_{t=1}^{n} d(x_i(t), \hat{x}_i(t))] \leq D + \epsilon$ to evaluate the error probability $Pr(x_i(t) \neq \hat{x}_i(t))$ with

$$d(x_i(t), \hat{x}_i(t)) = \begin{cases} 1, & \text{if } x_i(t) \neq \hat{x}_i(t), \\ 0, & \text{if } x_i(t) = \hat{x}_i(t), \end{cases}$$

and $\epsilon$ representing an arbitrarily small positive number.

Finally, the destination reconstructs the source information $x^n$ of which the estimate is denoted as $\hat{x}^n$ based on a decision rule from $\hat{x}_1^n$ and $\hat{x}_2^n$. Therefore, the distortion measure $E[\frac{1}{n} \sum_{t=1}^{n} d(x(t), \hat{x}(t))] \leq D + \epsilon$ can be formulated as a

---

$^1$In WMN applications, the nodes correspond to the transceivers in the multiple routes. In a WMN, a source communicates with a destination through multiple intermediate nodes if they are not within the communication coverage. If errors are allowed in the messages forwarded by the intermediate nodes, the WMN can be also modeled as the model shown in Fig. 1 [19].
function of $D_i$, $i = 1, 2$, as $D = f(D_1, D_2)$, where function $f(\cdot)$ is detailed in subsection IV-A. It should be emphasized here that function $D = f(D_1, D_2)$ limits the decoding scheme to which first reconstructs $x^n_1$ and $x^n_2$ and then makes the decision on $x^n$ from those reconstructions (it is referred to as sequential decoding), as shown in Fig. 1. The optimality of such a decoding scheme is an open problem, but it is definitely of interest for practical systems. Furthermore, $f(D_1, D_2)$ largely depends on the decision rule, i.e., there exists different function $f(D_1, D_2)$ for different decision rules.

According to the source-channel separation theorem for lossy source coding [23], distortions $D_1$ and $D_2$ can be achieved if the following inequalities hold:

$$
\begin{align*}
R_1(D_1) \cdot r_1 & \leq C(\gamma_1), \\
R_2(D_2) \cdot r_2 & \leq C(\gamma_2),
\end{align*}
$$

where $R_i(D_i)$ is the rate-distortion function for the source coding and $C(\gamma)$ is the Shannon capacity using Gaussian codebook$^3$ with the argument $\gamma$ denoting the signal-to-noise ratio (SNR) of the channel. As stated above, $r_1$ and $r_2$ represent end-to-end coding rates of two links. Our goal is to derive the theoretical lower bound on the HD for the system shown in Fig. 1. It is equivalent to minimizing the expected Hamming distortion $D$ through a function $f(D_1, D_2)$ under constraints shown in (3), as

$$
\min_{D_1, D_2} D = f(D_1, D_2)
$$

s.t. (3).

The minimization being performed in (4) is for a specific system which maps the average distortions $D_1$ and $D_2$ to $D$, since function $f(D_1, D_2)$ is defined for designated decision rules. To achieve this goal by solving (4), we turn to derive the rate-distortion function $R_i(D_i)$ for the problem shown in Fig. 1 and to establish the function $D = f(D_1, D_2)$ for the decision rule used at the destination.

### III. Rate-Distortion Region Analysis

#### A. Source Coding

In network information theory, the source coding of the communication system shown in Fig. 1 is modeled by the

$$
C(\gamma) = \frac{1}{2} \log_2(1 + 2\gamma), \quad \text{and for two dimensional signal, } C(\gamma) = \log_2(1 + \gamma) [24].
$$

Fig. 2. The abstract model of the binary CEO problem with two independent nodes.

Fig. 3. The binary multiterminal source coding problem for two correlated binary sources.

binary CEO problem. The abstract model of the binary CEO problem is illustrated in Fig. 2. In order to derive the rate-distortion function $R_i(D_i)$, we first reduce the binary CEO problem to a binary multiterminal source coding problem. An outer bound for the rate-distortion region which is determined by the rate-distortion function $R_i(D_i)$ is then derived for the binary multiterminal source coding problem through the converse proof, as in the Gaussian case [25].

The binary multiterminal source coding problem which we consider is depicted in Fig. 3. Since random sources $X^n_1$ and $X^n_2$ originate from the common source $X^n$, the random variable pair $(X_1, X_2)$ follows a joint probability distribution $Pr_{X_1, X_2}(x_1, x_2)$ given by

$$
Pr_{X_1, X_2}(x_1, x_2) = \begin{cases} 
\frac{1}{2} \cdot \rho, & \text{if } x_1 \neq x_2, \\
\frac{1}{2} \cdot (1 - \rho), & \text{otherwise},
\end{cases}
$$

where $\rho = Pr(x_1 \neq x_2)$ is the correlation parameter between the sources $X_1$ and $X_2$, i.e., $X_2$ can be seen as the output of a binary symmetric channel (BSC) with the crossover probability $\rho$ where $X_1$ is the input. Two encoders separately encode $X^n_1$ and $X^n_2$ at rates $R_1$ and $R_2$ as

$$
\begin{align*}
\varphi_1 : & X^n \rightarrow M_1 = \{1, 2, \cdots, 2^{nR_1}\}, \\
\varphi_2 : & X^n \rightarrow M_2 = \{1, 2, \cdots, 2^{nR_2}\}.
\end{align*}
$$

The encoder output sequences $U_1 = \varphi_1(X^n_1)$ and $U_2 = \varphi_2(X^n_2)$ are transmitted to a common receiver. It jointly decodes the received samples to construct the estimates $(X^n_1, X^n_2)$ of the source pair $(X^n_1, X^n_2)$ denoted as $(X^n_1, X^n_2) = \psi(\varphi_1(X^n_1), \varphi_2(X^n_2))$.

For given distortion values $D_1 \in [0, \frac{1}{2}]$ and $D_2 \in [0, \frac{1}{2}]$, the rate-distortion region $R(D_1, D_2)$ is defined as

$$
R(D_1, D_2) = \left\{(R_1, R_2) : (R_1, R_2) \text{ is admissible such that } E_1 \sum_{i=1}^{n} d(x_i(t), \hat{x}_i(t)) \leq D_i + \epsilon, i = 1, 2\right\}.
$$

#### B. Outer Bound for Rate-Distortion Region

We provide a bound $R_0(D_1, D_2)$ of the rate-distortion region $R(D_1, D_2)$.

**Definition 1:**

$$
R_0^\Gamma(D_1) = \left\{(R_1, R_2) : \forall R_2 \leq 1 \\
R_1 \geq H_b[\rho \cdot H_b^{-1}(1 - R_2)] - H_b(D_1)\right\},
$$

$^3$It has been assumed in this setup that 1) each encoder uses joint typicality encoding and binning based on random coding arguments, and the decoder performs joint typicality decoding with a sufficiently large $n$ to achieve the average distortion $D$, as in the Berger-Tung source coding problem [22]; 2) the errors occurring in each sequence $X^n$ are i.i.d. In the practical system, we use random interleavers to asymptotically make this assumption practical.

As shown in section V-B, the simulation results are consistent with the lower bound calculation based on majority logic decision.

$^3$For one dimensional signal, $C(\gamma) = \frac{1}{2} \log_2(1 + 2\gamma)$, and for two dimensional signal, $C(\gamma) = \log_2(1 + \gamma)$ [24].
with $H_b(a) = -a \log_2(a) - (1-a) \log_2(1-a)$ and $H_b^{-1}(a)$ representing the binary entropy function and its inverse function, respectively. The operator $*$ calculates the binary convolution of the two variables, i.e., $a * b = a(1-b) + b(1-a)$.

$$\mathcal{R}_2^o(D_2) = \left\{(R_1, R_2) : \forall R_1 \leq 1 \quad R_2 \geq H_b[\rho * H_b^{-1}(1 - R_1)] - H_b(D_2)\right\}, \quad (7)$$

$$\mathcal{R}_{12}^o(D_1, D_2) = \left\{(R_1, R_2) : \quad R_1 + R_2 \geq 1 + H_b(\rho) - H_b(D_1) - H_b(D_2)\right\}. \quad (8)$$

For every $D_1 \in [0, \frac{1}{2}]$ and $D_2 \in [0, \frac{1}{2}]$,

$$\mathcal{R}^o(D_1, D_2) = \mathcal{R}_1^o(D_1) \cap \mathcal{R}_2^o(D_2) \cap \mathcal{R}_{12}^o(D_1, D_2). \quad (9)$$

In what follows, we prove that $\mathcal{R}_1^o(D_1)$, $\mathcal{R}_2^o(D_2)$ and $\mathcal{R}_{12}^o(D_1, D_2)$ are the supersets of the regions of $\mathcal{R}(D_1, D_2)$. It means that the following theorem holds.

**Theorem 1:** $\mathcal{R}^o(D_1, D_2)$ is an outer bound for the rate-distortion region $\mathcal{R}(D_1, D_2)$; i.e., $\mathcal{R}(D_1, D_2) \subseteq \mathcal{R}^o(D_1, D_2)$.

C. Proof

**Proof of Theorem 1 (Converse):** To prove Theorem 1, the following three different cases of the binary multiterminal source coding problem are considered.

**Case 1.** In order to prove that $\mathcal{R}(D_1, D_2) \subseteq \mathcal{R}_2^o(D_1)$, we assume that the rate pair $(R_1, R_2) \in \mathcal{R}(D_1, D_2)$ and show that this implies that $(R_1, R_2) \in \mathcal{R}_2^o(D_1)$. In the proof, $X_2^n$ is first reconstructed without constraint on $D_2$ which results in (15), and then $X_2^n$ is regarded as the side information to recover $X_1^n$. Assume that a rate pair $(R_1, R_2)$ achieves distortion $D_1$, then

$$n \cdot (R_1 + \epsilon) \geq H(U_1) \geq H(U_1 | U_2) = I(X_1^n; U_1 | U_2)$$

$$= I(X_1^n; U_1) - I(X_1^n; U_2) \geq I(X_1^n; U_1) - I(X_1^n; U_2)$$

where the steps are justified, because (10) conditioning reduces entropy,

(11) $U_1$ is a function of $X_1^n$,

(12) the chain rule of mutual information,

(13) $X_1^n \rightarrow (U_1, U_2) \rightarrow X_2^n$ forms a Markov chain.

Now our aim is to lower bound $I(X_1^n; X_2^n)$ and upper bound $H(X_1^n | U_2)$. Since $I(X_1^n; X_2^n) = H(X_1^n) - H(X_1^n | X_2^n) = n - H(X_1^n | X_2^n)$, to lower bound $I(X_1^n; X_2^n)$ is equivalent to upper bound $H(X_1^n | X_2^n)$. According to the Fano’s inequality, we have

$$H(X_1^n | X_2^n) \leq n \cdot H_b(D_1) + n \cdot D_1 \cdot \log(|\mathcal{X}| - 1) = n \cdot H_b(D_1). \quad (14)$$

On the other hand, since $I(X_1^n; U_2) = H(X_1^n) - H(X_1^n | U_2) = n - H(X_1^n | U_2)$, an upper bound on $I(X_1^n; U_2)$ corresponds to the lower bound on $H(X_1^n | U_2)$. Observing that $X_1^n \rightarrow X_2^n \rightarrow U_2$ forms a Markov chain, it can be shown that $H(X_1^n | U_2) \geq n H_b(\rho * \beta) \beta$ [26, Corollary 4], where $\beta = \frac{1}{n} H_b^{-1} \left[ H_b(X_2^n | U_2) \right]$. A more detailed explanation is given in Appendix A. Since the binary convolution $*$ is monotonically increasing with respect to $\beta$ if $\rho$ is fixed, we need to find the minimizing value of $\beta$ to lower bound $H_b(\rho * \beta)$. We also have the rate constraint on $R_2$ as

$$n \cdot (R_2 + \epsilon) \geq H(U_2) = I(X_2^n; U_2). \quad (15)$$

The inverse function $H_b^{-1}(\alpha) : [0, 1] \rightarrow [0, \frac{1}{2}]$ only takes values from the interval $[0, \frac{1}{2}]$ since distortion is assumed to be within this range.
Letting \( \epsilon \to 0 \), we have \( n \cdot R_2 \geq I(X^n_1;U_2) = H(X^n_2) - H(X^n_2|U_2) = n - nH_0(\beta) \), and hence \( \beta \geq H^{-1}_b(1 - R_2) \). Therefore, the lower bound on \( H(X^n_1|U_2) \) is given by

\[
H(X^n_1|U_2) \geq n \cdot H_0[\rho * H^{-1}_b(1 - R_2)].
\]

(16)

From (13), (14) and (16), we can obtain

\[
n \cdot (R_1 + \epsilon) \geq H(U_1) + H(U_2) \\
\geq H(U_1, U_2) \\
= I(X^n_1;X^n_2;U_1, U_2) \quad (18) \\
= I(X^n_1;U_1, U_2) + I(X^n_2;U_1, U_2|X^n_1) \\
= I(X^n_1;U_1, U_2) + I(X^n_2;X^n_1, U_1, U_2) \\
- n \cdot I(X_1;X_2) \\
\geq I(X^n_1;X^n_1) + I(X^n_2;X^n_2) \\
- n \cdot I(X_1;X_2),
\]

(19)

where (18) holds since \( U_i \) is a function of \( X^n_i \), \( i = 1, 2 \). Similarly, by utilizing Fano’s inequality to lower bound \( I(X^n_1;X^n_1) \) and \( I(X^n_2;X^n_2) \), we have

\[
n \cdot (R_1 + R_2 + \epsilon) \geq n + nH_0(\rho) - nH_0(D_1) - nH_0(D_2).
\]

(20)

Letting \( \epsilon \to 0 \) in the above inequality, we conclude that (8) holds. That is, \( (R_1, R_2) \in \mathcal{R}^n_{12}(D_1, D_2) \).

Through these three cases, it can be concluded that the admissible rate pair \( (R_1 + \epsilon, R_2 + \epsilon) \in \mathcal{R}^n(D_1, D_2) \). Furthermore, the monotonicity of the outer bound \( \mathcal{R}^n(D_1, D_2) \) [27] implies that \( \mathcal{R}^n(D_1, D_2) \subseteq \mathcal{R}^n(D_1 + \epsilon, D_2 + \epsilon) \). Since \( (R_1, R_2) \) is admissible, we conclude that \( \mathcal{R}(D_1, D_2) \subseteq \mathcal{R}^n(D_1, D_2) \) by letting \( \epsilon \to 0 \).

Remark 1: If either \( R_1 = 0 \) or \( R_2 = 0 \), \( \mathcal{R}^n(D_1, D_2) \) is consistent with the rate-distortion function \( 1 - H_0(D_i) \) for the binary source.

Remark 2: \( \mathcal{R}^n(D_1, D_2) \) reduces to the Slepian-Wolf rate region [28] for correlated binary sources if we set \( D_1 \to 0 \) and \( D_2 \to 0 \). The Slepian-Wolf rate region and \( \mathcal{R}^n(D_1, D_2) \) are shown in Fig. 4. It can be found that by allowing nonzero distortion values, the sources can be further compressed compared to the Slepian-Wolf lossless case.

Remark 3: If we are interested in reconstructing only one source of the two sources, say \( X_1 \), and there is no rate limit on describing \( X^n_2 \), i.e., \( R_2 \geq \frac{1}{n}H(X^n_2) \), then it is equivalent to the Wyner-Ziv compression problem [22]. Fig. 5 plots the rate-distortion bound of the Wyner-Ziv source coding [29] and our derived outer bound. In this case, \( \mathcal{R}^n_1(D_1) \) is not tight, since it can be found from Fig. 5 that the rate-distortion region of the Wyner-Ziv problem lies inside of \( \mathcal{R}^n_1(D_1) \).

Remark 4: As it is known that the exact rate-distortion bound of lossy multiterminal source coding problem lies between the Berger-Tung inner and outer bounds [22]. We also derived the rate-distortion region \( \mathcal{R}(D_1, D_2) \) based on the Berger-Tung inner bound [30] after several steps of elementary calculation in information theory, as

\[
\mathcal{R}(D_1, D_2) = \mathcal{R}^n_1(D_1) \cap \mathcal{R}^n_2(D_2) \cap \mathcal{R}^n_{12}(D_1, D_2)
\]

(21)

with

\[
\mathcal{R}^n_1(D_1) = \{(R_1, R_2)|R_1 \geq H_b(\rho * D_1 * D_2) - H_b(D_1)\}
\]

\[
\mathcal{R}^n_2(D_2) = \{(R_1, R_2)|R_2 \geq H_b(\rho * D_1 * D_2) - H_b(D_2)\}
\]

\[
\mathcal{R}^n_{12}(D_1, D_2) = \{(R_1, R_2)|R_1 + R_2 \geq 1 + H_b(\rho * D_1 * D_2) - \sum_{i=1}^2 H_b(D_i)\},
\]

for every \( 0 \leq D_1, D_2 \leq \frac{1}{2} \). In Fig. 4, the Berger-Tung inner bound for binary case \( \mathcal{R}(D_1, D_2) \) is also presented as a reference to verify how close the bounds \( \mathcal{R}^n(D_1, D_2) \) and \( \mathcal{R}(D_1, D_2) \) are. It can be seen from the figure that they are very close to each other for small values of \( D_1 \) and \( D_2 \), i.e., the outer bound can be considered as a useful reference in the evaluation of the BER performance, even though there exists a small gap. However, to resolve this gap, further insightful discussions are still needed [25].

Remark 5: The outer bound \( \mathcal{R}^n(D_1, D_2) \) is obtained by assuming that the code length \( n \) is sufficiently large. However, the penalty on the rate-distortion function is unavoidable if the code length is finite [31], where the rate-distortion function taking into account \( n \) is derived for a conventional point-to-point communication system. According to [31], the impact of \( n \) is not significant when \( n \) is large enough. Therefore, this work focuses on the infinite code length, while the outer bound \( \mathcal{R}^n(D_1, D_2) \) of finite code length is left as a future study.

In summary, the rate-distortion function \( R_n(D_i) \) is given by

\[
R_1(D_1) \geq H_b(\rho * H^{-1}_b(1 - R_2(D_2))) - H_b(D_1),
\]

\[
R_2(D_2) \geq H_b(\rho * H^{-1}_b(1 - R_1(D_1))) - H_b(D_2),
\]

\[
\sum_{i=1}^2 R_i(D_i) \geq 1 + H_b(\rho) - H_b(D_i).
\]

(22)

IV. HAMMING DISTORTION LOWER BOUND

A. Function \( D = f(D_1, D_2) \)

As stated in Section II, distortion \( D \) is a function of distortions \( D_i, i = 1, 2 \). Function \( f(D_1, D_2) \) is obtained by evaluating the relationship between the binary CEO and the binary multiterminal source coding problems in terms of distortions, where the model of the relationship is shown in Fig. 6. The estimate \( X \) is obtained based on the decision rule from the outputs of two parallel BSCs with crossover
probabilities $p_1 \ast D_1$, $p_2 \ast D_2$ and input $X$. The distortion $D$ largely depends on the decision rule used by the destination. Here we only consider two decision rules. One is the weighted majority decision and the other the optimal decision.

1) Weighted majority decision: Distortion $D$ is obtained by evaluating the probability of an error event. Let $\theta_1 = p_1 \ast D_1$ and $\theta_2 = p_2 \ast D_2$. Without loss of generality, we assume that $\theta_1 \leq \theta_2$. Hence, the error event is composed of two independent events: node 1 makes a wrong decision and node 2 makes correct decision or both node 1 and node 2 make erroneous decisions. Therefore, the distortion $D$ in this case is approximated by $D \approx \theta_1 (1 - \theta_2) + \theta_1 \theta_2 = \theta_1$. It can be found that the corner point $\theta_1$ or $\theta_2$ in the rate-distortion region is achieved. Hence, the weighted majority decision rule can be seen as being equivalent to that derived from the time sharing method.

2) Optimal decision: Since the block length is assumed to be infinite and the code is random, an optimal lower bound on the distortion $D$ is determined by utilizing the rate-distortion function for the binary source [24], as

$$1 - H_b(\hat{d}) = I(X; \hat{X}) \leq I(X; \hat{X}_1, \hat{X}_2) = H(X) + H(\hat{X}_1, \hat{X}_2) - H(X, \hat{X}_1, \hat{X}_2) = 1 + H_b(\theta_1 \ast \theta_2) - [H(X) + H(\hat{X}_1 | X) + H(\hat{X}_2 | X)]$$

$$= 2 + H_b(\theta_1 + \theta_2) - [1 + H_b(\theta_1) + H_b(\theta_2)] = 1 + H_b(\theta_1 + \theta_2) - H_b(\theta_1) - H_b(\theta_2),$$

(25)

where $\hat{d}$ is the Hamming distortion measure between $X$ and $\hat{X}$, and the steps are justified as:

(23) rate-distortion function for the binary source,

(24) $\hat{X}_1 \rightarrow X \rightarrow \hat{X}_2$ forms a Markov chain.

Thus, it is obvious from (25) that for $0 \leq \hat{d} \leq \frac{1}{2}$, $\hat{d} \geq H_b^{-1}[H_b(\theta_1) + H_b(\theta_2) - H_b(\theta_1 + \theta_2)]$. Therefore, the distortion $D$ is the minimum value of $\hat{d}$, as

$$D = H_b^{-1}[H_b(\theta_1) + H_b(\theta_2) - H_b(\theta_1 + \theta_2)].$$

(26)

It should be emphasized here that the optimal decision acts as a universal lower bound on the HD for specific schemes which assume sequential decoding. However, in the design of practical encoding/decoding algorithms, we do not consider this decision rule.

In summary, the distortion level $D$ of the two decision rules described above is given as

$$D = \begin{cases} \min \{\theta_1, \theta_2\}, & \text{majority decision,} \\ H_b^{-1}[H_b(\theta_1) + H_b(\theta_2) - H_b(\theta_1 + \theta_2)], & \text{optimal.} \end{cases}$$

(27)

B. Distortion Minimization

By substituting the rate-distortion function (22) and (27) into the minimization problem (4), we have

$$\min_{D_1, D_2} D \quad \text{s.t.}$$

$$H_b[p + H_b^{-1}(1 - \frac{C(\gamma_2)}{r_2})] - H_b(D_1) \leq \frac{C(\gamma_1)}{r_1},$$

$$H_b[p + H_b^{-1}(1 - \frac{C(\gamma_1)}{r_1})] - H_b(D_2) \leq \frac{C(\gamma_2)}{r_2},$$

$$1 + H_b(\rho) - H_b(D_1) - H_b(D_2) \leq \frac{C(\gamma_1)}{r_1} + \frac{C(\gamma_2)}{r_2},$$

$$D_i \leq \frac{1}{2}, \quad i = 1, 2,$$

$$D_i \geq 0, \quad i = 1, 2.$$ 

(28)

The reason of using the derived outer bound, not the Berger-Tung inner bound is that, the outer bound can be easily formulated as a convex optimization. The Berger-Tung inner bound includes term $D_1 + D_2$ in the binary entropy function which cannot be easily handled in the minimization. It is found that distortion $D = f(D_1, D_2)$ is monotonically increasing function on the intervals $D_i \in [0, \frac{1}{2}], i = 1, 2$ for both the majority decision and optimal decision rules, and the proof is detailed in Appendix B. Furthermore, since the sequential decoding (first reconstructs $X_1^n$ and $X_2^n$, then makes decision on $X^n$) is applied, we first minimize the $\ell_2$-norm of a vector.
[\[D_1, D_2\]] instead of directly minimizing \( D \), as
\[
\min_{D_1, D_2} ||[D_1, D_2]||_2
\]
s.t.
\[
-H_b(D_1) - H_b(D_2) \leq \frac{C(\gamma_1)}{r_1} + \frac{C(\gamma_2)}{r_2} - 1 - H_b(\rho),
\]
\[
-H_b(D_1) \leq \frac{C(\gamma_1)}{r_1} - H_b[\rho * H_b^{-1}(1 - \frac{C(\gamma_2)}{r_2})],
\]
\[
-H_b(D_2) \leq \frac{C(\gamma_2)}{r_2} - H_b[\rho * H_b^{-1}(1 - \frac{C(\gamma_1)}{r_1})],
\]
\[
D_1 \leq \frac{1}{2}, \quad i = 1, 2,
\]
\[
D_i \leq -0, \quad i = 1, 2,
\]
to obtain the minimal values of \( D_1 \) and \( D_2 \), and then map them to \( D \) by using function \( f(D_1, D_2) \).

It is easily found that the problem (29) is convex since the objective function is convex and function \(-H_b(\cdot)\) is also convex. Therefore, it can be efficiently solved using convex optimization tools. Assume that the minimum values of \( D_1 \) and \( D_2 \) obtained through the convex optimization are denoted as \( D^*_1 \) and \( D^*_2 \), respectively. Substituting \( D^*_1 \) and \( D^*_2 \) into (27), the minimum distortion value \( D^* \) is then obtained through
\[
D^* = \begin{cases} \min\{\theta^*_1, \theta^*_2\}, & \text{majority decision}, \\ H_b^{-1}[H_b(\theta^*_1) + H_b(\theta^*_2) - H_b(\theta^*_1 * \theta^*_2)], & \text{optimal}, \end{cases}
\]

where \( \theta^*_1 \) and \( \theta^*_2 \) are \( p_1 * D^*_1 \) and \( p_2 * D^*_2 \), respectively. It should be emphasized here that the distortion \( D_1 \) or \( D_2 \) should be set to 0 in the optimization problem (28) if \( \frac{C(\gamma_1)}{r_1} \) or \( \frac{C(\gamma_2)}{r_2} \) is larger than or equal to 1, which is the binary entropy of the source \( X_1 \) and \( X_2 \). The reason is that a source can be reconstructed under an arbitrary small error probability if the source coding rate is larger than its entropy even in the case the helper does not exist [24].

V. BER EVALUATION

A. Coding and Decoding Algorithm

We briefly review the encoding/decoding algorithm [10], [16] which is illustrated in Fig. 7. This algorithm is used to verify the theoretical HD lower bound. As illustrated in Fig. 7, each node encodes its erroneous sequence by using a serially concatenated memory-1 convolutional code and an accumulator (ACC). The encoder output sequences are then modulated and transmitted to the destination over statistically independent AWGN and block Rayleigh fading channels, where the channel gain \( h_i \) is static within each block but varies independently block-by-block. At the destination, iterative decoding process is carried out between the decoders of the convolutional code and the ACC, as well as between the two decoders of the convolutional codes through the LLR updating function \( f_e \) to modify the extrinsic LLR, according to the error probabilities \( p_1 \) and \( p_2 \).

B. Numerical Results

The lower bounds\(^6\) on the HD for different SNR values \( \gamma_1, \gamma_2 \) are obtained through solving the convex optimization problem which we presented in Section IV. The results are shown in Figs. 8–11 for AWGN channels and Fig. 12 for block Rayleigh fading channels. The common parameters used in the simulations are:

- Frame length: \( n = 10000 \) bits for AWGN channels and \( n = 2048 \) bits for block Rayleigh fading channels.
- The number of frames: 1000 for AWGN channels and 10000 for block Rayleigh fading channels.
- Interleavers: random.
- Encoder \( C_1 \): half-rate nonrecursive systematic convolutional code with generator polynomial \( G = [03, 02]_8 \), where \([\cdot]_8\) represents the argument is an octal number.
- Modulation: binary phase-shift keying (BPSK) and quadrature phase-shift keying (QPSK) with coherent detection, where channel state information is assumed to be known to the receiver. Natural mapping is used as the mapping rule in QPSK [32, Example 18.2].
- Doping ratio of ACC: 1 for BPSK and 8 for QPSK.
- Decoding algorithm for \( DCC_i \) and \( ACC^{-1} \): log-maximum a posteriori (MAP).

\(^6\)The terminology “lower bound” used here is due to the HD is calculated based on the derived outer bound, even though the approximation of the objective functions is used.
the number of iterations: 30 times.

Fig. 8 shows the error probability lower bounds and the BER versus SNR when \( p_1, \ p_2 \) and SNRs of the two nodes are set identically; this is referred as the symmetric case. It can be found that, the BER curves obtained by simulations and the theoretical lower bounds on the HD (or BEP) exhibit a similar tendency. Furthermore, it is clearly found that the error floor of the BER obtained by the simulation and the lower bound on the HD based on majority decision match exactly. The reason is that if the SNRs of two nodes are large enough, the distortion levels \( D_1 \) and \( D_2 \) are almost 0, which results in the error floor being determined completely by the error probabilities \( p_1 \) and \( p_2 \). A gap clearly appears between the HD lower bounds using the majority and optimal decision rules. The reason is twofold: 1) the optimality of the majority decision can not be guaranteed; 2) optimal decision is derived based on the assumption of the binary rate-distortion function without considering any loss during processing the information. To find a better decision rule than majority decision rule is left as a future study. However, it is clear that the HD lower bound deriving from the optimal decision cannot be exceeded.

The impact of the variation of the error probabilities \( p_1, \ p_2 \) and the coding rates \( r_i \) are evaluated in AWGN channels. Fig. 9 shows the results for asymmetric \( p_1 \) and \( p_2 \) but symmetric SNRs. When the coding rates\(^7 \) \( r_1 \) and \( r_2 \) are set as \( \frac{1}{3} \) and \( \frac{1}{2} \), respectively, the BER performance shown in Fig. 10 is obtained. We further consider using different modulation schemes for the nodes to achieve different rates of the channel code in Fig. 11, where QPSK is used for node 1 and BPSK for node 2. Even in these asymmetric cases, the theoretical lower bounds on the HD can still provide us with a useful reference when we evaluate the BER performance of practical systems. Furthermore, the theoretical lower bounds on the HD obtained based on our derived outer bound exhibit similar behaviors to those of the BER curves found by simulations.

\(^7\)We simply transmit the output of ACC without doping to achieve rate \( \frac{1}{4} \). No optimized design of the channel code is considered.

In both the symmetric and asymmetric cases, the threshold SNR value at which turbo cliff in the BER obtained by the simulation is around 1.5 dB larger than that observed in the theoretical lower bounds in static AWGN channels. In addition, since the lower bounds on the HD plateaus at a certain level even if the power is increased at high SNR regime, increasing the number of nodes is a proper way to improve performance in the practical deployment.

In Fig. 12, the channels between two nodes and the destination experience independent block Rayleigh fading. Therefore, the instantaneous SNRs of two nodes are different while the average SNRs of the two channels are the same. The lower bounds on the HD shown in Fig. 12 are calculated as

\[
\text{HD}_{\text{fading}}^* = \int_0^{+\infty} \int_0^{+\infty} D^*(\gamma_1, \gamma_2) \cdot \Pr(\gamma_1) \cdot \Pr(\gamma_2) \ d\gamma_1 \ d\gamma_2,
\]

where \( D^*(\gamma_1, \gamma_2) \) is the result of (30), obtained for static AWGN channels. \( \Pr(\gamma_i) \) is the probability density function of the SNR \( \gamma_i \), which follows the Rayleigh distribution. We use Monte Carlo method to obtain the lower bounds on the average HD \( \text{HD}_{\text{fading}}^* \) instead of theoretically calculating (31). In the Rayleigh fading case, the shape of the BER curves and the lower bounds on the HD are almost the same.


VI. Conclusion

We examined theoretically the lower bound on the HD for the binary CEO problem, where two independent nodes forward the erroneous versions of a common binary source to the destination over static AWGN and block Rayleigh fading channels. The binary CEO problem was first formulated as the binary multiterminal source coding problem, which is the core part of the binary CEO problem. The outer bound for rate-distortion region of the binary multiterminal source coding problem was then derived based on the converse proof of the bound. The relationship between the binary CEO problem and the binary multiterminal source coding problem in terms of the distortion function has been established. According to the lossy source-channel separation theorem, the lower bound on
The joint probability mass function of this triple is given by
random variables, where $X$ we assumed and the output of a BSC with crossover probability $H$. If $Y$ with crossover probability $p$, In other words, $Y$ belongs to an arbitrary discrete set $A$. Then Corollary

**APPENDIX A**

**LOWER BOUNDED $H(X^n|U_2)$**

For the completeness of the paper, we briefly provide the core of [26, Corollary 4]. Let $(X^n, Y^n, W)$ be a triple of random variables, where $X^n$ and $Y^n$ take values from a set $X^n = \{0, 1\}^n$, and $W$ belongs to an arbitrary discrete set $W$. The joint probability mass function of this triple is given by

$$
Pr_{X,Y,W}(x, y, w) = Pr\{X = x, Y = y, W = w\} = \prod_{i=1}^{n} Pr(y(t)|x(t)) \cdot Pr_{X,W}(x, w).
$$

(32)

In other words, $Y^n$ can be seen as the output of a BSC channel with crossover probability $p_0$ when $X^n$ is the input, and $W$ is conditionally independent of $Y^n$ given $X^n$. Then Corollary 4:

If $H(X^n|W) \geq n \cdot v$, then $H(Y^n|W) \geq n \cdot H_0[p_0 \cdot H^{-1}(v)]$.

When we substitute $X^n_1$, $X^n_2$ and $U_2$ into the corollary, we have $H(X^n_1|U_2) \geq nH_0(\rho + \beta)$ since $X^n_1$ and $X^n_2$ is the input and the output of a BSC with crossover probability $\rho$, and as we assumed $H(X^n_2|U_2) \geq H_0(\beta)$.

**APPENDIX B**

**MONOTONICITY OF DISTORTION $D$**

Majority decision. $D = \min\{\theta_1, \theta_2\}$. Since $\theta_i$, $i = 1, 2$ is the result of the binary convolution on $p_i$ and $D$, $\theta_i$ is obviously increasing as $D$ is increasing, when $p_i$ is fixed. Optimal decision. $D = H^{−1}_b(\theta_1) + H^{−1}_b(\theta_2) - H_b(\theta_1 \ast \theta_2)$. In this case, $D$ is a composite function of $H^{−1}_b(\cdot)$ and $H_b(\theta_1) + H_b(\theta_2) - H_b(\theta_1 \ast \theta_2)$. Since the function $H^{−1}_b(\cdot)$ is monotonically increasing, we only need to prove that $g(\theta_1, \theta_2) = H_b(\theta_1) + H_b(\theta_2) - H_b(\theta_1 \ast \theta_2)$ is also an increasing function of $\theta_1$ and $\theta_2$.

Assume $\theta_2$ is fixed. The partial derivative $\frac{\partial g(\theta_1, \theta_2)}{\partial \theta_1}$ on $\theta_1$ is

$$
\frac{\partial g(\theta_1, \theta_2)}{\partial \theta_1} = \log \frac{1 - \theta_1}{1 - \theta_1 \ast \theta_2} - \log \frac{1 - \theta_1 * \theta_2}{\theta_1 \ast \theta_2}. \tag{33}
$$

In order to prove that (33) is nonnegative, we should prove

$$
1 - \frac{\theta_1}{\theta_1 \ast \theta_2} \geq \left(1 - \frac{\theta_1 * \theta_2}{\theta_1 \ast \theta_2}\right)^{(1 - 2\theta_2)}. \tag{34}
$$

Therefore, it is found that (33) can not take negative values according to (35). Symmetrically, we can assume $\theta_1$ is fixed, and show that the partial derivative $\frac{\partial g(\theta_1, \theta_2)}{\partial \theta_2}$ on $\theta_2$ is also nonnegative. Hence, $g(\theta_1, \theta_2)$ is increasing in the dimension of $\theta_1$ and $\theta_2$, respectively. Based on the above two cases, it is concluded that the distortion $D$ is increasing with respect to $D_1$ and $D_2$.

**REFERENCES**


