Wyner’s Common Information for Continuous Random Variables -
A Lossy Source Coding Interpretation

Ge Xu  
_Syracuse University_

Wei Liu  
_Syracuse University_

Biao Chen  
_Syracuse University, Department of Electrical Engineering and Computer Science_, bchien@ecs.syr.edu

Follow this and additional works at: [https://surface.syr.edu/eecs_techreports](https://surface.syr.edu/eecs_techreports)

Part of the Computer Sciences Commons

Recommended Citation
[https://surface.syr.edu/eecs_techreports/48](https://surface.syr.edu/eecs_techreports/48)
Wyner’s Common Information for Continuous Random Variables - A Lossy Source Coding Interpretation

Ge Xu  gexu@syr.edu
Wei Liu  wliu28@syr.edu
Biao Chen  bichen@syr.edu

ABSTRACT: Wyner’s common information can be easily generalized for continuous random variables. We provide an operational meaning for such generalization using the Gray-Wyner network with lossy source coding. Specifically, a Gray-Wyner network consists of one encoder and two decoders. A sequence of independent copies of a pair of random variables \((X, Y) \sim p(x, y)\) is encoded into three messages, one of them is a common input to both decoders. The two decoders attempt to reconstruct the two sequences respectively subject to individual distortion constraints. We show that Wyner’s common information equals the smallest common message rate when the total rate is arbitrarily close to the rate-distortion function with joint decoding. A surprising observation is that such equality holds independent of the values of distortion constraints as long as the distortions are less than certain thresholds. An interpretation for such thresholds is given for the symmetric case.

KEYWORDS: Wyner’s common information, lossy source coding, continuous random variables, Gaussian random variables
Abstract—Wyner’s common information can be easily generalized for continuous random variables. We provide an operational meaning for such generalization using the Gray-Wyner network with lossy source coding. Specifically, a Gray-Wyner network consists of one encoder and two decoders. A sequence of independent copies of a pair of random variables \((X, Y) \sim p(x, y)\) is encoded into three messages, one of them is a common input to both decoders. The two decoders attempt to reconstruct the two sequences respectively subject to individual distortion constraints. We show that Wyner’s common information equals the smallest common message rate when the total rate is arbitrarily close to the rate-distortion function with joint decoding. A surprising observation is that such equality holds independent of the values of distortion constraints as long as the distortions are less than certain thresholds. An interpretation for such thresholds is given for the symmetric case.

I. INTRODUCTION

Classical notions that characterize the amount of common information contained in a pair of dependent random variables \((X, Y) \sim p(x, y)\) include Shannon’s mutual information and Gács and Körner’s common randomness [1]. Each of these notions carries clearly defined operational meaning. Wyner provided yet another definition to quantify the common information for a pair of discrete random variables with finite alphabet [3]:

\[
C(X; Y) = \inf_{X \rightarrow W \rightarrow Y} I(X; Y; W). \tag{1}
\]

Here, the infimum is taken over all auxiliary random variables \(W\) such that \(X, W, Y\) form a Markov chain, i.e., \(X\) and \(Y\) are conditionally independent given \(W\).

Wyner’s common information also carries its own operational meanings, two of them were given in Wyner’s original paper [3]. The first approach is shown in Fig. 1. The encoder observes a pair of sequences \((X^n, Y^n)\) and outputs three messages \(W_0, W_1, W_2\) where

\[
W_i \in \{1, \cdots, 2^{nR_i}\}
\]

for \(i = 0, 1, 2\). Decoder 1 reconstructs \(X^n\) from messages \((W_0, W_1)\), and decoder 2 reconstructs \(Y^n\) from \((W_0, W_2)\). For any given \(\epsilon\), define \(C_1\) to be the minimum \(R_0\) for the system in Fig. 1 such that the total rate \(\sum_{i=0}^2 R_i < H(X, Y) + \epsilon\) and the probabilities of error at both decoders are bounded by \(\epsilon\).

Wyner’s second approach is shown in Fig. 2. In this model, a common message \(W\) uniformly distributed on \(W = \{1, \cdots, 2^{nR_0}\}\) is passed to two separate random number generators, whose outputs are generated independently according to distributions \(p_1(X|W)\) and \(p_2(Y|W)\). The output sequences of the two processors, denoted as \(\hat{X}^n\) and \(\hat{Y}^n\) respectively, have joint distribution

\[
p(\hat{X}^n, \hat{Y}^n) = \frac{1}{|W|} \sum_{w \in W} p_1(X^n|W)p_2(Y^n|W). \tag{2}
\]

Let \(C_2\) be the smallest \(R_0\) such that the distribution in (2) is arbitrarily close to \(Q(X^n, Y^n)\), in the sense that

\[
D_n(Q; P) = \frac{1}{n} \sum_{x^n \in X^n, y^n \in Y^n} P(x^n, y^n) \log \frac{P(x^n, y^n)}{Q(x^n, y^n)} \tag{3}
\]

can be made arbitrarily small. Wyner proved that

\[
C(X, Y) = C_1 = C_2, \tag{4}
\]

lending clear practical interpretations to \(C(X, Y)\) defined in equation (1). However, these interpretations only apply to random variables with finite alphabet sets. Indeed, Wyner’s common information was originally defined only for a pair of finite-alphabet discrete random variables. Its generalization to multiple dependent variables was first mentioned in [6] and further developed in [7]. The definition in (1) also applies.

\[\]

Department of EECS

Syracuse University

Email: gexu@syr.edu

Department of EECS

Syracuse University

Email: wliu28@syr.edu

Department of EECS

Syracuse University

Email: bichen@syr.edu

Fig. 1. Gray-Wyner source coding network.

Fig. 2. Distribution Approximation.
to random variables with continuous alphabet. However, it is
not clear what physical interpretation such quantity carries for
continuous random variables.

We provide such an interpretation using the rate distortion
result for the Gray-Wyner network as described in Fig. 1.
That is, instead of requiring the sources to be reproduced losslessly at the two decoders, we allow certain distortions
subject to given distortion constraints [9]. It turns out that
Wyner’s common information is precisely the smallest common
message rate for a certain range of distortion constraints
when the total rate is arbitrarily close to the rate distortion
function with joint decoding. A surprising result is that,
Wyner’s common information is only a function of the joint
distribution, this smallest common rate remains constant
even if the distortion constraints vary, as long as they are less than
certain thresholds.

The rest of the paper is organized as follows. Section II
gives the problem formulation and the main results. The proofs
are given in Appendix. In section III, two examples, the
doubly symmetric binary source studied in [3] and the bivariate
Gaussian source, are given. We conclude in Section IV.

II. PROBLEM FORMULATION AND MAIN RESULT

Let \( \{(X_k, Y_k)\}_{k=1}^{\infty} \) be independent copies of a pair of
dependent random variables \((X, Y) \sim Q(x, y)\) which take
values in some arbitrary (finite, countable, or continuous)
spaces \( \mathcal{X} \times \mathcal{Y} \). Here, we use \( Q(x, y) \) to denote the joint
distribution of \((X, Y)\), i.e., probability mass function if \((X, Y)\)
discrete and probability density function if \((X, Y)\) are
continuous. Thus the joint distribution of length \( n \) vectors
\((X^n, Y^n)\)

\[
Q^n(x^n, y^n) = \prod_{i=1}^{n} Q(x_i, y_i). (5)
\]

The common information of the pair \((X, Y)\) is a functional of
\( Q \) and is defined as

\[
C(X, Y) \triangleq \inf I(X; Y; W), (6)
\]

where the infimum is taken over all random variable triples
\( X, Y, W \) satisfying

- (C1) The marginal distribution for \( X, Y \) is \( Q(x, y) \),
- (C2) \( X \) and \( Y \) are conditionally independent given \( W \).

Let us consider the lossy source coding problem described
in Fig. 1. The encoder observes a pair of sequences \((X^n, Y^n)\),
and map them to three messages \( W_0, W_1, W_2 \) with

\[
W_i \in \{1, \cdots, 2^{nR_i}\},
\]

for \( i = 0, 1, 2 \). Let \( d_1(x, \hat{x}) \) and \( d_2(x, \hat{y}) \) be bounded single
letter distortion functions defined on \( \mathcal{X} \times \mathcal{X} \) and \( \mathcal{Y} \times \mathcal{Y} \) respectively.
Decoder 1 reproduces \( X^n \) from \((W_0, W_1)\) subject to
an average distortion constraint \( \Delta_1 \); decoder 2 reproduces \( Y^n \)
from \((W_0, W_2)\) subject to an average distortion constraint \( \Delta_2 \).
We now give a precise definition of the quantity \( C_3(\Delta_1, \Delta_2) \),
which is the smallest common rate \( R_0 \) such that the total rate
meets the rate distortion bound with joint decoding.

Definition 1: An \((n, M_0, M_1, M_2)\) rate distortion code consists of the following:

- One encoder mapping \( f_E \)
  \[
  f_E : \mathcal{X}^n \times \mathcal{Y}^n \to I_{M_0} \times I_{M_1} \times I_{M_2}, (7)
  \]
  where \( I_{M_i} = \{0, 1, 2, \cdots, M_i - 1\} \) for \( i = 0, 1, 2 \).
- Two decoder mappings \( f_D^{(X)}, f_D^{(Y)} \)
  \[
  f_D^{(X)} : I_{M_0} \times I_{M_1} \to \mathcal{X}^n, (8)
  \]
  \[
  f_D^{(Y)} : I_{M_0} \times I_{M_2} \to \mathcal{Y}^n. (9)
  \]

Let \( f_E(X^n, Y^n) = (W_0, W_1, W_2), 1 \leq W_i \leq M_i \) and

\[
\hat{X}^n = f_D^{(X)}(W_0, W_1), (10)
\]

\[
\hat{Y}^n = f_D^{(Y)}(W_0, W_2). (11)
\]

Denote by \( (\Delta_X, \Delta_Y) \) the average distortion between encoder
inputs and decoder outputs:

\[
\Delta_X = Ed_1(X^n, \hat{X}^n), (12)
\]

\[
\Delta_Y = Ed_2(Y^n, \hat{Y}^n), (13)
\]

where

\[
d_1(x^n, \hat{x}^n) = \frac{1}{n} \sum_{k=1}^{n} d_1(x_k, \hat{x}_k), (14)
\]

\[
d_2(y^n, \hat{y}^n) = \frac{1}{n} \sum_{k=1}^{n} d_2(y_k, \hat{y}_k). (15)
\]

An \((n, M_0, M_1, M_2)\) code with distortion \( (\Delta_X, \Delta_Y) \) is referred to as an \((n, M_0, M_1, M_2, \Delta_X, \Delta_Y)\) rate distortion code.

Definition 2: For any \( \Delta_1, \Delta_2 \geq 0 \), a number \( R_0 \) is said to be \((\Delta_1, \Delta_2)\)-achievable if for any \( \epsilon > 0 \) we can find \( n \) sufficiently large such that there exists a
\((n, M_0, M_1, M_2, \Delta_X, \Delta_Y)\) rate distortion code with

\[
M_0 \leq 2^{nR_0}, (16)
\]

\[
\frac{1}{n} \sum_{i=0}^{2} \log M_i \leq R_{XY}(\Delta_1, \Delta_2) + \epsilon, (17)
\]

\[
\Delta_X \leq \Delta_1 + \epsilon, \quad \Delta_Y \leq \Delta_2 + \epsilon. (18)
\]

where \( R_{XY}(\Delta_1, \Delta_2) \) is the rate distortion function for \((X, Y)\)
with joint encoding and decoding, i.e.,

\[
R_{XY}(\Delta_1, \Delta_2) = \min I(X; Y; \hat{X}, \hat{Y}), (19)
\]

where the minimum is taken over all the test channels
\( q(\hat{x}, \hat{y}|x, y) \) such that \( Ed_1(X, \hat{X}) \leq \Delta_1, Ed_2(Y, \hat{Y}) \leq \Delta_2 \).

Definition 3: \( C_3(\Delta_1, \Delta_2) \) is defined as the infimum of all
\( R_0 \) that is \((\Delta_1, \Delta_2)\)-achievable.

We now state the main results.

Theorem 1: The common information \( C(X; Y) = C_3(\Delta_1, \Delta_2) \) in some neighborhood of the origin
\( \{(\Delta_1, \Delta_2) : 0 \leq \Delta_1, \Delta_2 \leq \gamma\} \) provided that

\[
Q(x, y) > 0 \quad \text{all } x \in \mathcal{X}, y \in \mathcal{Y}, (20)
\]

and \( d_1, d_2 \) satisfy

\[
d_1(x, \hat{x}) > d_1(x, x) = 0, x \neq \hat{x}, (21)
\]

\[
d_2(y, \hat{y}) > d_2(y, y) = 0, y \neq \hat{y}. (22)
\]
A proof of Theorem 1 is given in Appendix A.

The condition on $d_1$ and $d_2$ set forth in the theorem amounts to requiring the distortion function be normal, as defined in [10].

If $(X, Y)$ are discrete random variables with finite alphabet, and $d_1 = d_2 = d_H$ are the Hamming distortion, defined as

$$d_H(u, \hat{u}) = \begin{cases} 0, & u = \hat{u} \\ 1, & u \neq \hat{u}, \end{cases}$$

then for $\Delta_1 = \Delta_2 = 0$, $C_3(\Delta_1, \Delta_2) = C_1 = C(X; Y)$.

Therefore, approach 1 in [3] is a special case of Theorem 1.

**Theorem 2:** For the symmetric case $\Delta_1 = \Delta_2 = \Delta$, $C_3(\Delta) = C(X, Y)$ if and only if $\Delta \leq \frac{1}{2} R_{XY}^{-1}(C(X, Y))$, where $R_{XY}^{-1}(\cdot)$ denotes the inverse function of $R_{XY}(\Delta, \Delta)$, i.e., the distortion rate function.

A proof of Theorem 2 is given in Appendix B.

### III. Examples

#### A. Doubly symmetric binary source (DSBS)

Consider a DSBS as in [3], [9]. That is, a binary source where $X = Y = \{0, 1\}$ and for $x, y = 0, 1$,

$$Q(x, y) = \frac{1}{2}(1 - a_0)\delta_{x,y} + \frac{1}{2}a_0(1 - \delta_{x,y}),$$

$$0 \leq a_0 \leq \frac{1}{2}$$

and $\delta_{x,y}$ is an indicator function of $x = y$. $X$ can be considered as an unbiased binary input to a binary symmetric channel (BSC) with crossover probability $a_0$ and $Y$ as the corresponding output, or vice versa.

It is shown in [3] that for the DSBS

$$C(X; Y) = 1 + h(a_0) - 2h(a_1),$$

where $h(a_0)$ is the binary entropy function for $0 \leq a_0 \leq 1$ and $a_1 = \frac{1}{2} - \frac{1}{2}(1 - 2a_0)^{\frac{1}{2}}$.

For a DSBS with Hamming distortion $d_1 = d_2 = d_H$ and symmetric distortion constraint $\Delta_1 = \Delta_2 = \Delta$, the joint rate distortion function [12] is given by (26), where $L(x) = -x \log x$. It can be seen that

$$R_{XY}(a_1, a_1) = 1 + h(a_0) - 2h(a_1) = C(X, Y).$$

Therefore, by Theorem 2 we have $\gamma = a_1$, $C_3(\Delta, \Delta) = 1 + h(a_0) - 2h(a_1)$ for any $0 \leq \Delta \leq a_1$.

**Remark:** $C_3(\Delta, \Delta)$ for any $0 \leq \Delta \leq a_1$ is achieved by $R_0 = R_{XY}(a_1, a_1) = 1 + h(a_0) - 2h(a_1)$, $R_1 = R_{X|\hat{X},\hat{Y}}(\Delta)$, and $R_2 = R_{Y|\hat{X},\hat{Y}}(\Delta)$, where $(\hat{X}, \hat{Y})$ are the random variables achieving $R_{XY}(a_1, a_1)$. The test channels are

$$P_r\{X = x|\hat{x}, \hat{y}\} = (1 - a_1)\delta_{x,\hat{x}} + a_1(1 - \delta_{x,\hat{x}}),$$

$$P_r\{Y = y|\hat{x}, \hat{y}\} = (1 - a_1)\delta_{y,\hat{y}} + a_1(1 - \delta_{u,\hat{y}}).$$

Hence, $R_1 = R_2 = h(a_1) - h(\Delta)$.

#### B. Gaussian source

In this section we consider the case when $X, Y$ are bivariate Gaussian with zero mean and covariance matrix

$$K = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$  

**Proposition 1:** For the Gaussian random variable $(X, Y)$ described above, the common information is

$$C(X; Y) = \frac{1}{2} \log\left[\frac{1 + \rho}{1 - \rho}\right].$$

The proof is given in Appendix C. Proposition 1 can be extended to multivariate Gaussian distributions.

**Corollary 1:** For $N$ joint Gaussian random variables $X_1, X_2, \ldots , X_N$ with covariance matrix

$$K_N = \begin{bmatrix} 1 & \rho & \ldots & \rho \\ \rho & 1 & \ldots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \ldots & 1 \end{bmatrix},$$

the common information is

$$C(X_1, X_2, \ldots , X_N) = \frac{1}{2} \log\left[\frac{1 + N\rho}{1 - \rho}\right].$$

**Proposition 2:** For bivariate Gaussian random variables $X, Y$ with zero mean and covariance matrix in (29) and squared error distortion $d_1(u, \hat{u}) = d_2(u, \hat{u}) = (u - \hat{u})^2$, we have

$$C_3(\Delta, \Delta) = C(X; Y).$$

for any $\Delta \leq 1 - \rho$.

**Proof:** The joint rate distortion function for Gaussian random variables with symmetric squared error distortion [12] is

$$R_{XY}(\beta, \beta) = \begin{cases} \frac{1}{2} \log\left\{\frac{1 + \rho^2}{1 - \rho}\right\} & 0 \leq \beta \leq 1 - \rho \\ \frac{1}{2} \log\left\{\frac{1 + \rho}{1 - \rho}\right\} - \frac{1}{2} \log\left\{\frac{1 + \rho}{1 - \rho}\right\} & 1 - \rho \leq \beta \leq 1 \\ 0 & \beta \geq 1 \end{cases}.$$  

Thus we have

$$R_{XY}(1 - \rho, 1 - \rho) = \frac{1}{2} \log\left\{\frac{1 + \rho}{1 - \rho}\right\} = C(X, Y).$$

By Theorem 2, $\gamma = 1 - \rho$. This means that $C_3(\Delta, \Delta) = C(X; Y)$ for any $\Delta \leq 1 - \rho$.

**Remark:** $C_3(\Delta, \Delta)$ for any $0 \leq \Delta \leq 1 - \rho$ is achieved by $R_0 = R_{XY}(1 - \rho, 1 - \rho)$, $R_1 = R_{X|\hat{X},\hat{Y}}(\Delta)$, $R_2 = R_{Y|\hat{X},\hat{Y}}(\Delta) = \frac{1}{2} \log\left\{\frac{1 + \rho}{1 - \rho}\right\}$, where $(X, Y)$ are the random variables achieving $R_{XY}(1 - \rho, 1 - \rho)$.

### IV. Conclusion

In this paper, we generalized Wyner’s common information to that of continuous random variables and provided a lossy source coding interpretation using the Gray-Wyner network. A surprising observation is that the minimum common rate for lossy source coding is invariant to the distortion constraint as long as it is less than a certain threshold.
\[ R_{XY}(\beta, \beta) = \begin{cases} 1 + h(a_0) - 2h(\beta) \\ L(1-a_0) - \frac{1}{2}L(2\beta - a_0) + L[2(1-\beta) - a_0] \end{cases} \text{if } 0 \leq \beta \leq a_1 \]
\[ a_1 \leq \beta \leq \frac{1}{2} \]  

(26)

**APPENDIX**

A. Proof of Theorem 1

We first introduce the following two lemmas. The first one is given by Gray [8].

**Lemma 1:** Given a two-dimensional source \( X, Y \) and a compound distortion measure, we have the following inequalities

\[ R_{XY}(\Delta_1, \Delta_2) \geq R_{XY}(\Delta_1) + R_Y(\Delta_2), \]
\[ R_{X|Y}(\Delta_1) \geq R_X(\Delta_1) - I(X; Y), \]

and equalities hold in some neighborhood of the origin \( \{(\Delta_1, \Delta_2) : 0 \leq \Delta_1, \Delta_2 \leq \gamma \} \), provided that

\[ Q(x, y) > 0 \text{ for all } x \in \mathcal{X}, y \in \mathcal{Y}, \]

and \( d_1, d_2 \) satisfy

\[ d_1(x, \hat{x}) > d_1(x, x) = 0, x \neq \hat{x}, \]
\[ d_2(y, \hat{y}) > d_2(y, y) = 0, y \neq \hat{y}. \]

Here \( R_{X|Y}(\Delta) \) is the conditional rate distortion function which is defined as

\[ R_{X|Y}(\Delta) = \min I(X; \hat{X}|Y), \]

(41)

where the minimum is taken with respect to all test channels \( q_2(\hat{x}|x, y) \) such that \( Ed(X, \hat{X}) \leq \Delta \).

The second lemma is given by Gray and Wyner [9].

**Lemma 2:** For the lossy source coding problem described in the previous section, for \( \Delta_1, \Delta_2 \geq 0 \), the rate distortion region is given by

\[ \mathcal{R}(\Delta_1, \Delta_2) = \{(R_0, R_1, R_2) : R_0 \geq I(X, Y; W), R_1 \geq R_{X|W}(\Delta_1), R_2 \geq R_{Y|W}(\Delta_2)\}, \]

(42)

for some distributions \( p(w|x, y)Q(x, y) \).

Note that Lemma 2 is valid for both the discrete case and the continuous case. Although Gray and Wyner only treated the discrete case in [9], the result can be generalized to the continuous case [11].

We now prove Theorem 1.

1) **Achievability:** For a given \( Q(x, y) > 0 \) for all \( x \in \mathcal{X}, y \in \mathcal{Y} \), let \( C(X; Y) = I(X; Y|W) \) where \( (X, Y, W) \) satisfies \( X - W - Y \) and \( \sum_w p(x, y, w) = Q(x, y) \), i.e., \( W \) is the auxiliary variable that achieves \( C(X, Y) \). Let \( (\Delta_1, \Delta_2) \) be in the range \( \{0 \leq \Delta_1, \Delta_2 \leq \gamma \} \) where \( \gamma \) is chosen such that the following equalities hold

\[ R_X(\Delta_1) = R_{X|W}(\Delta_1) + I(X; W), \]
\[ R_Y(\Delta_2) = R_{Y|W}(\Delta_2) + I(Y; W), \]
\[ R_{XY}(\Delta_1, \Delta_2) = R_X(\Delta_1) + R_Y(\Delta_2) - I(X; Y). \]

(43)
\[ (44) \]
\[ (45) \]

We now prove that \( C_3(\Delta_1, \Delta_2) \leq C(X, Y) \) in the range \( \{0 \leq \Delta_1, \Delta_2 \leq \gamma \} \). For any \( R_0 > C(X, Y) \) and \( \epsilon > 0 \) let

\[ \epsilon_1 = \min(\epsilon/3, R_0 - C(X, Y)). \]

(46)

Since \( \epsilon_1 > 0 \), we know from Lemma 2 that there exists a code \( (n, M_0, M_1, W, \Delta_X, \Delta_Y) \) with \( \Delta_X \leq \Delta_1 + \epsilon_1, \Delta_Y \leq \Delta_2 + \epsilon_1 \) and

\[ \frac{1}{n} \log M_0 \leq I(X, Y; W) + \epsilon_1 \]
\[ = C(X, Y) + \epsilon_1 \leq R_0, \]
\[ \frac{1}{n} \log M_1 \leq R_{X|W}(\Delta_1) + \epsilon_1, \]
\[ \frac{1}{n} \log M_2 \leq R_{Y|W}(\Delta_2) + \epsilon_1. \]

(47)
(48)
(49)

From (47)-(49), we have that

\[ \frac{1}{n} \sum_{i=0}^{2} \log M_i \]
\[ \leq I(X, Y; W) + R_{X|W}(\Delta_1) + R_{Y|W}(\Delta_2) + 3\epsilon_1, \]
\[ = I(X; W) + R_{X|W}(\Delta_1) + I(Y; W) + R_{Y|W}(\Delta_2) \]
\[ - I(X; Y) + 3\epsilon_1, \]
\[ = R_X(\Delta_1) + R_Y(\Delta_2) - I(X; Y) + 3\epsilon_1, \]
\[ \leq R_{XY}(\Delta_1, \Delta_2) + \epsilon. \]

(50)
(51)
(52)
(53)

where (51) follows from the chain rule and the Markov Chain \( X - W - Y \), (52) and (53) follow from (43)-(46).

This proves that the code satisfies (16)-(18), i.e., \( R_0 \) is \( (\Delta_1, \Delta_2) \)-achievable. This completes the proof of \( C_3(\Delta_1, \Delta_2) \leq C(X; Y) \).

2) **Converse:** Let \( \Delta_1, \Delta_2 \) be in the region \( \{0 \leq \Delta_1, \Delta_2 \leq \gamma \} \) such that

\[ R_X(\Delta_1) + R_Y(\Delta_2) - I(X; Y) = R_{XY}(\Delta_1, \Delta_2). \]

(54)

Let \( R_0 \) be \( (\Delta_1, \Delta_2) \)-achievable. We will show that \( R_0 \geq C(X; Y) \). The proof follows similar procedures as the proof of Theorem 5.1 in [3].

Since \( R_0 \) is \( (\Delta_1, \Delta_2) \)-achievable, there exists an \( (n, M_0, M_1, M_2, \Delta_X, \Delta_Y) \) code satisfying (16)-(18). Let \( f_E(X^n, Y^n) = (W_0, W_1, W_2) \), we have that

\[ R_0 \geq \frac{1}{n} \log M_0 \geq \frac{1}{n} H(W_0), \]
\[ \geq \frac{1}{n} I(X^n, Y^n; W_0), \]
\[ = \frac{1}{n} H(X^n, Y^n) - \frac{1}{n} H(X^n, Y^n|W_0), \]
\[ = H(X, Y) - \frac{1}{n} \sum_{k=1}^{n} H(X_k, Y_k|X^{k-1}, Y^{k-1}, W_0), \]
\[ \geq H(X, Y) - \frac{1}{n} \sum_{k=1}^{n} \Gamma_1(\delta(k)), \]

(55)
(56)
(57)
(58)
(59)
where (59) comes from the definition of $\Gamma_1(\cdot)$ (c.f. Corollary 4.5, [3]) and the definition of $\delta(k)$, where
\[
\delta(k) = I(X_k; Y_k|X^{k-1}, Y^{k-1}, W_0).
\]
Inequality (60) follows from the concavity of $\Gamma(\cdot)$.

Therefore, since $C(X; Y) = H(X, Y) - \Gamma_1(0)$ (c.f. equation (4.4) in [3]) and (60), to establish $R_0 \geq C(X; Y)$ we only need to prove that, for arbitrary $\epsilon > 0$,
\[
\frac{1}{n} \sum_{k=1}^{n} \delta(k) \leq \nu(\epsilon), \quad \lim_{\epsilon \to 0} \nu(\epsilon) = 0.
\]
(61) (62)

From (57), we have that
\[
\frac{1}{n} \log M_0 \geq \frac{1}{n} H(X^n; Y^n) - \frac{1}{n} H(X^n; Y^n|W_0),
\]
\[
= \frac{1}{n} H(X^n, Y^n) + \frac{1}{n} I(X^n; Y^n|W_0) - \frac{1}{n} H(X^n|W_0) - \frac{1}{n} H(Y^n|W_0).
\]
(63) (64)

Consider again the $(n, M_0, M_1, M_2, \Delta_X, \Delta_Y)$ code that satisfies (16)-(18) for arbitrary $\epsilon > 0$. Set $\hat{X}^n = f_D^{\hat{X}}(W_0, W_1)$ and $\hat{Y}^n = f_D^{\hat{Y}}(W_0, W_2)$, we have
\[
\frac{1}{n} \log M_1 \geq \frac{1}{n} H(W_1),
\]
\[
\geq \frac{1}{n} H(W_1|W_0),
\]
\[
\geq \frac{1}{n} I(X^n; W_1|W_0),
\]
\[
\geq \frac{1}{n} I(X^n; \hat{X}^n|W_0).
\]
(65) (66) (67) (68)

where inequality (68) follows from the Markov chain $X^n - W_0, W_1 - \hat{X}^n$. Similarly, we have
\[
\frac{1}{n} \log M_2 \geq \frac{1}{n} I(Y^n; \hat{Y}^n|W_0).
\]
(69)

Adding (64), (68) and (69), we obtain
\[
\sum_{i=0}^{2} \frac{1}{n} \log M_i \geq \frac{1}{n} (H(X^n, Y^n) + I(X^n; Y^n|W_0) - H(X^n|W_0) - H(Y^n|W_0) + I(X^n; \hat{X}^n|W_0) + I(Y^n; \hat{Y}^n|W_0)).
\]
(70)

\[
= \frac{1}{n} (I(X^n; W_0) + I(Y^n; W_0) - I(X^n; Y^n) + I(X^n; Y^n|W_0) + I(X^n; \hat{X}^n|W_0) + I(Y^n; \hat{Y}^n|W_0)).
\]
(71)

\[
\geq \frac{1}{n} (I(X^n; \hat{X}^n) + I(Y^n; \hat{Y}^n) - n I(X; Y) + I(X^n; Y^n|W_0)),
\]
(72)

\[
\geq R_X(\Delta_1) + R_Y(\Delta_2) - I(X; Y) + \frac{1}{n} I(X^n; Y^n|W_0),
\]
(73)

\[
\geq R_{XY}(\Delta_1, \Delta_2) + \frac{1}{n} I(X^n; Y^n|W_0).
\]
(74)

where (71), (72) follow from the chain rule, (73) follows from the fact that conditioning does not increase entropy, (74) follows from the definition of rate distortion function and (75) is from (54).

On the other hand, the code satisfies (17), so we have
\[
\sum_{i=0}^{2} \frac{1}{n} \log M_i \leq R_{XY}(\Delta_1, \Delta_2) + \epsilon.
\]
(76)

Combining (75) and (76) we will have that
\[
\frac{1}{n} I(X^n; Y^n|W_0) \leq \epsilon.
\]
(77)

Also, it is easy to check that the following inequality is true.
\[
\frac{1}{n} I(X^n; Y^n|W_0) \geq \frac{1}{n} \sum_{k=1}^{n} \delta(k).
\]
(78)

Combining (77) and (78), we obtain
\[
\frac{1}{n} \sum_{k=1}^{n} \delta(k) \leq \epsilon,
\]
(79)

which completes the proof.

B. Proof of Theorem 2

Before proving Theorem 2, we first introduce two lemmas.

**Lemma 3:** For any $\Delta_1, \Delta_2$,
\[
C_3(\Delta_1, \Delta_2) \leq R_{XY}(\Delta_1, \Delta_2).
\]
(80)

**Proof:** The lemma follows from the fact that $R_0 = R_{XY}(\Delta_1, \Delta_2)$ is $C(\Delta_1, \Delta_2)$-achievable.

**Lemma 4:** Let $\tau = R_{XY}^{-1}(C(X, Y))$, $\Delta \leq \tau$, if $R_0$ is $C$-achievable, then there exists a $W$ such that $X - W - Y$, $R_0 \geq I(X; Y; W)$, and
\[
I(X; Y; W) + R_{X|W}(\Delta) + R_{Y|W}(\Delta) = R_{XY}(\Delta, \Delta).\]
(81)

**Proof:** For $\Delta \leq \tau$, if $R_0$ is $C$-achievable, we have that for any $\epsilon > 0$, there exists a code $(n, M_0, M_1, M_2, \Delta, \Delta)$ that satisfies (16)-(18). Let $R_i = \frac{1}{n} \log M_i$ for $i = 0, 1, 2$, we have that
\[
\sum_{i=0}^{2} R_i \leq R_{XY}(\Delta, \Delta) + \epsilon.
\]
(82)

From the definition of rate distortion region [9], we know that $(R_0', R_1' - \epsilon/2, R_2' - \epsilon/2)$ is in the rate distortion region $R$. By Lemma 2, there exists a $W$ jointly distributed with $X, Y$ as $p(w|x, y)q(x, y)$ and satisfies
\[
R_0' + R_1' + R_2' - \epsilon \geq I(X; Y; W) + R_{X|W}(\Delta) + R_{Y|W}(\Delta),
\]
(83)

\[
\geq I(X; Y; W) + R_{XY}(\Delta, \Delta),
\]
(84)

\[
\geq R_{XY}(\Delta, \Delta),
\]
(85)

where inequalities (84) and (85) are from Theorem 3.1 in [8].

The equality in (84) holds only when $X$ is conditionally
independent of \( Y \) given \( W \) and equality in (85) holds only when \( 0 \leq \Delta \leq \gamma \). For \( \Delta = \tau \), combined with (82), we have that 
\[
I(X, Y; W) = R_{X|W}(\tau, \tau).
\]
Hence for any \( \Delta \leq \tau \),
\[
I(X, Y; W) + R_{X|W}(\Delta) + R_{Y|W}(\Delta) = R_{XY}(\Delta, \Delta). \tag{86}
\]
This completes the proof.

We now prove Theorem 2.

First we show that for any \( \Delta \) such that \( C_3(\Delta) = C(X, Y) \), we have \( \Delta \leq \tau \). From Lemma 3, \( R_{XY}(\Delta, \Delta) \geq C(X, Y) \).

\( R_{X|W}(\Delta) \) is a non increasing function of \( \Delta \), therefore, \( \Delta \leq R_{X|W}(\Delta) \).

Next we will show that for any distortion \( \Delta \leq \tau \), \( C_3(\Delta) = C(X, Y) \).

Then use the same proof as the achievability part of Theorem 1, to consider \( (\tau, \tau) \).

Straightforward calculation yields
\[
\begin{align*}
C(X, Y) + R_{X|W}(\tau) + R_{Y|W}(\tau) &= R_{XY}(\tau, \tau),
\end{align*}
\]
where \( W \) is the random variable such that \( C(X, Y) = I(X, Y; W) \). Thus, by Lemma 1, for any \( \Delta \leq \tau \),
\[
C(X, Y) + R_{X|W}(\Delta) + R_{Y|W}(\Delta) = R_{XY}(\Delta, \Delta).
\]
Then use the same proof as the achievability part of Theorem 1, to prove that when the distortion \( \Delta \leq \tau \), any rate \( R_0 > C(X, Y) \) is \( \tau \)-achievable. By Lemma 4, we have
\[
C(X, Y) + R_{X|W}(\tau) + R_{Y|W}(\tau) = R_{XY}(\tau, \tau),
\]
where \( W \) is the random variable such that \( C(X, Y) = I(X, Y; W) \). Thus, by Lemma 1, for any \( \Delta \leq \tau \),
\[
C(X, Y) + R_{X|W}(\Delta) + R_{Y|W}(\Delta) = R_{XY}(\Delta, \Delta).
\]
Without loss of generality, let \( W \) be a Gaussian random variable with zero mean and variance \( \sigma^2 \), and
\[
\begin{align*}
X &= \rho_1 W + \sqrt{1 - \rho_1^2} \sigma N_1, \quad \tag{90} \\
Y &= \rho_2 W + \sqrt{1 - \rho_2^2} \sigma N_2, \quad \tag{91}
\end{align*}
\]
where \( N_1 \) and \( N_2 \) are standard Gaussian random variables and \( W, N_1, N_2 \) are mutually independent with each other.

Since \( EXY = \rho \), we have
\[
\rho = \rho_1 \rho_2 \sigma^2, \quad \tag{92}
\]
and due to the Markov chain \( X - W - Y \), we have \( H(X|W) = H(X|Y) \), i.e.,
\[
1 - \rho_1^2 = 1 + 2 \rho_1 \rho_2 - \rho_1^2 - \rho_2^2. \tag{93}
\]
Combining (92) and (93), we get \( \sigma^2 = 1 \). Therefore, we can lower bound \( I(X, Y; W) \) by
\[
\begin{align*}
I(X, Y; W) &= h(X, Y) - h(X|W) - h(Y|W), \\
&= \frac{1}{2} \log \frac{1 - \rho^2}{(1 - \rho_1^2)(1 - \rho_2^2)}, \\
&= \frac{1}{2} \log \frac{1}{1 + \rho^2 - \rho_1^2 - \rho_2^2}, \\
&\geq \frac{1}{2} \log \frac{1 - \rho^2}{1 + \rho^2 - 2 \rho}, \\
&= \frac{1}{2} \log \frac{1 + \rho}{1 - \rho},
\end{align*}
\]
where we use the facts that \( \rho_1 \rho_2 = \rho \) and \( \rho_1^2 + \rho_2^2 \geq 2 \rho_1 \rho_2 \).

**REFERENCES**


