Extended Gray–Wyner System with Complementary Causal Side Information

Cheuk Ting Li and Abbas El Gamal
Department of Electrical Engineering, Stanford University
Email: ctl@stanford.edu, abbas@ee.stanford.edu

Abstract

We establish the rate region of an extended Gray–Wyner system for 2-DMS \((X, Y)\) with two additional decoders having complementary causal side information. This extension is interesting because in addition to the operationally significant extreme points of the Gray–Wyner rate region, which include Wyner’s common information, Gács-Körner common information and information bottleneck, the rate region for the extended system also includes the Körner graph entropy, the privacy funnel and excess functional information, as well as three new quantities of potential interest, as extreme points. To simplify the investigation of the 5-dimensional rate region of the extended Gray–Wyner system, we establish an equivalence of this region to a 3-dimensional mutual information region that consists of the set of all triples of the form \(\{I(X;U), I(Y;U), I(X,Y;U)\}\) for some \(P_{U|X,Y}\).

We further show that projections of this mutual information region yield the rate regions for many settings involving a 2-DMS, including lossless source coding with causal side information, distributed channel synthesis, and lossless source coding with a helper.

Index Terms

Gray–Wyner system, side information, complementary delivery, Körner graph entropy, privacy funnel.

I. INTRODUCTION

The lossless Gray–Wyner system [1] is a multi-terminal source coding setting for two discrete memoryless source (2-DMS) \((X, Y)\) with one encoder and two decoders. This setup draws some of its significance from providing operational interpretation for several information theoretic quantities of interest, namely Wyner’s common information [2], the Gács-Körner common information [3], the necessary conditional entropy [4], and the information bottleneck [5].

In this paper, we consider an extension of the Gray-Wyner system (henceforth called the EGW system), which includes two new individual descriptions and two decoders with causal side information as depicted in Figure 1. The encoder maps sequences from a 2-DMS \((X, Y)\) into five indices \(M_i \in [1 : 2^{nR_i}], i = 0, \ldots, 4\). Decoders 1 and 2 correspond to those of the Gray–Wyner system, that is, decoder 1 recovers \(X^n\) from \((M_0, M_1)\) and decoder 2 recovers \(Y^n\) from \((M_0, M_2)\). At time \(i \in [1 : n]\), decoder 3 recovers \(X_i\) causally from \((M_0, M_3, Y^n)\) and decoder 4 similarly recovers \(Y_i\) causally from \((M_0, M_4, X^n)\).

Note that decoders 3 and 4 correspond to those of the complementary delivery setup studied in [6], [7] and [8] with causal (instead of noncausal) side information and with two additional private indices \(M_3\) and \(M_4\). This extended Gray-Wyner system setup is lossless, that is, the decoders recover their respective source sequences with probability of error that vanishes as \(n\) approaches infinity.

The rate region \(\mathcal{R}\) of the EGW system is defined in the usual way as the closure of the set of achievable rate tuples \((R_0, R_1, R_2, R_3, R_4)\).

The first contribution of this paper is to establish the rate region of the EGW system. Moreover, to simplify the study of this rate region and its extreme points, we show that it is equivalent to the 3-dimensional mutual information region for \((X, Y)\) defined as

\[
\mathcal{I}_{XY} = \bigcup_{P_{U|XY}} \{I(X;U), I(Y;U), I(X,Y;U)\} \subseteq \mathbb{R}^3
\]

in the sense that we can express \(\mathcal{R}\) using \(\mathcal{I}\) and vice versa. As a consequence and of particular interest, the extreme points of the rate region \(\mathcal{R}\) (and its equivalent mutual information region \(\mathcal{I}_{XY}\)) for the EGW system include, in addition to the aforementioned extreme points of the Gray–Wyner system, the Körner graph entropy [8], privacy funnel [9] and excess functional information [10], as well as three new quantities with interesting operational meaning, which we refer to as the maximal interaction information, the asymmetric private interaction information, and the symmetric private interaction information. These extreme points can be cast as maximizations of the interaction information \(\{I(X;Y|U) - I(X;U)\}\) under various constraints. They can be considered as distances from extreme dependency, as they are equal to zero only under certain conditions of extreme dependency. In addition to providing operational interpretations to these information theoretic quantities, projections of the mutual information region yield the rate regions for many settings involving a 2-DMS, including lossless source coding with causal side information [12], distributed channel synthesis [13], [14], and lossless source coding with a helper [15], [16], [17].

A related extension of lossy Gray–Wyner system with two decoders with causal side information was studied by Timo and Vellambi [18]. If we only consider decoders 3 and 4 in EGW, then it can be considered as a special case of their setting.
(where the side information does not need to be complementary). Other related source coding setups to the EGW can be found in [19], [12], [20], [21], [22]. A related 3-dimensional region, called the region of tension, was investigated by Prabhakaran and Prabhakaran [23], [24]. We show that this region can be obtained from the mutual information region, but the other direction does not hold in general.

In the following section, we establish the rate region of the EGW system, relate it to the mutual information region, and show that the region of the original Gray–Wyner system and the region of tension can be obtained from the mutual information region. In Section III, we study the extreme points of the mutual information region. In Section IV, we establish the rate region for the same setup as the EGW system but with noncausal instead of causal side information at decoders 3 and 4. We show that the rate region of the noncausal EGW can be expressed in terms of the Gray–Wyner region, hence it does not contain as many interesting extreme points as the causal EGW. Moreover, we show that this region is equivalent to the closure of the limit of the mutual information region for \((X^n, Y^n)\) as \(n\) approaches infinity.

A. Notation

Throughout this paper, we assume that \(\log\) is base 2 and the entropy \(H\) is in bits. We use the notation: \(X^b_a = (X_a, \ldots, X_b)\), \(X^n = X^n_1\) and \([a:b] = [a, b] \cap \mathbb{Z}\).

For discrete \(X\), we write the probability mass function as \(p_X\). For \(A \subseteq \mathbb{R}^n\), we write the closure of \(A\) as \(\text{cl}(A)\) and the convex hull as \(\text{conv}(A)\). We write the support function as

\[
\psi_A(b) = \sup \{ a^T b : a \in A \}.
\]

We write the one-sided directional derivative of the support function as

\[
\psi'_A(b; c) = \lim_{t \to 0^+} \frac{1}{t} (\psi_A(b + tc) - \psi_A(b)).
\]

Note that if \(A\) is compact and convex, then

\[
\psi'_A(b; c) = \max \left\{ d^T c : d \in \arg \max_{a \in A} a^T b \right\}.
\]

II. RATE REGION OF EGW AND THE MUTUAL INFORMATION REGION

The rate region of the EGW system is given in the following.

**Theorem 1.** The rate region the EGW system \(\mathcal{R}\) is the set of rate tuples \((R_0, R_1, R_2, R_3, R_4)\) such that

\[
R_0 \geq I(X, Y; U),
R_1 \geq H(X|U),
R_2 \geq H(Y|U),
R_3 \geq H(X|Y, U),
R_4 \geq H(Y|X, U)
\]
for some $p_{U|XY}$, where $|U| \leq |X| \cdot |Y| + 2$.

Note that if we ignore decoders 3 and 4, i.e., let $R_3, R_4$ be sufficiently large, then this region reduces to the Gray–Wyner region.

**Proof:** The converse proof is quite straightforward and is given in Appendix A for completion. We now prove the achievability.

**Codebook generation.** Fix $p_{U|XY}$ and randomly and independently generate $2^{nR_0}$ sequences $u^n(m_0), m_0 \in [1 : 2^{nR_0}]$, each according to $\prod_{i=1}^{n} p_U(u_i)$. Given $u^n(m_0)$, assign indices $m_1 \in [1 : 2^{nR_1}], m_2 \in [1 : 2^{nR_2}]$ to the sequences in the conditional typical sets $\mathcal{T}_e(n)(X|u^n(m_0))$ and $\mathcal{T}_e(n)(Y|u^n(m_0))$, respectively. For each $y \in Y, u \in U$, assign indices $m_{3,y,u} \in [1 : 2^{nR_{3,y,u}} p_{Y|U}(y,u)]$ to the sequences in $\mathcal{T}_e^{(n+\epsilon)} p_{Y|U}(y,u)(X|y,u)$, where $\sum_{y,u} R_{3,y,u} p_{Y|U}(y,u) \leq R_3$. Define $m_{4,x,u}$ similarly.

**Encoding.** To encode the sequence $x^n, y^n$, find $m_0$ such that $(u^n(m_0), x^n, y^n) \in \mathcal{T}_e(n)$ is jointly typical, and find indices $m_1, m_2$ of $x^n, y^n$ in $\mathcal{T}_e(n)(X|u^n(m_0))$ and $\mathcal{T}_e(n)(Y|u^n(m_0))$ given $u^n(m_0)$. For each $x, y$, let $x_{y,u}^n$ be the subsequence of $x^n$ where $x_i$ is included if and only if $y_i = y$ and $u_i(m_0) = u$. Note that since $(u^n(m_0), y^n) \in \mathcal{T}_e(n)$, the length of $x_{y,u}^n$ is not greater than $n(1+\epsilon) p_{Y|U}(y,u)$. We then find an index $m_{3,y,u}$ of $x_{y,u}^{n(1+\epsilon)} p_{Y|U}(y,u) \in \mathcal{T}_e^{(n+\epsilon)} p_{Y|U}(y,u)(X|y,u)$ such that $x_{y,u}^{n}$ is a prefix of $x_{y,u}^{n(1+\epsilon)} p_{Y|U}(y,u)$, and output $m_3$ as the concatenation of $m_{3,y,u}$ for all $y, u$. Similar for $m_4$.

**Decoding.** Decoder 1 outputs the sequence corresponding to the index $m_1$ in $\mathcal{T}_e(n)(X|u^n(m_0))$. Decoder 2 performs similarly using $(m_0, m_2)$. Decoder 3, upon observing $y_i$, finds the sequence $\hat{x}_{y,u}^{n(1+\epsilon)} p_{Y|U}(y,u)$ at the index $m_{3,y,u}(m_0)$ in $\mathcal{T}_e^{(n+\epsilon)} p_{Y|U}(y,u)(X|y,u)$, and output the next symbol in the sequence that is not previously used. Decoder 4 performs similarly using $(m_0, m_4)$.

**Analysis of the probability of error.** By the covering lemma, the probability that there does not exist $m_0$ such that $(u^n(m_0), x^n, y^n) \in \mathcal{T}_e(n)$ tends to 0 if $R_0 > I(X;Y;U)$. Also $|\mathcal{T}_e(n)(X|u^n(m_0))| \leq 2^{nR_1}$ for large $n$ if $R_1 > H(X|U) + \delta(\epsilon)$ (similar for $R_2 > H(Y|U) + \delta(\epsilon)$). Note that $(u^n(m_0), x^n, y^n) \in \mathcal{T}_e(n)$ implies

$$\frac{|\{i : x_i = x, y_i = y, u_i(m_0) = u\}|}{n(1+\epsilon) p_{Y|U}(y,u)} \leq \frac{(1+\epsilon) p_{Y|U}(y,u)}{n(1+\epsilon)} \leq p_{X|YU}(x|y,u)$$

for all $(y,u)$. Hence there exists $\hat{x}_{y,u}^{n(1+\epsilon)} p_{Y|U}(y,u) \in \mathcal{T}_e^{(n+\epsilon)} p_{Y|U}(y,u)(X|y,u)$ such that $x_{y,u}^{n}$ is a prefix of $\hat{x}_{y,u}^{n(1+\epsilon)} p_{Y|U}(y,u)$. And $|\mathcal{T}_e^{(n+\epsilon)} p_{Y|U}(y,u)(X|y,u)| \leq 2^{nR_{3,y,u}} p_{Y|U}(y,u)$ for large $n$ if $R_{3,y,u} > (1+\epsilon) H(X|Y = y, U = u) + \delta(\epsilon)$. Hence we can assign suitable $R_{3,y,u}$ for each $y, u$ if $R_3 > (1+\epsilon) H(X|Y;U) + \delta(\epsilon)$.

Although $\mathcal{R}$ is 5-dimensional, the bounds on the rates can be expressed in terms of three quantities: $I(X;U), I(Y;U)$ and $I(X,Y;U)$ together with other constant quantities that involve only the given $(X,Y)$. This leads to the following equivalence of $\mathcal{R}$ to the mutual information region $\mathcal{I}_{XY}$ defined in (1). We denote the components of a vector $v \in \mathcal{I}_{XY}$ by $v = (v_X, v_Y, v_{XY})$.

**Proposition 1.** The rate region for the EGW system can be expressed as

$$\mathcal{R} = \bigcup_{v \in \mathcal{I}_{XY}} \left\{ (v_X, H(X) - v_X, H(Y) - v_Y, H(X|Y) - v_{XY} + v_Y, H(Y|X) - v_{XY} + v_X) \right\} + [0,\infty)^5,$$

(2)

where the last “+” denotes the Minkowski sum. Moreover, the mutual information region for $(X,Y)$ can be expressed as

$$\mathcal{I}_{XY} = \left\{ v \in \mathbb{R}^3 : (v_{XY}, H(X) - v_X, H(Y) - v_Y, H(X|Y) - v_{XY} + v_Y, H(Y|X) - v_{XY} + v_X) \in \mathcal{R} \right\}.$$

(3)

**Proof:** Note that (2) follows from the definitions of $\mathcal{R}$ and $\mathcal{I}_{XY}$. We now prove (3). The $\subseteq$ direction follows from (2). For the $\supseteq$ direction, let $v \in \mathbb{R}^3$ satisfy

$$(v_{XY}, H(X) - v_X, H(Y) - v_Y, H(X|Y) - v_{XY} + v_Y, H(Y|X) - v_{XY} + v_X) \in \mathcal{R}.$$ 

Then by Theorem 1 there exists $U$ such that

$$v_{XY} \geq I(X,Y;U),$$

(4)

$$H(X) - v_X \geq H(X|U),$$

(5)

$$H(Y) - v_Y \geq H(Y|U),$$

(6)

$$H(X|Y) - v_{XY} + v_Y \geq H(X|Y,U),$$

(7)

$$H(Y|X) - v_{XY} + v_X \geq H(Y|X,U).$$

(8)
Adding (4) and (8), we have \( v_X \geq I(X; U) \). Combining this with (3), we have \( v_Y = I(Y; U) \). Substituting this into (7), we have \( v_{XY} \leq I(X, Y; U) \). Combining this with (4), we have \( v_{XY} = I(X, Y; U) \). Hence \( v \in \mathcal{I}_{XY} \). 

In the following we list several properties of \( \mathcal{I}_{XY} \).

**Proposition 2.** The mutual information region \( \mathcal{I}_{XY} \) satisfies:

1) Compactness and convexity. \( \mathcal{I}_{XY} \) is compact and convex.

2) Outer bound. \( \mathcal{I}_{XY} \subseteq \mathcal{I}_{XY}^c \), where \( \mathcal{I}_{XY}^c \) is the set of \( v \) such that

\[
\begin{align*}
    v_X, v_Y &\geq 0, \\
v_X + v_Y - v_{XY} &\leq I(X; Y), \\
0 &\leq v_{XY} - v_Y \\
0 &\leq v_{XY} - v_X \\
H(Y|X) &\leq 0.
\end{align*}
\]

3) Inner bound. \( \mathcal{I}_{XY} \supseteq \mathcal{I}_{XY}^I \), where \( \mathcal{I}_{XY}^I \) is the convex hull of the points \((0, 0, 0), (H(X), I(X; Y), H(X)), (I(X; Y), H(Y), H(Y)), (H(X), H(Y), H(X), H(X), H(Y), H(X) + H(Y)), H(Y|X) \).

Moreover, there exists \( 0 \leq \epsilon_1, \epsilon_2 \leq \log I(X; Y) + 4 \) such that

\[
(0, H(Y|X) - \epsilon_1, H(Y|X)), (H(X|Y) - \epsilon_2, 0, H(X|Y)) \in \mathcal{I}_{XY}.
\]

4) Superadditivity. If \( (X_1, Y_1) \) is independent of \( (X_2, Y_2) \), then

\[
\mathcal{I}_{X_1,Y_1} + \mathcal{I}_{X_2,Y_2} \subseteq \mathcal{I}_{(X_1,X_2),(Y_1,Y_2)},
\]

where + denotes the Minkowski sum. As a result, if \( (X_i, Y_i) \sim p_{XY} \) i.i.d. for \( i = 1, \ldots, n \), \( \mathcal{I}_{XY} \subseteq \mathcal{I}_{XY,n,y^n} \).

5) Data processing. If \( X_2 - X_1 \rightarrow Y_1 - Y_2 \) forms a Markov chain, then for any \( v \in \mathcal{I}_{X_1,Y_1} \), there exists \( w \in \mathcal{I}_{X_2,Y_2} \) such that \( w_X \leq v_X, w_Y \leq v_Y \)

\[
I(X_2; Y_2) - w_X - w_Y + w_{XY} \leq I(X_2; Y_2) - v_X - v_Y + v_{XY}.
\]

6) Cardinality bound.

\[
\mathcal{I}_{XY} = \bigcup_{p_U|X,Y: |I(U|X),|Y|+2} \{(I(X; U), I(Y; U), I(X, Y; U))\}.
\]

7) Relation to Gray–Wyner region and region of tension. The Gray–Wyner region can be obtained from \( \mathcal{I}_{XY} \) as

\[
\mathcal{R}_{GW} = \bigcup_{p_U|X,Y} \{(I(X, Y; U), H(X|U), H(Y|U))\} + [0, \infty)^3
\]

\[
= \bigcup_{v \in \mathcal{I}_{XY}} \{(v_{XY}, H(X) - v_X, H(Y) - v_Y)\} + [0, \infty)^3.
\]

The region of tension can be obtained from \( \mathcal{I}_{XY} \) as

\[
\mathcal{T} = \bigcup_{p_U|X,Y} \{(I(X; U|X), I(X; U|Y), I(X; Y|U))\} + [0, \infty)^3
\]

\[
= \bigcup_{v \in \mathcal{I}_{XY}} \{(v_{XY} - v_X, v_{XY} - v_Y, I(X; Y) - v_X - v_Y + v_{XY})\} + [0, \infty)^3.
\]

The proof of this proposition is given in Appendix [B].

## III. Extreme Points of the Mutual Information Region

Many interesting information-theoretic quantities can be expressed as optimizations over \( \mathcal{I}_{XY} \) (and \( \mathcal{R} \)). Since \( \mathcal{I}_{XY} \) is convex and compact, some of these quantities can be represented in terms of the support function \( \psi_{\mathcal{I}_{XY}}(x) \) and its one-sided directional derivative, which provides a representation of those quantities using at most 6 coordinates. To avoid conflicts and for consistency, we use different notation for some of these quantities from the original literature. We use semicolons, e.g., \( G(X; Y) \), for symmetric quantities, and arrows, e.g., \( G(X \rightarrow Y) \), for asymmetric quantities.

Figures [2] and [3] illustrate the mutual information region \( \mathcal{I}_{XY} \) and its extreme points, and Table [1] lists the extreme points and their corresponding optimization problems and support function representations.

We first consider the extreme points of \( \mathcal{I}_{XY} \) that correspond to previously known quantities.

**Wyner’s common information** [2]

\[
J(X; Y) = \min_{X-U-Y} I(X, Y; U)
\]
Information bottleneck tradeoff
\[ J(X; Y) = \min_{U: H(U|X) = H(U|Y)} H(U) = \max_{U: X \rightarrow U, Y \rightarrow U} I(X; Y; U) \]

\[ K(X; Y) = \max \{ v_{XY} : v \in \mathcal{I}_{XY}, v_X + v_Y - v_{XY} = I(X; Y) \} \]
\[ = \max \{ R_0 : R_0^4 \in \mathcal{R}, R_0 + R_1 + R_2 = H(X, Y) \} \]
\[ = -\psi_0'(1, 1, -1; 0, 0, -1). \]

Gács-Körner common information \[ [3], [25] \]
\[ K(X; Y) = \max \{ v_{XY} : v \in \mathcal{I}_{XY}, v_X = v_Y = v_{XY} \} \]
\[ = \max \{ R_0 : R_0^4 \in \mathcal{R}, R_0 + R_1 = H(X), R_0 + R_2 = H(Y) \} \]
\[ = \psi_0'(1, 1, -2; 0, 0, 1). \]

Körner graph entropy \[ [8], [26] \]
Let \( G_{XY} \) be a graph with a set of vertices \( X \) and edges between confusable symbols upon observing \( Y \), i.e., there is an edge \((x_1, x_2)\) if \( p(x_1, y) \), \( p(x_2, y) > 0 \) for some \( y \). The Körner graph entropy
\[ H_K(G_{XY}, X) = \min_{U: U \rightarrow X-Y; H(U|Y,U) = 0} I(X; U) \]
Necessary conditional entropy source coding with causal side information setting.

In the Gray–Wyner system with causal complementary side information, $H_{K}(G_{XY}, X)$ corresponds to the setting with only decoders 1, 3 and $M_{3} = 0$, and we restrict the sum rate $R_{0} + R_{1} = H(X)$. This is in line with the lossless source coding setting with causal side information [12], where the optimal rate is also given by $H_{K}(G_{XY}, X)$. An intuitive reason of this equality is that $R_{0} + R_{1} = H(X)$ and the recovery requirement of decoder 1 forces $M_{0}$ and $M_{1}$ to contain negligible information outside $X^{n}$, hence the setting is similar to the case in which the encoder has access only to $X^{n}$. This corresponds to lossless source coding with causal side information setting.

**Necessary conditional entropy** [4] (also see $H(Y \mid X)$ in [27], $G(Y \to X)$ in [28], private information in [29] and [30])

$$H(Y \mid X) = \min_{U: H(U|Y)=0, X-U-Y} H(U|X) = \min_{U: X-Y-U, X-U-Y} I(Y; U) - I(X; Y)$$
can be expressed as
\[ H(Y \rightarrow X) = \min \{ v_{XY} : v \in \mathcal{I}_{XY}, \nu_Y = v_{XY}, \nu_X = I(X;Y) \} - I(X;Y) \]
\[ = \min \{ R_0 : R_0^t \in \mathcal{R}, R_0 + R_2 = H(Y), R_1 = H(X|Y) \} \]
\[ = -\psi^t_{\mathcal{I}_{XY}}(1, 2, -2; 1, 0, -1). \]

**Information bottleneck** \[5\]
\[ G_{IB}(t, X \rightarrow Y) = \min_{U: X \rightarrow Y-U, I(X;U) \geq t} I(Y;U) \]
can be expressed as
\[ G_{IB}(t, X \rightarrow Y) = \min \{ v_Y : v \in \mathcal{I}_{XY}, \nu_Y = v_{XY}, \nu_X \geq t \} \]
\[ = \min \{ R_0 : R_0^t \in \mathcal{R}, R_0 + R_2 = H(Y), R_1 \leq H(X) - t \}. \]

Note that the same tradeoff also appears in common randomness extraction on a 2-DMS with one-way communication \[31\], lossless source coding with a helper \[15, 16, 17\], and a quantity studied by Witsenhausen and Wyner \[32\]. It is shown in \[33\] that its slope is given by the chordal slope of the hypercontractivity of Markov operator \[34\]
\[ s^*(Y \rightarrow X) = \sup_{U: X \rightarrow Y-U} \frac{I(Y;U)}{I(X;U)} = \sup \{ v_X/v_Y : v \in \mathcal{I}_{XY}, \nu_Y = v_{XY} \}. \]

**Privacy funnel** \[9\] (also see the rate-privacy function defined in \[29\])
\[ G_{PF}(t, X \rightarrow Y) = \min_{U: X \rightarrow Y-U, I(X;U) \geq t} I(X;U) \]
can be expressed as
\[ G_{PF}(t, X \rightarrow Y) = \min \{ v_X : v \in \mathcal{I}_{XY}, \nu_Y = v_{XY}, \nu_X \geq t \} \]
\[ = \min \{ R_0 + R_4 - H(Y|X) : R_0^t \in \mathcal{R}, R_0 + R_2 = H(Y), R_0 \geq t \}. \]

In particular, the maximum \( R \) for perfect privacy (written as \( g_0(X;Y) \) in \[29\], also see \[35\]) is
\[ G_{R^*}(X \rightarrow Y) = \max \{ t \geq 0 : G_{PF}(t, X \rightarrow Y) = 0 \} \]
\[ = \max \{ \nu_Y : v \in \mathcal{I}_{XY}, \nu_Y = v_{XY}, \nu_X = 0 \} \]
\[ = \max \{ R_0 : R_0^t \in \mathcal{R}, R_0 + R_2 = H(Y), R_0 + R_4 = H(Y|X) \} \]
\[ = -\psi^t_{\mathcal{I}_{XY}}(-1, 1, -1; 0, 1, 0). \]

The optimal privacy-utility coefficient \[35\] is
\[ v^*(X \rightarrow Y) = \inf_{U: X \rightarrow Y-U} \frac{I(X;U)}{I(Y;U)} = \inf \{ v_X/v_Y : v \in \mathcal{I}_{XY}, \nu_Y = v_{XY} \}. \]

**Excess functional information** \[10\]
\[ \Psi(X \rightarrow Y) = \min_{U: X \rightarrow Y-U} H(Y|U) - I(X;Y) \]
is closely related to one-shot channel simulation \[36\] and lossless source coding, and can be expressed as
\[ \Psi(X \rightarrow Y) = H(Y|X) - \max \{ \nu_Y : v \in \mathcal{I}_{XY}, \nu_X = 0 \} \]
\[ = \min \{ R_2 : R_2^t \in \mathcal{R}, R_0 + R_4 = H(Y|X) \} - I(X;Y) \]
\[ = \min \{ R_2 : R_2^t \in \mathcal{R}, R_4 = 0, R_0 = H(Y|X) \} - I(X;Y) \]
\[ = -\psi^t_{\mathcal{I}_{XY}}(-2, 0; 0, 1, -1). \]

In the EGW system, \( \Psi(X \rightarrow Y) \) corresponds to the setting with only decoders 2, 4 and \( M_4 = \emptyset \) (since it is better to allocate the rate to \( R_0 \) instead of \( R_4 \)), and we restrict \( R_0 = H(Y|X) \). The value of \( \Psi(X \rightarrow Y) + I(X;Y) \) is the rate of the additional information \( M_2 \) that decoder 2 needs, in order to compensate the lack of side information compared to decoder 4.

**Minimum communication rate for distributed channel synthesis with common randomness rate** \( t \) \[13, 14\]
\[ C(t, X \rightarrow Y) = \min_{U: X \rightarrow Y-U} \max \{ I(X;U), I(X,Y;U) - t \} \]
can be expressed as
\[ C(t, X \rightarrow Y) = \min \{ \max \{ v_X, v_{XY} - t \} : v \in \mathcal{I}_{XY}, v_X + v_Y = v_{XY} = I(X;Y) \} \]
\[ = \min \{ \max \{ H(X) - R_1, R_0 - t \} : R_0^t \in \mathcal{R}, R_0 + R_1 + R_2 = H(X,Y) \}. \]
A. New information theoretic quantities

We now present three new quantities which arise as extreme points of $\mathcal{I}_{XY}$. These extreme points concern the case in which decoders 3 and 4 are active in the EGW system. Note that they are all maximizations of the interaction information $I(X:Y|U) - I(X;Y)$ under various constraints. They can be considered as distances from extreme dependency, in the sense that they are equal to zero only under certain conditions of extreme dependency.

**Maximal interaction information** is defined as

$$G_{\text{NN1}}(X;Y) = \max_{p_U|XY} I(X;Y|U) - I(X;Y).$$

It can be shown that

$$G_{\text{NN1}}(X;Y) = H(X|Y) + H(Y|X) - \min_{U: H(Y|X,U) = H(Y|U) = 0} I(X;Y|U)$$

$$= \max \{v_{XY} - v_X - v_Y : v \in \mathcal{I}_{XY}\}$$

$$= H(X|Y) + H(Y|X) - \min \{R_0 + R_3 + R_4 : R_0^4 \in \mathcal{R}\}$$

$$= H(X|Y) + H(Y|X) - \min \{R_0 : R_0^4 \in \mathcal{R}, R_3 = R_4 = 0\}$$

$$= \psi_{\mathcal{I}_{XY}}(-1, -1, 1).$$

The maximal interaction information concerns the sum-rate of the EGW system with only decoders 3,4. Note that it is always better to allocate the rates $R_3, R_4$ to $R_0$ instead, hence we can assume $R_3 = R_4 = 0$ (which corresponds to $H(Y|X,U) = H(X|Y,U) = 0$). The quantity $H(X|Y) + H(Y|X) - G_{\text{NN1}}(X;Y)$ is the maximum rate in the lossless causal version of the complementary delivery setup [7].

**Asymmetric private interaction information** is defined as

$$G_{\text{PNI}}(X \rightarrow Y) = \max_{U: U \perp X} I(X;Y|U) - I(X;Y).$$

It can be shown that

$$G_{\text{PNI}}(X \rightarrow Y) = H(Y|X) - \min_{U: U \perp X, H(Y|X,U) = 0} I(Y;U)$$

$$= H(Y|X) - \min \{v_Y : v \in \mathcal{I}_{XY}, v_X = 0, v_{XY} = H(Y|X)\}$$

$$= H(Y|X) - \min \{R_3 : R_0^4 \in \mathcal{R}, R_0 + R_4 = H(Y|X)\}$$

$$= H(Y|X) - \min \{R_3 : R_0^4 \in \mathcal{R}, R_4 = 0, R_0 = H(Y|X)\}$$

$$= \psi_{\mathcal{I}_{XY}}(-1, 0, 0; 0, -1, 1).$$

The asymmetric private interaction information is the opposite of excess functional information defined in [10] in which $I(Y;U)$ is maximized instead. Another operational meaning of $G_{\text{PNI}}$ is the generation of random variables with a privacy constraint. Suppose Alice observes $X$ and wants to generate $Y \sim p_{Y|X}(\cdot|X)$. However, she does not have any private randomness and can only access public randomness $W$, which is also available to Eve. Her goal is to generate $Y$ as a function of $X$ and $W$, while minimizing Eve’s knowledge of $Y$ measured by $I(Y;W)$. The minimum $I(Y;W)$ is $H(Y|X) - G_{\text{PNI}}(X \rightarrow Y)$.

**Symmetric private interaction information** is defined as

$$G_{\text{PPI}}(X:Y) = \max_{U: U \perp X, U \perp Y} I(X;Y|U) - I(X;Y).$$

It can be shown that

$$G_{\text{PPI}}(X:Y) = \max_{U: U \perp X, U \perp Y} I(X;Y|U)$$

$$= \max \{v_{XY} : v \in \mathcal{I}_{XY}, v_X = v_Y = 0\}$$

$$= \max \{R_0 : R_0^4 \in \mathcal{R}, R_0 + R_4 = H(X|Y), R_0 + R_4 = H(Y|X)\}$$

$$= \psi_{\mathcal{I}_{XY}}(-1, -1, 0; 0, 0, 1).$$

Intuitively, $G_{\text{PPI}}$ captures the maximum amount of information one can disclose about $(X,Y)$, such that an eavesdropper who only has one of $X$ or $Y$ would know nothing about the disclosed information. Another operational meaning of $G_{\text{PNI}}$ is the generation of random variables with a privacy constraint (similar to that for $G_{\text{PNI}}$). Suppose Alice observes $X$ and wants to generate $Y \sim p_{Y|X}(\cdot|X)$. She has access to public randomness $W$, which is also available to Eve. She also has access to private randomness. Her goal is to generate $Y$ using $X$, $W$ and her private randomness such that Eve has no knowledge on $Y$ (i.e., $I(Y;W) = 0$), while minimizing the amount of private randomness used measured by $H(Y|X,W)$ (note that if Alice
| Active decoders in EGW | Information quantity | Objective and constraints in EGW | Support fcn. rep. $(\psi = \psi_{X|Y})$ |
|-----------------------|----------------------|---------------------------------|---------------------------------|
| 1, 2                  | Wyner’s CI [2]       | min $R_0 : R_0 + R_1 + R_2 = H(X, Y)$ | $\psi'(1, 1, -1; 0, 0, -1)$ |
|                      | Gács-Körner CI [3], [23] | max $R_0 : R_0 + R_1 = H(X), R_0 + R_2 = H(Y)$ | $\psi'(1, 1, -2; 0, 0, 1)$ |
|                      | Necessary conditional entropy [4], [27] | min $R_0 : R_0 + R_2 = H(Y), R_1 = H(X|Y)$ | $\psi'(1, 2, -2; 1, 0, -1)$ |
|                      | Info. bottleneck [5] | min $R_0 : R_0 + R_2 = H(Y), R_1 \leq H(X) - t$ | none |
|                      | Comm. rate for channel synthesis [13], [14] | min $\max\{H(X) - R_1, R_0 - t\} : R_0 + R_1 + R_2 = H(X, Y)$ | none |
| 1, 3 or 2, 4          | Körner graph entropy [5] | min $R_0 : R_0 + R_1 = H(X), R_3 = 0$ | $\psi'(1, -1, 0; -1, 0, 0)$ |
|                      | Excess functional info. [19] | min $R_2 - I(X; Y) : R_4 = 0, R_0 = H(Y|X)$ | $\psi'(-2, 0, 1; 0, 1, -1)$ |
|                      | Max. rate for perfect privacy [9], [29] | max $R_0 : R_0 + R_2 = H(Y), R_0 + R_4 = H(Y|X)$ | $\psi'(-1, 1, -1; 0, 1, 0)$ |
|                      | Privacy funnel [9] | min $R_0 : R_0 + R_4 = H(Y|X) : R_0 + R_2 = H(Y|X), R_0 \geq t$ | none |
| 3, 4                  | Maximal interaction info. | max $H(Y|X) + H(Y|X) - R_0 : R_3 = 0$ | $\psi'(-1, -1)$ |
|                      | Asymm. private interaction info. | max $H(X|Y) - R_3 : R_4 = 0, R_0 = H(Y|X)$ | $\psi'(-1, 0, 0; 0, -1, 1)$ |
|                      | Symm. private interaction info. | max $R_0 : R_0 + R_4 = H(Y|X), R_0 + R_4 = H(Y|X)$ | $\psi'(-1, -1; 0, 0, 1)$ |

Table I
EXTREME POINTS OF $\mathcal{F}_{X|Y}$ AND THE CORRESPONDING EXTREME POINTS IN THE EGW, AND THEIR SUPPORT FUNCTION REPRESENTATIONS.

can flip fair coins for the private randomness, then by Knuth-Yao algorithm [23] the expected number of flips is bounded by $H(Y|X, W = 2)$. The minimum $H(Y|X, W)$ is $H(Y|X) - G_{PPI}(X; Y)$.

We now list several properties of $G_{PNI}, G_{PPI}$ and $G_{PPI}$.

**Proposition 3.** $G_{PNI}, G_{PPI}$ and $G_{PPI}$ satisfies

1) **Bounds.**

$$0 \leq G_{PPI}(X; Y) \leq G_{PNI}(X \rightarrow Y) \leq G_{PNI}(X; Y) \leq \min \{H(X|Y), H(Y|X)\}.$$

2) **Conditions for zero.**

- $G_{PNI}(X; Y) = 0$ if and only if the characteristic bipartite graph of $X, Y$ (i.e. vertices $X \cup Y$ with edge $(x, y)$ if $p(x, y) > 0$) does not contain paths of length 3, or equivalently, $p(x|y) = 1$ or $p(y|x) = 1$ for all $x, y$ such that $p(x, y) > 0$.
- $G_{PNI}(X \rightarrow Y) = 0$ if and only if $G_{PNI}(X; Y) = 0$.
- $G_{PPI}(X; Y) = 0$ if and only if the characteristic bipartite graph of $X, Y$ does not contain cycles.

3) **Condition for maximum.** If $H(X) = H(Y)$, then the following statements are equivalent:

- $G_{PNI}(X; Y) = H(Y|X)$.
- $G_{PNI}(X \rightarrow Y) = H(Y|X)$.
- $G_{PPI}(X; Y) = H(Y|X)$.
- $p(x) = p(y)$ for all $x, y$ such that $p(x, y) > 0$.

4) **Lower bound for independent $X, Y$.** If $X \perp Y$,

$$G_{PPI}(X; Y) \geq E[\log \max\{p(X), p(Y)\}] - 1.$$

5) **Superadditivity.** If $(X_1, Y_1)$ is independent of $(X_2, Y_2)$, then

$$G_{PNI}(X_1, X_2; Y_1, Y_2) \geq G_{PNI}(X_1; Y_1) + G_{PNI}(X_2; Y_2).$$

Similar for $G_{PNI}$ and $G_{PPI}$.

The proof of this proposition is given in Appendix $\Box$.
IV. EXTENDED GRAY–WYNER SYSTEM WITH NONCAUSAL COMPLEMENTARY SIDE INFORMATION

In this section we establish the rate region $\mathcal{R}'$ for the EGW system with complementary noncausal side information at decoders 3 and 4 (noncausal EGW), that is, decoder 3 recovers $X^n$ from $(M_0, M_3, X^n)$ and decoder 4 similarly recovers $Y^n$ from $(M_0, M_4, X^n)$. We show that $\mathcal{R}'$ can be expressed in terms of the Gray-Wyner region $\mathcal{R}_{GW}$, hence it contains fewer interesting extreme points compared to $\mathcal{R}$. This is the reason we emphasized the causal side information in this paper. We further show that $\mathcal{R}'$ is related to the asymptotic mutual information region defined as

$$\mathcal{I}_{X,Y}^\infty = \bigcup_{n=1}^{\infty} \frac{1}{n} \mathcal{I}_{X^n,Y^n},$$

where $(X^n, Y^n)$ is i.i.d. with $(X_1, Y_1) \sim p_{X,Y}$. Note that $\mathcal{I}_{X,Y}^\infty$ may not be closed (unlike $\mathcal{I}_{X,Y}$ which is always closed).

The following gives the rate region for the noncausal EGW.

**Theorem 2.** The optimal rate region $\mathcal{R}'$ for the extended Gray–Wyner system with noncausal complementary side information is the set of rate tuples $(R_0, R_1, R_2, R_3, R_4)$ such that

$$R_0 \geq I(X,Y; U),$$
$$R_1 \geq H(X|U),$$
$$R_2 \geq H(Y|U),$$
$$R_3 \geq H(X|U) - H(Y),$$
$$R_4 \geq H(Y|U) - H(X),$$
$$R_0 + R_2 \geq H(X,Y),$$
$$R_1 + R_4 \geq H(Y|X),$$
$$R_2 + R_3 \geq H(X|U),$$
$$R_0 + R_1 + R_2 + R_3 \geq H(X,Y),$$
$$R_0 + R_1 + R_4 \geq H(X,Y)$$

for some $p_{U|X,Y}$, where $|U| \leq |X| \cdot |Y| + 2$.

The proof is given in Appendix E. Then we characterize the closure of $\mathcal{I}_{X,Y}^\infty$. We show that $\text{cl}(\mathcal{I}_{X,Y}^\infty)$, $\mathcal{R}'$ and the Gray–Wyner region $\mathcal{R}_{GW}$ can be expressed in terms of each other.

**Proposition 4.** The closure of $\mathcal{I}_{X,Y}^\infty$, the rate region $\mathcal{R}'$ for the noncausal EGW and the Gray–Wyner region $\mathcal{R}_{GW}$ satisfy:

1) Characterization of $\text{cl}(\mathcal{I}_{X,Y}^\infty)$.

$$\text{cl}(\mathcal{I}_{X,Y}^\infty) = (\mathcal{I}_{X,Y} + (-\infty, 0] \times (-\infty, 0] \times [0, \infty)) \cap \mathcal{I}_{X,Y}^\infty = (\mathcal{I}_{X,Y} + \{(t, t, t) : t \leq 0\}) \cap ([0, \infty) \times [0, \infty) \times \mathbb{R}).$$

2) Equivalence between $\text{cl}(\mathcal{I}_{X,Y}^\infty)$ and $\mathcal{R}'$.

$$\mathcal{R}' = \bigcup_{v \in \text{cl}(\mathcal{I}_{X,Y}^\infty)} \{(v_{XY}, H(X) - v_X, H(Y) - v_Y, H(X|Y) - v_{XY} + v_Y, H(Y|X) - v_{XY} + v_X) + [0, \infty)^3,$$

and

$$\text{cl}(\mathcal{I}_{X,Y}^\infty) = \{v \in \mathbb{R}^3 : (v_{XY}, H(X) - v_X, H(Y) - v_Y, H(X|Y) - v_{XY} + v_Y, H(Y|X) - v_{XY} + v_X) \in \mathcal{R}'\}.$$

3) Equivalence between $\text{cl}(\mathcal{I}_{X,Y}^\infty)$ and $\mathcal{R}_{GW}$.

$$\mathcal{R}_{GW} = \bigcup_{v \in \text{cl}(\mathcal{I}_{X,Y}^\infty)} \{(v_{XY}, H(X) - v_X, H(Y) - v_Y) + [0, \infty)^3,$$

and

$$\text{cl}(\mathcal{I}_{X,Y}^\infty) = \{v \in \mathcal{I}_{X,Y}^\infty : (v_{XY}, H(X) - v_X, H(Y) - v_Y) \in \mathcal{R}_{GW}\}.$$

The proof is given in Appendix E. Note that Proposition 4 does not characterize $\mathcal{I}_{X,Y}^\infty$ completely since it does not specify which boundary points are in $\mathcal{I}_{X,Y}^\infty$. 
A. Proof of the converse of Theorem 1

To prove the converse, let $U_i = (M_0, X^{i-1}, Y^{i-1})$. Consider

$$nR_0 \geq I(X^n, Y^n; M_0)$$

$$= \sum_{i=1}^n I(X_i, Y_i; M_0 | X^{i-1}, Y^{i-1})$$

$$= \sum_{i=1}^n I(X_i, Y_i; M_0, X^{i-1}, Y^{i-1})$$

$$= \sum_{i=1}^n I(X_i, Y_i; U_i),$$

$$nR_1 \geq H(M_1 | M_0)$$

$$\geq I(X^n; M_1 | M_0)$$

$$= H(X^n | M_0) - H(X^n | M_0, M_1)$$

$$\geq \sum_{i=1}^n H(X_i | M_0, X^{i-1}, Y^{i-1}) - H(X^n | M_0, M_1)$$

$$\geq \sum_{i=1}^n H(X_i | M_0, X^{i-1}, Y^{i-1}) - \log |\mathcal{X}| \sum_{i=1}^n \mathbb{P} \left\{ X_i \neq \hat{X}_{3,i} \right\} - 1$$

(9)

$$\geq \sum_{i=1}^n H(X_i | U_i) - o(n),$$

where the last inequality follows by Fano’s inequality. Similarly $nR_2 \geq \sum_i H(Y_i | U_i) - o(n)$. Next, consider

$$nR_3 \geq H(M_3 | M_0)$$

$$\geq I(X^n, Y^n; M_3 | M_0)$$

$$= \sum_{i=1}^n I(X_i, Y_i; M_3 | M_0, X^{i-1}, Y^{i-1})$$

$$\geq \sum_{i=1}^n I(X_i; M_3 | M_0, X^{i-1}, Y^i)$$

$$= \sum_{i=1}^n \left( H(X_i | M_0, X^{i-1}, Y^i) - H(X_i | M_0, M_3, X^{i-1}, Y^i) \right)$$

$$= \sum_{i=1}^n H(X_i | Y_i, U_i) - \sum_{i=1}^n H(X_i | M_0, M_3, X^{i-1}, Y^i)$$

$$\geq \sum_{i=1}^n H(X_i | Y_i, U_i) - \log |\mathcal{X}| \sum_{i=1}^n \mathbb{P} \left\{ X_i \neq \hat{X}_{3,i} \right\} - 1$$

(10)

$$= \sum_{i=1}^n H(X_i | Y_i, U_i) - o(n),$$

where the last inequality follows by Fano’s inequality since $\hat{X}_{3,i}$ is a function of $M_0, M_3, Y^i$. Similarly $nR_4 \geq \sum_i H(Y_i | X_i, U_i) - o(n)$. Hence the point $(R_0 + \epsilon, \ldots, R_4 + \epsilon)$ is in the convex hull of $\mathcal{R}$ for any $\epsilon > 0$. From (2), $\mathcal{R}$ is the increasing hull of an affine transformation of $\mathcal{R}_{XY}$, and thus is convex.

To prove the cardinality bound, we apply Fenchel-Eggleston-Caratheodory theorem [38], [39] on the $(|\mathcal{X}| |\mathcal{Y}| + 2)$-dimensional vectors with entries $H(X | U = u)$, $H(Y | U = u)$, $H(X, Y | U = u)$ and $p(x, y | u)$ for $u \in \{1, \ldots, |U|\}$, $(x, y) \in \{1, \ldots, |\mathcal{X}| \times \{1, \ldots, |\mathcal{Y}| \} \setminus (|\mathcal{X}|, |\mathcal{Y}|)$; see [16], [19].
B. Proof of Proposition 2

1) To see that $\mathcal{F}_{XY}$ is convex, for any $U_0, U_1$ and $\lambda \in [0,1]$, let $Q \sim \text{Bern}(\lambda)$ be independent of $X, Y, U_0, U_1$, and let $U = (Q, U_Q)$. Then $I(X; U) = (1-\lambda)I(X; U_0) + \lambda I(X; U_1)$ (similarly for the other two quantities). Compactness will be proved later.

2) The outer bound follows directly from the properties of entropy and mutual information.

3) For the inner bound, the first 4 points can be obtained by substituting $U = \emptyset, X, Y, (X, Y)$ respectively. For the last point, by the functional representation lemma [10] p. 626], let $V \perp X$ such that $H(Y|X, V) = 0$. Again by the functional representation lemma, let $W \perp (Y, V)$ such that $H(Y|X, V, W) = 0$. Let $U = (V, W)$, then $I(X; Y|U) = I(X; Y) = H(X|Y) - H(X|Y, U) = H(X|Y)$, and

$$I(X, Y; U) = H(X, Y) - H(X|Y) = H(Y|X) - H(Y|X, U) - I(Y; U) = H(Y|X) - H(X|Y, U) = H(X|Y),$$

$$= I(X, Y; U) = I(X, Y; V, W)$$

$$= I(X, Y; V) + I(X, Y; W|V)$$

$$= H(Y|X) + H(X|Y).$$

Hence there exists $t \leq H(Y|X) + H(X|Y)$ such that $(t - H(Y|X), t - H(X|Y), t) \in \mathcal{F}_{XY}$ (by substituting $t = I(X, Y; U)$). Taking convex combination of this point and $(H(X), H(Y), H(X|Y)) \in \mathcal{F}_{XY}$, we have $(H(X), H(Y), H(X|Y) + H(Y|X)) \in \mathcal{F}_{XY}$.

The existence of $0 \leq \epsilon_1 \leq \log I(X; Y) + 4$ such that $(0, H(Y|X) - \epsilon_1, H(Y|X)) \in \mathcal{F}_{XY}$ can be proved by substituting $\epsilon_1 = \varphi(X \to Y)$ and invoking the strong functional representation lemma [10].

4) The superadditivity property can be obtained from considering $U = (U_1, U_2)$, where $(I(X_1; U_1), I(Y_1; U_1), I(X_1, Y_1; U_1)) \in \mathcal{F}_{X_1, Y_1}$. The data processing property can be obtained from considering $U$ where $I(X; U), I(Y; U), I(X, Y; U) \in \mathcal{F}_{X_1, Y_1}$.

5) The cardinality bound can be proved using Fenchel-Eggleston-Carathéodory theorem using the same arguments as in the converse proof of Theorem 1. Compactness follows from the fact that mutual information is a continuous function, and the set of conditional pmfs $p_{U|X,Y}$ with $|U| \leq |X| \cdot |Y| + 2$ is a compact set.

7) The relation to Gray–Wyner region and region of tension follows from the definitions of the regions.

C. Proof of Proposition 3

1) To prove the bound, note that $I(X; Y|U) \leq H(X)$, hence $I(X; Y|U) - I(X; Y) \leq H(X|Y)$. Let $Q$ achieves which connected component the edge $(X, Y)$ lies in. If the bipartite graph does not contain length 3 paths, every connected component is a star, i.e., for each $q$, either $H(X|Q = q) = 0$ or $H(Y|Q = q) = 0$. Then $I(X; Y) = H(Q) + I(X; Y|Q) = H(Q)$, and $I(X; Y|U) = H(Q|U) + I(X; Y|Q, U) = H(Q|U) \leq H(Q)$ for any $U$. Hence $G_{NN1}(X; Y) = G_{PPI}(X; Y) = 0$.

2) We then prove that if there exist a length 3 path in the bipartite graph, then $G_{NN1}(X; Y) \geq G_{PPI}(X; Y) > 0$. Assume $p(x_1, y_1), p(x_1, y_2), p(x_2, y_1) > 0$. Let $U \in \{1, 2\}$,

$$p(u|x, y) = \begin{cases} 1/2 + \epsilon/p(x_1, y_1) & \text{if } (x, y, u) = (x_1, y_1, 1) \\ 1/2 - \epsilon/p(x_1, y_1) & \text{if } (x, y, u) = (x_1, y_1, 2) \\ 1/2 - \epsilon/p(x_1, y_2) & \text{if } (x, y, u) = (x_1, y_2, 1) \\ 1/2 + \epsilon/p(x_1, y_2) & \text{if } (x, y, u) = (x_1, y_2, 2) \\ 1/2 & \text{otherwise,} \end{cases}$$

where $\epsilon > 0$ is small enough such that the above is a valid conditional pmf. One can verify that $U \perp X$. Since $p_{U|XY}(1|x_1, y_1) = 1/2 + \epsilon/p(x_1, y_1) \neq 1/2 = p_{U|XY}(1|x_2, y_1)$, $X$ and $U$ are not conditionally independent given $Y$. Hence $I(X; Y|U) = I(X; Y) > 0$.

We then prove that if $G_{PPI}(X; Y) > 0$, then there exists a cycle in the bipartite graph. Let $U \perp X, U \perp Y$ and $I(X; Y|U) > I(X; Y)$. Since $U$ is not independent of $X, Y$, there exists $x_1, y_1, u$ such that $p(x_1, y_1|u) > p(x_1, y_1)$. Since $\sum_y p(x_1, y|u) = p(x_1|u) = p(x_1) = \sum_y p(x_1, y')$, there exists $y_2 \neq y_1$ such that $p(x_1, y_2|u) < p(x_1, y_2)$. Since $\sum_x p(x, y_2|u) = p(y_2|u) = p(y_2) = \sum_x p(x, y_2)$, there exists $x_2 \neq x_1$ such that $p(x_2, y_2|u) > p(x_2, y_2)$. Continue this process until we return to a visited $x, y$ pair, i.e., $(x_a, y_a) = (x_b, y_b)$ for $a < b$. Then $y_a, x_a, y_{a+1}, x_{a+1}, \ldots, x_{b-1}, y_b$ forms a cycle.

We then prove that if there exist a cycle in the bipartite graph, then $G_{PPI}(X; Y) > 0$. Let $x_1, x_2, x_3, \ldots, x_a, y_{a+1} = y_1$
be a cycle. Let \( U \in \{1, 2\} \),

\[
p(u|x, y) = \begin{cases} 
1/2 + \epsilon/p(x, y) & \text{if } (x, y, u) = (x_i, y_i, 1) \\
1/2 - \epsilon/p(x, y) & \text{if } (x, y, u) = (x_i, y_i, 2) \\
1/2 - \epsilon/p(x, y_i+1) & \text{if } (x, y, u) = (x_i, y_i+1, 1) \\
1/2 + \epsilon/p(x, y_i+1) & \text{if } (x, y, u) = (x_i, y_i+1, 2) \\
1/2 & \text{otherwise},
\end{cases}
\]

where \( \epsilon > 0 \) is small enough such that the above is a valid conditional pmf. One can verify that \( U \perp X \) and \( U \perp Y \).

Since \( p_{U|XY}(1|x_1, y_1) > 1/2 > p_{U|XY}(1|x_1, y_2) \), \( U \) is not independent of \( X, Y \). Hence \( I(X; Y|U) - I(X; Y) = I(X, Y; U) > 0 \).

3) We then prove that if \( H(X) = H(Y) \) and \( p(x) = p(y) \) for all \( x, y \) such that \( p(x, y) > 0 \), then \( G_{\text{NNI}}(X; Y) = H(Y|X) \). Let \( Q \) achieves the Gács-Körner common information, and let \( X_q = \{ x : p(x|q) > 0 \}; Y_q = \{ y : p(y|q) > 0 \}; \), then \( X_q \{ Q = q \} \sim \text{Unif}(X_q) \), \( Y_q \{ Q = q \} \sim \text{Unif}(Y_q) \) and \( |X_q| = |Y_q| \) for all \( q \). Applying Birkhoff-von Neumann theorem on the submatrix of \( p(x, y) \) with rows \( X_q \) and columns \( Y_q \), there exists \( U_q \) such that \( p(x, y|q) = \sum_u p_{U_q}(u)p_{X,Y|U_q}(x,y|u,q) \), \( p_{X|U_q}(x|u,q) = p_{Y|U_q}(y|u,q) = 1/|X_q| \), and \( p_{X,Y|U_q}(x,y|u,q) \in \{0, 1/|X_q|\} \) for all \( x, y, u \). Let \( U = \{ U_q \}_{q \in \mathbb{Q}} \), where \( U_q \) are assumed to be independent across \( q \). Then for any \( x \) and \( u = \{ u_q \} \),

\[
p(x|\{ u_q \}) = p(x, q|\{ u_q \}) = p(q)p(x|u_q, q) = p(q)/|X_q| = p(x),
\]

where \( q = q(x) \) since \( H(Q|X) = 0 \). Hence \( U \perp X \). Similarly \( U \perp Y \). Also since there is only one non-zero in \( p_{X,Y|U_q}(x,y|u,q) \) for different \( x \), we have \( H(X|Y, U) = 0 \). Similarly \( H(Y|X, U) = 0 \). Hence \( I(X; Y|U) - I(X; Y) = I(Y; U|X) - I(Y; U) = H(Y|X) \).

We then prove that if \( H(X) = H(Y) \) and \( G_{\text{NNI}}(X; Y) = H(Y|X) \), then \( p(x) = p(y) \) for all \( x, y \) such that \( p(x, y) > 0 \). Let \( U \) satisfies \( I(X; Y|U) = I(X; Y) + H(Y|X) = H(Y) \), then one can check that \( U \perp X \), \( U \perp Y \), \( H(Y|X, U) = 0 \) and \( H(Y|X, U) = 0 \). For any \( x, y \) such that \( p(x, y) > 0 \), let \( u \) such that \( p(x, y, u) > 0 \), then

\[
p(x) = p(x|u) = p(x|u)p(y|x, u) = p(y|u)p(x|y, u) = p(y).
\]

4) We then prove the lower bound when \( X, Y \) independent. Assume \( X \perp Y \). Assume \( X = \{1, \ldots, |X|\}, Y = \{1, \ldots, |Y|\}, X = F_X^{-1}(V), Y = F_Y^{-1}(W), V, W \sim \text{Unif}(0, 1) \) independent. Let \( U = V + W \mod 1 \), then \( U \perp X, U \perp Y \).

\[
H(Y|X, U) = \sum_x p(x) \int_0^1 H(Y|U = u, X = x)du = \sum_x p(x) \int_0^1 H(Y|W \in ([u - F_X(x), u - F_X(x - 1)] \mod 1))du = \sum_x p(x) \int_0^1 H(Y|W \in ([u, u + p(x)] \mod 1))du = \sum_x p(x) \int_0^1 \sum_y l(P\{Y = y|W \in ([u, u + p(x)] \mod 1)\})du = \sum_x p(x) \int_0^1 \sum_y l(p(x)^{-1}([F_Y(y - 1), F_Y(y)] \cap ([u, u + p(x)] \mod 1))du = \sum_x p(x) \sum_y \int_0^1 l(p(x)^{-1}([0, p(y)] \cap ([u, u + p(x)] \mod 1))du = -H(X) + \sum_{x,y} \int_0^1 l([0, p(y)] \cap ([u, u + p(x)] \mod 1))du
\]
where we write $A \mod 1 = \{a \mod 1 : a \in A\}$ and $|A|$ for the Lebesgue measure for $A \subseteq \mathbb{R}$, $l(t) = -t \log t$. Consider

$$f(a, b) = \int_0^1 l \left( |0, b| \cap ([u, u + a] \mod 1) \right) du.$$ 

If $b \leq a \leq 1$ and $a + b \leq 1$,

$$f(a, b) = (a - b)l(b) + 2 \int_0^b l(u) du \leq (a - b)l(b) + 2bl(b/2) = a(b) + b^2 = ab \log \frac{1}{b} + b^2 \leq ab \log \frac{1}{b} + ab.$$ 

If $b \leq a$ and $a + b > 1$,

$$f(a, b) = (a - b)l(b) + (a + b - 1)l(a + b - 1) + 2 \int_{b+a-1}^b l(u) du \leq (a - b)l(b) + (a + b - 1)l(a + b - 1) + 2(1 - a)l \left( b - \frac{1-a}{2} \right) \leq (a - b)l(b) + (a + b - 1)l \left( \frac{b^2}{1 + b - a} \right) = (a - b)l(b) + b^2 \log \frac{1 + b - a}{b^2} \leq (a - b)l(b) + 2b \log \frac{2b}{b^2} = al(b) + b^2 \leq ab \log \frac{1}{b} + ab.$$ 

Hence

$$I(X, Y; U) = H(X, Y) - H(Y | U, X) - H(X | U) = H(X, Y) - \sum_{x, y} f(p(x), p(y)) \geq H(X, Y) - \sum_{x, y} \left( p(x)p(y) \log \frac{1}{\min\{p(x), p(y)\}} + p(x)p(y) \right) = \mathbb{E} \left[ \log \frac{1}{\max\{p(X), p(Y)\}} \right] - 1.$$ 

5) The superadditivity property follows from the superadditivity of mutual information region.

D. Proof of Theorem 2

We first prove the achievability. Without loss of generality assume $H(X) \geq H(Y)$. Fix any point $v = (v_X, v_Y, v_{XY}) \in \mathcal{I}_{XY}$. Consider the region

$$\mathcal{J}(v) = ((-\infty, v_X] \times (-\infty, v_Y] \times [v_{XY}, \infty)) \cap \mathcal{I}_{XY}^v.$$
It can be seen from Figure 2 that $\mathcal{F}(v)$ is a subset of the convex hull of the following 9 points:

$$v,$$

$$p_1 = (0, 0, 0),$$

$$p_2 = (H(X), I(X; Y), H(X)),$$

$$p_3 = (I(X; Y), H(Y), H(Y)),$$

$$p_4 = (H(X), H(Y), H(X, Y)),$$

$$p_5 = (H(X|Y), 0, H(X|Y)),$$

$$p_6 = (0, H(Y|X), H(Y|X)),$$

$$p_7 = (0, 0, H(Y|X)),$$

$$p_8 = (H(X) − H(Y), 0, H(X|Y)),$$

i.e., $v$ together with the corner points of $\mathcal{F}_{XY}$ except $(I(X; Y), I(X; Y), I(X; Y))$. We will prove that for any $w = (w_X, w_Y, w_{XY}) \in \mathcal{F}(v)$, the rate tuple $R(w) = (R_0(w), \ldots, R_4(w))$,

$$R_0(w) = w_{XY} + \epsilon,$$

$$R_1(w) = H(X) − w_X + \epsilon,$$

$$R_2(w) = H(Y) − w_Y + \epsilon,$$

$$R_3(w) = H(X|Y) − w_{XY} + w_Y + \epsilon,$$

$$R_4(w) = H(Y|X) − w_{XY} + w_X + \epsilon$$

is achievable in the extended Gray–Wyner system with noncausal complementary side information for $\epsilon > 0$. It suffices to prove the corner points $R(v), R(p_1), \ldots, R(p_8)$ are achievable.

$R(v)$ is achievable using the causal scheme in Theorem 1. To achieve $R(p_1), R(p_2), R(p_3)$ and $R(p_4)$, apply the causal scheme in Theorem 1 on $U \leftarrow \emptyset, U \leftarrow X, U \leftarrow Y$ and $U \leftarrow (X, Y)$, respectively.

To achieve $R(p_5), R(p_6), R(p_7)$, applying the strong functional representation lemma [10], there exists $V_n \subseteq Y^n$ such that $H(X^n|Y^n, V_n) = 0$ and $I(V_n; Y^n|X^n) \leq \epsilon n/2$ for $n$ large enough. We then apply the causal scheme on $X \leftarrow X^n, Y \leftarrow Y^n$ and $U \leftarrow V_n$. Similar for $R(p_8)$.

We now prove the achievability of $R(p_9)$. To generate the codebook, randomly partition $\mathcal{T}^{(n)}_{\epsilon}(X, Y)$ into bins $B_0(m_0)$ of size $2^{n(H(X,Y)+\epsilon/2-R_0)}$ for $m_0 \in [1 : 2^nR_0]$. Further randomly partition the bin $B_0(m_0)$ into $B_3(m_0, m_3)$ of size $2^{n(H(X,Y)+\epsilon/2-R_0-R_3)}$ for $m_3 \in [1 : 2^nR_3]$.

To encode $x^n, y^n$, find $m_0, m_3$ such that $(x^n, y^n) \in B_3(m_0, m_3)$. Directly encode $x^n, y^n$ into $m_1$ and $m_2$ respectively.

Decoder 3 receives $m_0, m_3, y^n$ and output the unique $\hat{x}$ such that $(\hat{x}, y^n) \in B_3(m_0, m_3)$. The probability of error vanishes if $H(Y) > H(X, Y) + \epsilon/2 - R_0 - R_3$, which is guaranteed by the definition of $R(p_7)$. Decoder 4 receives $m_0, x^n$ and output the unique $\hat{y}$ such that $(x^n, \hat{y}) \in B_0(m_0)$. The probability of error vanishes if $H(X) > H(X, Y) + \epsilon/2 - R_0$, which is guaranteed by the definition of $R(p_8)$.

The achievability of $R(p_9)$ is similar to that of $R(p_7)$. To generate the codebook, randomly partition $\mathcal{T}^{(n)}_{\epsilon}(X, Y)$ into bins $B_0(m_0)$ of size $2^{n(H(X,Y)+\epsilon/2-R_0)}$ for $m_0 \in [1 : 2^nR_0]$. Given $m_0$, assign indices $m_1$ to the sequences in the bin $B_0(m_0)$ for $m_1 \in [1 : 2^nR_1]$. This is possible if $R_1 \geq H(X,Y) + \epsilon/2 - R_0$, which is guaranteed by the definition of $R(p_8)$.

To encode $x^n, y^n$, find $m_0$ such that $(x^n, y^n) \in B_0(m_0)$ and find the index $m_1$. Directly encode $y^n$ into $m_2$.

Decoder 1 receives $x^n, m_1$ and output the unique $\hat{y}$ where $(\hat{x}, \hat{y}) \in B_0(m_0)$ with index $m_1$. Decoder 3 receives $m_0, y^n$ and output the unique $\hat{x}$ such that $(\hat{x}, y^n) \in B_0(m_0)$. The probability of error vanishes if $H(Y) > H(X, Y) + \epsilon/2 - R_0$, which is guaranteed by the definition of $R(p_8)$. Decoder 4 receives $m_0, x^n$ and output the unique $\hat{y}$ such that $(x^n, \hat{y}) \in B_0(m_0)$. The probability of error vanishes if $H(X) > H(X, Y) + \epsilon/2 - R_0$, which follows from the definition of $R(p_9)$ and $H(X) \geq H(Y)$.

Hence we have proved that for any point $v \in \mathcal{F}_{XY}$ and

$$w \in \mathcal{F}(v) = ((-\infty,v_X] \times (-\infty,v_Y] \times [v_{XY}, \infty)) \cap \mathcal{F}_{XY},$$

the rate tuple $R(w)$ is achievable. In other words, the region

$$R(\mathcal{F}_{XY} + (-\infty, 0] \times (-\infty, 0] \times [0, \infty)) \cap \mathcal{F}_{XY} + [0, \infty)^5$$
is achievable. The region can be written as

\[ w_{XY} \geq I(X,Y;U), \]
\[ w_X \leq I(X;U), \]
\[ w_Y \leq I(Y;U), \]
\[ w_X \geq 0, \]
\[ w_Y \geq 0, \]
\[ w_{XY} - w_X \leq H(X|Y), \]
\[ w_{XY} - w_Y \leq H(X|Y), \]
\[ R_0 \geq w_{XY} + \epsilon, \]
\[ R_1 \geq H(X) - w_X + \epsilon, \]
\[ R_2 \geq H(Y) - w_Y + \epsilon, \]
\[ R_3 \geq H(X|Y) - w_{XY} + w_Y + \epsilon, \]
\[ R_4 \geq H(Y|X) - w_{XY} + w_X + \epsilon \]

for some \( U, w_X, w_Y, w_{XY} \). The final rate region can be obtained by eliminating \( w_X, w_Y, w_{XY} \) using Fourier-Motzkin elimination.

We then prove the converse. Since decoder 3 observes \( M_0, M_3, Y^n \) and has to recover \( X^n \) with vanishing error probability, \( R_0 + R_3 \geq H(X|Y) \). Similarly \( R_0 + R_4 \geq H(Y|X) \). Note that decoders 2 and 3 together can recover \( X^n, Y^n \) with vanishing error probability (decoder 3 uses the output of decoder 2 as the side information), and hence \( R_0 + R_2 + R_3 \geq H(X,Y) \). Similarly \( R_0 + R_1 + R_4 \geq H(X,Y) \).

Let \( U_i = (M_0, X^{i-1}, Y^{i-1}) \). Using the same arguments in the proof of Theorem 1 we have \( R_0 \geq I(X,Y;U), R_1 \geq H(X|U), R_2 \geq H(Y|U) \).

\[
nR_3 \geq H(M_3|M_0) \\
\geq I(X^n; M_3|M_0) \\
= H(X^n|M_0) - H(X^n|M_0, M_3) \\
= \sum_{i=1}^{n} H(X_i|M_0, X^{i-1}) - H(X^n|M_0, M_3) \\
\geq \sum_{i=1}^{n} H(X_i|M_0, X^{i-1}, Y^{i-1}) - H(Y^n) - H(X^n|M_0, M_3, Y^n) \\
\geq \sum_{i=1}^{n} H(X_i|U_i) - H(Y^n) - o(n),
\]

where the last inequality is due to Fano’s inequality. Similarly \( nR_4 \geq \sum_i H(Y_i|U_i) - H(X^n) - o(n) \).

\[
n(R_2 + R_3) \\
\geq H(M_2, M_3|M_0) \\
\geq I(X^n; M_2, M_3|M_0) \\
= H(X^n|M_0) - H(X^n|M_0, M_2, M_3) \\
= \sum_{i=1}^{n} H(X_i|M_0, X^{i-1}) - H(X^n|M_0, M_2, M_3) \\
\geq \sum_{i=1}^{n} H(X_i|M_0, X^{i-1}, Y^{i-1}) - H(Y^n|M_0, M_2, M_3) - H(X^n|M_0, M_2, M_3, Y^n) \\
\geq \sum_{i=1}^{n} H(X_i|U_i) - o(n),
\]

where the last inequality follows by Fano’s inequality. Similarly \( n(R_1 + R_4) \geq \sum_i H(Y_i|U_i) - o(n) \). Hence the point \((R_0 + \epsilon, \ldots, R_4 + \epsilon)\) is in the convex hull of \( R' \) for any \( \epsilon > 0 \). We have seen in the achievability proof that (for \( \epsilon = 0 \))

\[
R' = R((I_{XY} + (-\infty,0] \times (-\infty,0] \times [0,\infty)) \cap J_{XY}^o) + [0,\infty)^5
\]

is the increasing hull of an affine transformation of a convex set. Therefore \( R' \) is convex.
E. Proof of Proposition 4

1) Since the Gray–Wyner region tensorizes, \( \text{cl}(\mathcal{F}_{XY}^\infty) \subseteq (\mathcal{F}_{XY} + (-\infty, 0] \times (-\infty, 0] \times [0, \infty)) \cap \mathcal{F}_{XY}^\infty \). To prove the other direction, let \( w \in (\mathcal{F}_{XY} + (-\infty, 0] \times (-\infty, 0] \times [0, \infty)) \cap \mathcal{F}_{XY}^\infty \), then by Theorem 2 the following rate tuple is achievable

\[
\begin{align*}
R_0(w) &= w_{XY} + \epsilon, \\
R_1(w) &= H(X) - w_X + \epsilon, \\
R_2(w) &= H(Y) - w_Y + \epsilon, \\
R_3(w) &= H(X|Y) - w_{XY} + w_Y + \epsilon, \\
R_4(w) &= H(Y|X) - w_{XY} + w_X + \epsilon,
\end{align*}
\]

i.e. for the source \( X^l, Y^l \), the probability of error \( P_e(l) \to 0 \) as \( l \to \infty \). Apply this scheme \( n \) times on the source \( X^{nl}, Y^{nl} \). This can be considered as a causal scheme on the source sequence \( (X_1^1, Y_1^1), (X_{l+1}^{2l}, Y_{l+1}^{2l}), \ldots, (X_{(n-1)l+1}^{nl}, Y_{(n-1)l+1}^{nl}) \) with rate tuple \( IR(w) \) and symbol error probability \( P_e(l) \). Hence by (2) and (10) in the proof of Theorem 1

\[
R(l) + \log ((|X| \cdot |Y|)) P_e(l) \cdot 1 \in (1/l)\mathcal{R}(X^l, Y^l).
\]

Let \( \epsilon' = \epsilon + \log (|X| \cdot |Y|) P_e(l) \). Since

\[
\frac{1}{l} \mathcal{R}(X^l, Y^l) = \bigcup_{v \in (1/l)\mathcal{R}^{(X^l, Y^l)}} [v_{XY}, \infty) \times [H(X) - v_X, \infty) \times [H(Y) - v_Y, \infty),
\]

\[
\times [H(X|Y) - v_{XY} + v_Y, \infty) \times [H(Y|X) - v_{XY} + v_X, \infty),
\]

there exists \( v \in (1/l)\mathcal{R}^{(X^l, Y^l)} \subseteq \mathcal{F}_{XY}^\infty \) such that \( v_{XY} \leq w_{XY} + \epsilon' \), \( H(X) - v_X \leq H(X) - w_X + \epsilon' \), and similar for the other 3 dimensions, which implies \( ||v - w|| \leq 2\epsilon' \). The result follows from taking \( l \to \infty \), \( \epsilon \to 0 \).

To show

\[
(\mathcal{F}_{XY} + (-\infty, 0] \times (-\infty, 0] \times [0, \infty)) \cap \mathcal{F}_{XY}^\infty = (\mathcal{F}_{XY} + \{(t, t, t) : t \leq 0\}) \cap ([0, \infty) \times [0, \infty) \times \mathbb{R}),
\]

note that they are both equal to the union of the convex hulls of \( \{v, p_1, \ldots, p_8\} \) for \( v \in \mathcal{F}_{XY} \) (as in the proof of Theorem 2).

2) The equivalence between \( \text{cl}(\mathcal{F}_{XY}^\infty) \) and \( \mathcal{R}' \) is proved in the Fourier-Motzkin elimination step in the proof of Theorem 2.

3) By Proposition 2

\[
\mathcal{R}_{GW} = \bigcup_{v \in \mathcal{F}_{XY}} \{(v_{XY}, H(X) - v_X, H(Y) - v_Y) + [0, \infty) \times [0, \infty) \times [0, \infty) \times [0, \infty) \times \mathbb{R}
\]

For the other direction,

\[
\text{cl}(\mathcal{F}_{XY}^\infty) = (\mathcal{F}_{XY} + (-\infty, 0] \times [0, \infty)) \cap \mathcal{F}_{XY}^\infty
\]

\[
= \{ v \in \mathcal{F}_{XY}^\infty : v_X \leq w_X, v_Y \leq w_Y, v_{XY} \geq w_{XY} \text{ for some } w \in \mathcal{F}_{XY} \}
\]

\[
= \{ v \in \mathcal{F}_{XY}^\infty : v_X \leq I(X; U), v_Y \leq I(Y; U), v_{XY} \geq I(X, Y; U) \text{ for some } U \}
\]

\[
= \{ v \in \mathcal{F}_{XY}^\infty : H(X) - v_X \geq H(X|U), H(Y) - v_Y \geq H(Y|U), v_{XY} \geq I(X, Y; U) \text{ for some } U \}
\]

\[
= \{ v \in \mathcal{F}_{XY}^\infty : (v_{XY}, H(X) - v_X, H(Y) - v_Y) \in \mathcal{R}_{GW} \}
\]

REFERENCES

[1] R. M. Gray and A. D. Wyner, “Source coding for a simple network,” Bell Syst. Tech. J., vol. 53, no. 9, pp. 1681–1721, 1974.
[2] A. D. Wyner, “The common information of two dependent random variables,” IEEE Trans. Inf. Theory, vol. 21, no. 2, pp. 163–179, Mar. 1975.
[3] P. Gács and J. Körner, “Common information is far less than mutual information,” Probl. Control Inform. Theory, vol. 2, no. 2, pp. 149–162, 1973.
[4] P. Cuff, H. Permuter, and T. M. Cover, “Coordination capacity,” IEEE Trans. Info. Theory, vol. 56, no. 9, pp. 4181–4206, 2010.
[5] N. Tishby, F. C. Pereira, and W. Bialek, “The information bottleneck method,” arXiv preprint physics/0004057, 2000.
[6] A. D. Wyner, J. K. Wolf, and F. M. J. Willems, “Communicating via a processing broadcast satellite,” IEEE Transactions on Information Theory, vol. 48, no. 6, pp. 1243–1249, Jun 2002.
