Extended Gray–Wyner System with Complementary Causal Side Information

Cheuk Ting Li and Abbas El Gamal
Department of Electrical Engineering, Stanford University
Email: ctli@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, 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 interpretations 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^i) and decoder 4 similarly recovers Y_i causally from (M_0, M_4, X^i). Note that decoders 3 and 4 correspond to those of the complementary delivery setup studied in [6], [7] 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 $\mathscr{R}$ of the EGW system is defined in the usual way as the closure of the set of achievable rate tuples $(R_0, R_1, R_2, R_3, R_4)$.

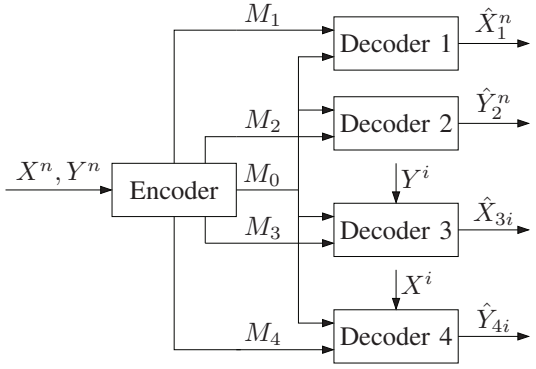

Figure 1. Extended Gray–Wyner system.

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

$$\mathscr{I}_{XY} = \bigcup_{p_{U|XY}} \{(I(X;U),\, I(Y;U),\, I(X,Y;U))\} \subseteq \mathbb{R}^3 \quad (1)$$

in the sense that we can express $\mathscr{R}$ using $\mathscr{I}$ and vice versa. As a consequence and of particular interest, the extreme points of the rate region $\mathscr{R}$ (and its equivalent mutual information region $\mathscr{I}_{XY}$) 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 [11] $I(X;Y|U) - I(X;Y)$ 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

978-1-5090-4096-4/17/$31.00 ©2017 IEEE

information [12], distributed channel synthesis [13], [14], and lossless source coding with a helper [15], [16].

A related extension of lossy Gray–Wyner system with two decoders with causal side information was studied by Timo and Vellambi [17]. 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 [18], [12], [19], [20], [21]. A related 3-dimensional region, called the region of tension, was investigated by Prabhakaran and Prabhakaran [22], [23]. This region can be obtained from the mutual information region, but the other direction does not hold in general.

The omitted proofs and derivations can be found in [24].

A. Notation

We assume log is base 2 and the entropy H is in bits. We write $X_a^b = (X_a, \ldots, X_b)$, $X^n = X_1^n$ and $[a:b] = [a,b] \cap \mathbb{Z}$.

For $A \subseteq \mathbb{R}^n$, we write the closure of A as $\mathrm{cl}(A)$. We write the support function as $\psi_A(b) = \sup\left\{a^T b : a \in A\right\}$. We write the one-sided directional derivative of the support function as

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

Note that if A is compact and convex, then $\psi'_A(b;c)$ is the solution of a constrained optimization problem

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

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 $\mathscr{R}$ is the set of rate tuples $(R_0, R_1, R_2, R_3, R_4)$ such that*

$$\begin{aligned} R_0 &\geq I(X,Y;U), \\ R_1 &\geq H(X|U), \\ R_2 &\geq H(Y|U), \\ R_3 &\geq H(X|Y,U), \\ R_4 &\geq H(Y|X,U) \end{aligned}$$

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

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 can be found in [24]. 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 $\mathcal{T}_\epsilon^{(n)}(X|u^n(m_0))$ and $\mathcal{T}_\epsilon^{(n)}(Y|u^n(m_0))$, respectively. For each $y \in \mathcal{Y}$, $u \in \mathcal{U}$, assign indices $m_{3,y,u} \in [1 : 2^{nR_{3,y,u}p_{YU}(y,u)}]$ to the sequences in $\mathcal{T}_\epsilon^{n(1+\epsilon)p_{YU}(y,u)}(X|y,u)$, where $\sum_{y,u} R_{3,y,u} p_{YU}(y,u) \leq R_3$. Define $m_{4,x,u}$ similarly.

Encoding. To encode the sequence x^n, y^n, find m_0 such that $(u^n(m_0), x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}$ is jointly typical, and find indices m_1, m_2 of x^n, y^n in $\mathcal{T}_\epsilon^{(n)}(X|u^n(m_0))$ and $\mathcal{T}_\epsilon^{(n)}(Y|u^n(m_0))$ 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 $(u^n(m_0), y^n) \in \mathcal{T}_\epsilon^{(n)}$, the length of $x_{y,u}^n$ is not greater than $n(1+\epsilon)p_{YU}(y,u)$. We then find an index $m_{3,y,u}$ of $\hat{x}_{y,u}^{n(1+\epsilon)p_{YU}(y,u)} \in \mathcal{T}_\epsilon^{n(1+\epsilon)p_{YU}(y,u)}(X|y,u)$ such that $x_{y,u}^n$ is a prefix of $\hat{x}_{y,u}^{n(1+\epsilon)p_{YU}(y,u)}$, 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 index m_1 in $\mathcal{T}_\epsilon^{(n)}(X|u^n(m_0))$. Similar for Decoder 2. Decoder 3, upon observing y_i, finds the sequence $\hat{x}_{y_i,u_i(m_0)}^{n(1+\epsilon)p_{YU}(y_i,u_i(m_0))}$ at index $m_{3,y_i,u_i(m_0)}$ in $\mathcal{T}_\epsilon^{n(1+\epsilon)p_{YU}(y_i,u_i(m_0))}(X|y_i,u_i(m_0))$, and output the next symbol in the sequence that is not previously used. Similar for Decoder 4.

The analysis of the probability of error is straightforward and can be found in [24].

■

Although $\mathscr{R}$ 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 $\mathscr{R}$ to the mutual information region $\mathscr{I}_{XY}$ defined in (1). We denote the components of a vector $v \in \mathscr{I}_{XY}$ by $v = (v_X, v_Y, v_{XY})$. The proof can be found in [24].

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

$$\begin{aligned} \mathscr{R} = \bigcup_{v \in \mathscr{I}_{XY}} \big\{\big(&v_{XY},\, H(X) - v_X,\, H(Y) - v_Y, \\ &H(X|Y) - v_{XY} + v_Y,\, H(Y|X) - v_{XY} + v_X\big)\big\} + [0,\infty)^5, \end{aligned} \tag{2}$$

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

$$\begin{aligned} \mathscr{I}_{XY} = \Big\{ v \in \mathbb{R}^3 : \big(&v_{XY},\, H(X) - v_X,\, H(Y) - v_Y, \\ &H(X|Y) - v_{XY} + v_Y,\, H(Y|X) - v_{XY} + v_X\big) \in \mathscr{R} \Big\}. \end{aligned} \tag{3}$$

Some properties about $\mathscr{I}_{XY}$ and its relation to the Gray–Wyner region and the region of tension can be found in [24].

III. Extreme Points of the Mutual Information Region

Many interesting information theoretic quantities can be expressed as optimizations over $\mathscr{I}_{XY}$ (and $\mathscr{R}$). It can be shown that $\mathscr{I}_{XY}$ is convex and compact, hence some of these quantities can be represented in terms of the support function $\psi_{\mathscr{I}_{XY}}(x)$ 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., $G(X \to Y)$, for asymmetric quantities. Figure 2 illustrate the mutual information region $\mathscr{I}_{XY}$ and its extreme points. We first consider the extreme points of $\mathscr{I}_{XY}$ that correspond to previously known quantities.

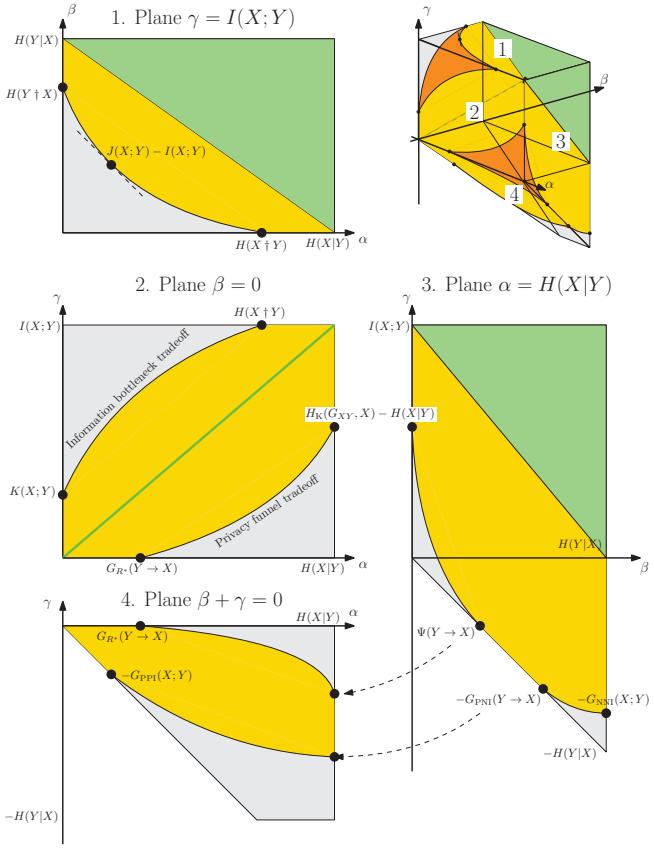

Figure 2. Illustration of $\mathscr{I}_{XY}$ (yellow) restricted to different planes. The green and grey regions are the inner bound $\mathscr{I}^{\mathrm{i}}_{XY}$ and outer bound $\mathscr{I}^{\mathrm{o}}_{XY}$ respectively (see [24]). The axes are $\alpha = I(X;U|Y) = v_{XY} - v_Y$, $\beta = I(Y;U|X) = v_{XY} - v_X$ and $\gamma = v_X + v_Y - v_{XY}$. We assume $H(X) \geq H(Y)$.

Wyner's common information [2]

$$J(X;Y) = \min_{X-U-Y} I(X,Y;U)$$

can be expressed as

$$\begin{aligned}&\min\{v_{XY} : v \in \mathscr{I}_{XY},\, v_X + v_Y - v_{XY} = I(X;Y)\}\\ &= \min\{R_0 : R_0^4 \in \mathscr{R},\, R_0 + R_1 + R_2 = H(X,Y)\}\\ &= -\psi'_{\mathscr{I}_{XY}}(1,1,-1;\, 0,0,-1).\end{aligned}$$

Gács-Körner common information [3], [25]

$$K(X;Y) = \max_{U:\, H(U|X)=H(U|Y)=0} H(U)$$

can be expressed as

$$\begin{aligned}&\max\{v_{XY} : v \in \mathscr{I}_{XY},\, v_X = v_Y = v_{XY}\}\\ &= \max\{R_0 :\, R_0^4 \in \mathscr{R},\, R_0 + R_1 = H(X),\, R_0 + R_2 = H(Y)\}\\ &= \psi'_{\mathscr{I}_{XY}}(1,1,-2;\, 0,0,1).\end{aligned}$$

Körner graph entropy [8], [26]. Let G_{XY} be a graph with a set of vertices $\mathcal{X}$ 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

$$H_{\mathrm{K}}(G_{XY}, X) = \min_{U:\, U-X-Y,\, H(X|Y,U)=0} I(X;U)$$

can be expressed as

$$\begin{aligned}&\min\{v_X :\, v \in \mathscr{I}_{XY},\, v_X = v_{XY},\, v_{XY} - v_Y = H(X|Y)\}\\ &= \min\{R_0 :\, R_0^4 \in \mathscr{R},\, R_0 + R_1 = H(X),\, R_3 = 0\}\\ &= -\psi'_{\mathscr{I}_{XY}}(1,-1,0;\, -1,0,0).\end{aligned}$$

In the Gray–Wyner system with causal complementary side information, $H_{\mathrm{K}}(G_{XY}, X)$ corresponds to the setting with only decoders 1, 3 and $M_3 = \emptyset$, 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 $H_{\mathrm{K}}(G_{XY}, X)$. 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 $H(Y \searrow X|X)$ in [27] and $G(Y \to X)$ in [28])

$$H(Y \dagger X) = \min_{U:\, H(U|Y)=0,\, X-U-Y} H(U|X)$$

can be expressed as

$$\begin{aligned}&\min\{v_{XY} : v \in \mathscr{I}_{XY}, v_Y = v_{XY}, v_X = I(X;Y)\} - I(X;Y)\\ &= \min\{R_0 :\, R_0^4 \in \mathscr{R},\, R_0 + R_2 = H(Y),\, R_1 = H(X|Y)\}\\ &= -\psi'_{\mathscr{I}_{XY}}(1,2,-2;\, 1,0,-1).\end{aligned}$$

Information bottleneck [5]

$$G_{\mathrm{IB}}(t,\, X \to Y) = \min_{U:\, X-Y-U,\, I(X;U)\geq t} I(Y;U)$$

can be expressed as

$$\begin{aligned}&\min\{v_Y :\, v \in \mathscr{I}_{XY},\, v_Y = v_{XY},\, v_X \geq t\}\\ &= \min\{R_0 :\, R_0^4 \in \mathscr{R},\, R_0 + R_2 = H(Y),\, R_1 \leq H(X) - t\}.\end{aligned}$$

Note that the same tradeoff also appears in common randomness extraction on a 2-DMS with one-way communication [29], lossless source coding with a helper [15], [16], and a quantity studied by Witsenhausen and Wyner [30].

Privacy funnel [9] (also see the rate-privacy function in [31])

$$G_{\mathrm{PF}}(t,\, X \to Y) = \min_{U:\, X-Y-U,\, I(Y;U)\geq t} I(X;U)$$

can be expressed as

$$\begin{aligned}&\min\{v_X :\, v \in \mathscr{I}_{XY},\, v_Y = v_{XY},\, v_Y \geq t\}\\ &= \min\{R_0 + R_4 - H(Y|X) :\, R_0^4 \in \mathscr{R},\, R_0 + R_2 = H(Y),\, R_0 \geq t\}.\end{aligned}$$

In particular, the maximum R for perfect privacy (written as $g_0(X;Y)$ in [31], also see [32]) is

$$\begin{aligned}&G_{R^*}(X \to Y) = \max\{t \geq 0 :\, G_{\mathrm{PF}}(t,\, X \to Y) = 0\}\\ &= \max\{v_Y :\, v \in \mathscr{I}_{XY},\, v_Y = v_{XY},\, v_X = 0\}\end{aligned}$$

$$= \max\{R_0 \colon R_0^4 \in \mathscr{R}, R_0+R_2=H(Y), R_0+R_4=H(Y|X)\}$$
$$= \psi'_{\mathscr{I}_{XY}}(-1,1,-1;\,0,1,0).$$

Excess functional information [10]

$$\Psi(X \to Y) = \min_{U\colon U \perp\!\!\!\perp X} H(Y|U) - I(X;Y)$$

is closely related to one-shot channel simulation [33] and lossy source coding, and can be expressed as

$$H(Y|X) - \max\{v_Y : v \in \mathscr{I}_{XY},\, v_X = 0\}$$
$$= \min\{R_2 : R_0^4 \in \mathscr{R},\, R_4 = 0,\, R_0 = H(Y|X)\} - I(X;Y)$$
$$= -\psi'_{\mathscr{I}_{XY}}(-2,0,1;\,0,1,-1).$$

In the EGW system, $\Psi(X \to Y)$ 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 $\Psi(X \to Y) + I(X;Y)$ is the rate of the additional information M_2 that decoder 2 needs, in order to compensate the lack of side information.

Minimum communication rate for distributed channel synthesis with common randomness rate t [13], [14]

$$C(t, X \to Y) = \min_{U\colon X-U-Y} \max\{I(X;U),\, I(X,Y;U) - t\}$$

can be expressed as

$$\min\{\max\{v_X, v_{XY}-t\} \colon v \in \mathscr{I}_{XY}, v_X+v_Y-v_{XY}=I(X;Y)\}$$
$$= \min\{\max\{H(X) - R_1,\, R_0 - t\} :\, R_0^4 \in \mathscr{R},$$
$$R_0 + R_1 + R_2 = H(X,Y)\}.$$

A. New information theoretic quantities

We now present three new quantities which arise as extreme points of $\mathscr{I}_{XY}$. 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.

Maximal interaction information is defined as

$$G_{\mathrm{NNI}}(X;\,Y) = \max_{p_{U|XY}} I(X;Y|U) - I(X;Y).$$

It can be shown that

$$G_{\mathrm{NNI}}(X;\,Y)$$
$$= H(X|Y)+H(Y|X)- \min_{U\colon H(Y|X,U)=H(X|Y,U)=0} I(X,Y;U)$$
$$= H(X|Y)+H(Y|X) - \min\{R_0 :\, R_0^4 \in \mathscr{R},\, R_3 = R_4 = 0\}$$
$$= \max\{v_{XY} - v_X - v_Y :\, v \in \mathscr{I}_{XY}\} = \psi_{\mathscr{I}_{XY}}(-1,-1,1).$$

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 $R_3 = R_4 = 0$. The quantity $H(X|Y) + H(Y|X) - G_{\mathrm{NNI}}(X;\,Y)$ is the maximum rate in the lossless causal version of the complementary delivery setup [7].

Asymmetric private interaction information is defined as

$$G_{\mathrm{PNI}}(X \to Y) = \max_{U\colon U \perp\!\!\!\perp X} I(X;Y|U) - I(X;Y).$$

It can be shown that

$$G_{\mathrm{PNI}}(X \to Y)$$
$$= H(Y|X) - \min_{U\colon U \perp\!\!\!\perp X,\, H(Y|X,U)=0} I(Y;U)$$
$$= H(Y|X) - \min\{v_Y : v \in \mathscr{I}_{XY}, v_X = 0, v_{XY} = H(Y|X)\}$$
$$= H(X|Y) - \min\{R_3 :\, R_0^4 \in \mathscr{R},\, R_4 = 0,\, R_0 = H(Y|X)\}$$
$$= \psi'_{\mathscr{I}_{XY}}(-1,0,0;\,0,-1,1).$$

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_{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 on Y measured by $I(Y;W)$. The minimum $I(Y;W)$ is $H(Y|X) - G_{\mathrm{PNI}}(X \to Y)$.

Symmetric private interaction information is defined as

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

It can be shown that

$$G_{\mathrm{PPI}}(X;\,Y)$$
$$= \max\{v_{XY} :\, v \in \mathscr{I}_{XY},\, v_X = v_Y = 0\}$$
$$= \max\{R_0 \colon R_0^4 \in \mathscr{R}, R_0+R_3=H(X|Y), R_0+R_4=H(Y|X)\}$$
$$= \psi_{\mathscr{I}_{XY}}(-1,-1,1).$$

Intuitively, G_{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_{PNI} is the generation of random variables with a privacy constraint (similar to that for G_{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 can flip fair coins for the private randomness, then by Knuth-Yao algorithm [34] the expected number of flips is bounded by $H(Y|X,W) + 2$). The minimum $H(Y|X,W)$ is $H(Y|X) - G_{\mathrm{PPI}}(X;Y)$.

Some properties of these quantities can be found in [24].

IV. Extended Gray–Wyner system with Noncausal Complementary Side Information

In this section we establish the rate region $\mathscr{R}'$ 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, Y^n) and decoder 4 similarly recovers Y^n from (M_0, M_4, X^n). We show that $\mathscr{R}'$ can be expressed in terms of the Gray-Wyner region $\mathscr{R}_{\mathrm{GW}}$, hence it

contains fewer interesting extreme points compared to $\mathscr{R}$. This is the reason we emphasized the causal side information in this paper. We further show that $\mathscr{R}'$ is related to the *asymptotic mutual information region* defined as

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

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

The following gives the rate region for the noncausal EGW. The proof of this theorem can be found in [24].

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

$$\begin{aligned}
&R_0 \geq I(X,Y;U),\\
&R_1 \geq H(X|U),\ R_2 \geq H(Y|U),\\
&R_3 \geq H(X|U) - H(Y),\ R_4 \geq H(Y|U) - H(X),\\
&R_0 + R_3 \geq H(X|Y), R_0 + R_4 \geq H(Y|X),\\
&R_2 + R_3 \geq H(X|U), R_1 + R_4 \geq H(Y|U),\\
&R_0 + R_2 + R_3 \geq H(X,Y), R_0 + R_1 + R_4 \geq H(X,Y)
\end{aligned}$$

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

The proof that $\mathscr{R}'$ can be expressed in terms of $\mathscr{R}_{\mathrm{GW}}$ can be found in [24]. The intuitive reason is that in both $\mathscr{R}'$ and $\mathscr{R}_{\mathrm{GW}}$, the only quantities involving U appearing in the lower bounds on R_i's are $I(X,Y;U)$, $H(X|U)$ and $H(Y|U)$. Hence they are both about how small these quantities can get. Using Theorem 2, we can characterize the closure of $\mathscr{I}_{XY}^{\infty}$ as

$$\mathrm{cl}(\mathscr{I}_{XY}^{\infty}) = (\mathscr{I}_{XY} + \{(t,t,t) : t \leq 0\}) \cap \big([0,\infty) \times [0,\infty) \times \mathbb{R}\big).$$

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 Inf. 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.

[7] A. Kimura and T. Uyematsu, "Multiterminal source coding with complementary delivery," in *Proc. IEEE Int. Symp. Inf. Theory Appl.*, Seoul, Korea, October 2006, pp. 189–194.

[8] J. Körner, "Coding of an information source having ambiguous alphabet and the entropy of graphs," in *6th Prague conference on information theory*, 1973, pp. 411–425.

[9] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Medard, "From the information bottleneck to the privacy funnel," in *Information Theory Workshop (ITW), 2014 IEEE*, Nov 2014, pp. 501–505.

[10] C. T. Li and A. El Gamal, "Strong functional representation lemma and applications to coding theorems," *arXiv preprint*, 2017. [Online]. Available: http://arxiv.org/abs/1701.02827

[11] W. J. McGill, "Multivariate information transmission," *Psychometrika*, vol. 19, no. 2, pp. 97–116, 1954. [Online]. Available: http://dx.doi.org/10.1007/BF02289159

[12] T. Weissman and A. El Gamal, "Source coding with limited-look-ahead side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5218–5239, Dec. 2006.

[13] P. Cuff, "Distributed channel synthesis," *IEEE Trans. Info. Theory*, vol. 59, no. 11, pp. 7071–7096, 2013.

[14] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, "The quantum reverse shannon theorem and resource tradeoffs for simulating quantum channels," *IEEE Trans. Info. Theory*, vol. 60, no. 5, pp. 2926–2959, May 2014.

[15] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. 21, no. 6, pp. 629–637, 1975.

[16] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 21, no. 3, pp. 294–300, 1975.

[17] R. Timo and B. N. Vellambi, "Two lossy source coding problems with causal side-information," in *2009 IEEE International Symposium on Information Theory*, June 2009, pp. 1040–1044.

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

[19] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Transactions on Information Theory*, vol. 31, no. 6, pp. 727–734, Nov 1985.

[20] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," in *International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings.*, 2004, pp. 364–364.

[21] C. Tian and S. N. Diggavi, "Side-information scalable source coding," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5591–5608, Dec 2008.

[22] V. M. Prabhakaran and M. M. Prabhakaran, "Assisted common information with an application to secure two-party sampling," *IEEE Transactions on Information Theory*, vol. 60, no. 6, pp. 3413–3434, June 2014.

[23] M. M. Prabhakaran and V. M. Prabhakaran, "Tension bounds for information complexity," *arXiv preprint arXiv:1408.6285*, 2014.

[24] C. T. Li and A. El Gamal, "Extended Gray–Wyner system with complementary causal side information," *arXiv preprint*, 2017. [Online]. Available: http://arxiv.org/abs/1701.03207

[25] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM Journal on Applied Mathematics*, vol. 28, no. 1, pp. 100–113, 1975.

[26] N. Alon and A. Orlitsky, "Source coding and graph entropies," *IEEE Transactions on Information Theory*, vol. 42, no. 5, pp. 1329–1339, Sep 1996.

[27] S. Wolf and J. Wullschleger, "New monotones and lower bounds in unconditional two-party computation," *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2792–2797, June 2008.

[28] S. Kamath and V. Anantharam, "A new dual to the Gács-Körner common information defined via the Gray-Wyner system," in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, Sept 2010, pp. 1340–1346.

[29] R. Ahlswede and I. Csiszar, "Common randomness in information theory and cryptography. II. CR capacity," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 225–240, Jan 1998.

[30] H. Witsenhausen and A. Wyner, "A conditional entropy bound for a pair of discrete random variables," *IEEE Transactions on Information Theory*, vol. 21, no. 5, pp. 493–501, Sep 1975.

[31] S. Asoodeh, F. Alajaji, and T. Linder, "Notes on information-theoretic privacy," in *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*, Sept 2014, pp. 1272–1278.

[32] F. P. Calmon, A. Makhdoumi, and M. Medard, "Fundamental limits of perfect privacy," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 1796–1800.

[33] P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan, "The communication complexity of correlation," *IEEE Trans. Info. Theory*, vol. 56, no. 1, pp. 438–449, Jan 2010.

[34] D. E. Knuth and A. C. Yao, "The complexity of nonuniform random number generation," *Algorithms and complexity: new directions and recent results*, pp. 357–428, 1976.