Sufficient Conditions for the Equality of Exact and Wyner Common Information

Badri N. Vellambi, Jörg Kliewer
New Jersey Institute of Technology
Newark, NJ 07102
Email: badri.n.vellambi@ieee.org, jkliewer@njit.edu

***Abstract*—The minimum common randomness required for the approximate and separate generation of a pair of correlated discrete memoryless sources is quantified by Wyner's notion of common information. Recently, Kumar, Li, and El Gamal introduced the notion of *exact common information* as the minimum common randomness required for the *exact* and separate generation of a pair of correlated discrete memoryless sources. This new notion of common information, which does not have a general single-letter characterization, was shown to match Wyner's notion for the symmetric binary erasure source. In this work, we present two conditions on the joint statistics of the pair of sources under either of which the exact and Wyner's notions of common information coincide. Though the conditions are implicit, we prove the equality of Wyner and exact common information for the generalized binary Z-source, generalized erasure source and the noisy typewriter source by establishing that these sources meet either of these conditions.**

***Index Terms*—Distributed source generation, Wyner common information, exact common information, channel resolvability, letter typicality.**

I. Introduction

One of the fundamental issues in source coding is to quantify the information common to two (or more) random variables. Significant research efforts during the last sixty years have uncovered a great deal of knowledge in answering this question, and in particular, one observation has become evident. The answer to this fundamental question is *subjective*, and depends on the actual context and/or application involved, partly because many notions of information exist in information theory.

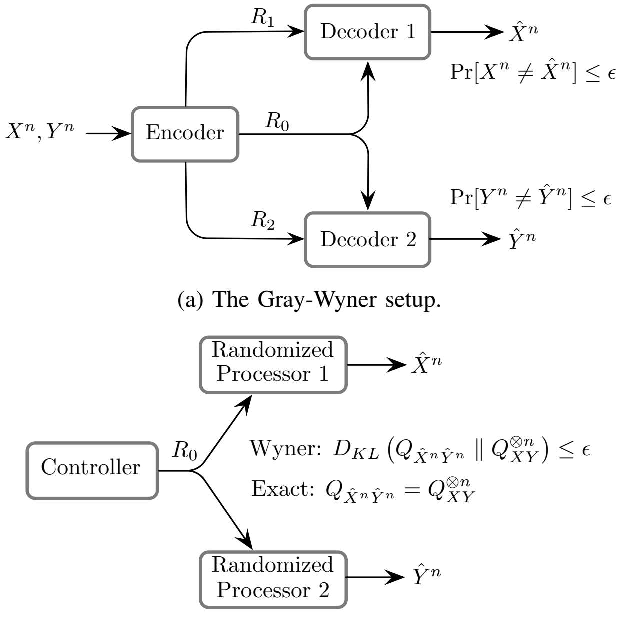
The primary and most ubiquitous of all notions is that of mutual information, which quantifies the reduction in the entropy of a random variable due to the knowledge of a correlated random variable. Gács and Körner formulated a notion of common information of two DMSs as the rate of randomness that can be simultaneously extracted from either of the two correlated sources [1]. While it was hoped that this notion will agree with the notion of mutual information, it was proven that the Gács-Körner common information between a pair of sources is more restrictive than and distinct from mutual information. Despite being restrictive, Gács-Körner common information plays a critical role in the optimal or best-known schemes for several multi-user source coding problems (see e.g. [2]–[4]).

This work builds a deeper understanding between the problem of distributed exact generation of a pair of correlated sources and another notion of common information introduced by Wyner. The origin of Wyner common information lies in part in the Gray-Wyner problem [5]. As depicted in Fig. 1a, the Gray-Wyner problem corresponds to the characterization of the rates of communication required to communicate a pair of correlated sources to two receivers with each requiring one of the sources. The rate region is characterized in [5] to be

$$\mathscr{R} \triangleq \mathsf{cl}\left[\cup \left\{ (R_0, R_1, R_2) : \begin{smallmatrix} R_0 \geq I(X,Y;W) \\ R_1 \geq H(X|W) \\ R_2 \geq H(Y|W) \end{smallmatrix} \right\} \right], \qquad (1)$$

where cl denotes the topological closure, and the union is over all joint probability mass functions (pmfs) Q_{XYW} such that their marginal Q_{XY} equal the pmf of the pair of sources. A particularly interesting operating point in the Gray-Wyner rate region is the one with the least common channel rate on the Pangloss plane $R_0 + R_1 + R_2 = H(X, Y)$, i.e., the least rate conveyed on the common channel so as to be sum-rate optimal. This rate was characterized to be

$$\mathscr{W}(X;Y) \triangleq \min_{X \leftrightarrow W \leftrightarrow Y} I(X,Y;W). \qquad (2)$$


R_1
Decoder 1
$\hat{X}^n$
$\Pr[X^n \neq \hat{X}^n] \leq \epsilon$
X^n, Y^n
Encoder
R_0
$\Pr[Y^n \neq \hat{Y}^n] \leq \epsilon$
R_2
Decoder 2
$\hat{Y}^n$
Randomized Processor 1
$\hat{X}^n$
Controller
R_0
Wyner: $D_{KL}\left(Q_{\hat{X}^n\hat{Y}^n} \,\|\, Q_{XY}^{\otimes n}\right) \leq \epsilon$
Exact: $Q_{\hat{X}^n\hat{Y}^n} = Q_{XY}^{\otimes n}$
Randomized Processor 2
$\hat{Y}^n$

(a) The Gray-Wyner setup.

(b) Separate generation of correlated sources.

Fig. 1: Two setups for defining Wyner common information.

In [6], Wyner presented yet another operational interpretation for the quantity $\mathscr{W}(X;Y)$ – commonly known as the Wyner common information (between X and Y). In [6], Wyner considered the setup in Fig. 1b to show that $\mathscr{W}(X;Y)$ is the smallest rate of a uniform random seed that must be supplied to two independent processors in order for them to *approximately* generate the two DMSs X and Y separately. The Kullback-Leibler divergence was the metric chosen for quantifying the precision of the approximation of the generated sources to the design distribution of the correlated discrete memoryless sources. Owing to this operational interpretation, Wyner common information and the achievability scheme in [6] play a central role in strong coordination problems, which can be broadly termed as problems of generating distributed correlated sources with a specified distribution (see e.g. [7]–[10]). Other connections of Wyner common information to lossy reconstruction problems, and an extension to multiple random variables were explored in [11, 12]. It must be remarked here that despite the innocuous-looking formulation in (2), the computation of Wyner common information is known only for a few joint pmfs [13].

Recently, in [14], Kumar et al. proposed the notion of exact common information using the setup in Fig. 1b with the exact reconstruction requirement to match the distribution $Q_{\hat{X}^n,\hat{Y}^n}$ precisely to $Q_{XY}^{\otimes n}$ – the n-fold product of Q_{XY}, which is the pmf of n i.i.d. RVs each distributed according to Q_{XY}. In other words, exact common information is the smallest rate of a common message that must be shared by two processors to *separately* generate DMSs Q_X and Q_Y correlated jointly according to Q_{XY}.

In [14], the authors derived the fundamental properties of exact common information and proved that for the symmetric binary erasure source, the exact common information rate matches the Wyner common information rate. Since the reconstruction requirement is more stringent than the one Wyner imposed in [6], the exact common information between two random variables is lower bounded by Wyner common information. However, it is unknown if this inequality is strict.

In this work, we present two sufficient conditions under which the notions of exact and Wyner common information coincide. The conditions are implicit in the sense that they depend on the auxiliary random variable W that results from the optimization of Wyner common information in (2). We then present a few examples of pairs of sources, namely the generalized binary Z-source, the generalized erasure source and the noisy typewriter source, that satisfy these conditions.

II. Notation

Given a finite set S, $\mathsf{unif}(S)$ denotes the uniform pmf on S, and $\mathbb{1}_S$ denotes the indicator function on S . For a vector $a^n \in \mathcal{A}^n$ and $\tilde{a} \in \mathcal{A}$, $\#_{a^n}(\tilde{a}) \triangleq \{i : a_i = \tilde{a}\}$. Given a joint pmf Q_{AB} over an alphabet $\mathcal{A} \times \mathcal{B}$, we define the following letter-typical sets.

$$T_\varepsilon^n[Q_A] \triangleq \left\{ a^n : \sup_{\tilde{a}\in\mathsf{S}(Q_A)} \left| \frac{\#_{a^n}(\tilde{a})}{nQ_A(a)} - 1 \right| \le \varepsilon \right\},$$

$$T_\varepsilon^n[Q_{A|B}; b^n] \triangleq \left\{ a^n : \sup_{\substack{(\tilde{a},\tilde{b})\in\mathsf{S}(Q_{AB}) \\ \#_{b^n}(\tilde{b})>0}} \left| \frac{\frac{\#_{a^n,b^n}(\tilde{a},\tilde{b})}{\#_{b^n}(\tilde{b})}}{Q_{A|B}(\tilde{a}|\tilde{b})} - 1 \right| \le \varepsilon \right\}.$$

For $\ell \in \mathbb{N}$ and $1 \le i \le \ell$, $\mathsf{e}_{i,\ell}$ denotes a unit column vector of length ℓ with a single 1 at the i^{th} component. Given random variables A, B, C, we let $A \leftrightarrow B \leftrightarrow C$ to indicate the conditional independence of A and C given B. The support of a random variable X is denoted by $\mathsf{S}(X)$. For two $m\times n$ matrices A, B, we let $A \preceq B$ if $A_{ij} \le B_{ij}$ for all $i = 1, \dots, m$ and $j = 1, \dots, n$. Lastly, $\mathtt{h}(x) \triangleq -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function.

The following remarks can be now made.

Remark 1: For any $0 < \varepsilon < 1$, there exists $n_0 \in \mathbb{N}$ such that for $n > n_0$ and $A^n \sim Q_A^{\otimes n}$,

$$\mathbb{P}[A^n \in T_\varepsilon^n[Q_A]] > 1 - \varepsilon. \tag{3}$$

Note that unlike the oft-used definition of conditional (letter) typicality [15], the above definition of $T_\varepsilon^n[Q_{A|B}; b^n]$ does not require b^n to meet any marginal typicality requirement. Consequently, the following seemingly surprising result holds even though the infinitesimal for conditional typicality is smaller than that for marginal typicality for B.

Remark 2: For any $0 < \varepsilon < 1$, there exists $n_0 \in \mathbb{N}$ such that for $n > n_0$, $b^n \in T_\varepsilon^n[Q_B]$, and $A^n \sim \prod_{i=1}^{n} Q_{A|B=b_i}$,

$$\mathbb{P}[A^n \in T_{\frac{\varepsilon}{2}}^n[Q_{A|B}; b^n]] > 1 - \varepsilon. \tag{4}$$

III. Problem Definition and Basic Results

We define the exact common information $\mathscr{E}(X;Y)$ through Fig. 1b with the Kullback-Leibler divergence term equated to zero instead. Given a joint pmf Q_{XY}, we say that exact generation is possible at a rate of R if for every $\varepsilon > 0$, there exists an $n \in \mathbb{N}$, and random variable W_n such that $X^n \leftrightarrow W \leftrightarrow Y^n$ and $H(W_n) \le n(\mathsf{R} + \varepsilon)$. In other words, if W_n is conveyed to two randomized processors, they can individually use the realization of W_n to separately generate X^n and Y^n, respectively, and the reconstructions together will have a joint pmf matching $Q_{XY}^{\otimes n}$. The exact common information is then infimum of all such achievable rates, summarized by the following functional definition.

Definition 1: Given pmf Q_{XY}, the exact common information is defined to be

$$\mathscr{E}(X;Y) \triangleq \lim_{n\to\infty} \left(\inf_{X^n \leftrightarrow W_n \leftrightarrow Y^n} \frac{H(W_n)}{n} \right). \tag{5}$$

Notice that the alphabet size of W_n is allowed to grow with n, and hence, this is not a computable form of exact common information. Several basic properties of $\mathscr{E}(X;Y)$ can be found in [14] where $\overline{G}(X;Y)$ is used to indicate the same quantity. For the sake of completeness, we denote the Wyner common information by the following.

Definition 2: Given pmf Q_{XY}, the Wyner common information is defined to be

$$\mathscr{W}(X;Y) \triangleq \inf_{\substack{X \leftrightarrow W \leftrightarrow Y \\ |\mathcal{W}| \le |\mathcal{X}||\mathcal{Y}|}} I(X,Y;W). \tag{6}$$

The following properties are in order. Since the setups for exact and Wyner common information are identical, and the generation requirement for the former is more stringent, the following ordering of the different notions of common information holds.

Remark 3: [14, Prop. 3] Given $(X,Y) \sim Q_{XY}$, let $\mathscr{G}(X;Y)$ denote the Gács-Körner common information between random variables X and Y. Then,

$$\mathscr{G}(X;Y) \le I(X;Y) \le \mathscr{W}(X;Y) \le \mathscr{E}(X;Y) \le H(X,Y).$$

The next property argues that concatenation increases exact common information.

Lemma 1: Let $(A,B,C,D) \sim Q_{AB}Q_{CD}$, $X=(A,C)$ and $Y=(B,D)$. Then,

$$\begin{aligned} \max\{\mathscr{E}(A;B), \mathscr{E}(C;D), \mathscr{W}(X;Y)\} &\le \mathscr{E}(X;Y) \\ &= \mathscr{E}(A;B) + \mathscr{E}(C;D). \end{aligned}$$

Further, if $\mathscr{E}(A;B) = \mathscr{W}(A;B)$ and $\mathscr{E}(C;D) = \mathscr{W}(C;D)$, then $\mathscr{E}(X;Y) = \mathscr{E}(A;B) + \mathscr{E}(C;D) = \mathscr{W}(X;Y)$.

Proof: The lower bound is trivial since the separate generation of X and Y involves the generation of A and B, and the generation of C and D. For the upper bound, let $\{W_n^{(1)} : n \in \mathbb{N}\}$ and $\{W_n^{(2)} : n \in \mathbb{N}\}$ be sequences of random variables such that $A^n \leftrightarrow W_n^{(1)} \leftrightarrow B^n$ and $C^n \leftrightarrow W_n^{(2)} \leftrightarrow D^n$ for each $n \in \mathbb{N}$, $\frac{H(W_n^{(1)})}{n} \to \mathscr{E}(A;B)$ and $\frac{H(W_n^{(2)})}{n} \to \mathscr{E}(C;D)$. Define for each $n \in \mathbb{N}$, joint pmf

$$Q_{A^nB^nC^nD^nW_n^{(1)}W_n^{(2)}} \triangleq \begin{pmatrix} Q_{W_n^{(1)}}Q_{A^n\mathcal{W}_n^{(1)}}Q_{B^n|W_n^{(1)}} \\ \times Q_{W_n^{(2)}}Q_{C^n\mathcal{W}_n^{(2)}}Q_{D^n|W_n^{(2)}} \end{pmatrix}.$$

Then, the upper bound follows from the following argument,

$$\begin{aligned} \mathscr{E}(X;Y) &= \lim_{n\to\infty}\left(\inf_{X^n\leftrightarrow W_n\leftrightarrow Y^n} \frac{H(W_n)}{n}\right) \\ &\le \lim_{n\to\infty}\left(\frac{H(W_n^{(1)})+H(W_n^{(2)})}{n}\right) \\ &= \mathscr{E}(A;B)+\mathscr{E}(C;D). \qquad (7) \end{aligned}$$

Note that

$$\begin{aligned} &\mathscr{W}(A;B)+\mathscr{W}(C;D) \\ &= \inf_{A\leftrightarrow W\leftrightarrow B} I(A,B;W) + \inf_{C\leftrightarrow W\leftrightarrow D} I(C,D;W) \qquad (8) \\ &\le \inf_{A\leftrightarrow W\leftrightarrow B} I(A,B;W) + \inf_{C\leftrightarrow W\leftrightarrow D} I(C,D;W,A,B) \\ &\overset{(a)}{=} \inf_{A\leftrightarrow W\leftrightarrow B} I(A,B;W) + \inf_{C\leftrightarrow W\leftrightarrow D} I(C,D;W|A,B) \\ &\le \inf_{X\leftrightarrow W\leftrightarrow Y} I(A,B;W) + \inf_{X\leftrightarrow W\leftrightarrow Y} I(C,D;W|A,B) \\ &\le \inf_{X\leftrightarrow W\leftrightarrow Y} \Big(I(A,B;W)+I(C,D;W|A,B)\Big) \\ &\le \inf_{X\leftrightarrow W\leftrightarrow Y} I(X,Y;W) = \mathscr{W}(X;Y), \qquad (9) \end{aligned}$$

where (a) follows from the independence of (A,B) and (C,D). Now, define a joint pmf

$$Q_{A,B,C,D,W_1W_2} = Q_{W_1}Q_{W_2}Q_{A|W_1}Q_{B|W_1}Q_{C|W_2}Q_{D|W_2},$$

where Q_{AW_1B} and Q_{CW_2D} are chosen to attain the infima in (8). Then, we see that

$$\begin{aligned} \mathscr{W}(X;Y) &= \inf_{X\leftrightarrow W\leftrightarrow Y} I(X,Y;W) \\ &\le I(A,B,C,D;W_1,W_2) \\ &= I(A,B;W_1) + I(C,D;W_2) \\ &= \mathscr{W}(A;B) + \mathscr{W}(C;D). \qquad (10) \end{aligned}$$

Hence, it follows that $\mathscr{W}(A;B) + \mathscr{W}(C;D) = \mathscr{W}(X;Y)$. Now, if $\mathscr{E}(A;B) = \mathscr{W}(A;B)$ and $\mathscr{E}(C;D) = \mathscr{W}(C;D)$, then

$$\begin{aligned} \mathscr{E}(A;B)+\mathscr{E}(C;D) &= \mathscr{W}(A;B)+\mathscr{W}(C;D) \\ &= \mathscr{W}(X;Y) \le \mathscr{E}(X;Y). \qquad (11) \end{aligned}$$

The claim then follows from (7). ■

Another property is analogous to the data processing inequality and characterizes the monotonicity of exact common information with respect to channel degradedness.

Remark 4: $\mathscr{E}(X;Y) \le \mathscr{E}(X';Y')$ for any pmf $Q_{XX'YY'}$ such that $X \leftrightarrow X' \leftrightarrow Y' \leftrightarrow Y$

IV. New Results

In this section, we present the two conditions that ensure the equality of exact and Wyner common information.

Theorem 1: Let $(X,Y) \sim Q_{XY}$ be given. Let random variable W be such that $X \leftrightarrow W \leftrightarrow Y$, $I(X,Y;W) = \mathscr{W}(X;Y)$, and $H(W|X,Y) = 0$, then

$$\mathscr{E}(X;Y) = \mathscr{W}(X;Y). \qquad (12)$$

Proof: Note that this theorem is applicable only when W that satisfies $X \leftrightarrow W \leftrightarrow Y$ also meets $I(X,Y;W) = \mathscr{W}(X;Y) = H(W)$, i.e., the Wyner-optimal auxiliary RV W is a function of X, or Y or (X,Y). In this setting, from Definition 1, we see that

$$\begin{aligned} \mathscr{E}(X;Y) &\triangleq \lim_{n\to\infty}\left(\inf_{X^n\leftrightarrow Z\leftrightarrow Y^n} \frac{H(Z)}{n}\right) \overset{(a)}{\le} \frac{H(W^n)}{n} \qquad (13) \\ &= H(W) = \mathscr{W}(X;Y) \overset{(b)}{\le} \mathscr{E}(X;Y), \qquad (14) \end{aligned}$$

where (a) follows by setting $Z = W^n$ and (b) follows from Remark 3. The achievability follows simply by using a variable-length code (such as a Huffman code [16]) to compress the source W^n that, on average, uses no more than $nH(W)+1$ bits. The two sources are then generated using the channels $Q_{X|W}^{\otimes n}$ and $Q_{Y|W}^{\otimes n}$, respectively. ■

Theorem 2: Given $(X,Y) \sim Q_{XY}$, if there exists a random variable W such that $I(X,Y;W) = \mathscr{W}(X;Y)$ and

$$\sum_{w\in\mathcal{W}} H(X|W=w)\cdot H(Y|W=w) = 0, \qquad (15)$$

then $\mathscr{E}(X;Y) = \mathscr{W}(X;Y)$.

Proof: Let $\mathcal{W}_x = \{x : H(X|W=w) > 0\}$ and $\mathcal{W}_y = \{w : H(Y|W=w) > 0\}$. Then, by (15), we must have $\mathcal{W}_x \cap \mathcal{W}_y = \emptyset$. Hence,

$$H(X,Y|W) = \sum_{w\in\mathcal{W}} Q_W(w) H(X,Y|W=w)$$

$$
\begin{aligned}
&\geq \sum_{w \in \mathcal{W}_x \cup \mathcal{W}_y} Q_W(w) H(X,Y|W=w) \\
&\geq \Bigg(\sum_{w \in \mathcal{W}_x} Q_W(w) H(X|W=w) \\
&\qquad + \sum_{w \in \mathcal{W}_y} Q_W(w) H(Y|W=w) \Bigg) \\
&= H(X|W) + H(Y|W) \geq H(X,Y|W) \qquad (16)
\end{aligned}
$$

Hence, it follows that $X \leftrightarrow W \leftrightarrow Y$.

We now devise a two-stage scheme to exactly match the output statistics to $Q^{\otimes n}_{XY}$. The first stage, which does the bulk of approximating the joint pmf is a modification of Wyner's achievability scheme that is also used in the problem of approximating the statistics of a channel output, also known as the channel resolvability problem [7, 17, 18]. Wyner's approach and that in [7] only require the sources to match in the sense of an asymptotically vanishing variational distance (or Kullback-Leibler divergence) constraint. The modification we impose will allow for a *nearly-uniform*-type convergence constraint. The second stage then refines the shortcomings of the first to match the exact generation requirement.

Let auxiliary random variable W and joint pmf Q_{XWY} be such that (a) $I(X,Y;W) = \mathscr{W}(X;Y)$ and (b) (15) is met. To devise a scheme to exactly generate the two sources, we first pick $\varepsilon > 0$, and pick n sufficiently large so that (3) of Remark 1 holds for Q_W, and (4) of Remark 2 holds for both $Q_{X|W}$ and $Q_{Y|W}$. Consequently,

$$
\mathbb{P}\big[W^n \in T^n_\varepsilon[Q_W]\big] > 1 - \varepsilon, \qquad (17)
$$

and for any $w^n \in T^n_\varepsilon[Q_W]$, $X^n \sim Q^{\otimes n}_{X|W}(\cdot|w^n)$ and $Y^n \sim Q^{\otimes n}_{Y|W}(\cdot|w^n)$,

$$
\mathbb{P}\big[X^n \in T^n_{\frac{\varepsilon}{2}}[Q_{X|W};w^n]\big] > 1 - \varepsilon, \qquad (18)
$$

$$
\mathbb{P}\big[Y^n \in T^n_{\frac{\varepsilon}{2}}[Q_{Y|W};w^n]\big] > 1 - \varepsilon. \qquad (19)
$$

Next, suppose that

$$
R \triangleq I(X,Y;W) + 7\varepsilon \log_2 |\mathcal{X}||\mathcal{Y}|. \qquad (20)
$$

Now, let $\mathcal{C} \triangleq \{W^n(i)\}_{i=1}^{2^{nR}}$ be a codebook with each codeword $W^n(i)$, $i = 1, \ldots, 2^{nR}$, generated i.i.d. using $\tilde{Q}^n_W$ defined by

$$
\tilde{Q}^n_W(w^n) = \frac{Q^{\otimes n}_W(w^n) \mathbb{1}_{T^n_\varepsilon[Q_W]}(w^n)}{Q^{\otimes n}_W\left(T^n_\varepsilon[Q_W]\right)}. \qquad (21)
$$

Now, define a channel $\tilde{Q}^n_{X|W}(\cdot|\cdot)$ with input alphabet $T^n_\varepsilon[Q_W]$ and output alphabet $\mathcal{X}^n$ by

$$
\tilde{Q}^n_{X|W}(x^n|w^n) = \frac{Q^{\otimes n}_{X|W}(x^n|w^n) \mathbb{1}_{T^n_{\frac{\varepsilon}{2}}[Q_{X|W};w^n]}(x^n)}{Q^{\otimes n}_{X|W}(T^n_{\frac{\varepsilon}{2}}[Q_{X|W};w^n]|w^n)}, \qquad (22)
$$

and similarly define the channel $\tilde{Q}^n_{Y|W}$ with input alphabet $T^n_\varepsilon[Q_W]$ and output alphabet $\mathcal{Y}^n$ by

$$
\tilde{Q}^n_{Y|W}(y^n|w^n) = \frac{Q^{\otimes n}_{Y|W}(y^n|w^n) \mathbb{1}_{T^n_{\frac{\varepsilon}{2}}[Q_{Y|W};w^n]}(y^n)}{Q^{\otimes n}_{Y|W}(T^n_{\frac{\varepsilon}{2}}[Q_{Y|W};w^n]|w^n)}. \qquad (23)
$$

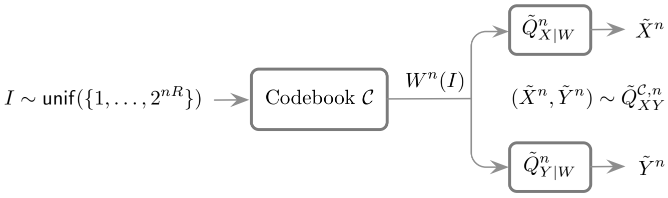

$I \sim \mathsf{unif}(\{1,\ldots,2^{nR}\})$
Codebook $\mathcal{C}$
$W^n(I)$
$\tilde{Q}^n_{X|W}$
$\tilde{X}^n$
$(\tilde{X}^n, \tilde{Y}^n) \sim \tilde{Q}^{\mathcal{C},n}_{XY}$
$\tilde{Q}^n_{Y|W}$
$\tilde{Y}^n$

Fig. 2: The codebook setup for generating $(\tilde{X}^n, \tilde{Y}^n)$.

For each $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$, define 2^{nR} i.i.d. RVs $Z_i(x^n, y^n)$ by

$$
Z_i(x^n, y^n) = \tilde{Q}^n_{X|W}(x^n|W^n(i)) \tilde{Q}^n_{Y|W}(y^n|W^n(i)). \qquad (24)
$$

As a consequence of Lemma 2 of Appendix A and (16)-(19), we have for any $i = 1, \ldots, 2^{nR}$, and $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$,

$$
Z_i(x^n, y^n) \in \left[0, \frac{2^{-n(H(X,Y|W)(1-2\varepsilon)}}{(1-\varepsilon)^2}\right] \qquad (25)
$$

Let $\mu(x^n, y^n) \triangleq \mathbb{E}[Z_i(x^n, y^n)]$. Since we have chosen n appropriately large, we can easily see that

$$
\begin{aligned}
\mu(x^n, y^n) &= \mathbb{E}[Z_i(x^n, y^n)] \\
&= \sum_{w^n} \tilde{Q}^n_W(w^n) \tilde{Q}^n_{X|W}(x^n|w^n) \tilde{Q}^n_{Y|W}(y^n|w^n) \\
&\overset{(17)-(19)}{\leq} \sum_{w^n} \frac{Q^{\otimes n}_W(w^n)}{1-\varepsilon} \frac{Q^{\otimes n}_{X|W}(x^n|w^n)}{1-\varepsilon} \frac{Q^{\otimes n}_{Y|W}(y^n|w^n)}{1-\varepsilon} \\
&= \frac{Q^{\otimes n}_{XY}(x^n, y^n)}{(1-\varepsilon)^3}. \qquad (26)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[Z_i^2(x^n, y^n)] &\overset{(25)}{\leq} \frac{2^{-nH(X,Y|W)(1-2\varepsilon)}}{(1-\varepsilon)^2} \mathbb{E}[Z_i(x^n, y^n)] \qquad (27) \\
&= \frac{2^{-nH(X,Y|W)(1-2\varepsilon)}}{(1-\varepsilon)^2} \mu(x^n, y^n). \qquad (28)
\end{aligned}
$$

From the above, it also follows that

$$
\mathsf{var}(Z_i(x^n, y^n)) \leq \frac{2^{-nH(X,Y|W)(1-2\varepsilon)}}{(1-\varepsilon)^2} \mu(x^n, y^n) \qquad (29)
$$

Now, define a (random) pmf $\tilde{Q}^{\mathcal{C},n}_{XY}$ on $\mathcal{X}^n \times \mathcal{Y}^n$ that is determined by the random codebook $\mathcal{C}$ by

$$
\begin{aligned}
\tilde{Q}^{\mathcal{C},n}_{XY}(x^n, y^n) &\triangleq \sum_{i=1}^{2^{nR}} \frac{Z_i(x^n, y^n)}{2^{nR}} \\
&= \sum_{i=1}^{2^{nR}} \frac{\tilde{Q}^n_{X|W}(x^n|W^n(i)) \tilde{Q}^n_{Y|W}(y^n|W^n(i))}{2^{nR}}.
\end{aligned}
$$

As shown in Fig. 2, the pmf $\tilde{Q}^{\mathcal{C},n}_{XY}$ can be seen to be the output pmf when a codeword from the channel resolvability codebook $\mathcal{C}$ is passed through two parallel channels $\tilde{Q}^n_{X|W}$ and $\tilde{Q}^n_{Y|W}$, respectively. Note that $\tilde{p}^{\mathcal{C},n}_{XY}$ is a weighted sum of i.i.d. RVs, and hence, we can use well-known concentration inequalities to bound tail events of interest. In this work, the inequality of choice will be Bernstein's inequality [19]. We now make the following observation. For any realization of the codebook, all W-codewords lie in $T_{\varepsilon}^{n}[Q_W]$, and for each $w^n \in T_{\varepsilon}^{n}[Q_W]$, the support of the pair of outputs from the combined parallel channels is

$$\begin{aligned}\mathsf{S}\left(\tilde{Q}_{X|W}^{n}(\cdot|w^n)\tilde{Q}_{Y|W}^{n}(\cdot|w^n)\right) &= \begin{pmatrix} T_{\frac{\varepsilon}{2}}^{n}[Q_{X|W};w^n] \\ \times T_{\frac{\varepsilon}{2}}^{n}[Q_{Y|W};w^n]\end{pmatrix} \\ &= T_{\frac{\varepsilon}{2}}^{n}[Q_{XY|W};w^n], \qquad (30)\end{aligned}$$

where the last equality follows from Lemma 3 of Appendix A. Therefore, from (63) of Lemma 2 of Appendix A, for any $\mathcal{C}$,

$$\mathsf{S}(\tilde{Q}_{XY}^{\mathcal{C},n}) \subseteq T_{2\varepsilon}^{n}[Q_{XY}]. \qquad (31)$$

In other words, no matter what the realization of the W-codebook, the support of the (random) code-induced distribution $\tilde{Q}_{XY}^{\mathcal{C},n}$, which itself is a random subset of $\mathcal{X}^n \times \mathcal{Y}^n$ is *always* subset of the typical set $T_{2\varepsilon}^{n}[Q_{XY}]$. Thus, it then follows that $\mu(x^n,y^n)=0$ for any $(x^n,y^n) \notin T_{2\varepsilon}^{n}[Q_{XY}]$, and

$$\mathsf{S}(\mu) = \bigcup_{\mathcal{C}} \mathsf{S}(\tilde{Q}_{XY}^{\mathcal{C},n}) \subseteq T_{2\varepsilon}^{n}[Q_{XY}]. \qquad (32)$$

Now, fix $(x^n,y^n) \in \mathsf{S}(\mu)$ and suppose that $\eta \triangleq \frac{1}{(1-\varepsilon)^3}+\varepsilon$. Then, from (26) and (32), we see that

$$\begin{aligned}\Delta(x^n,y^n) &\triangleq \eta\, Q_{XY}^{\otimes n}(x^n,y^n) - \mu(x^n,y^n) \\ &\geq \varepsilon\, Q_{XY}^{\otimes n}(x^n,y^n) \overset{(32)}{\geq} \varepsilon\, 2^{-nH(XY)(1+2\varepsilon)}, \qquad (33)\\ \Delta_1(x^n,y^n) &\triangleq \tfrac{\eta}{3} Q_{XY}^{\otimes n}(x^n,y^n) + \tfrac{2}{3}\mu(x^n,y^n) \\ &\leq \eta\, Q_{XY}^{\otimes n}(x^n,y^n) \overset{(32)}{\leq} \eta\, 2^{-nH(XY)(1-2\varepsilon)}. \qquad (34)\end{aligned}$$

We proceed with the computation of the following tail event for this choice of (x^n,y^n).

$$\begin{aligned}&\mathbb{P}\left[\left|\tilde{Q}_{XY}^{\mathcal{C},n}(x^n,y^n) - \mu(x^n,y^n)\right| > \Delta(x^n,y^n)\right] \\ &= \mathbb{P}\left[\left|\sum_{i=1}^{2^{nR}} \frac{Z_i(x^n,y^n)}{2^{nR}} - \mu(x^n,y^n)\right| > \Delta(x^n,y^n)\right] \\ &\overset{(25)}{\leq} 2\exp\left[\frac{-\frac{1}{2}\left(2^{nR}\Delta(x^n,y^n)\right)^2 2^{-nR}}{\mathsf{var}(Z_1(x^n,y^n)) + \frac{2^{-nH(X,Y|W)(1-2\varepsilon)}}{(1-\varepsilon)^2}\frac{\Delta(x^n,y^n)}{3}}\right],\end{aligned}$$

where we have used Bernstein's inequality to bound the tail event. Note that the above bound quantifies the probability that a random code $\mathcal{C}$ meets the upper bound for $\tilde{Q}_{XY}^{\mathcal{C},n}$ evaluated at the chosen (x^n,y^n). We can now use the upper bounds and lower bounds of $\Delta(x^n,y^n)$ in (33), (34), and the upper bound for variance from (29) to show that

$$\begin{aligned}&\mathbb{P}\left[\left|\tilde{Q}_{XY}^{\mathcal{C},n}(x^n,y^n) - \mu(x^n,y^n)\right| > \Delta(x^n,y^n)\right] \\ &\overset{(29)}{\leq} 2\exp\left[\frac{-\frac{1}{2}2^{nR}\left(\eta\, Q_{XY}^{\otimes n}(x^n,y^n) - \mu(x^n,y^n)\right)^2}{\frac{2^{-nH(X,Y|W)(1-2\varepsilon)}}{(1-\varepsilon)^2}\frac{\left(\eta\, Q_{XY}^{\otimes n}(x^n,y^n)+2\mu(x^n,y^n)\right)}{3}}\right] \\ &\overset{(33),(34)}{\leq} 2\exp\left[\frac{-\frac{1}{2}\varepsilon^2 2^{nR}2^{-nH(XY)(2+4\varepsilon)}}{\frac{2^{-nH(X,Y|W)(1-2\varepsilon)}}{(1-\varepsilon)^2}\eta\, 2^{-nH(X,Y)(1-2\varepsilon)}}\right] \\ &\leq 2\exp\left[-\frac{(1-\varepsilon)^2\varepsilon^2}{2\eta}2^{n(R-I(W;X,Y)(1+6\varepsilon))}\right] \\ &\overset{(20)}{\leq} 2\exp\left[-\frac{(1-\varepsilon)^2\varepsilon^2}{2\eta}2^{n\varepsilon|\mathcal{X}||\mathcal{Y}|}\right]. \qquad (35)\end{aligned}$$

Note that the above bound holds for any $(x^n,y^n) \in \mathsf{S}(\mu)$. Let $\bigwedge$ denote the logical "and" operator. Then,

$$\begin{aligned}&\mathbb{P}\left[\bigwedge_{(x^n,y^n)\in\mathsf{S}\left(\tilde{Q}_{XY}^{\mathcal{C},n}\right)}\left(\frac{\tilde{Q}_{XY}^{\mathcal{C},n}(x^n,y^n)}{Q_{XY}^{\otimes n}(x^n,y^n)} \leq \eta\right)\right] \\ &\geq \mathbb{P}\left[\bigwedge_{(x^n,y^n)\in\mathsf{S}(\mu)}\left(\frac{\tilde{Q}_{XY}^{\mathcal{C},n}(x^n,y^n)}{Q_{XY}^{\otimes n}(x^n,y^n)} \leq \eta\right)\right] \\ &= \mathbb{P}\left[\bigwedge_{(x^n,y^n)\in\mathsf{S}(\mu)}\left(\frac{\tilde{Q}_{XY}^{\mathcal{C},n}(x^n,y^n) - \mu(x^n,y^n)}{\Delta(x^n,y^n)} \leq 1\right)\right] \\ &\geq \mathbb{P}\left[\bigwedge_{(x^n,y^n)\in\mathsf{S}(\mu)}\left(\frac{\left|\tilde{Q}_{XY}^{\mathcal{C},n}(x^n,y^n) - \mu(x^n,y^n)\right|}{\Delta(x^n,y^n)} \leq 1\right)\right] \\ &\overset{(a)}{\geq} 1 - |\mathsf{S}(\mu)|\left(2\exp\left[-\frac{(1-\varepsilon)^2\varepsilon^2}{2\eta}2^{n\varepsilon|\mathcal{X}||\mathcal{Y}|}\right]\right) \\ &\geq 1 - 2\exp\left[-\frac{(1-\varepsilon)^2\varepsilon^2}{2\eta}2^{n\varepsilon|\mathcal{X}||\mathcal{Y}|} - n\log_e|\mathcal{X}||\mathcal{Y}|\right], \qquad (36)\end{aligned}$$

where (a) follows from (35) and the union bound. Thus, by choosing n large enough, we can ensure that the quantity in (36) is positive. Hence, there exists a realization of the random codebook $\overline{\mathcal{C}} = \{\bar{w}^n(i) : i=1,\ldots,2^{nR}\}$ such that for any $(x^n,y^n) \in \mathsf{S}\big(\tilde{Q}_{XY}^{\overline{\mathcal{C}},n}\big)$,

$$\begin{aligned}\tilde{Q}_{XY}^{\overline{\mathcal{C}},n}(x^n,y^n) &\triangleq \frac{1}{2^{nR}}\sum_{i=1}^{2^{nR}}\tilde{Q}_{X|W}^{n}(x^n|\bar{w}^n(i))\tilde{Q}_{Y|W}^{n}(y^n|\bar{w}^n(i)) \\ &\leq \eta\, Q_{XY}^{\otimes n}(x^n,y^n). \qquad (37)\end{aligned}$$

Since $\tilde{Q}_{XY}^{\overline{\mathcal{C}},n}(x^n,y^n)=0$ if $(x^n,y^n) \notin \mathsf{S}\big(\tilde{Q}_{XY}^{\overline{\mathcal{C}},n}\big)$, the above bound holds for all $(x^n,y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$. Thus, it must be true that

$$r_{XY}^{n} \triangleq \frac{\eta\, Q_{XY}^{\otimes n} - \tilde{Q}_{XY}^{\overline{\mathcal{C}},n}}{\eta - 1}. \qquad (38)$$

defines a pmf over $\mathcal{X}^n \times \mathcal{Y}^n$. We can then rewrite $Q_{XY}^{\otimes n}$ as an convex combination of $\tilde{Q}_{XY}^{\overline{\mathcal{C}},n}$ and r_{XY}^{n} that highlights the two stages for generation of $Q_{XY}^{\otimes n}$.

$$Q_{XY}^{\otimes n} = \frac{1}{\eta}\tilde{Q}_{XY}^{\overline{\mathcal{C}},n} + \left(1-\frac{1}{\eta}\right)r_{XY}^{n}, \qquad (39)$$

In the above, the first term that encompasses all but an infinitesimal portion of the probability measure indicates the first-stage approximation of $Q_{XY}^{\otimes n}$ using a modified Wyner-style codebook. This is followed by the second-stage approximation, which carries an infinitesimal portion of the probability measure, but is indispensable in meeting the exact generation constraint. To generate n copies of the two sources exactly distributed according to Q_{XY}, the controller first generates an instance of a binary random variable V with

$$Q_V(0) \triangleq \left\|Q_{XY}^{\otimes n} - \frac{1}{\eta}\tilde{Q}_{XY}^{\overline{\mathcal{C}},n}\right\|_1 = 1 - \frac{1}{\eta}. \qquad (40)$$

The controller conveys the realization of V to both terminals. If $V = 0$, the controller additionally generates an instance of $(\tilde{X}^n, \tilde{Y}^n) \sim r^n_{XY}$, and conveys $\phi_r(\tilde{X}^n, \tilde{Y}^n)$ to the two terminals, where $\phi_r : \mathcal{X}^n \times \mathcal{Y}^n \to \{1, \ldots, 2^{\lceil n \log_2 |\mathcal{X}||\mathcal{Y}| \rceil}\}$ is a bijective map. Note that conveying the outcome of this bijective map requires no more than $\lceil n \log_2 |\mathcal{X}||\mathcal{Y}| \rceil$ bits. The two terminals then map the bits back to the corresponding instances using ϕ_r^{-1}, and output the respective components. Note that when $V = 0$, each terminal knows exactly the realization of the other source as well.

Now, if $V = 1$, the controller generates nR bits uniformly at random, and conveys it to both terminals. The terminals use the bits to identify the appropriate codeword from $\overline{\mathcal{C}}$, and generate their source realizations using the chosen codeword and the respective channels $\tilde{Q}^n_{X|W}$ or $\tilde{Q}^n_{Y|W}$. On average, this scheme uses no more than

$$\frac{1}{n} + \frac{R}{\eta} + \left(1 - \frac{1}{\eta}\right)\left(\log_2 |\mathcal{X}||\mathcal{Y}| + \frac{1}{n}\right) \text{ bits/symbol.} \tag{41}$$

By allowing n to grow unbounded and then ε to vanish, we can see that η approaches unity and the above quantity approaches the required limit of $I(X, Y; W) = \mathscr{W}(X; Y)$. Thus we can build schemes for separate but exact generation of the pair of sources at rates arbitrarily close to but larger than $\mathscr{W}(X; Y)$. Combined with Remark 3, the claim follows. ■

V. Some Examples

The two main conditions derived in the previous section characterize when the exact and Wyner notions of common information coincide. However, they are implicit in the sense that a given pmf Q_{XY} can be verified to meet either of these criteria only after solving for the Wyner-optimal auxiliary RV W corresponding to Q_{XY}. Since the characterization of Wyner common information in (6) involves the optimization of convex function over a non-convex set, we are stymied when attempting to quantify the Wyner-optimal auxiliary RV W, and hence in verifying if a pmf Q_{XY} meets either of these implicit conditions of Theorems 1 and 2. Despite this obstacle, we can establish three classes of pmfs that meet the implicit condition of Theorem 2.

A. The General Noisy Typewriter Source

Definition 3: A general noisy typewriter channel is k-input k-output channel (for some integer $k \geq 2$) whose output $B \in \{1, \ldots, k\}$ is related to the input $A \in \{1, \ldots, k\}$ as follows. For $i, j \in \{1, \ldots, k\}$,

$$Q_{B|A}(j|i) \geq 0 \Leftrightarrow (j - i) \in \{0, 1, 1 - k\}. \tag{42}$$

A pair $(X, Y) \sim Q_{XY}$ is a general noisy typewriter source if either $Q_{X|Y}$ or $Q_{Y|X}$ is a general noisy typewriter channel.

Theorem 3: If pmf Q_{XY} is a general noisy typewriter source, then $\mathscr{E}(X; Y) = \mathscr{W}(X; Y)$.

Proof: Let Q_{XY} be a general noisy typewriter source with $|\mathcal{X}| = |\mathcal{Y}| = k$. Then, Q_{XY} takes the form

$$Q_{XY} \equiv \begin{bmatrix} * & * & 0 & 0 & \cdots & 0 & 0 \\ 0 & * & * & 0 & \cdots & 0 & 0 \\ & & & \vdots & & & \\ 0 & 0 & 0 & 0 & \cdots & * & * \\ * & 0 & 0 & 0 & \cdots & 0 & * \end{bmatrix}, \tag{43}$$

where $*$ indicates a possible non-zero entry. Now, let W be such that $X \leftrightarrow W \leftrightarrow Y$. Let $w \in \mathsf{S}(W)$, $Q_{X|W}(\cdot|w) = [\alpha_1 \cdots \alpha_k]$, and $Q_{Y|W}(\cdot|w) = [\beta_1 \cdots \beta_k]$. Then,

$$\begin{aligned} Q_{X,Y|W=w} &= \begin{bmatrix} \alpha_1\beta_1 & \alpha_1\beta_2 & \cdots & \alpha_1\beta_k \\ \alpha_2\beta_1 & \alpha_2\beta_2 & \cdots & \alpha_2\beta_k \\ \vdots & \vdots & & \vdots \\ \alpha_{k-1}\beta_1 & \alpha_{k-1}\beta_2 & \cdots & \alpha_{k-1}\beta_k \\ \alpha_k\beta_1 & \alpha_k\beta_2 & \cdots & \alpha_k\beta_k \end{bmatrix} \\ &\preceq \frac{Q_{XY}}{p_W(w)} \equiv \begin{bmatrix} * & * & 0 & 0 & \cdots & 0 & 0 \\ 0 & * & * & 0 & \cdots & 0 & 0 \\ & & & \vdots & & & \\ 0 & 0 & 0 & 0 & \cdots & * & * \\ * & 0 & 0 & 0 & \cdots & 0 & * \end{bmatrix} \end{aligned} \tag{44}$$

Let us now characterize all matrices $Q_{X,Y|W=w}$ that meet the above requirement. To this end, we may assume that the entries in positions indicated by $*$ are positive. We proceed by simply comparing the entries of the two matrices in positions (i, j) where $Q_{XY}(i, j)$ is guaranteed to be zero. Let us first identify matrices with $\alpha_1\beta_1 > 0$ satisfying (44). Comparing the zero entries in the first row and column, we infer that

$$\beta_j = 0, \quad 2 < j \leq k, \tag{45}$$

$$\alpha_i = 0, \quad 2 \leq i < k. \tag{46}$$

Now, $(\alpha_1\beta_2)(\alpha_k\beta_1) \leq \alpha_k\beta_2 \leq Q_{XY}(k, 2) = 0$. Hence, both the $(1, 2)^{\text{th}}$ and $(k, 1)^{\text{st}}$ entries of $Q_{X,Y|W}(\cdot, \cdot|w)$ in (44) cannot be positive. Hence, the only possible solutions to (44) with $\alpha_1\beta_1 > 0$ are $\mathsf{e}_{1,k} \cdot (\beta_1\mathsf{e}^T_{1,k} + \beta_2\mathsf{e}^T_{2,k})$ or $(\alpha_1\mathsf{e}_{1,k} + \alpha_k\mathsf{e}_{k,k}) \cdot \mathsf{e}^T_{1,k}$.

Now, to identify solutions with $\alpha_1\beta_2 > 0$, we compare the first row and the second column on both sides to infer that

$$\beta_j = 0, \quad 2 < j \leq k, \tag{47}$$

$$\alpha_i = 0, \quad 2 < i \leq k. \tag{48}$$

Also, as before, $(\alpha_1\beta_1)(\alpha_2\beta_2) \leq \alpha_2\beta_1 \leq Q_{XY}(2, 1) = 0$. Hence, at most one of the $(1, 1)^{\text{st}}$ and $(2, 2)^{\text{nd}}$ entries of $Q_{X,Y|W}(\cdot, \cdot|w)$ in (44) can be positive. Hence, the only possible solution other than $\mathsf{e}_{1,k} \cdot (\beta_1\mathsf{e}^T_{1,k} + \beta_2\mathsf{e}^T_{2,k})$ is $(\alpha_1\mathsf{e}_{1,k} + \alpha_2\mathsf{e}_{2,k}) \cdot \mathsf{e}^T_{2,k}$. Note that in the above three solutions, either

$$H(X|W = w) = -\sum_{i=1}^{k} \alpha_i \log_2 \alpha_i = 0, \quad \text{or} \tag{49}$$

$$H(Y|W = w) = -\sum_{i=1}^{k} \beta_i \log_2 \beta_i = 0. \tag{50}$$

By a simple argument that involves renaming indices, we can show that the possible solutions are as follows.

$$\mathsf{e}_{i,k} \cdot (\beta_i \mathsf{e}_{i,k}^T + \beta_{\phi(i+1)} \mathsf{e}_{\phi(i+1),k}^T) \quad \text{or} \tag{51}$$

$$(\alpha_i \mathsf{e}_{i,k} + \alpha_{\phi(i-1)} \mathsf{e}_{\phi(i-1),k}) \cdot \mathsf{e}_{i,k}^T, \quad i = 1, \ldots, k, \tag{52}$$

where $\phi(0) = k$, $\phi(i) = i$ for $i \in \{1, \ldots, k\}$ and $\phi(k+1) = 1$. Note that for each of the solution, (50) holds. Hence, all solutions to (44), and hence for any $X \leftrightarrow W \leftrightarrow Y$, we must have $H(X|W = w) = 0$ or $H(Y|W = w) = 0$ for all $w \in \mathsf{S}(W)$. Since W is any auxiliary RV that meets $X \leftrightarrow W \leftrightarrow Y$, (15) must also be met by the Wyner-optimal auxiliary RV of Q_{XY}. Hence, from Theorem 2, the claim follows. ■

B. General Erasure Source

Definition 4: A general erasure source is a pair of RVs (X, Y) with $\mathcal{X} = \{1, \ldots, k\}$, $\mathcal{Y} = \{1, \ldots, k+1\}$, and

$$Q_{XY}(i,j) \geq 0 \Leftrightarrow j \in \{i, k+1\}. \tag{53}$$

Theorem 4: If pmf Q_{XY} is a general erasure source, then $\mathscr{E}(X;Y) = \mathscr{W}(X;Y)$.

Proof: The proof is similar to that of Theorem 3. Here, for any $X \leftrightarrow W \leftrightarrow Y$, and $w \in \mathsf{S}(W)$, we must have

$$\begin{aligned} Q_{XY|W=w} &= \begin{bmatrix} \alpha_1\beta_1 & \alpha_1\beta_2 & \cdots & \alpha_1\beta_k \\ \alpha_2\beta_1 & \alpha_2\beta_2 & \cdots & \alpha_2\beta_k \\ \vdots & \vdots & & \vdots \\ \alpha_k\beta_1 & \alpha_k\beta_2 & \cdots & \alpha_k\beta_k \end{bmatrix} \\ &\preceq \frac{Q_{XY}}{p_W(w)} \equiv \begin{bmatrix} * & 0 & 0 & \cdots & 0 & * \\ 0 & * & 0 & \cdots & 0 & * \\ & & \vdots & & & \\ 0 & 0 & 0 & \cdots & * & * \end{bmatrix}. \end{aligned} \tag{54}$$

Suppose we seek solutions to (54) with $\alpha_i\beta_i > 0$, $1 \leq i \leq k$. By comparing the i^{th} column, we see that $\alpha_{i'} = 0$ if $i' \neq i$. Hence, the only solution with with $\alpha_i\beta_i > 0$, $1 \leq i \leq k$, is $\mathsf{e}_{i,k} \cdot (\beta_i \mathsf{e}_{i,k+1}^T + \beta_{k+1} \mathsf{e}_{k+1,k+1}^T)$. It is straightforward to see that $(\alpha_1 \mathsf{e}_{1,k} + \cdots + \alpha_k \mathsf{e}_{k,k}) \cdot \mathsf{e}_{k+1,k+1}^T$ is the only other possible solution to (54). As before, it follows that for any $X \leftrightarrow W \leftrightarrow Y$, we must have $H(X|W = w) = 0$ or $H(Y|W = w) = 0$ for all $w \in \mathcal{W}$. Since W is any auxiliary RV that meets $X \leftrightarrow W \leftrightarrow Y$, (15) must also be met by the Wyner-optimal auxiliary RV corresponding to Q_{XY}, the claim follows from Theorem 2. ■

Theorem 5: For an erasure source $Q_{XY} = \begin{bmatrix} \mathsf{a} & 0 & \mathsf{b} \\ 0 & \mathsf{c} & \mathsf{d} \end{bmatrix}$,

$$\begin{aligned} \mathscr{E}(X;Y) &= \mathscr{W}(X;Y) \\ &= \begin{cases} (\mathsf{a}+\mathsf{d})\,\mathrm{h}\big(\frac{\mathsf{a}}{\mathsf{a}+\mathsf{d}}\big) + (\mathsf{b}+\mathsf{c})\,\mathrm{h}\big(\frac{\mathsf{b}}{\mathsf{b}+\mathsf{c}}\big) & \mathsf{ac} \leq \mathsf{bd} \\ \mathrm{h}\,(\mathsf{a}+\mathsf{b}) & \mathsf{ac} > \mathsf{bd} \end{cases}. \end{aligned}$$

Proof: From the proof of Theorem 4, we see that the optimal decomposition of Q_{XY} in terms of its Wyner common information auxiliary RV must be

$$\begin{aligned} Q_{XY} = \Bigg(&(\mathsf{a}+x) \begin{bmatrix} \frac{\mathsf{a}}{\mathsf{a}+x} & 0 & \frac{x}{\mathsf{a}+x} \\ 0 & 0 & 0 \end{bmatrix} \\ &+ (\mathsf{b}-x+\mathsf{d}-y) \begin{bmatrix} 0 & 0 & \frac{\mathsf{b}-x}{\mathsf{b}-x+\mathsf{d}-y} \\ 0 & 0 & \frac{\mathsf{d}-y}{\mathsf{b}-x+\mathsf{d}-y} \end{bmatrix} \\ &+ (\mathsf{c}+y) \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{\mathsf{c}}{\mathsf{c}+y} & \frac{y}{\mathsf{c}+y} \end{bmatrix} \Bigg) \end{aligned} \tag{55}$$

for some suitable x and y. In the above decomposition, for a fixed $0 \leq x \leq \mathsf{b}$, the optimal choice of y must correspond to an optimal decomposition of a pmf $Q_{X'Y'}$ in terms of its Wyner common information auxiliary RV, where

$$Q_{X'Y'} = \begin{bmatrix} 0 & 0 & \frac{\mathsf{b}-x}{1-\mathsf{a}-x} \\ 0 & \frac{\mathsf{c}}{1-\mathsf{a}-x} & \frac{\mathsf{d}}{1-\mathsf{a}-x} \end{bmatrix}, \tag{56}$$

which is an L-shaped pmf. From [13], we see that the optimal y for a given x equals

$$y = \frac{\mathsf{dc}}{\mathsf{b}-x+\mathsf{c}}. \tag{57}$$

Hence, the optimal decomposition of Q_{XY} should take the following form:

$$\begin{aligned} Q_{XY} = \Bigg(&\mathsf{a}+x) \begin{bmatrix} \frac{\mathsf{a}}{\mathsf{a}+x} & 0 & \frac{x}{\mathsf{a}+x} \\ 0 & 0 & 0 \end{bmatrix} \\ &+ \frac{(\mathsf{b}-x)(1-\mathsf{a}-x)}{\mathsf{b}-x+\mathsf{c}} \begin{bmatrix} 0 & 0 & \frac{\mathsf{b}-x+\mathsf{c}}{1-\mathsf{a}-x} \\ 0 & 0 & \frac{\mathsf{d}}{1-\mathsf{a}-x} \end{bmatrix} \\ &+ \frac{\mathsf{c}(1-\mathsf{a}-x)}{\mathsf{b}-x+\mathsf{c}} \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{\mathsf{b}-x+\mathsf{c}}{1-\mathsf{a}-x} & \frac{\mathsf{d}}{1-\mathsf{a}-x} \end{bmatrix} \Bigg). \end{aligned} \tag{58}$$

The optimal x can then be found by maximizing $H(X, Y|W)$ as a function of x. The optimal x is given by

$$x^* = \arg\max_{0 \leq x \leq \mathsf{b}} \Big((\mathsf{a}+x)\,\mathrm{h}\big(\tfrac{\mathsf{a}}{\mathsf{a}+x}\big) + (1-\mathsf{a}-x)\,\mathrm{h}\big(\tfrac{\mathsf{d}}{1-\mathsf{a}-x}\big) \Big).$$

Due to the concavity of the binary entropy function, the unconstrained optimum is attained when

$$\tfrac{\mathsf{a}}{\mathsf{a}+x^*} = \tfrac{\mathsf{d}}{1-\mathsf{a}-x^*} = \mathsf{a}+\mathsf{d}, \tag{59}$$

which occurs at $x^* = \frac{\mathsf{a}(\mathsf{b}+\mathsf{c})}{\mathsf{a}+\mathsf{d}}$. Since the derivative is positive at $x = 0$, we can conclude that the optimal choice for the constrained optimization is

$$x^* = \min\left\{ \mathsf{b}, \frac{\mathsf{a}(\mathsf{b}+\mathsf{c})}{\mathsf{a}+\mathsf{d}} \right\}. \tag{60}$$

Computing $I(X, Y; W)$ for this choice yields the result. ■

Remark 5: For the symmetric binary erasure source

$$Q_{XY} = \begin{bmatrix} \frac{1}{2}(1-p) & \frac{1}{2}p & 0 \\ 0 & \frac{1}{2}p & \frac{1}{2}(1-p) \end{bmatrix},$$

$$\mathscr{E}(X;Y) = \mathscr{W}(X;Y) = \mathrm{h}(p)\mathbb{1}_{[0.5,1]}(p) + \mathbb{1}_{[0,0.5)}(p). \tag{61}$$

While the above result was specifically proved for the symmetric binary erasure source in [14], through our approach we are able to establish the equality of exact and Wyner common information generally for any erasure source Q_{XY} according to Definition 4.

Remark 6: Let $\mathsf{a}, \mathsf{b}, \mathsf{d} \in [0,1]$ with $\mathsf{a}+\mathsf{b}+\mathsf{d} = 1$. For the generalized binary Z-channel $Q_{XY} = \begin{bmatrix} \mathsf{a} & \mathsf{b} \\ 0 & \mathsf{d} \end{bmatrix}$,

$$\mathscr{E}(X;Y) = \mathscr{W}(X;Y) = (\mathsf{a}+\mathsf{d})\,\mathrm{h}\left(\frac{\mathsf{a}}{\mathsf{a}+\mathsf{d}}\right). \tag{62}$$

Appendix A
Required Results

Lemma 2: Let $b^n \in T^n_{\varepsilon}[Q_B]$ and $a^n \in T^n_{\frac{\varepsilon}{2}}[Q_{A|B}; b^n]$ for some $\varepsilon \in (0,1)$. Then,

$$(a^n, b^n) \in T^n_{2\varepsilon}[Q_{A,B}], \tag{63}$$

$$Q_{A|B}(a^n|b^n) \leq 2^{-nH(A|B)(1-2\varepsilon)}. \tag{64}$$

Proof: Let $(\tilde{a}, \tilde{b}) \in \mathsf{S}(Q_{AB})$. Since $b^n \in T^n_{\varepsilon}[Q_B]$,

$$0 < n(1-\varepsilon)Q_B(\tilde{b}) \leq \#_{b^n}(\tilde{b}) \leq n(1+\varepsilon)Q_B(\tilde{b}). \tag{65}$$

Next, since $a^n \in T^n_{\frac{\varepsilon}{2}}[p_{A|B}; b^n]$, it follows that

$$1 - \frac{\varepsilon}{2} \leq \frac{\#_{a^n,b^n}(\tilde{a},\tilde{b})}{\#_{b^n}(\tilde{b})Q_{A|B}(\tilde{a}|\tilde{b})} \leq 1 + \frac{\varepsilon}{2}. \tag{66}$$

Combining the two above equations, we see that

$$1 - 2\varepsilon < \frac{\#_{a^n,b^n}(\tilde{a},\tilde{b})}{nQ_{AB}(\tilde{a},\tilde{b})} < 1 + 2\varepsilon, \tag{67}$$

which establishes that $(a^n, b^n) \in T^n_{2\varepsilon}[Q_{A,B}]$. To prove the next result, we see that

$$\begin{aligned} Q_{A|B}(a^n|b^n) &= \prod_{(\tilde{a},\tilde{b}) \in \mathsf{S}(Q_{AB})} \left(Q_{A|B}(\tilde{a}|\tilde{b})\right)^{\#_{a^n,b^n}(\tilde{a},\tilde{b})} \\ &\overset{(67)}{\leq} \prod_{(\tilde{a},\tilde{b}) \in \mathsf{S}(Q_{AB})} \left(Q_{A|B}(\tilde{a}|\tilde{b})\right)^{n(1-2\varepsilon)Q_{AB}(\tilde{a},\tilde{b})} \\ &= 2^{-nH(A|B)(1-2\varepsilon)}. \end{aligned} \tag{68}$$

■

Lemma 3: Let $(X,Y) \sim Q_{XY}$ be given. Let auxiliary random variable W be such that for any $w \in \mathsf{S}(\mathcal{W})$,

$$H(X \mid W = w) \cdot H(Y \mid W = w) = 0. \tag{69}$$

Then, for any $\varepsilon, \delta > 0$ and $w^n \in T^n_{\delta}[Q_W]$,

$$T^n_{\varepsilon}[Q_{X|W}; w^n] \times T^n_{\varepsilon}[Q_{Y|W}; w^n] = T^n_{\varepsilon}[Q_{X,Y|W}; w^n].$$

Proof: Note that (69) implies the conditional independence of X and Y given W, i.e., $X \leftrightarrow W \leftrightarrow Y$. Thus, we only need to show that the Cartesian product of marginally conditional typical sets is a subset of the jointly conditional typical set. The reverse direction is trivial. Let $w^n \in T^n_{\delta}[Q_W]$, $x^n \in T^n_{\varepsilon}[Q_{X|W}; w^n]$ and $y^n \in T^n_{\varepsilon}[Q_{Y|W}; w^n]$. Pick $\tilde{w} \in \mathsf{S}(W)$, and $(\tilde{x}, \tilde{y}, \tilde{w}) \in \mathsf{S}(Q_{XYW})$. For this choice of $\tilde{w}$, let us assume that $H(X|W = \tilde{w}) = 0$. Since $(\tilde{x}, \tilde{y}, \tilde{w}) \in \mathsf{S}(Q_{XYW})$, it must follow that $Q_{X|W}(\tilde{x}|\tilde{w}) = 1$. Consequently,

$$\#_{x^n,y^n,w^n}(\tilde{x},\tilde{y},\tilde{w}) = \#_{y^n,w^n}(\tilde{y},\tilde{w}), \tag{70}$$

$$Q_{XY|W}(\tilde{x},\tilde{y}|\tilde{w}) = Q_{Y|W}(\tilde{y}|\tilde{w}). \tag{71}$$

Then,

$$\begin{aligned} &\left|\frac{\#_{x^n,y^n,w^n}(\tilde{x},\tilde{y},\tilde{w})}{\#_{w^n}(\tilde{w})Q_{XY|W}(\tilde{x},\tilde{y}|\tilde{w})} - 1\right| \\ &\overset{(70),(71)}{=} \left|\frac{\#_{y^n,w^n}(\tilde{y},\tilde{w})}{\#_{w^n}(\tilde{w})Q_{Y|W}(\tilde{x},\tilde{y}|\tilde{w})} - 1\right| \leq \varepsilon, \end{aligned} \tag{72}$$

where the last inequality follows since $y^n \in T^n_{\varepsilon}[Q_{Y|W}; w^n]$. For $\tilde{w}$ such that $H(Y|W = \tilde{w}) = 0$ we simply reverse the roles of X and Y in the above argument. Thus,

$$\sup_{\substack{(\tilde{x},\tilde{y},\tilde{w}) \in \mathsf{S}(Q_{XYW}) \\ \#_{w^n}(\tilde{w}) > 0}} \left|\frac{\#_{x^n,y^n,w^n}(\tilde{x},\tilde{y},\tilde{w})}{\#_{w^n}(\tilde{w})Q_{XY|W}(\tilde{x},\tilde{y}|\tilde{w})} - 1\right| \leq \varepsilon. \tag{73}$$

Consequently, $(x^n, y^n) \in T^n_{\varepsilon}[Q_{XY|W}; w^n]$. ■

References

[1] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems of Control and Information Theory*, vol. 2, no. 2, pp. 149–162, 1973.

[2] A. B. Wagner, B. G. Kelly, and Y. Altug, "The lossy one-helper conjecture is false," in *47th Annual Allerton Conference on Communication, Control, and Computing*, Sept 2009, pp. 716–723.

[3] B. N. Vellambi and R. Timo, "The Heegard-Berger problem with common receiver reconstructions," in *IEEE Information Theory Workshop (ITW 2013)*, Sept 2013, pp. 1–5.

[4] ——, "Successive refinement with common receiver reconstructions," in *2014 IEEE International Symposium on Information Theory*, June 2014, pp. 2664–2668.

[5] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *The Bell System Technical Journal*, vol. 53, no. 9, pp. 1681–1721, Nov 1974.

[6] A. D. Wyner, "The common information of two dependent random variables," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 163–179, Mar 1975.

[7] P. Cuff, "Communication requirements for generating correlated random variables," in *2008 IEEE International Symposium on Information Theory*, July 2008, pp. 1393–1397.

[8] P. Cuff, H. Permuter, and T. M. Cover, "Coordination capacity," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4181–4206, Sep. 2010.

[9] B. N. Vellambi, J. Kliewer, and M. Bloch, "Strong coordination over multi-hop line networks," in *2015 IEEE Information Theory Workshop*, Oct. 2015, pp. 192–196.

[10] ——, "Strong coordination over a line when actions are Markovian," in *50th Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2015.

[11] G. Xu, W. Liu, and B. Chen, "A lossy source coding interpretation of Wyner's common information," *IEEE Transactions on Information Theory*, vol. 62, no. 2, pp. 754–768, Feb 2016.

[12] K. B. Viswanatha, E. Akyol, and K. Rose, "The lossy common information of correlated sources," *IEEE Transactions on Information Theory*, vol. 60, no. 6, pp. 3238–3253, June 2014.

[13] H. S. Witsenhausen, "Values and bounds for the common information of two discrete random variables," *SIAM Journal on Applied Mathematics*, vol. 31, no. 2, pp. 313–333, 1976.

[14] G. R. Kumar, C. T. Li, and A. E. Gamal, "Exact common information," in *2014 IEEE International Symposium on Information Theory*, June 2014, pp. 161–165.

[15] G. Kramer, "Topics in multi-user information theory," *Found. Trends Commun. Inf. Theory*, vol. 4, no. 4-5, pp. 265–444, 2007.

[16] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the IRE*, vol. 40, no. 9, pp. 1098–1101, Sept 1952.

[17] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[18] T. S. Han, *Information-Spectrum Methods in Information Theory*, 1st ed. Springer, 2003.

[19] S. Boucheron, G. Lugosi, P. Massart, and M. Ledoux, *Concentration inequalities : a nonasymptotic theory of independence*. Oxford University Press, 2013.