New Results on the Equality of Exact and Wyner Common Information Rates

Badri N. Vellambi
Australian National University
Acton, ACT 2601, Australia
Email: badri.vellambi@anu.edu.au

Jörg Kliewer
New Jersey Institute of Technology
Newark, NJ 07102
Email: jkliewer@njit.edu

Abstract—**Recently, Kumar, Li, and El Gamal proposed a notion of common information using a variation of a setup used to define Wyner common information rate. This notion, known as the *exact common information*, is the minimum common randomness required for the *exact* and separate generation of a pair of correlated discrete memoryless sources. While exact common information rate is not known to have a single-letter characterization, it was shown to equal the Wyner common information rate for the symmetric binary erasure source in Kumar - Li - El Gamal - ISIT 2014. The authors extended this result to establish the equality of the two notions of common information for general noisy typewriter, Z- and erasure sources in Vellambi - Kliewer - Allerton 2016. In this work, we investigate the connection between exact and Wyner common information rates to derive two new implicit conditions (on the joint source distribution) that ensure the equality of the two notions.**

I. Introduction

One of the fundamental aims of source coding is to quantify the amount and role of information in various multi-user communication problems. Research efforts into quantifying the information common to two (or more) correlated random variables indicate that the amount and nature of information depends on the actual setup and/or application, and that there is no one universal notion of information. The most common notion is Shannon's mutual information, which is the reduction in the entropy of a random variable due to the knowledge of a correlated random variable. Gács and Körner envisaged common information between two discrete memoryless sources (DMSs) as the rate of randomness that can be simultaneously extracted from either of the two correlated sources [1]. It was proven that the Gács-Körner common information between a pair of sources is more restrictive than and distinct from mutual information; this notion of common information plays an important role in the optimal and/or best-known schemes in many multi-user source coding problems [2]–[4].

Another well known notion of common information can be traced to Wyner's seminar work [6] on the Gray-Wyner problem [5]. The Gray-Wyner problem (see Fig. 1) 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 Wyner common information rate between the two sources is the smallest communication rate R_0 on the common channel such that the sum rate is kept at its absolute minimum, i.e., $R_0 + R_1 + R_2 = H(X, Y)$. Wyner common information rate is given by the following single-letter characterization.

$$\mathscr{W}(X;Y) \triangleq \min_{\substack{X \leftrightarrow W \leftrightarrow Y \\ |\mathcal{W}| \leq |\mathcal{X}||\mathcal{Y}|}} I(X,Y;W). \qquad (1)$$

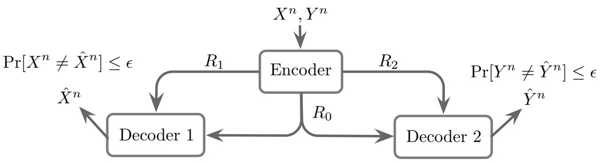
X^n, Y^n
Encoder
R_1
R_2
R_0
$\Pr[X^n \neq \hat{X}^n] \leq \epsilon$
$\Pr[Y^n \neq \hat{Y}^n] \leq \epsilon$
$\hat{X}^n$
$\hat{Y}^n$
Decoder 1
Decoder 2

Fig. 1. The Gray-Wyner setup.

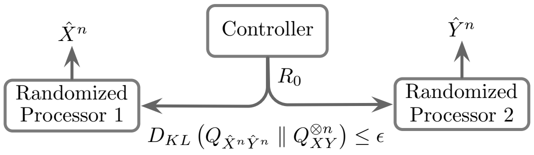
Controller
$\hat{X}^n$
$\hat{Y}^n$
R_0
Randomized Processor 1
Randomized Processor 2
$D_{KL}\left(Q_{\hat{X}^n\hat{Y}^n} \parallel Q_{XY}^{\otimes n}\right) \leq \epsilon$

Fig. 2. Separate generation of correlated sources.

In addition to the Gray-Wyner problem setup in [6], Wyner used the setup in Fig. 2 to show that the Wyner common information rate $\mathscr{W}(X;Y)$ is also the smallest rate of a uniform random seed that must be supplied to two independent processors to *approximately* generate the two DMSs X and Y separately. In this setup, a Kullback-Leibler-divergence-based metric was chosen to quantify 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] feature commonly in achievable schemes in strong coordination problems, e.g. [7]–[9]. Other connections of Wyner common information to lossy reconstruction problems, and an extension to multiple random variables were explored in [10], [11]. It must be remarked that despite the simple formulation in (1), explicit formula/value for Wyner common information rate is known only for a few joint probability mass functions (pmfs) [12].

Recently, in [13], Kumar et al. proposed the notion of exact common information using the setup in Fig. 2 by requiring the distribution $Q_{\hat{X}^n,\hat{Y}^n}$ to *equal* the n-fold product $Q_{XY}^{\otimes n}$ of

This work was supported by NSF grants CCF-1440014 and CNS-1526547.

Q_{XY}, which is the joint 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 and precisely according to a given Q_{XY}.

In [13], Kumar et al. derived the fundamental properties of exact common information and proved that for the symmetric binary erasure source, the exact and Wyner common information rates coincide. In [14], we presented two sufficient conditions under which the notions of exact and Wyner common information coincide. Using these conditions, it was shown that the exact and Wyner common information rates are equal for the binary Z-, the general erasure and the general noisy typewriter sources.

In this work, we derive new connections between exact and Wyner common information, and establish two new implicit conditions under which the two notions coincide.

II. Notation

All random variables (RVs) in this work are assumed to be over finite alphabets. Given RVs A, B, C, the conditional independence of A and C given B is denoted by $A \leftrightarrow B \leftrightarrow C$. The support of a RV $X \sim Q_X$ is denoted by $\mathsf{S}(Q_X)$. Given a finite set S, $\mathbb{1}_S$ 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 $\mathcal{A} \times \mathcal{B}$, we define the following letter-typical sets [15].

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

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

III. Problem Definition and Known Results

Exact common information $\mathscr{E}(X;Y)$ is defined via Fig. 2 with the Kullback-Leibler divergence term set 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 an RV W_n such that $X^n \leftrightarrow W_n \leftrightarrow Y^n$ and $H(W_n) \le n(\mathsf{R} + \varepsilon)$. That is, if W_n is conveyed to two randomized processors, they can separately generate X^n and Y^n jointly correlated according to $Q_{XY}^{\otimes n}$. The exact common information is then the infimum of all such achievable rates.

Definition 1 ([13]): Given pmf Q_{XY}, the exact common information rate between X and Y is defined to be

$$\mathscr{E}(X;Y) \triangleq \lim_{n\to\infty} \left(\inf_{X^n \leftrightarrow W_n \leftrightarrow Y^n} \frac{H(W_n)}{n} \right). \tag{2}$$

Notice that the size of the alphabet of W_n is allowed to grow with n, and hence, (2) is not a computable form of exact common information. We summarily present three basic properties (Remarks 1-3) of the exact common information between two random variables. Details of the proof of these results can be found in [13], [14]. The first property is an ordering of different notions of common information. The position of the exact common information is justified by the fact that while the setups for exact and Wyner common information are identical, the source generation requirement for the former is more stringent. The second remark details the expected result that concatenation cannot decrease exact common information. Lastly, the third parallels data processing inequality and characterizes the monotonicity of exact common information with respect to stochastic degradedness.

Remark 1: [13, Prop. 3] Given $(X, Y) \sim Q_{XY}$, let $\mathscr{G}(X;Y)$ denote the Gács-Körner common information between RVs X and Y. Then, $\mathscr{G}(X;Y)$, $I(X;Y)$, $\mathscr{W}(X;Y)$, $\mathscr{E}(X;Y)$, and $\min\{H(X), H(Y)\}$ is a list of non-decreasing non-negative real numbers.

Remark 2: Let $(A, B, C, D) \sim Q_{AB}Q_{CD}$, $X = (A, C)$ and $Y = (B, D)$. Then,

$$\mathscr{E}(X;Y) \ge \max\{\mathscr{E}(A;B), \mathscr{E}(C;D), \mathscr{W}(X;Y)\} \tag{3}$$
$$\mathscr{E}(X;Y) \le \mathscr{E}(A;B) + \mathscr{E}(C;D). \tag{4}$$

Further, (4) holds with equality if $\mathscr{E}(A;B) = \mathscr{W}(A;B)$ and $\mathscr{E}(C;D) = \mathscr{W}(C;D)$.

Remark 3: $\mathscr{E}(X;Y) \le \mathscr{E}(X';Y')$ if $X \leftrightarrow X' \leftrightarrow Y' \leftrightarrow Y$.

In addition to the above properties, in [14], we established the equality of Wyner and exact common information rates provided the pmf Q_{XY} is such that there *either* exists:

- a RV W such that $X \leftrightarrow W \leftrightarrow Y$, $I(X, Y; W) = \mathscr{W}(X;Y)$, and $H(W|X, Y) = 0$; or
- a RV W such that $I(X, Y; W) = \mathscr{W}(X;Y)$, and

$$\sum_{w \in \mathcal{W}} H(X|W = w) \cdot H(Y|W = w) = 0. \tag{5}$$

IV. New Results

This section contains three new results. The first result presents a bound on the exact common information between two random variables based on a decomposition using a third correlated random variable. The next two results present implicit conditions on the pmf Q_{XY} under which the Wyner and exact common information notions coincide.

Theorem 1: Let $(X, Y, G) \sim Q_{X,Y,G}$. Let for $g \in \mathcal{G}$, $(X_g, Y_g) \sim Q_{X_g,Y_g} \triangleq Q_{XY|G=g}$. Then,

$$\mathscr{E}(X;Y) \le H(G) + \sum_{g \in \mathcal{G}} Q_G(g) \mathscr{E}(X_g;Y_g).$$

Further, if the RV G is such that $H(G|X) = H(G|Y) = 0$, and for each $g \in \mathcal{G}$, $\mathscr{E}(X_g;Y_g) = \mathscr{W}(X_g;Y_g)$, then

$$\begin{aligned} \mathscr{E}(X;Y) &= H(G) + \textstyle\sum_{g \in \mathcal{G}} Q_G(g) \mathscr{E}(X_g;Y_g) \\ &= H(G) + \textstyle\sum_{g \in \mathcal{G}} Q_G(g) \mathscr{W}(X_g;Y_g) = \mathscr{W}(X;Y). \end{aligned}$$

Proof: Let $\varepsilon > 0$. Let n_0 be a positive integer such that for $\ell > n_0$, $\mathbb{P}[G^\ell \notin T_\varepsilon^\ell[Q_G]] \le \varepsilon$. Let $n > n_0$ then be chosen. Let $\mathcal{G} \triangleq \{\mathsf{g}_1, \ldots, \mathsf{g}_k\}$, and for the sake of notational ease, let $n_{\ell,\varepsilon} \triangleq \lfloor n(1+\varepsilon) Q_G(\mathsf{g}_\ell) \rfloor$ for $\ell = 1, \ldots, k$. Further, define

$$\boldsymbol{X}_{(n,\varepsilon)} \triangleq [G^n, (X_{\mathsf{g}_1})^{n_{1,\varepsilon}}, \ldots, (X_{\mathsf{g}_k})^{n_{k,\varepsilon}}],$$
$$\boldsymbol{Y}_{(n,\varepsilon)} \triangleq [G^n, (Y_{\mathsf{g}_1})^{n_{1,\varepsilon}}, \ldots, (Y_{\mathsf{g}_k})^{n_{k,\varepsilon}}],$$

where $\boldsymbol{X}_{(n,\varepsilon)}$ contains n i.i.d. copies of G, and $n_{i,\varepsilon}$ i.i.d. copies of X_{g_i}, $i = 1, \ldots, k$. Similarly, $\boldsymbol{Y}_{(n,\varepsilon)}$ contains n i.i.d. copies of G, and $n_{i,\varepsilon}$ i.i.d. copies of Y_{g_i}, $i = 1, \ldots, k$. Let the joint distribution between these two sources be

$$Q_{\boldsymbol{X}_{(n,\varepsilon)},\boldsymbol{Y}_{(n,\varepsilon)}} \triangleq Q_G^{\otimes n} \prod_{\ell=1}^{k} Q_{X_{\mathsf{g}_\ell},Y_{\mathsf{g}_\ell}}^{\otimes n_{\ell,\varepsilon}}. \tag{6}$$

From Remark 2, it follows that

$$\mathscr{E}(\boldsymbol{X}_{(n,\varepsilon)}; \boldsymbol{Y}_{(n,\varepsilon)}) \leq nH(G) + \textstyle\sum_\ell n_{\ell,\varepsilon} \mathscr{E}(X_{\mathsf{g}_\ell}; Y_{\mathsf{g}_\ell}). \tag{7}$$

Suppose that we are given a scheme for generating N symbols of $Q_{\boldsymbol{X}_{(n,\varepsilon)},\boldsymbol{Y}_{(n,\varepsilon)}}$ using a message M_N that the controller conveys at no more than $N(\mathscr{E}(\boldsymbol{X}_{(n,\varepsilon)}; \boldsymbol{Y}_{(n,\varepsilon)}) + \varepsilon)$ bits conveying a message M_N. Since the nN symbols of G are common to both terminals, it must follow that these nN symbols are a function of the common message conveyed by the controller. Using this fact, we can devise a scheme to generate nN copies of Q_{XY} in the following way.

A: If the realization of G^{nN} corresponding to M_N is an element of $T_\varepsilon^{nN}[Q_G]$, the X-terminal can generate $n_{\ell,\varepsilon} N$ copies of X_{g_ℓ}, $\ell = 1, \ldots, k$, which it can reposition according to the realization of G^{nN} while maintaining the same order of realization for indices with the same g_ℓ, $\ell = 1, \ldots, k$. The Y-terminal operates similarly. An illustration for a source with $k = 2$ is given in Fig. 3.

B: If the realization of G^{nN} corresponding to M_N is not in $T_\varepsilon^{nN}[Q_G]$, the controller simply generates nN symbols of the two sources and forwards the realizations to both terminals using at most $nN \log_2 |\mathcal{X}||\mathcal{Y}| + 1$ bits.

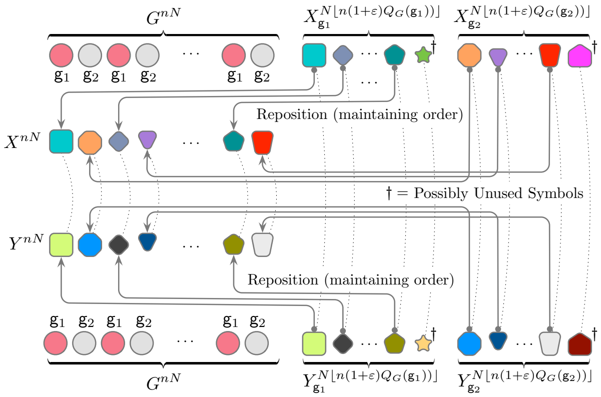
G^{nN}
$X_{\mathsf{g}_1}^{N\lfloor n(1+\varepsilon)Q_G(\mathsf{g}_1))\rfloor}$
$X_{\mathsf{g}_2}^{N\lfloor n(1+\varepsilon)Q_G(\mathsf{g}_2))\rfloor}$
g_1 g_2 g_1 g_2 $\cdots$ g_1 g_2
†
†
Reposition (maintaining order)
X^{nN}
† = Possibly Unused Symbols
Y^{nN}
Reposition (maintaining order)
g_1 g_2 g_1 g_2 $\cdots$ g_1 g_2
†
†
G^{nN}
$Y_{\mathsf{g}_1}^{N\lfloor n(1+\varepsilon)Q_G(\mathsf{g}_1))\rfloor}$
$Y_{\mathsf{g}_2}^{N\lfloor n(1+\varepsilon)Q_G(\mathsf{g}_2))\rfloor}$

Fig. 3. Repositioning the realizations of $\boldsymbol{X}_{(n,\varepsilon)}, \boldsymbol{Y}_{(n,\varepsilon)}$ according to the generated G_{nN} to generate nN copies of Q_{XY}.

This scheme generates nN copies of the source Q_{XY} using N copies of $Q_{\boldsymbol{X}_{(n,\varepsilon)},\boldsymbol{Y}_{(n,\varepsilon)}}$ with an additional overhead (due to the atypicality of G^{nN}). Hence, we have

$$nN\mathscr{E}(X;Y) \leq \left(\begin{array}{r} N(\mathscr{E}(\boldsymbol{X}_{(n,\varepsilon)}; \boldsymbol{Y}_{(n,\varepsilon)}) + \varepsilon) \\ +\, \varepsilon\,(nN \log_2 |\mathcal{X}||\mathcal{Y}| + 1) \end{array} \right), \tag{8}$$

where we have used the fact that $\mathbb{P}[G^{nN} \notin T_\varepsilon^{nN}[Q_G]] \leq \varepsilon$. Replacing $\mathscr{E}(\boldsymbol{X}_{(n,\varepsilon)}; \boldsymbol{Y}_{(n,\varepsilon)})$ with the bound in (7), taking n to infinity, and then ε to zero, we obtain

$$\mathscr{E}(X;Y) \leq H(G) + \textstyle\sum_\ell Q_G(\mathsf{g}_\ell) \mathscr{E}(X_{\mathsf{g}_\ell}; Y_{\mathsf{g}_\ell}), \tag{9}$$

which is the upper bound we set to establish.

Now, if $\mathscr{E}(X_{g_\ell}; Y_{g_\ell}) = \mathscr{W}(X_{g_\ell}; Y_{g_\ell})$ for each component pmf $Q_{X_{\mathsf{g}_\ell} Y_{\mathsf{g}_\ell}}$, then from Remark 2, we are guaranteed that (9) holds with equality, i.e., for any $\varepsilon > 0$ and $n \in \mathbb{N}$,

$$\mathscr{E}(\boldsymbol{X}_{(n,\varepsilon)}; \boldsymbol{Y}_{(n,\varepsilon)}) = nH(G) + \textstyle\sum_\ell n_{\ell,\varepsilon} \mathscr{E}(X_{\mathsf{g}_\ell}; Y_{\mathsf{g}_\ell}). \tag{10}$$

In this setting, we can adapt a scheme for exact generation of Q_{XY}. Fix $\varepsilon \in (0, \frac{1}{n_0})$. We can then find a suitably large $m > \frac{1}{\varepsilon} > n_0$ and an exact generation scheme for generating m copies of the joint source Q_{XY}, where the controller conveys a message M_m with an average length of no more that $m(\mathscr{E}(X;Y) + \varepsilon)$ bits. Since $X^m \leftrightarrow M_m \leftrightarrow Y^m$ it must follow that $I(X^m; Y^m | M_m) = H(G^m | M_m) = 0$, i.e., the controller knows G^m. We will use this fact to devise a scheme for the source $Q_{\boldsymbol{X}_{(m',\varepsilon)},\boldsymbol{Y}_{(m',\varepsilon)}}$, where $m' = \frac{m(1-2\varepsilon)}{1+\varepsilon}$.

To derive a scheme that outputs *one* symbol of $Q_{\boldsymbol{X}_{(m',\varepsilon)},\boldsymbol{Y}_{(m',\varepsilon)}}$, the controller generates a realization of M_m. The controller uses the realization to verify if the corresponding realization of G^m is ε-letter typical. If so, it conveys the typicality of G^m by a bit, and then conveys the realization of M_m to the two terminals. The two terminals use the received message and generate X^m and Y^m respectively. The terminals then identify (at least) $G^m = f(X^m) = g(Y^m)$. Since G^m is typical, the terminals can definitely identify $\lfloor m(1-\varepsilon) Q_G(\mathsf{g}_\ell) \rfloor > m'(1+\varepsilon)$ copies of X_{g_ℓ} and Y_{g_ℓ}, $\ell = 1, \ldots, k$, respectively. Employing the repositioning technique in Fig. 3 in the reverse direction, we can then identify one symbol of $\boldsymbol{X}_{(m',\varepsilon)}$ and $\boldsymbol{Y}_{(m',\varepsilon)}$ at the two terminals.

On the other hand, if G^m corresponding to the realization of M_m is atypical, the controller conveys the atypicality using a bit, and then generates a realization of $(\boldsymbol{X}_{(m',\varepsilon)}, \boldsymbol{Y}_{(m',\varepsilon)})$ and conveys that in no more than $m' \log_2 |\mathcal{X}||\mathcal{Y}| + 1$ bits.

The above scheme is then an exact generation scheme for one symbol of $Q_{\boldsymbol{X}_{(m',\varepsilon)},\boldsymbol{Y}_{(m',\varepsilon)}}$, and hence,

$$\mathscr{E}(\boldsymbol{X}_{(m',\varepsilon)}; \boldsymbol{Y}_{(m',\varepsilon)}) \leq \left(\begin{array}{r} 1 + m(\mathscr{E}(X;Y) + \varepsilon) \\ +\varepsilon(m' \log_2 |\mathcal{X}||\mathcal{Y}| + 1) \end{array} \right),$$

where we have used $\mathbb{P}[G^m \notin T_\varepsilon^m[Q_G]] \leq \varepsilon$. Combining the above with (9), and letting $\varepsilon \to 0$, we see that

$$\mathscr{E}(X;Y) \geq H(G) + \textstyle\sum_\ell Q_G(\mathsf{g}_\ell) \mathscr{E}(X_{\mathsf{g}_\ell}; Y_{\mathsf{g}_\ell}), \tag{11}$$

thereby establishing the second part of the claim. ■

Theorem 2: If $Q_{XY} = \frac{\mathbb{1}_{\mathsf{S}(Q_{XY})}}{|\mathsf{S}(Q_{XY})|}$, then

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

Proof: Just as in the proof of [14, Theorem 2], we devise a two-stage scheme to exactly match the output statistics of the two sources to $Q_{XY}^{\otimes n}$. The first stage uses a modified channel resolvability codebook [7], [16] so that the joint pmf of the source outputs satisfies a *nearly-uniform*-type convergence constraint. The second stage then refines the shortcomings of the first to match the exact generation requirement.

Pick auxiliary RV W and pmf Q_{XWY} such that (a) $I(X,Y;W) = \mathscr{W}(X;Y)$ and (b) $X \leftrightarrow W \leftrightarrow Y$. To exactly generate the sources, pick $\varepsilon > 0$, and n large so that

$$\mathbb{P}\big[W^n \in T_\varepsilon^n[Q_W]\big] > 1 - \varepsilon, \tag{13}$$

and for every $w^n \in T_\varepsilon^n[Q_W]$, $X^n \sim Q_{X|W}^{\otimes n}(\cdot|w^n)$ and $Y^n \sim Q_{Y|W}^{\otimes n}(\cdot|w^n)$,

$$\mathbb{P}\big[X^n \in T_{\frac{\varepsilon}{2}}^n[Q_{X|W}; w^n]\big] > 1-\varepsilon, \tag{14}$$

$$\mathbb{P}\big[Y^n \in T_{\frac{\varepsilon}{2}}^n[Q_{Y|W}; w^n]\big] > 1-\varepsilon. \tag{15}$$

Let $\ell = |\mathbf{S}(Q_{XY})|$. Then, due to the structure of Q_{XY}, $H(X,Y) = \log_2 \ell$. Let rate R be given by

$$R = I(X,Y;W) + 3\varepsilon|\mathcal{X}||\mathcal{Y}| \tag{16}$$

$$= \log_2 \ell - H(X,Y|W) + 3\varepsilon|\mathcal{X}||\mathcal{Y}|. \tag{17}$$

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 $\tilde{Q}_W^{\otimes n}$ defined by

$$\tilde{Q}_W^n(w^n) = \frac{Q_W^{\otimes n}(w^n)\mathbb{1}_{T_\varepsilon^n[Q_W]}(w^n)}{Q_W^{\otimes n}\left(T_\varepsilon^n[Q_W]\right)}. \tag{18}$$

By construction, *every* codeword in the codebook is ε-strongly letter typical. Now, define a channel $\tilde{Q}_{X|W}^n(\cdot|\cdot)$ with input alphabet $T_\varepsilon^n[Q_W]$ and output alphabet $\mathcal{X}^n$ by

$$\tilde{Q}_{X|W}^n(x^n|w^n) = \frac{Q_{X|W}^{\otimes n}(x^n|w^n)\mathbb{1}_{T_{\frac{\varepsilon}{2}}^n[Q_{X|W};w^n]}(x^n)}{Q_{X|W}^{\otimes n}(T_{\frac{\varepsilon}{2}}^n[Q_{X|W}; w^n]|w^n)}, \tag{19}$$

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

$$\tilde{Q}_{Y|W}^n(y^n|w^n) = \frac{Q_{Y|W}^{\otimes n}(y^n|w^n)\mathbb{1}_{T_{\frac{\varepsilon}{2}}^n[Q_{Y|W};w^n]}(y^n)}{Q_{Y|W}^{\otimes n}(T_{\frac{\varepsilon}{2}}^n[Q_{Y|W}; w^n]|w^n)}. \tag{20}$$

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

$$Z_i(x^n, y^n) = \tilde{Q}_{X|W}^n(x^n|W^n(i))\tilde{Q}_{Y|W}^n(y^n|W^n(i)), \tag{21}$$

and let $\mu(x^n, y^n) \triangleq \mathbb{E}[Z_i(x^n, y^n)]$. Using the properties of letter-typical sequences, and (13)-(15), one can show that for any $i = 1, \ldots, 2^{nR}$, and $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$,

$$Z_i(x^n, y^n) \in \left[0, \frac{2^{-n(H(X,Y|W)(1-2\varepsilon)}}{(1-\varepsilon)^2}\right], \tag{22}$$

$$\mu(x^n, y^n) \leq \frac{Q_{XY}^{\otimes n}(x^n, y^n)}{(1-\varepsilon)^3} \leq \frac{\ell^{-n}}{(1-\varepsilon)^3}, \tag{23}$$

$$\mathbb{E}[Z_i^2(x^n, y^n)] \leq \frac{2^{-nH(X,Y|W)(1-2\varepsilon)}}{(1-\varepsilon)^2}\mu(x^n, y^n). \tag{24}$$

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

$$\tilde{Q}_{XY}^{\mathcal{C},n}(x^n, y^n) \triangleq 2^{-nR}\textstyle\sum_{i=1}^{2^{nR}} Z_i(x^n, y^n),$$

which is the joint pmf of the outputs when a randomly selected codeword from the random channel resolvability codebook $\mathcal{C}$ is passed through two parallel channels $\tilde{Q}_{X|W}^n$ and $\tilde{Q}_{Y|W}^n$, respectively. Unlike in the proof of [14, Theorem 2], the support of the pmfs $\tilde{Q}_{XY}^{\mathcal{C},n}$ and $\mathbb{E}[\tilde{Q}_{XY}^{\mathcal{C},n}] = \mu$ need not lie within the $T_{2\varepsilon}^n[Q_{XY}]$. Instead, we are only guaranteed that

$$\mathbf{S}(\tilde{Q}_{XY}^{\mathcal{C},n}) \subseteq \mathbf{S}(\mu) \subseteq (T_{2\varepsilon}^n[Q_X] \times T_{2\varepsilon}^n[Q_Y]) \cap \mathbf{S}(Q_{XY}^{\otimes n}). \tag{25}$$

Thus, it is possible that the realizations of the pair of sources emitted by a random codebook are atypical. However, this will not cause a severe problem because of the specific form of the joint pmf Q_{XY}. Now, let $\eta \triangleq \frac{1}{(1-\varepsilon)^3} + \varepsilon$. Then, we see that for any $(x^n, y^n) \in \mathbf{S}(\mu)$,

$$\Delta(x^n, y^n) \triangleq \eta\, Q_{XY}^{\otimes n}(x^n, y^n) - \mu(x^n, y^n) \overset{(23)}{\geq} \varepsilon\, \ell^{-n}, \tag{26}$$

Similarly, for any $(x^n, y^n) \in \mathbf{S}(\mu)$,

$$\eta\, Q_{XY}^{\otimes n}(x^n, y^n) + 2\mu(x^n, y^n) \overset{(23)}{\leq} 3\eta\, \ell^{-n}. \tag{27}$$

Note that the random pmf $\tilde{Q}_{XY}^{\mathcal{C},n}$ is a weighted sum of i.i.d. RVs, and hence one can use standard bounding techniques to establish concentration properties of tail events. Now consider the following tail event for $(x^n, y^n) \in \mathbf{S}(\mu)$.

$$\begin{aligned}
&\mathbb{P}\left[\left|\, \tilde{Q}_{XY}^{\mathcal{C},n}(x^n, y^n) - \mu(x^n, y^n)\right| > \Delta(x^n, y^n)\right] \\
&= \mathbb{P}\left[\left|\sum_{i=1}^{2^{nR}} \frac{Z_i(x^n, y^n)}{2^{nR}} - \mu(x^n, y^n)\right| > \Delta(x^n, y^n)\right] \\
&\overset{(22)}{\leq} 2\exp\left[\frac{-\frac{2^{-nR}}{2}\left(2^{nR}\Delta(x^n, y^n)\right)^2}{\mathsf{var}(Z_1(x^n, y^n)) + \frac{2^{-n(H(X,Y|W)(1-2\varepsilon)}}{(1-\varepsilon)^2}\frac{\Delta(x^n,y^n)}{3}}\right] \\
&\overset{(24)}{\leq} 2\exp\left[\frac{-\frac{2^{-nR}}{2}\left(2^{nR}\left(\eta\, Q_{XY}^{\otimes n}(x^n, y^n) - \mu(x^n, y^n)\right)\right)^2}{\frac{2^{-n(H(X,Y|W)(1-2\varepsilon)}}{(1-\varepsilon)^2}\frac{\left(\eta\, Q_{XY}^{\otimes n}(x^n,y^n) + 2\mu(x^n,y^n)\right)}{3}}\right] \\
&\overset{(26),(27)}{\leq} 2\exp\left[-\frac{(1-\varepsilon)^2\varepsilon^2}{2\eta}2^{n(R-\log_2 \ell + H(X,Y|W)(1-2\varepsilon)}\right] \\
&\leq 2\exp\left[-\frac{(1-\varepsilon)^2\varepsilon^2}{2\eta}2^{n\varepsilon|\mathcal{X}||\mathcal{Y}|}\right].
\end{aligned} \tag{28}$$

Note that we have used Bernstein's inequality [17] to bound the tail event at the third step above. We can then proceed with a union bound over all pairs $(x^n, y^n) \in \mathbf{S}(\mu)$ as follows.

$$\begin{aligned}
&\mathbb{P}\left[\bigcap_{(x^n,y^n)\in\mathbf{S}\left(\tilde{Q}_{XY}^{\mathcal{C},n}\right)} \left(\tilde{Q}_{XY}^{\mathcal{C},n}(x^n, y^n) \leq \eta\, Q_{XY}^{\otimes n}(x^n, y^n)\right)\right] \\
&\geq \mathbb{P}\left[\bigcap_{(x^n,y^n)\in\mathbf{S}(\mu)} \left(\tilde{Q}_{XY}^{\mathcal{C},n}(x^n, y^n) \leq \eta\, Q_{XY}^{\otimes n}(x^n, y^n)\right)\right] \\
&\geq \mathbb{P}\left[\bigcap_{(x^n,y^n)\in\mathbf{S}(\mu)} \left(\frac{|\tilde{Q}_{XY}^{\mathcal{C},n}(x^n, y^n) - \mu(x^n, y^n)|}{\Delta(x^n, y^n)} \leq 1\right)\right] \\
&\geq 1 - 2\,|\mathbf{S}(\mu)| \cdot \exp\left[-\frac{(1-\varepsilon)^2\varepsilon^2}{2\eta}2^{n\varepsilon|\mathcal{X}||\mathcal{Y}|}\right] \\
&\geq 1 - 2\exp\left[-\frac{(1-\varepsilon)^2\varepsilon^2}{2\eta}2^{n\varepsilon|\mathcal{X}||\mathcal{Y}|} - n\log_e|\mathcal{X}||\mathcal{Y}|\right],
\end{aligned} \tag{29}$$

which approaches unity as n diverges. Thus, for sufficiently large n, there exists a codebook $\overline{\mathcal{C}}$ with

$$\tilde{Q}_{XY}^{\overline{\mathcal{C}},n}(x^n, y^n) \leq \eta\, Q_{XY}^{\otimes n}(x^n, y^n) \text{ for all } (x^n, y^n) \in \mathbf{S}(\tilde{Q}_{XY}^{\mathcal{C},n}).$$

Then $r_{XY}^n \triangleq \frac{\eta\, Q_{XY}^{\otimes n} - \tilde{Q}_{XY}^{\overline{\mathcal{C}},n}}{\eta - 1}$ defines a pmf over $\mathcal{X}^n \times \mathcal{Y}^n$. Further, $Q_{XY}^{\otimes n}$ is expressible as a convex combination of two pmfs $\tilde{Q}_{XY}^{\overline{\mathcal{C}},n}$ and r_{XY}^n, i.e.,

$$Q_{XY}^{\otimes n} = \eta^{-1}\tilde{Q}_{XY}^{\overline{\mathcal{C}},n} + \left(1 - \eta^{-1}\right) r_{XY}^n. \tag{30}$$

The above equation essentially yields the required two-stage exact generation scheme. To generate n copies of the X and Y distributed according to Q_{XY}, the controller first generates an instance of a binary RV V with $Q_V(0) = 1 - \eta^{-1}$. The controller conveys 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 them in $\lceil n \log_2 |\mathcal{X}||\mathcal{Y}| \rceil$ bits. Note that in this case, each terminal knows both source realizations.

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 exact source-generation scheme uses no more than

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

By allowing n to grow unbounded and then allowing ε to vanish, we can see that the above quantity approaches the required limit of $I(X,Y;W) = \mathscr{W}(X;Y)$. Thus, we can build schemes for separate, exact generation of the pair of sources at rates arbitrarily close to but larger than $\mathscr{W}(X;Y)$. When combined with Remark 1, the claim then follows. ■

Theorem 3: Suppose Q_{XY} (when viewed as a matrix) consists only of rational entries and that in each row, the non-zero entries are identical (but the non-zero entries in different rows can be distinct). Then $\mathscr{E}(X;Y) = \mathscr{W}(X;Y)$.

Proof: Let $\frac{k_i}{l_i}$ be the non-zero entry appearing in the i^{th} row of Q_{XY}. Let $\ell = \mathsf{lcm}\{l_i : i = 1, \ldots, |\mathcal{X}|\}$. Let $\kappa = \mathsf{gcd}\{k_i : i = 1, \ldots |\mathcal{X}|\}$. Let $m_i \triangleq \frac{k_i}{\kappa}\frac{\ell}{l_i} \in \mathbb{N}$ and $S_i = \mathsf{S}(Q_{Y|X=i})$ for $i = 1, \ldots, |\mathcal{X}|$. Let $\mathcal{X}' = \{1, \ldots, \sum_{i=1}^{|\mathcal{X}|} m_i\}$. Define a *noiseless* but *random* channel $Q_{X'|X}$ with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{X}'$ such that: (a) $|\mathsf{S}(Q_{X'|X=i})| = m_i$; (b) $H(X'|X = i) = \log_2 m_i$; and (c) $H(X|X') = 0$. An illustration for such a channel is given in Fig. 4.

Let $Q_{X'XY} \triangleq Q_{X'|X}Q_{XY}$. Then, by definition, $X' \leftrightarrow X \leftrightarrow Y$. Further, since $H(X|X') = 0$, it is also true that $X \leftrightarrow X' \leftrightarrow Y$. Thus by Remark 3 and its analogue for Wyner common information, it follows that

$$\mathscr{E}(X';Y) = \mathscr{E}(X;Y), \tag{31}$$
$$\mathscr{W}(X';Y) = \mathscr{W}(X;Y). \tag{32}$$

For $m \in \mathbb{N}$ and $S \subseteq \{1, \ldots, |\mathcal{Y}|\}$, let $\mathbb{I}_{m,S}$ be a $m \times |\mathcal{Y}|$-dimensional $\{0,1\}$-matrix with ones at locations (i,j) if and only if $j \in S$. Then, by definition, it can be seen that

$$Q_{X'Y} \equiv \frac{\kappa}{\ell}\begin{bmatrix} \mathbb{I}_{m_1,S_1} \\ \mathbb{I}_{m_2,S_2} \\ \vdots \\ \mathbb{I}_{m_{|\mathcal{X}|},S_{|\mathcal{X}|}} \end{bmatrix}. \tag{33}$$

The claim then follows from (31), (32), and by noticing that $Q_{X'Y}$ meets the structural requirement of Theorem 2. ■

Since X and Y are interchangeable, the following also holds.

Remark 4: Suppose Q_{XY} (when viewed as a matrix) consists only of rational entries and that in each column, the non-zero entries are identical (but the non-zero entries in different columns can differ). Then $\mathscr{E}(X;Y) = \mathscr{W}(X;Y)$.

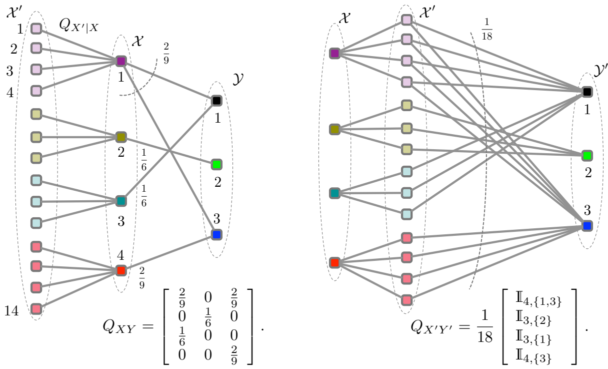
$\mathcal{X}'$
$Q_{X'|X}$
$\mathcal{X}$
$\mathcal{Y}$
1
2
3
4
14
$\frac{2}{9}$
$\frac{1}{6}$
$\frac{1}{6}$
$\frac{2}{9}$
$Q_{XY} = \begin{bmatrix} \frac{2}{9} & 0 & \frac{2}{9} \\ 0 & \frac{1}{6} & 0 \\ \frac{1}{6} & 0 & 0 \\ 0 & 0 & \frac{2}{9} \end{bmatrix}.$
$\mathcal{X}$
$\mathcal{X}'$
$\frac{1}{18}$
$\mathcal{Y}'$
$Q_{X'Y'} = \frac{1}{18}\begin{bmatrix} \mathbb{I}_{4,\{1,3\}} \\ \mathbb{I}_{3,\{2\}} \\ \mathbb{I}_{3,\{1\}} \\ \mathbb{I}_{4,\{3\}} \end{bmatrix}.$

Fig. 4. An illustration of the splitting of the source alphabet.

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] 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.

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

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] B. N. Vellambi and J. Kliewer, "Sufficient conditions for the equality of exact and Wyner common information," in 54th Annual Allerton Conference on Communication, Control, and Computing, Sep. 2016, pp. 370-377.

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

[16] 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.

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