On Lossless Coding With Coded Side Information

Daniel Marco, *Member, IEEE*, and Michelle Effros, *Fellow, IEEE*

***Abstract*—This paper considers the problem, first introduced by Ahlswede and Körner in 1975, of lossless source coding with coded side information. Specifically, let X and Y be two random variables such that X is desired losslessly at the decoder while Y serves as side information. The random variables are encoded independently, and both descriptions are used by the decoder to reconstruct X. Ahlswede and Körner describe the achievable rate region in terms of an auxiliary random variable. This paper gives a partial solution for an optimal auxiliary random variable, thereby describing part of the rate region explicitly in terms of the distribution of X and Y.**

***Index Terms*—Auxiliary random variables, coded side information, common information, lossless coding, rate region.**

I. Introduction

In 1975, Ahlswede and Körner [1] introduced the following coding problem (see Fig. 1). Random variables X and Y are independently encoded and jointly decoded. The decoder wishes to reconstruct almost losslessly only X, and so the description of Y serves as side information. Letting R_X and R_Y denote the rates used to encode X and Y, respectively, the question becomes: What rate pairs R_X and R_Y are achievable. The answer is given in terms of an auxiliary random variable in [1]. Specifically, X can be reconstructed with arbitrarily small probability of error if and only if

$$R_X \geq H(X|U)$$
$$R_Y \geq I(Y;U)$$

for some random variable U such that $X \rightarrow Y \rightarrow U$ is a Markov chain and $|\mathcal{U}| \leq |\mathcal{Y}| + 2$, where $|\mathcal{U}|$ and $|\mathcal{Y}|$ are the alphabet sizes of U and Y, respectively. The bound on $|\mathcal{U}|$ is tightened to $|\mathcal{U}| \leq |\mathcal{Y}|$ for points on the lower boundary of the rate region in [2].

The intuition behind this solution is quite simple. Random variable U can be thought of as the encoded version of Y; thus, $R_Y \geq I(Y;U)$. Since the useful part of U is then known to the decoder, the description of X requires rate $H(X|U)$. The Markov condition is quite straightforward, and the original bound on the alphabet size of U derives from Carathéodory's theorem.

Manuscript received December 26, 2006; revised January 04, 2009. Current version published June 24, 2009. This work was supported by the Center for the Mathematics of Information at California Institute of Technology. The material in this paper was presented in part at the IEEE Information Threory Workshop, Punta del Este, Uruguay, March 2006.

D. Marco was with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: danielmarco@gmail.com).

M. Effros is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: effros@caltech.edu), 626 395 3729.

Communicated by W. Szpankowski, Associate Editor for Source Coding.

Digital Object Identifier 10.1109/TIT.2009.2021309

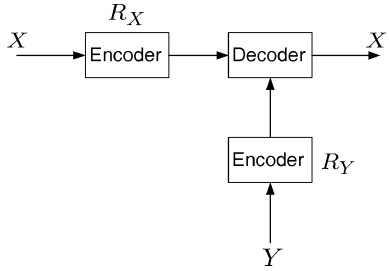

Fig. 1. Random variables X and Y are independently encoded and jointly decoded. The decoder wishes to reconstruct almost losslessly only X.

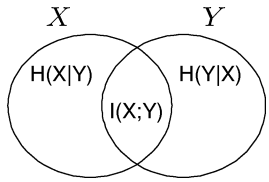

Fig. 2. The relationship between entropies and mutual information of random variables X and Y.

The above method for describing a rate region in terms of auxiliary random variables is rather common, for example [3]–[7]. These characterizations give great intuition into the basic form of a solution. Unfortunately, precise calculation of the rate region requires solution of the optimal auxiliary random variable, which is surprisingly difficult, even for simple sources [2]. For any $\epsilon > 0$, it is possible to use this characterization to approximate the rate region to within a multiplicative factor $(1+\epsilon)$ in time polynomial in $(1/\epsilon)$ by [8]. Unfortunately, numerical solutions of this type fail to provide much insight into basic questions of theoretical interest. For example, is the point $R_X = H(X|Y), R_Y = I(X;Y)$ always in the achievable rate region? Is it ever in the achievable rate region? Does achieving $R_X = H(X|Y)$ ever require $R_Y \geq H(Y)$? Furthermore, no intuition is provided as to how one should go about designing optimal auxiliary random variables. Ideally, we would like an explicit description comparable to the one given by Slepian and Wolf [9] for their famous problem.

In this paper, we give a partial solution for an optimal auxiliary random variable in Ahlswede and Körner's coding with side information problem. Thus, we describe part of the achievable rate region explicitly in terms of the distribution of X and the conditional distribution of Y given X. As a byproduct of this effort, we are able to provide answers to some of our fundamental questions regarding the relationships between random variables. For example, it is tempting to interpret the Venn diagram of [10, p. 20] (reproduced in Fig. 2) to mean that describing the information that Y holds about X at rate $R_Y = I(X;Y)$ and describing the remaining uncertainty about X at rate $H(X|Y)$ should always suffice for a complete description of X. This

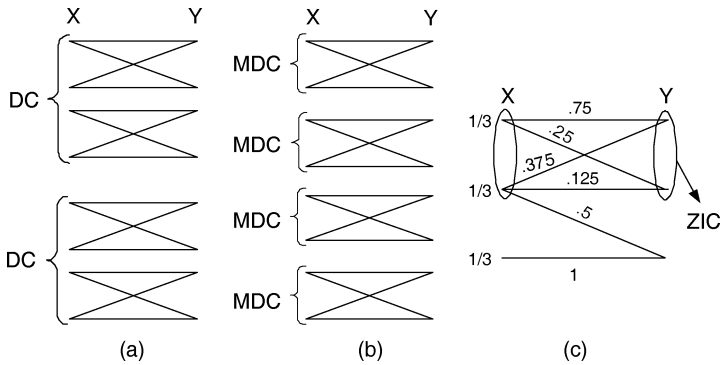

Fig. 3. The three types of components are illustrated: (a) DC, (b) MDC, (c) ZIC. A number next to a line represents the transition probability. A line with no number represents a positive transition probability $p(y|x)$.

turns out not to be the case. In fact, we show that there exist simple examples where $I(X;Y)$ is arbitrarily small, $H(Y)$ is arbitrarily large, and yet in order to make full use of the information that Y holds about X, one needs to fully describe Y, giving $R_Y \geq H(Y) \gg I(X;Y)$. This shows that the information contained in Y about X cannot be separated from the other information that Y contains, in general.

In the process of deriving the partial solution for the coding with side information problem we define two functionals of the joint distribution of X and Y. One of these functionals, $K(X;Y)$, turns out to equal the common information defined by Gács and Körner [11]. See Section V for more details.

The remainder of this paper is organized as follows. Section II introduces notation and definitions. Section III provides the main results, namely, a partial explicit description of the achievable rate region for which the structure of optimal auxiliary random variables is found. Section IV provides additional results that are useful for constructing optimal auxiliary random variables. Additionally, it outlines open questions that need to be resolved in order to obtain a complete explicit solution. In Section V, a connection is made between the functional $K(X;Y)$ defined in Section III and common information. Section VI offers concluding remarks. Finally, the Appendix A contains certain lemmas and proofs.

II. Notation and Definitions

Let X, Y, and U denote discrete random variables with finite alphabets $\mathcal{X}, \mathcal{Y}$, and $\mathcal{U}$, respectively. Set $R_X = H(X|U)$ and $R_Y = I(Y;U)$. Let $\mathcal{X}_i \subseteq \mathcal{X}$, $\mathcal{Y}_i \subseteq \mathcal{Y}$, and $\mathcal{U}_i \subseteq \mathcal{U}$ denote subsets of the possible outcomes of X, Y, and U, respectively. The index i allows us to distinguish between distinct (but possibly overlapping) subsets. Pairs $(\mathcal{X}_i, \mathcal{Y}_i)$ and $(\mathcal{Y}_i, \mathcal{U}_i)$, and triplet $(\mathcal{X}_i, \mathcal{Y}_i, \mathcal{U}_i)$ are called *components*. The functions $p(x)$, $p(y)$, $p(u)$, $p(x|y)$, $p(x|u)$, $p(y|x)$, $p(y|u)$, $p(u|x)$, and $p(u|y)$ are naturally defined marginal and conditional probabilities on the alphabets $\mathcal{X}, \mathcal{Y}$, and $\mathcal{U}$. Additionally, $p(\mathcal{X}_i) \triangleq \sum_{x \in \mathcal{X}_i} p(x)$, and $p(\mathcal{Y}_i)$ and $p(\mathcal{U}_i)$ are similarly defined. We define $\mathcal{H}(q) = -q \log_2 q$ and use the convention $\mathcal{H}(0) = 0$.

Next, we provide three definitions, which are key in the derivations that follow.

Definition 1: $(\mathcal{X}_i, \mathcal{Y}_i)$ is a *disjoint component* (DC) if

1. $\forall x \in \mathcal{X}_i,\ y \notin \mathcal{Y}_i, \quad p(y|x) = 0$
2. $\forall x \notin \mathcal{X}_i,\ y \in \mathcal{Y}_i, \quad p(y|x) = 0.$

Definition 2: $(\mathcal{X}_i, \mathcal{Y}_i)$ is a *minimal disjoint component* (MDC) if it is a DC that contains no DCs other than itself.

Definition 3: $(\mathcal{X}_i, \mathcal{Y}_i)$ is a *zero information component* (ZIC) if

1. $\forall x \in \mathcal{X}_i,\ y, y' \in \mathcal{Y}_i, \quad p(x|y) = p(x|y')$
2. $\forall x \notin \mathcal{X}_i,\ y \in \mathcal{Y}_i, \quad p(y|x) = 0.$

We call $|\mathcal{Y}_i|$ the size of the ZIC.

Fig. 3 illustrates DCs, MDCs, and ZICs. Note that in order to check that a component is a ZIC, one needs to translate from the transition probabilities $p(y|x)$ to $p(x|y)$. Note further that unlike DCs and MDCs, ZICs are not symmetrical. Specifically, $(\mathcal{X}_i, \mathcal{Y}_i)$ is a ZIC does not imply $(\mathcal{Y}_i, \mathcal{X}_i)$ is a ZIC. (In fact, if $(\mathcal{X}_i, \mathcal{Y}_i)$ is a ZIC, then $(\mathcal{Y}_i, \mathcal{X}_i)$ is a ZIC if and only if $(\mathcal{X}_i, \mathcal{Y}_i)$ is an MDC.)

MDCs are useful because they allow us to break a large problem into smaller subproblems. The importance of ZICs stems from the fact that for a ZIC, say $(\mathcal{X}_i, \mathcal{Y}_i)$, knowing that $Y \in \mathcal{Y}_i$ tells us that $x \in \mathcal{X}_i$ and specifies a conditional distribution on X that is unchanged by the knowledge of *which* $y \in \mathcal{Y}_i$ has occurred; that is, for all $x \in \mathcal{X}$

$$\Pr(X = x \,|\, Y \in \mathcal{Y}_i) = \Pr(X = x \,|\, Y = y), \quad y \in \mathcal{Y}_i.$$

(Note that this does not imply that the conditional distribution of X given $Y \in \mathcal{Y}_i$ is uniform, which ordinarily is not the case.)

Two properties of MDCs and ZICs are useful to the ensuing discussion. First, every (X,Y) imposes a unique decomposition of $(\mathcal{X}, \mathcal{Y})$ into MDCs, e.g., $(\mathcal{X}, \mathcal{Y}) = \{(\mathcal{X}_1, \mathcal{Y}_1), \ldots, (\mathcal{X}_k, \mathcal{Y}_k)\}$, where for any sets $\mathcal{A}$ and $\mathcal{B}$, $(\mathcal{A}, \mathcal{B}) = \{(\mathcal{A}_1, \mathcal{B}_1), \ldots, (\mathcal{A}_n, \mathcal{B}_n)\}$ implies $\mathcal{A} = \cup_{i=1}^n \mathcal{A}_i$ and $\mathcal{B}_1, \ldots, \mathcal{B}_n$ partitions $\mathcal{B}$. Secondly, an MDC $(\mathcal{X}_i, \mathcal{Y}_i)$ can be uniquely decomposed into largest ZICs. Specifically, $(\mathcal{X}_i, \mathcal{Y}_i) = \{(\mathcal{X}_{i1}, \mathcal{Y}_{i1}), \ldots, (\mathcal{X}_{in}, \mathcal{Y}_{in})\}$, where for each $j \in \{1, \ldots, n\}$, $(\mathcal{X}_{ij}, \mathcal{Y}_{ij})$ is a ZIC, and there does not exist a ZIC $(\bar{\mathcal{X}}, \bar{\mathcal{Y}})$ that strictly contains $(\mathcal{X}_{ij}, \mathcal{Y}_{ij})$. This can be shown by identifying ZICs and enlarging them as much as possible.

We proceed with two more definitions.

Definition 4: Let $\{(\mathcal{X}_1, \mathcal{Y}_1), \ldots, (\mathcal{X}_l, \mathcal{Y}_l)\}$ be a decomposition of $(\mathcal{X}, \mathcal{Y})$ into DCs or into largest ZICs (in both cases $\mathcal{Y}_i \cap \mathcal{Y}_j = \emptyset$, for $i \neq j$). We say that random variable U satisfies the *decomposition property* if there exits a partition

$\{\mathcal{U}_1,\ldots,\mathcal{U}_l\}$ of $\mathcal{U}$ such that for all $i \in \{1,\ldots,l\}$, $u \in \mathcal{U}_i$, and $y \notin \mathcal{Y}_i$, $p(u|y) = 0$.

Definition 5: Suppose that $\{(\mathcal{X}_1,\mathcal{Y}_1),\ldots,(\mathcal{X}_l,\mathcal{Y}_l)\}$ is a decomposition of $(\mathcal{X},\mathcal{Y})$ into DCs and that U satisfies the decomposition property. Then we define X_i, Y_i, and U_i to be the restrictions of X, Y, and U to $\mathcal{X}_i$, $\mathcal{Y}_i$, and $\mathcal{U}_i$, respectively, and call them *component random variables*. (Note that if $l > 1$, then $p(\mathcal{X}_i) = p(\mathcal{Y}_i) = p(\mathcal{U}_i) < 1$, so component random variables need not be random variables.)

It is convenient to define the operations of mutual information, entropy, and conditional entropy for component random variables. These operations are defined analogously to their standard definitions. Specifically

$$H(X_i) = \sum_{x\in\mathcal{X}_i} \mathcal{H}(p(x))$$
$$H(X_i|Y_i) = \sum_{x\in\mathcal{X}_i}\sum_{y\in\mathcal{Y}_i} p(y)\mathcal{H}(p(x|y))$$
$$I(X_i;Y_i) = \sum_{x\in\mathcal{X}_i}\sum_{y\in\mathcal{Y}_i} p(x,y)\log_2\frac{p(x,y)}{p(x)p(y)}.$$

Because the components are disjoint, the following properties hold: $H(X) = \sum_{i=1}^{l} H(X_i)$, $H(X|Y) = \sum_{i=1}^{l} H(X_i|Y_i)$, $I(X;Y) = \sum_{i=1}^{l} I(X_i;Y_i)$, and $I(X_i;U_i) \le I(X_i;Y_i)$, where the last property is the Data Processing Inequality, which holds for component random variables as well.

Finally, we define $R_{Y_i} = I(Y_i;U_i)$ to be the rate designated by the Y encoder for the i^{th} DC. It follows that $R_Y = \sum_{i=1}^{l} R_{Y_i}$.

III. Results

We focus on identifying key points in the achievable rate region. The point $(R_Y,R_X) = (0,H(X))$[1] is clearly in the achievable rate region. Likewise, $R_X = H(X|Y)$ and $R_Y = H(Y)$ is achievable. It is the auxiliary random variables $U = 0$ and $U = Y$, respectively, that attain these points. The straight line connecting these two points is an upper bound to the lower convex hull of the achievable rate region, as immediately follows from a time sharing argument. A more interesting question raised in Section I is whether one can operate at rate $R_Y < H(Y)$ while maintaining $R_X = H(X|Y)$. As noted, and is shown, the answer is sometimes yes. We define $J(X;Y)$ to be the minimal rate R_Y for which $R_X = H(X|Y)$ is achievable and note that $I(X;Y) \le J(X;Y) \le H(Y)$. Theorem 2 provides a formula for computing $J(X;Y)$. Corollary 3 and Theorem 4 give necessary and sufficient conditions under which $J(X;Y) = H(Y)$ and $J(X;Y) = I(X;Y)$, respectively.

The following lemma shows that when $H(X|U) = H(X|Y)$, U must satisfy the decomposition property. This is needed in the proofs of Theorems 2 and 4.

Lemma 1: Let $\{(\mathcal{X}_1,\mathcal{Y}_1),\ldots,(\mathcal{X}_l,\mathcal{Y}_l)\}$ be either the unique decomposition of $(\mathcal{X},\mathcal{Y})$ into largest ZICs or the decomposition into MDCs. Any auxiliary random variable U for which $H(X|U) = H(X|Y)$ must satisfy the decomposition property.

[1]Note that Fig. 4 draws R_X on the vertical axis and R_Y on the horizontal axis. We therefore report rate points as (R_Y, R_X) for consistency.

Proof: We show the lemma by showing the contrapositive. Namely, we show that if there exists a symbol $\bar{u} \in \mathcal{U}$ and symbols y_i and y_j from distinct partition sets $\mathcal{Y}_i$ and $\mathcal{Y}_j$, respectively, for which $p(\bar{u}|y_i) > 0$ and $p(\bar{u}|y_j) > 0$, then $H(X|U) > H(X|Y)$. To do so, we construct an auxiliary random variable V and show that $H(X|Y) \le H(X|V)$ and $H(X|V) < H(X|U)$. The alphabet of V is $\mathcal{V} = (\mathcal{U}\setminus\{\bar{u}\}) \cup \{\bar{v}_1,\ldots,\bar{v}_l\}$. For each $v \in \mathcal{V}\setminus\{\bar{v}_1,\ldots,\bar{v}_l\}$ and $y \in \mathcal{Y}$, set $p(v|y)$ to the probability of the corresponding symbol under the conditional distribution of U given Y. For each $k \in \{1,\ldots l\}$, set $p(\bar{v}_k|y) = p(\bar{u}|y)$ for all $y \in \mathcal{Y}_k$ and $p(\bar{v}_k|y) = 0$ for all $y \notin \mathcal{Y}_k$. (Notice that there may be $k \notin \{i,j\}$ for which $p(\bar{v}_k) = 0$, however, $p(\bar{v}_i) > 0$ and $p(\bar{v}_j) > 0$ are guaranteed by the construction since $p(\bar{u}|y_i) > 0$ and $p(\bar{u}|y_j) > 0$.) Since $X \to Y \to V$ forms a Markov chain, the Data Processing Inequality implies that $H(X|V) \ge H(X|Y)$. Thus, it suffices to show that $H(X|U) > H(X|V)$. Note that

$$\begin{aligned}
&H(X|U) - H(X|V)\\
&\quad= \Bigg(H(X|U=\bar{u})p(\bar{u}) + \sum_{u\in\mathcal{U}\setminus\{\bar{u}\}} H(X|U=u)p(u)\Bigg)\\
&\qquad- \Bigg(\sum_{k=1}^{l} H(X|V=\bar{v}_k)p(\bar{v}_k) + \sum_{u\in\mathcal{U}\setminus\{\bar{u}\}} H(X|V=u)p(u)\Bigg)\\
&\quad= H(X|U=\bar{u})p(\bar{u}) - \sum_{k=1}^{l} H(X|V=\bar{v}_k)p(\bar{v}_k)
\end{aligned}$$

by the construction of V.

Next, we consider two cases. The first case is that $\{(\mathcal{X}_1,\mathcal{Y}_1),\ldots,(\mathcal{X}_l,\mathcal{Y}_l)\}$ is the decomposition of $(\mathcal{X},\mathcal{Y})$ into largest ZICs. Notice that in this case there exists $\bar{x} \in \mathcal{X}_i \cup \mathcal{X}_j$ such that $p(\bar{x}|y_i) \ne p(\bar{x}|y_j)$ (otherwise, $p(x|y_i) = p(x|y_j)$ for all $x \in \mathcal{X}_i \cup \mathcal{X}_j$ would imply that we could form a larger ZIC by combining $(\mathcal{X}_i,\mathcal{Y}_i)$ with $(\mathcal{X}_j,\mathcal{Y}_j)$, which would give a contradiction). Therefore

$$\begin{aligned}
H(X|U=\bar{u})p(\bar{u}) &= \sum_{x\in\mathcal{X}} \mathcal{H}(p(x|\bar{u}))p(\bar{u})\\
&= \sum_{x\in\mathcal{X}} \mathcal{H}\Big(\sum_{y\in\mathcal{Y}} p(x|y)p(y|\bar{u})\Big)p(\bar{u})\\
&\overset{(a)}{>} \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \mathcal{H}(p(x|y))p(y|\bar{u})p(\bar{u})\\
&\overset{(b)}{=} \sum_{x\in\mathcal{X}}\sum_{k=1}^{l}\sum_{y\in\mathcal{Y}_k} \mathcal{H}(p(x|y))p(y,\bar{u})\\
&\overset{(c)}{=} \sum_{x\in\mathcal{X}}\sum_{k=1}^{l} \mathcal{H}(p(x|y_k)) \sum_{y\in\mathcal{Y}_k} p(y,\bar{u})\\
&\overset{(d)}{=} \sum_{x\in\mathcal{X}}\sum_{k=1}^{l} \mathcal{H}(p(x|\bar{v}_k))p(\bar{v}_k)
\end{aligned}$$

$$= \sum_{k=1}^{l} H(X|V=\bar{v}_k)p(\bar{v}_k)$$

where the strict inequality in (a) follows from the strict concavity of $\mathcal{H}$, the fact that $p(y_i|\bar{u})$ and $p(y_j|\bar{u})$ are both nonzero, and the previous observation that $p(\bar{x}|y_i) \neq p(\bar{x}|y_j)$; (b) follows since $\{\mathcal{Y}_1, \ldots, \mathcal{Y}_l\}$ is a partition of $\mathcal{Y}$; (c) derives from the fact that $p(x|y)$ is the same for all $y \in \mathcal{Y}_k$ by definition of a ZIC; and (d) follows since

$$p(x|\bar{v}_k) = \sum_{y\in\mathcal{Y}_k} p(x|y)p(y|\bar{v}_k) = p(x|y_k)\sum_{y\in\mathcal{Y}_k} p(y|\bar{v}_k) = p(x|y_k).$$

Thus, $H(X|U) - H(X|V) > 0$ when the decomposition is into largest ZICs.

When the decomposition of (X, Y) is into MDCs, there exists $x_i \in \mathcal{X}_i$ such that $p(x_i|\mathcal{Y}_i, \bar{u}) > 0$ while $p(x_i|\mathcal{Y}_j, \bar{u}) = 0$. Therefore

$$\begin{aligned}
H(X|U=\bar{u})p(\bar{u}) &= \sum_{x\in\mathcal{X}} \mathcal{H}(p(x|\bar{u}))p(\bar{u}) \\
&= \sum_{x\in\mathcal{X}} \mathcal{H}\Big(\sum_{k=1}^{l} p(x|\mathcal{Y}_k, \bar{u})p(\mathcal{Y}_k|\bar{u})\Big)p(\bar{u}) \\
&\overset{(a)}{>} \sum_{x\in\mathcal{X}}\sum_{k=1}^{l} \mathcal{H}\big(p(x|\mathcal{Y}_k, \bar{u})\big)p(\mathcal{Y}_k|\bar{u})p(\bar{u}) \\
&\overset{(b)}{=} \sum_{x\in\mathcal{X}}\sum_{k=1}^{l} \mathcal{H}\big(p(x|\mathcal{Y}_k, \bar{u})\big)p(\bar{v}_k) \\
&\overset{(c)}{=} \sum_{x\in\mathcal{X}}\sum_{k=1}^{l} \mathcal{H}\big(p(x|\bar{v}_k)\big)p(\bar{v}_k) \\
&= \sum_{k=1}^{l} H(X|V=\bar{v}_k)p(\bar{v}_k)
\end{aligned}$$

where the strict inequality in (a) derives from the strict concavity of $\mathcal{H}$, the fact that $p(\mathcal{Y}_i|\bar{u})$ and $p(\mathcal{Y}_j|\bar{u})$ are both nonzero, and the observation that $p(x_i|\mathcal{Y}_i, \bar{u}) \neq p(x_i|\mathcal{Y}_j, \bar{u})$; (b) follows since $p(\bar{u}|\mathcal{Y}_k)p(\mathcal{Y}_k) = p(\bar{v}_k|\mathcal{Y}_k)p(\mathcal{Y}_k) = p(\bar{v}_k)$; and (c) holds trivially if $p(\mathcal{Y}_k, \bar{u}) = 0$, and if $p(\mathcal{Y}_k, \bar{u}) > 0$, then

$$\begin{aligned}
p(x|\mathcal{Y}_k, \bar{u}) &= \frac{p(\bar{u}|\mathcal{Y}_k, x)p(\mathcal{Y}_k|x)p(x)}{p(\mathcal{Y}_k, \bar{u})} \\
&= \frac{p(\bar{v}_k|\mathcal{Y}_k, x)p(\mathcal{Y}_k|x)p(x)}{p(\bar{v}_k)} \\
&= \frac{p(\bar{v}_k|x)p(x)}{p(\bar{v}_k)} = p(x|\bar{v}_k). \qquad \square
\end{aligned}$$

Theorem 2: Let $\{(\mathcal{X}_1, \mathcal{Y}_1), \ldots, (\mathcal{X}_n, \mathcal{Y}_n)\}$ be the unique decomposition of $(\mathcal{X}, \mathcal{Y})$ into largest ZICs. Then

$$J(X;Y) = \sum_{i=1}^{n} \mathcal{H}(p(\mathcal{Y}_i)).$$

Proof: We obtain $J(X;Y)$ constructively. Let $\mathcal{U} = \{u_1, \ldots, u_n\}$ and set $p(u_i|y) = 1$ for all $y \in \mathcal{Y}_i$ and $p(u_i|y) = 0$ for all $y \notin \mathcal{Y}_i$. We need to show that $R_X = H(X|U) = H(X|Y)$, $R_Y = I(Y;U) = \sum_{i=1}^{n} \mathcal{H}(p(\mathcal{Y}_i))$, and R_Y can be no smaller when $R_X = H(X|Y)$. To show that $H(X|U) = H(X|Y)$ we let y_i be an arbitrary element of $\mathcal{Y}_i$ and obtain

$$\begin{aligned}
R_X = H(X|U) &= \sum_{i=1}^{n}\sum_{x\in\mathcal{X}} \mathcal{H}(p(x|u_i))p(u_i) \\
&\overset{(a)}{=} \sum_{i=1}^{n}\sum_{x\in\mathcal{X}_i} \mathcal{H}\Big(\sum_{y\in\mathcal{Y}} p(x|y, u_i)p(y|u_i)\Big)p(u_i) \\
&\overset{(b)}{=} \sum_{i=1}^{n}\sum_{x\in\mathcal{X}_i} \mathcal{H}\Big(\sum_{y\in\mathcal{Y}_i} p(x|y)p(y|u_i)\Big)p(u_i) \\
&\overset{(c)}{=} \sum_{i=1}^{n}\sum_{x\in\mathcal{X}_i} \mathcal{H}(p(x|y_i))p(u_i) \\
&= \sum_{i=1}^{n}\sum_{x\in\mathcal{X}_i} \mathcal{H}(p(x|y_i)) \sum_{y\in\mathcal{Y}_i} p(y) \\
&\overset{(d)}{=} \sum_{i=1}^{n}\sum_{y\in\mathcal{Y}_i}\sum_{x\in\mathcal{X}_i} \mathcal{H}(p(x|y))p(y) \\
&\overset{(e)}{=} \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} \mathcal{H}(p(x|y))p(y) \\
&= H(X|Y)
\end{aligned}$$

where (a) derives from the fact that $p(x|u_i) = 0$ for all $x \notin \mathcal{X}_i$; (b) follows since $X \to Y \to U$ is a Markov chain and $p(y|u_i) = 0$ for all $y \notin \mathcal{Y}_i$; (c) and (d) follow since $(\mathcal{X}_i, \mathcal{Y}_i)$ is a ZIC, and thus for any $x \in \mathcal{X}_i$, $p(x|y)$ is independent of $y \in \mathcal{Y}_i$; and (e) also follows from $(\mathcal{X}_i, \mathcal{Y}_i)$ being a ZIC, since that implies that for any $x \notin \mathcal{X}_i$, $p(x|y) = 0$ for all $y \in \mathcal{Y}_i$.

The fact that $R_Y = \sum_{i=1}^{n} \mathcal{H}(p(\mathcal{Y}_i))$ follows since U is a deterministic function of Y, giving

$$\begin{aligned}
R_Y &= I(Y;U) = H(U) - H(U|Y) = H(U) \\
&= \sum_{i=1}^{n} \mathcal{H}(p(\mathcal{Y}_i)).
\end{aligned}$$

To show that U minimizes R_Y given that $R_X = H(X|Y)$, recall that Lemma 1 shows that any auxiliary random variable W for which $H(X|W) = H(X|Y)$ has to satisfy the decomposition property. Let W be such an auxiliary random variable and use $\mathcal{W} = \{\mathcal{W}_1, \ldots, \mathcal{W}_n\}$ to denote its decomposition. Then

$$\begin{aligned}
I(Y;W) &= H(Y) - H(Y|W) \\
&\overset{(a)}{=} H(Y) - \sum_{i=1}^{n} H(Y_i\,|\,W_i) \\
&\overset{(b)}{\geq} H(Y) - \sum_{i=1}^{n} H(Y_i\,|\,U_i) \\
&\overset{(c)}{=} I(Y;U)
\end{aligned}$$

where (a) and (c) follow since $\{(\mathcal{Y}_i, \mathcal{W}_i)\}_{i=1}^{n}$ and $\{(\mathcal{Y}_i, \mathcal{U}_i)\}_{i=1}^{n}$ are DCs and thus Y_i, W_i, and U_i are component random variables; (b) follows from Lemma A1 of the Appendix, which shows that $H(Y_i|W_i)$ is maximized when $\mathcal{W}_i = \{w_i\}$, which is precisely how U is defined. $\square$

Theorem 2 enables us to improve the previous upper bound to the lower convex hull of the achievable rate region. Specifically,

the improved upper bound is the connecting line between the rate points $(0, H(X))$ and $(J(X;Y), H(X|Y))$.

Below are a direct corollary to Theorem 2 and a theorem that uses Theorem 2.

Corollary 3: $J(X;Y) = H(Y)$ if and only if (X,Y) contains no ZICs of size greater than one.

Theorem 4: Let $\{(\mathcal{X}_1,\mathcal{Y}_1),\ldots,(\mathcal{X}_k,\mathcal{Y}_k)\}$ be the decomposition of $(\mathcal{X},\mathcal{Y})$ into MDCs. Then $J(X;Y) = I(X;Y)$ if and only if each $(\mathcal{X}_i,\mathcal{Y}_i)$ is a ZIC.

Proof: We begin by showing the "if" part. Let $\{(\mathcal{X}_1,\mathcal{Y}_1),\ldots,(\mathcal{X}_k,\mathcal{Y}_k)\}$ be the decomposition of $(\mathcal{X},\mathcal{Y})$ into MDCs that are also ZICs. We show that $J(X;Y) = I(X;Y)$. We notice that the given decomposition is also into largest ZICs. Thus, applying Theorem 2 we obtain that $J(X;Y) = \sum_{i=1}^{k} \mathcal{H}(p(\mathcal{Y}_i))$. Next since $I(X;Y) = \sum_{i=1}^{k} I(X_i;Y_i)$, where X_i and Y_i are component random variables, it suffices to show that $I(X_i;Y_i) = \mathcal{H}(p(\mathcal{Y}_i))$ for all $i \in \{1,\ldots,k\}$.

Letting x_i be an arbitrary element of $\mathcal{X}_i$, we have

$$\begin{aligned} I(X_i;Y_i) &= \sum_{y\in\mathcal{Y}_i}\sum_{x\in\mathcal{X}_i} p(y|x)p(x)\log\frac{p(y|x)}{p(y)} \\ &\overset{(a)}{=} \sum_{x\in\mathcal{X}_i} p(x)\sum_{y\in\mathcal{Y}_i} p(y|x_i)\log\frac{p(y|x_i)}{p(y)} \\ &\overset{(b)}{=} p(\mathcal{X}_i)\sum_{y\in\mathcal{Y}_i}\frac{p(y)}{p(\mathcal{Y}_i)}\log\frac{1}{p(\mathcal{Y}_i)} \\ &\overset{(c)}{=} \mathcal{H}(p(\mathcal{Y}_i)). \end{aligned}$$

First (a) follows since $(\mathcal{X}_i,\mathcal{Y}_i)$ is a ZIC and an MDC implies that $(\mathcal{Y}_i,\mathcal{X}_i)$ is also a ZIC, and thus for any $y \in \mathcal{Y}_i$, $p(y|x)$ is independent of $x \in \mathcal{X}_i$. Next (b) derives from the fact that $p(y|x_i) = \frac{p(y)}{p(\mathcal{Y}_i)}$ for all $y \in \mathcal{Y}_i$, which can be seen as follows. Let y_i be an arbitrary element of $\mathcal{Y}_i$. Then

$$\begin{aligned} 1 &= \sum_{y\in\mathcal{Y}_i} p(y|x_i) = \sum_{y\in\mathcal{Y}_i}\frac{p(x_i|y)p(y)}{p(x_i)} \\ &= p(\mathcal{Y}_i)\frac{p(x_i|y_i)}{p(x_i)} = p(\mathcal{Y}_i)\frac{p(y_i|x_i)}{p(y_i)} \end{aligned}$$

where the third equality follows since $(\mathcal{X}_i,\mathcal{Y}_i)$ is a ZIC. Finally, (c) is due to $(\mathcal{X}_i,\mathcal{Y}_i)$ being an MDC, which implies $p(\mathcal{X}_i) = p(\mathcal{Y}_i)$.

Next, we let $J(X;Y) = I(X;Y)$ and show that $(\mathcal{X},\mathcal{Y})$ must decompose as given in the theorem statement. Let $\{(\mathcal{X}_1,\mathcal{Y}_1),\ldots,(\mathcal{X}_k,\mathcal{Y}_k)\}$ be the decomposition of $(\mathcal{X},\mathcal{Y})$ into MDCs. We wish to show that each $(\mathcal{X}_i,\mathcal{Y}_i)$ is a ZIC. First, observe that if $J(X;Y) = I(X;Y)$, then for an (optimal) auxiliary random variable U that achieves $R_Y = J(X;Y)$, $I(Y;U) + H(X|U) = H(X)$. Second, observe that if U is an optimal auxiliary random variable for which $R_Y = J(X;Y)$, then $R_X = H(X|U) = H(X|Y)$, and thus Lemma 1 implies that U satisfies the decomposition property. Let U be an optimal auxiliary random variable such that $R_Y = J(X;Y) = I(X;Y)$, let $\{(\mathcal{Y}_1,\mathcal{U}_1),\ldots,(\mathcal{Y}_k,\mathcal{U}_k)\}$ be the induced decomposition of $(\mathcal{Y},\mathcal{U})$ into DCs, and let X_i, Y_i, and U_i be the corresponding component random variables. The two observations above imply

$$\begin{aligned} \sum_{i=1}^{k} H(X_i) &= H(X) = I(Y;U) + H(X|U) \\ &= \sum_{i=1}^{k}\Big(I(Y_i;U_i) + H(X_i|U_i)\Big). \end{aligned}$$

The last equation together with Lemma A2 (Part B) of the Appendix, which shows that $I(Y_i;U_i) + H(X_i|U_i) \geq H(X_i)$ for all i, implies that $I(Y_i;U_i) + H(X_i|U_i) = H(X_i)$ for all i. It then follows from the condition for equality in Lemma A2 (Part B) that $(\mathcal{U}_i,\mathcal{Y}_i)$ is a ZIC, which implies via Lemma A2 (Part A) that $I(Y_i;U_i) = \mathcal{H}(p(\mathcal{Y}_i))$. Consequently, $J(X;Y) = I(Y;U) = \sum_{i=1}^{k} I(Y_i;U_i) = \sum_{i=1}^{k} \mathcal{H}(p(\mathcal{Y}_i))$. Thus

$$\sum_{i=1}^{k} I(Y_i;X_i) = I(X;Y) = J(X;Y) = \sum_{i=1}^{k} \mathcal{H}(p(\mathcal{Y}_i)).$$

Finally, letting X_i play the role of U_i in Lemma A2 (Part A), we have that $I(Y_i;X_i) \geq \mathcal{H}(p(\mathcal{Y}_i))$ for all i. This together with the last equation implies that $I(Y_i;X_i) = \mathcal{H}(p(\mathcal{Y}_i))$ for all i. Thus, it follows from the condition for equality in Lemma A2 (Part A) that $(\mathcal{X}_i,\mathcal{Y}_i)$ is a ZIC for all i, which gives the desired result. □

Corollary 3 and Theorem 4 give conditions under which $J(X;Y)$ reaches its highest and lowest possible values, respectively. Corollary 3 demonstrates that when (X,Y) lacks the special structure required to form ZICs of size larger than one, we cannot transmit the useful part of Y without describing Y completely. The following is an example where $H(Y) \gg I(X;Y)$ and yet $J(X;Y) = H(Y)$. This shows that there are cases where Y contains very little information about X and yet extracting this minuscule amount of information requires a complete description of Y.

Example 1 ($H(Y) \gg I(X;Y)$, $J(X;Y) = H(Y)$): Let $\mathcal{X} = \{x_1,x_2\}$ and suppose $p(x_1) = p(x_2) = \frac{1}{2}$. Let $\mathcal{Y} = \{y_1,\ldots,y_m\}$. For all $i \in \{1,\ldots,m\}$, let $p(y_i|x_1) = 1/m$ and $p(y_i|x_2) = q_i$, where $q_i \neq q_j$ for all $i \neq j$, and each q_i is very close, but not equal, to $1/m$. Then, since $p(y)$ and $p(y|x)$ are both approximately uniform, it follows that $H(Y) \approx \log_2 m$ and $I(X;Y) \approx 0$. Since $q_i \neq 1/m$, (X,Y) induces no ZICs of size greater than one. Consequently, Corollary 3 implies that $J(X;Y) = H(Y) \gg I(X;Y)$. ♢

Fig. 4 illustrates $J(X;Y)$ for an example pair of random variables (X,Y) for which $I(X;Y) < J(X;Y) < H(Y)$. It also shows the lower bound $R_Y + R_X \geq H(X)$, which follows from the source coding theorem. It is interesting to ask how much of this lower bound (beyond the obvious $(0, H(X))$ point) can actually be achieved and what auxiliary random variables achieve points on this lower bound. We define $K(X;Y)$ to be the maximal value of R_Y for which this lower bound is achieved with equality. Thus, $K(X;Y) = 0$ when $(R_Y, R_X) = (0, H(X))$ is the only achievable point on that lower bound and $K(X;Y) = I(X;Y)$ under the conditions of Theorem 4. Theorem 5 characterizes $K(X;Y)$ precisely.

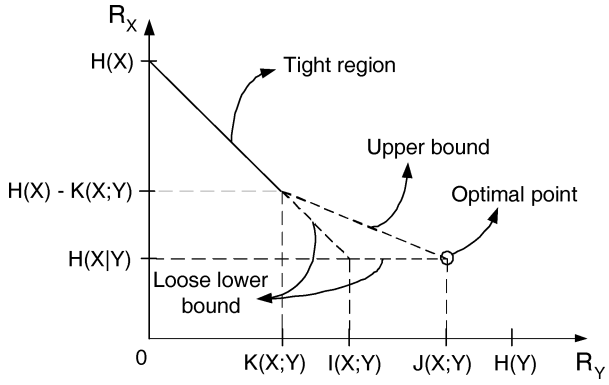

Fig. 4. The achievable rate region as known thus far.

Theorem 5: Let $\{(\mathcal{X}_1,\mathcal{Y}_1),\ldots,(\mathcal{X}_k,\mathcal{Y}_k)\}$ be the decomposition of $(\mathcal{X},\mathcal{Y})$ into MDCs. Then

$$K(X;Y)=\sum_{i=1}^{k}\mathcal{H}(p(\mathcal{Y}_i)).$$

Proof: The proof proceeds in two steps. First, we design an auxiliary random variable for which $R_Y=K(X;Y)$, as given in the theorem statement, and show that the sum–rate equals $H(X)$. Secondly, we show that if $R_Y > K(X;Y)$, then the sum–rate must be strictly greater than $H(X)$.

Let $\mathcal{U}=\{u_1,\ldots,u_k\}$ and set $p(u_i|y)=1$ for all $y\in\mathcal{Y}_i$, and $p(u_i|y)=0$ for all $y\notin\mathcal{Y}_i$. Then U is a deterministic function of X and Y, which implies

$$\begin{aligned}R_Y&=I(Y;U)=H(U)-H(U|Y)=H(U)\\&=\sum_{i=1}^{k}\mathcal{H}(p(\mathcal{Y}_i))=K(X;Y)\\R_Y+R_X&=I(Y;U)+H(X|U)\\&=H(U)+[H(X)+H(U|X)-H(U)]\\&=H(X).\end{aligned}$$

This completes the first step of the proof.

We show next that for any auxiliary random variable U, if $I(Y;U)>K(X;Y)$, then $I(Y;U)+H(X|U)>H(X)$. Theorem 7 of Section IV shows that if U is optimal and $I(Y;U)>K(X;Y)$, then U must have the decomposition property. Let X_i, Y_i, and U_i be component random variables. We write

$$\sum_{i=1}^{k}I(Y_i;U_i)=I(Y;U)>K(X;Y)=\sum_{i=1}^{k}\mathcal{H}(p(\mathcal{Y}_i))$$

where the inequality is our case assumption. Then Lemma A2 (Part A) implies that there exists $j\in\{1\ldots,k\}$ for which $I(Y_j;U_j)>\mathcal{H}(p(\mathcal{Y}_j))$ and that $(\mathcal{U}_j,\mathcal{Y}_j)$ is not a ZIC. Lemma A2 (Part B) implies that $I(Y_j;U_j)+H(X_j|U_j)>H(X_j)$. Using this and the observation that $I(Y_i;U_i)+H(X_i|U_i)\geq H(X_i)$ for all i, as follows from the Data Processing Inequality for component random variables, gives

$$\begin{aligned}I(Y;U)+H(X|U)&=\sum_{i=1}^{k}\Big(I(Y_i;U_i)+H(X_i|U_i)\Big)\\&>\sum_{i=1}^{k}H(X_i)=H(X).\end{aligned}$$

This completes the proof of the second step and of the theorem as a whole. □

From Theorem 5 we find that $K(X;Y)=0$ if and only if $(\mathcal{X},\mathcal{Y})$ is an MDC (i.e., $k=1$ in the construct of the proof, see Example 2 below) and, using Theorem 4, $K(X;Y)=I(X;Y)$ if and only if all MDCs are ZICs (see Example 3 below). The second observation follows from Theorem 4 since $K(X;Y)=I(X;Y)$ implies $(R_Y,R_X)=(I(X;Y),H(X|Y))$ is achievable. Thus, $0\leq K(X;Y)\leq I(X;Y)\leq J(X;Y)$ where the last two inequalities are either both strict or both equalities. Notice that all MDCs are ZICs if and only if X and Y are conditionally independent given the auxiliary random variable U defined in Theorem 5. Essentially the same random variable was defined in the context of studying common information in [11], which shows [11, Corollary 1] that $H(U)$ (which equals $K(X;Y)$) equals $I(X;Y)$ if and only if the conditional independence mentioned above holds. We briefly discuss the relationship between common information and coding with side information in Section V.

Examples 2 and 3 below illustrate these concepts. The first is the case where $R_Y+R_X=H(X)$ can only be achieved when $R_X=H(X)$. The second is the case where $R_Y+R_X=H(X)$ is achievable for all interesting R_X values (i.e., $H(X|Y)\leq R_X\leq H(X)$).

Example 2 ($K(X;Y)=0$): The distribution $p(y|x)$ is a binary-symmetric channel with crossover probability that is strictly between zero and one. In this case, there is only one MDC, thus $K(X;Y)=0$. ♢

Example 3 ($K(X;Y)=I(X;Y)$): The distribution $p(y|x)$ is defined by k binary-symmetric channels, each with crossover probability one half. Specifically, let $\mathcal{X}=\{x_1,x_2,\ldots,x_{2k}\}$, $\mathcal{Y}=\{y_1,y_2,\ldots,y_{2k}\}$, and for all $i\in\{1,\ldots,k\}$, let $p(y_{2i-1}|x_{2i-1})=p(y_{2i-1}|x_{2i})=p(y_{2i}|x_{2i-1})=p(y_{2i}|x_{2i})=\frac{1}{2}$ and $p(y_{2i-1}|x_m)=p(y_{2i}|x_m)=0$ for all $m\notin\{2i-1,2i\}$. In this case, $K(X;Y)=I(X;Y)=k$. ♢

Corollary 6: Any point on the line connecting $(0,H(X))$ and $(K(X;Y),H(X)-K(X;Y))$ is an optimal and achievable rate point.

Corollary 6 follows immediately from a time sharing argument. Example 4 below shows that any point on the line connecting $(0,H(X))$ and $(K(X;Y),H(X)-K(X;Y))$, can also be achieved via a direct construction.

Example 4 (Achieving Directly Any Point Between $(0,H(X))$ and $(K(X;Y),H(X)-K(X;Y))$): Let $\{(\mathcal{X}_1,\mathcal{Y}_1),\ldots,(\mathcal{X}_k,\mathcal{Y}_k)\}$ be the decomposition of $(\mathcal{X},\mathcal{Y})$ into MDCs. Let $q\in[0,\frac{1}{k}]$ be arbitrary. We define auxiliary random variable U_q with alphabet $\mathcal{U}=\{u_1,\ldots,u_k\}$ by setting $p(u_i|y)=1-(k-1)q$ for all $y\in\mathcal{Y}_i$, and $p(u_i|y)=q$ for all $y\notin\mathcal{Y}_i$. Let y_i be an arbitrary element of $\mathcal{Y}_i$. Then

$$\begin{aligned}H(U_q|X)&=\sum_{i=1}^{k}\sum_{x\in\mathcal{X}_i}\sum_{u\in\mathcal{U}}\mathcal{H}(p(u|x))p(x)\\&\overset{(a)}{=}\sum_{i=1}^{k}\sum_{x\in\mathcal{X}_i}\sum_{u\in\mathcal{U}}\mathcal{H}\Big(\sum_{y\in\mathcal{Y}_i}p(u|y)p(y|x)\Big)p(x)\end{aligned}$$

$$
\begin{aligned}
&\overset{(b)}{=} \sum_{i=1}^{k} \sum_{x\in\mathcal{X}_i} \sum_{u\in\mathcal{U}} \mathcal{H}\Big(p(u|y_i) \sum_{y\in\mathcal{Y}_i} p(y|x)\Big) p(x) \\
&= \sum_{i=1}^{k} \sum_{x\in\mathcal{X}_i} \sum_{u\in\mathcal{U}} \mathcal{H}(p(u|y_i)) p(x) \\
&= \sum_{i=1}^{k} \sum_{u\in\mathcal{U}} \mathcal{H}(p(u|y_i)) p(\mathcal{Y}_i) \\
&= \sum_{i=1}^{k} \sum_{u\in\mathcal{U}} \mathcal{H}(p(u|y_i)) \sum_{y\in\mathcal{Y}_i} p(y) \\
&\overset{(c)}{=} \sum_{i=1}^{k} \sum_{u\in\mathcal{U}} \sum_{y\in\mathcal{Y}_i} \mathcal{H}(p(u|y)) p(y) \\
&= \sum_{u\in\mathcal{U}} \sum_{y\in\mathcal{Y}} \mathcal{H}(p(u|y)) p(y) \\
&= H(U_q|Y)
\end{aligned}
$$

where (a) follows since $p(y|x) = 0$ for all $x \in \mathcal{X}_i$ and $y \notin \mathcal{Y}_i$, and (b) and (c) both follow since for any fixed u, $p(u|y)$ is the same for all $y \in \mathcal{Y}_i$. Thus

$$
\begin{aligned}
&I(Y;U_q) + H(X|U_q) \\
&\quad = [H(U_q) - H(U_q|Y)] + [H(X) + H(U_q|X) - H(U_q)] \\
&\quad = H(X).
\end{aligned}
$$

Since $I(Y;U_q)$ is a continuous function of q that equals 0 when $q = \frac{1}{k}$ and equals $K(X;Y)$ when $q = 0$, for any value $R_Y \in (0, K(X;Y))$, there exists $q \in (0, \frac{1}{k})$, for which $I(Y;U_q) = R_Y$ as follows from the Intermediate Value Theorem. ♢

IV. The $R_Y > K(X;Y)$ Region

The lower convex hull of the achievable rate region is still not known for $K(X;Y) < R_Y < J(X;Y)$. The following theorem characterizes what we know so far about optimal auxiliary random variables in that region. The optimal U used in Theorem 5 to achieve the point $(K(X;Y), H(X) - K(X;Y))$ satisfies the decomposition property. This is the lowest rate R_Y achievable by any auxiliary random variable U with this property. Theorem 7 below shows that any optimal U for any rate $R_Y > K(X;Y)$ must satisfy this property as well.

Theorem 7: Let $\{(\mathcal{X}_1, \mathcal{Y}_1), \ldots, (\mathcal{X}_k, \mathcal{Y}_k)\}$ be the decomposition of $(\mathcal{X}, \mathcal{Y})$ into MDCs. Any optimal auxiliary random variable U operating at rate $I(Y;U) > K(X;Y)$ must satisfy the decomposition property.

Proof: We show the contrapositive. Namely, we show that if U satisfies $I(Y;U) > K(X;Y)$ and U does not satisfy the decomposition property, then there exists an auxiliary random variable W satisfying the decomposition property for which $I(Y;W) = I(Y;U)$ and $H(X|W) < H(X|U)$, thus implying that U is not optimal.

We construct an auxiliary random variable V as an intermediate step in constructing W. Essentially, V is constructed from U by duplicating each u_i that is connected to several components so that each of its copies is connected to a single component. More precisely, let $\mathcal{U} = \{u_1, \ldots, u_l\}$ and define $B(u) = \{j \in \{1, \ldots, k\} : \exists y \in \mathcal{Y}_j$ such that $p(u|y) > 0\}$. The alphabet of V is $\mathcal{V} = \{v_{ij} : i \in \{1, \ldots, l\}, j \in B(u_i)\}$. For all $v_{ij} \in \mathcal{V}$, $p(v_{ij}|y) = p(u_i|y)$ for all $y \in \mathcal{Y}_j$, and $p(v_{ij}|y) = 0$ for all $y \notin \mathcal{Y}_j$. Fig. 5 illustrates this construction. Note that V satisfies the decomposition property and that

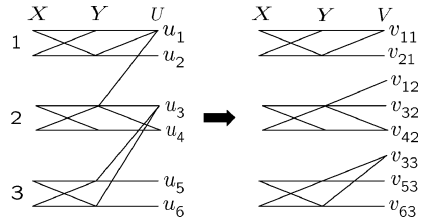

Fig. 5. Construction of V from U, where $B(u_1) = \{1, 2\}$, $B(u_2) = \{1\}$, $B(u_3) = \{2, 3\}$, $B(u_4) = \{2\}$, $B(u_5) = \{3\}$, and $B(u_6) = \{3\}$; and where v_{11} and v_{12} are generated from u_1, and v_{32} and v_{33} are generated from u_3.

$$
\begin{aligned}
H(V) &= \sum_{i=1}^{l} \sum_{j\in B(u_i)} \mathcal{H}(p(v_{ij})) \overset{(a)}{>} \sum_{i=1}^{l} \mathcal{H}\Big(\sum_{j\in B(u_i)} p(v_{ij})\Big) \\
&= \sum_{i=1}^{l} \mathcal{H}(p(u_i)) = H(U)
\end{aligned}
$$

where (a) follows from the strict concavity of $\mathcal{H}$ since U violates the decomposition property and therefore there exists some u_i for which $B(u_i)$ has more than one element. Let $\Delta = H(V) - H(U)$. While $H(V) > H(U)$, note that $H(V|Y) = H(U|Y)$ since

$$
\begin{aligned}
H(V|Y) &= \sum_{i=1}^{l} \sum_{j\in B(u_i)} \sum_{y\in\mathcal{Y}} \mathcal{H}(p(v_{ij}|y)) p(y) \\
&= \sum_{i=1}^{l} \sum_{y\in\mathcal{Y}} \mathcal{H}(p(u_i|y)) p(y) \\
&= H(U|Y).
\end{aligned}
$$

Similarly, $H(V|X) = H(U|X)$. Thus

$$
\begin{aligned}
I(Y;V) &= H(V) - H(V|Y) = H(U) + \Delta - H(U|Y) \\
&= I(Y;U) + \Delta, \\
H(X|V) &= H(X) + H(V|X) - H(V) \\
&= H(X) + H(U|X) - H(U) - \Delta \\
&= H(X|U) - \Delta.
\end{aligned}
$$

We next build W from V in a manner that maintains the decomposition property while decreasing R_Y from $I(Y;V) = I(Y;U) + \Delta$ to $I(Y;W) = I(Y;U)$ and increasing R_X from $H(X|V) = H(X|U) - \Delta$ to $H(X|W) < H(X|U)$.

We construct W by constructing a component random variable W_i for each MDC $(\mathcal{X}_i, \mathcal{Y}_i)$, $i \in \{1, \ldots, k\}$. We choose each W_i so that $I(Y_i;W_i) = I(Y_i;V_i) - \delta_i$, where $\delta_i \geq 0$ for all i and $\sum_{i=1}^{k} \delta_i = \Delta$, which gives

$$
\begin{aligned}
I(Y;W) &= \sum_{i=1}^{k} I(Y_i;W_i) = \sum_{i=1}^{k} \big(I(Y_i;V_i) - \delta_i\big) \\
&= I(Y;V) - \Delta = I(Y;U)
\end{aligned}
$$

as desired. To prove that such a W exists, we must show that the given values $I(Y_i;W_i)$ are achievable for all i. The following argument shows that for each i we can design a component random variable W_i to achieve $I(Y_i;W_i)$ equal to any value between $\mathcal{H}(p(\mathcal{Y}_i))$ and $H(Y_i)$. The upper bound is achieved when $H(Y_i|W_i)=0$, which occurs when $W_i=Y_i$ with probability one. The lower bound is achieved when $H(Y_i|W_i) = p(\mathcal{Y}_i)\sum_{y\in\mathcal{Y}_i}\mathcal{H}(\frac{p(y)}{p(\mathcal{Y}_i)})$, which occurs when $\mathcal{W}_i=\{w_i\}$ by Lemma A1. Any value between these bounds can be achieved by a time sharing argument. As a result, we can design component random variables to achieve any value of $I(Y;W)$ between $\sum_{i=1}^k \mathcal{H}(p(\mathcal{Y}_i)) = K(X;Y)$ and $\sum_{i=1}^k H(Y_i)=H(Y)$. Since by assumption $K(X;Y) < I(Y;U) \le H(Y)$, $I(Y;W) = I(Y;U)$ is achievable.

For each i we now fix $I(Y_i;W_i)$ and assume that each W_i is optimal in the sense that it minimizes $H(X_i|W_i)$ subject to the given constraint on $I(Y_i;W_i)$. It remains to show that $H(X|W)<H(X|U)$. For all $r\in[\mathcal{H}(p(\mathcal{Y}_i)),H(Y_i)]$ let $g_i(r)$ be the curve representing the lower convex hull of the achievable rate region for the ith MDC. Lemma A2 shows that for any component random variable Z_i

$$I(Y_i;Z_i)+H(X_i|Z_i)\ge H(X_i)$$

with equality if and only if $I(Y_i;Z_i)=\mathcal{H}(p(\mathcal{Y}_i))$. This implies that the curve $g_i(r)$, for $r\in(\mathcal{H}(p(\mathcal{Y}_i)),H(Y_i)]$, lies strictly above the line of slope -1 that originates at the point $(R_{Y_i},R_{X_i})=(\mathcal{H}(p(\mathcal{Y}_i)),H(X_i))$. Thus for any $\delta_i>0$ and any r satisfying $r\in(\mathcal{H}(p(\mathcal{Y}_i)),H(Y_i)-\delta_i]$

$$\frac{g_i(r)-g_i(r+\delta_i)}{\delta_i}<1 \tag{1}$$

by the convexity of g_i. See Fig. 6 for an illustration. Thus, for each i with $\delta_i>0$

$$\begin{aligned} H(X_i|W_i)-H(X_i|V_i) &\overset{(a)}{=} g_i(I(Y_i;W_i))-H(X_i|V_i)\\ &\overset{(b)}{\le} g_i(I(Y_i;W_i))-g_i(I(Y_i;V_i))\\ &\overset{(c)}{=} g_i(I(Y_i;W_i))-g_i(I(Y_i;W_i)+\delta_i)\\ &\overset{(d)}{<} \delta_i \end{aligned}$$

where (a) and (b) follow since W_i is an optimal component random variable, while V_i might not be an optimal component random variable, (c) follows from the definition of δ_i, and (d) is obtained from (1).

Finally, $H(X_i|W_i)<H(X_i|V_i)+\delta_i$ implies

$$\begin{aligned} H(X|W)&=\sum_{i=1}^k H(X_i|W_i)<\sum_{i=1}^k\big(H(X_i|V_i)+\delta_i\big)\\ &=H(X|V)+\Delta=H(X|U). \end{aligned}$$ □

Theorem 7 implies that we can find an optimal U for any $R_Y>K(X;Y)$ by allocating the rate R_Y among the MDCs and then independently finding an optimal U_i at the given rate for each MDC.

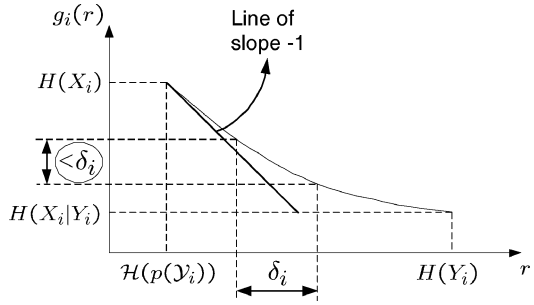

Fig. 6. The lower convex hull of the ith MDC, denoted by g_i, and a representation of (1).

Lemma 8 below, whose proof is left to the Appendix, provides necessary conditions on the structure of an optimal auxiliary random variable. Theorem 9 builds on this result, showing that an optimal auxiliary random variable for an arbitrary random pair (X,Y) can be solved by collapsing any ZIC of size greater than one in (X,Y) into a ZIC of size one in some new random pair $(\bar{X},\bar{Y})$ and then finding an optimal auxiliary random variable $\bar{U}$ for $(\bar{X},\bar{Y})$; auxiliary random variable $\bar{U}$ can be easily transformed into an optimal auxiliary random variable U for (X,Y) that achieves the same rate as $\bar{U}$ for $(\bar{X},\bar{Y})$.

Lemma 8: Let $\{(\mathcal{X}_1,\mathcal{Y}_1),\dots,(\mathcal{X}_l,\mathcal{Y}_l)\}$ be a decomposition of $(\mathcal{X},\mathcal{Y})$ into ZICs. If U is an optimal auxiliary random variable, then it must satisfy for all $u\in\mathcal{U}$ and all $i\in\{1,\dots,l\}$, $p(u|y)=p(u|y')$ for all $y,y'\in\mathcal{Y}_i$.

The necessary conditions of Lemma 8 are not sufficient. It is not difficult to construct examples of suboptimal auxiliary random variables that satisfy Lemma 8.

Theorem 9: Given an arbitrary random pair (X,Y) with alphabet $(\mathcal{X},\mathcal{Y})$ and a decomposition into ZICs $\{(\mathcal{X}_1,\mathcal{Y}_1),\dots,(\mathcal{X}_l,\mathcal{Y}_l)\}$, construct another random pair $(\bar{X},\bar{Y})$ with alphabet $(\bar{\mathcal{X}},\bar{\mathcal{Y}})=(\mathcal{X},\{\bar{y}_1,\dots,\bar{y}_l\})$ and let $\bar{X}$ have the same marginal distribution as X and $p(\bar{y}_i|x)=\sum_{y\in\mathcal{Y}_i}p(y|x)$ for all $i\in\{1,\dots,l\}$ and all $x\in\mathcal{X}$. If the auxiliary random variable $\bar{U}$ with alphabet $\bar{\mathcal{U}}=\{\bar{u}_1,\dots,\bar{u}_s\}$ is optimal for $(\bar{X},\bar{Y})$ at rate $(R_{\bar{Y}},R_{\bar{X}})$, then the auxiliary random variable U with alphabet $\mathcal{U}=\{u_1,\dots,u_s\}$ and conditional distribution $p(u_j|y)=p(\bar{u}_j|\bar{y}_i)$ for all $j\in\{1,\dots,s\}$ and all $y\in\mathcal{Y}_i$ is optimal for (X,Y) at rate $(R_Y,R_X)=(R_{\bar{Y}},R_{\bar{X}})$.

Proof: We must show both that U is optimal and that $(R_Y,R_X)=(R_{\bar{Y}},R_{\bar{X}})$ or, equivalently, that $I(Y;U)=I(\bar{Y};\bar{U})$ and $H(X|U)=H(\bar{X}|\bar{U})$. We begin by showing that $I(Y;U)=I(\bar{Y};\bar{U})$. For any $j\in\{1,\dots,s\}$

$$\begin{aligned} p(\bar{u}_j)&=\sum_{i=1}^l p(\bar{u}_j|\bar{y}_i)p(\bar{y}_i)\\ &=\sum_{i=1}^l p(\bar{u}_j|\bar{y}_i)\sum_{x\in\bar{\mathcal{X}}}p(\bar{y}_i|x)p(x)\\ &\overset{(a)}{=}\sum_{i=1}^l p(\bar{u}_j|\bar{y}_i)\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_i}p(y|x)p(x)\\ &=\sum_{i=1}^l p(\bar{u}_j|\bar{y}_i)\sum_{y\in\mathcal{Y}_i}p(y) \end{aligned}$$

$$\stackrel{(b)}{=} \sum_{i=1}^{l} \sum_{y\in\mathcal{Y}_i} p(u_j|y)p(y)$$
$$= p(u_j). \tag{2}$$

Here (a) follows since $p(\bar{y}_i|x) = \sum_{y\in\mathcal{Y}_i} p(y|x)$, $\bar{\mathcal{X}} = \mathcal{X}$, and X and $\bar{X}$ have the same marginal distribution. Then (b) follows since $p(u_j|y) = p(\bar{u}_j|\bar{y}_i)$ for all $y \in \mathcal{Y}_i$. Thus $H(U) = H(\bar{U})$. Next

$$\begin{aligned} H(\bar{U}|\bar{Y}) &= \sum_{j=1}^{s}\sum_{i=1}^{l} \mathcal{H}(p(\bar{u}_j|\bar{y}_i))p(\bar{y}_i) \\ &\stackrel{(a)}{=} \sum_{j=1}^{s}\sum_{i=1}^{l} \mathcal{H}(p(\bar{u}_j|\bar{y}_i)) \sum_{y\in\mathcal{Y}_i} p(y) \\ &\stackrel{(b)}{=} \sum_{j=1}^{s}\sum_{i=1}^{l}\sum_{y\in\mathcal{Y}_i} \mathcal{H}(p(u_j|y))p(y) \\ &= \sum_{j=1}^{s}\sum_{y\in\mathcal{Y}} \mathcal{H}(p(u_j|y))p(y) \\ &= H(U|Y) \end{aligned} \tag{3}$$

where (a) follows since

$$\begin{aligned} p(\bar{y}_i) &= \sum_{x\in\bar{\mathcal{X}}} p(\bar{y}_i|x)p(x) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_i} p(y|x)p(x) \\ &= \sum_{y\in\mathcal{Y}_i} p(y) \end{aligned}$$

and (b) derives from the fact that $p(u_j|y) = p(\bar{u}_j|\bar{y}_i)$ for all $y \in \mathcal{Y}_i$. Thus, since $H(U) = H(\bar{U})$, it follows that $I(Y;U) = I(\bar{Y};\bar{U})$.

Next, to prove that $H(X|U) = H(\bar{X}|\bar{U})$, note that

$$\begin{aligned} H(U|X) &= \sum_{j=1}^{s}\sum_{x\in\mathcal{X}} \mathcal{H}(p(u_j|x))p(x) \\ &= \sum_{j=1}^{s}\sum_{x\in\mathcal{X}} \mathcal{H}\Big(\sum_{i=1}^{l}\sum_{y\in\mathcal{Y}_i} p(u_j|y)p(y|x)\Big)p(x) \\ &\stackrel{(a)}{=} \sum_{j=1}^{s}\sum_{x\in\mathcal{X}} \mathcal{H}\Big(\sum_{i=1}^{l} p(\bar{u}_j|\bar{y}_i) \sum_{y\in\mathcal{Y}_i} p(y|x)\Big)p(x) \\ &\stackrel{(b)}{=} \sum_{j=1}^{s}\sum_{x\in\bar{\mathcal{X}}} \mathcal{H}\Big(\sum_{i=1}^{l} p(\bar{u}_j|\bar{y}_i)p(\bar{y}_i|x)\Big)p(x) \\ &= \sum_{j=1}^{s}\sum_{x\in\bar{\mathcal{X}}} \mathcal{H}(p(\bar{u}_j|x))p(x) \\ &= H(\bar{U}|\bar{X}) \end{aligned}$$

where (a) and (b) are obtained in the same way that (b) and (a), respectively, are obtained in (2). Since $H(U) = H(\bar{U})$, and X and $\bar{X}$ have the same marginal distribution, $H(U|X) = H(\bar{U}|\bar{X})$ implies $H(X|U) = H(\bar{X}|\bar{U})$.

It remains to show that U is optimal for (X,Y). The proof is by contradiction. Suppose that U is not optimal. Then there exists an auxiliary random variable W such that $I(Y;W) < I(Y;U)$ and $H(X|W) \le H(X|U)$. We use W to construct an auxiliary random variable $\bar{W}$ for $(\bar{X},\bar{Y})$ such that $I(\bar{Y};\bar{W}) = I(Y;W) < I(Y;U) = I(\bar{Y};\bar{U})$ and $H(\bar{X}|\bar{W}) = H(X|W) \le H(X|U) = H(\bar{X}|\bar{U})$. This contradicts the optimality of $\bar{U}$, and thereby gives the desired result.

We construct $\bar{W}$ from W as follows. Index the alphabet of W as $\mathcal{W} = \{w_1,\dots,w_m\}$. Since W is an optimal auxiliary random variable, Lemma 8 implies that for all $j \in \{1,\dots,m\}$, $p(w_j|y) = p(w_j|y')$ for all $y, y' \in \mathcal{Y}_i$ and all $i \in \{1,\dots,l\}$. Set $\bar{\mathcal{W}} = \{\bar{w}_1,\dots,\bar{w}_m\}$ and $p(\bar{w}_j|\bar{y}_i) = p(w_j|y)$ for $j \in \{1,\dots,m\}$, $i \in \{1,\dots,l\}$, and $y \in \mathcal{Y}_i$.

We show that $H(\bar{W}) = H(W)$ by noting that

$$\begin{aligned} p(\bar{w}_j) &= \sum_{i=1}^{l} p(\bar{w}_j|\bar{y}_i)p(\bar{y}_i) \stackrel{(a)}{=} \sum_{i=1}^{l} p(\bar{w}_j|\bar{y}_i) \sum_{y\in\mathcal{Y}_i} p(y) \\ &\stackrel{(b)}{=} \sum_{i=1}^{l}\sum_{y\in\mathcal{Y}_i} p(w_j|y)p(y) = \sum_{y\in\mathcal{Y}} p(w_j|y)p(y) \\ &= p(w_j) \end{aligned}$$

where (a) follows since $p(\bar{y}_i) = \sum_{y\in\mathcal{Y}_i} p(y)$ as shown below (3), and (b) follows from the definition of $p(\bar{w}_j|\bar{y}_i)$. By similar arguments

$$\begin{aligned} H(\bar{W}|\bar{Y}) &= \sum_{j=1}^{m}\sum_{i=1}^{l} \mathcal{H}(p(\bar{w}_j|\bar{y}_i))p(\bar{y}_i) \\ &= \sum_{j=1}^{m}\sum_{i=1}^{l}\sum_{y\in\mathcal{Y}_i} \mathcal{H}(p(w_j|y))p(y) \\ &= H(W|Y), \\ H(\bar{W}|\bar{X}) &= \sum_{j=1}^{m}\sum_{x\in\bar{\mathcal{X}}} \mathcal{H}(p(\bar{w}_j|x))p(x) \\ &= \sum_{j=1}^{m}\sum_{x\in\bar{\mathcal{X}}} \mathcal{H}\Big(\sum_{i=1}^{l} p(\bar{w}_j|\bar{y}_i)p(\bar{y}_i|x)\Big)p(x) \\ &= \sum_{j=1}^{m}\sum_{x\in\mathcal{X}} \mathcal{H}\Big(\sum_{i=1}^{l}\sum_{y\in\mathcal{Y}_i} p(w_j|y)p(y|x)\Big)p(x) \\ &= \sum_{j=1}^{m}\sum_{x\in\mathcal{X}} \mathcal{H}(p(w_j|x))p(x) \\ &= H(W|X). \end{aligned}$$

Thus, $I(\bar{Y};\bar{W}) = I(Y;W)$ and $H(\bar{X}|\bar{W}) = H(X|W)$, which gives the desired result. □

The remainder of this section focuses on the minimum alphabet size of optimal auxiliary random variables. The solution of the rate region provided in [1] bounds the alphabet size of the auxiliary random variable by the alphabet size of Y plus 2 (i.e., $|\mathcal{U}| \le |\mathcal{Y}| + 2$). In [2] it is shown that $|\mathcal{U}| \le |\mathcal{Y}|$ suffices for *optimal* auxiliary random variables. Corollary 10 combines this tighter bound with Theorem 9 and shows that $|\mathcal{U}|$ need never exceed the number of largest ZICs imposed by (X,Y). The bound is sometimes tight (for example when $R_Y = J(X;Y)$). This further reduces the space of possible optimal auxiliary random variables.

Corollary 10: Let $\{(\mathcal{X}_1,\mathcal{Y}_1),\dots,(\mathcal{X}_n,\mathcal{Y}_n)\}$ be the decomposition of $(\mathcal{X},\mathcal{Y})$ into largest ZICs. For any achievable rate

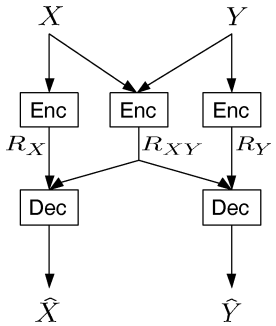

Fig. 7. The rates of the encoders are $R_X = H(X) - R$, $R_{XY} = R$, and $R_Y = H(Y) - R$ in the case of $C_{GK}(X;Y)$ and $R_X = R_X$, $R_{XY} = R$ and $R_Y = R_Y$ in the case of $C_W(X;Y)$.

point (R_Y, R_X), there exists an auxiliary random variable U such that $I(Y;U) \leq R_Y$, $H(X|U) \leq R_X$, and $|\mathcal{U}| \leq n$.

Together, Theorems 7 and 9 and Corollary 10 significantly restrict the space of possible optimal auxiliary random variables by reducing the problem to that of finding an optimal auxiliary component random variable for an MDC that has no ZICs of size greater than one, and whose alphabet size is no larger than the alphabet size of the Y component random variable.

V. Common Information and $K(X;Y)$

The notion of common information of two random variables has been addressed in [11]–[14], where various definitions have been proposed. In [11], Gács and Körner define common information by defining functions f and g for which $f(X) = g(Y)$ with probability one, and the number of values taken by f (or g) with positive probability is largest possible; the Gács–Körner common information, here denoted by $C_{GK}(X;Y)$, is then given by $C_{GK}(X;Y) = H(f(X))$. By [13, p. 404], $C_{GK}(X;Y)$ equals the largest rate R for which $(R_X, R_{XY}, R_Y) = (H(X) - R, R, H(Y) - R)$ is an achievable rate triple for the network given in Fig. 7. Furthermore, $C_{GK}(X;Y) \leq I(X;Y)$ always holds [11], [13, p. 405]. In [12], Wyner defines common information, here denoted by $C_W(X;Y)$, as the least rate R for which there exist R_X and R_Y such that $(R_X, R_{XY}, R_Y) = (R_X, R, R_Y)$ is an achievable rate triple for the same network and $R_X + R + R_Y = H(X,Y)$. Wyner shows that $C_W(X;Y) = \inf I(X,Y;W)$, where the infimum is taken over auxiliary random variables W that satisfy the Markov chain $X \to W \to Y$. Wyner also shows that $I(X;Y) \leq C_W(X;Y) \leq \min\{H(X), H(Y)\}$. It follows that

$$\begin{aligned} C_{GK}(X;Y) &\leq I(X;Y) \leq C_W(X;Y) \\ &\leq \min\{H(X), H(Y)\}. \end{aligned}$$

While both $C_{GK}(X;Y)$ and $C_W(X;Y)$ consider the rates associated with the scheme in Fig. 7, they answer different questions. Specifically, the Gács–Körner interpretation minimizes the sum–rate into each *decoder* while carrying as much of the load as possible with the central encoder. (Achieving sum–rates of $H(X)$ and $H(Y)$ into the decoders is trivial when the central encoder has rate 0 but more difficult when the central encoder is involved.) In contrast, Wyner's interpretation minimizes the sum–rate out of the three *encoders* while carrying the minimal load at the central encoder. (Achieving sum–rate $H(X,Y)$ out of the three encoders is trivial when the side encoders have rate 0 but more difficult otherwise.) Thus, $C_{GK}(X;Y)$ describes the maximal amount of shared information that is useful in describing both X and Y individually, while $C_W(X;Y)$ describes the minimal amount of shared information needed to describe X and Y jointly.

It is interesting to note that the auxiliary random variable U used to achieve $R_Y = K(X;Y)$ meets the definition of the random variable $f(X) = g(Y)$ in the definition of $C_{GK}(X;Y)$. Thus, $K(X;Y) = C_{GK}(X;Y)$ and $C_{GK}(X;Y) = 0$ if and only if $(\mathcal{X}, \mathcal{Y})$ is an MDC. Csiszár and Körner call such a distribution "indecomposable" [13, p. 403]. Further, $C_{GK}(X;Y) = I(X;Y) = C_W(X;Y)$ if and only if all MDCs are ZICs or, equivalently, if and only if there exist functions f and g such that $f(X) = g(Y)$ with probability one and X and Y are conditionally independent given f (or g) [13, p. 405].

The fact that $K(X;Y) = C_{GK}(X;Y)$ might be a little surprising since the middle encoder in Fig. 7 has access to both X and Y, while the side information in Fig. 1 has access to Y only. This shows that indeed the common information is "common" in the sense that it can be fully extracted from either X or Y separately.

VI. Conclusion

This paper considers the problem of lossless source coding with coded side information. Specifically, X and Y are two random variables that are independently encoded and jointly decoded, and only X needs to be reconstructed (losslessly). The solution to this problem, namely, the achievable rate region, is given in [1] in terms of an auxiliary random variable. In this paper, we obtain a partial solution for an optimal auxiliary random variable, thus providing part of the rate region explicitly in terms of the distribution of X and the conditional distribution of Y given X. An explicit solution of the rate region remains elusive for rates $K(X;Y) < R_Y < J(X;Y)$. Solution in this region hinges on finding a construction for an optimal auxiliary random variable for a single MDC that is not a ZIC. We also show that the alphabet size for this optimal auxiliary random variable need not exceed the number of largest ZICs in the decomposition of this MDC.

Appendix

Lemma A1: Let $(\mathcal{Y},\mathcal{U})$ be a DC and (Y,U) the corresponding component random variables. Then $H(Y|U) \leq p(\mathcal{Y}) \sum_{y \in \mathcal{Y}} \mathcal{H}\big(\frac{p(y)}{p(\mathcal{Y})}\big)$ with equality if and only if $(\mathcal{Y},\mathcal{U})$ is a ZIC, for example, when $\mathcal{U} = \{u\}$.

Proof:

$$\begin{aligned} H(Y|U) &= \sum_{y \in \mathcal{Y}} \sum_{u \in \mathcal{U}} \mathcal{H}\big(p(y|u)\big) p(u) \\ &= p(\mathcal{Y}) \sum_{y \in \mathcal{Y}} \sum_{u \in \mathcal{U}} \mathcal{H}\big(p(y|u)\big) \frac{p(u)}{p(\mathcal{Y})} \\ &\overset{(a)}{\leq} p(\mathcal{Y}) \sum_{y \in \mathcal{Y}} \mathcal{H}\Big(\sum_{u \in \mathcal{U}} \frac{p(y|u)p(u)}{p(\mathcal{Y})} \Big) \\ &= p(\mathcal{Y}) \sum_{y \in \mathcal{Y}} \mathcal{H}\Big(\frac{p(y)}{p(\mathcal{Y})} \Big) \end{aligned}$$

where (a) follows from the strict concavity of $\mathcal{H}$ and the fact that $\sum_{u\in\mathcal{U}} \frac{p(u)}{p(\mathcal{Y})} = 1$. The inequality is satisfied with equality if and only if for each $y \in \mathcal{Y}$, $p(y|u) = p(y|u')$ for all $u, u' \in \mathcal{U}$ for which $p(u), p(u') > 0$. This is equivalent to saying that $(\mathcal{Y}, \mathcal{U})$ is a ZIC. □

Lemma A2: Let $\{(\mathcal{X}_1, \mathcal{Y}_1), \ldots, (\mathcal{X}_k, \mathcal{Y}_k)\}$ be the decomposition of $(\mathcal{X}, \mathcal{Y})$ into MDCs, and let U be an auxiliary random variable that has the decomposition property. For each $i \in \{1, \ldots, k\}$ let X_i, Y_i, and U_i be component random variables for $\mathcal{X}_i, \mathcal{Y}_i$, and $\mathcal{U}_i$. Then

A. $I(Y_i; U_i) \geq \mathcal{H}(p(\mathcal{Y}_i))$,
with equality if and only if $(\mathcal{U}_i, \mathcal{Y}_i)$ is a ZIC

B. $I(Y_i; U_i) + H(X_i|U_i) \geq H(X_i)$,
with equality if and only if $(\mathcal{U}_i, \mathcal{Y}_i)$ is a ZIC.

Proof: We begin with the proof of Part A.

$$\begin{aligned} I(Y_i;U_i) &= H(U_i) - H(U_i|Y_i) \\ &= \sum_{u\in\mathcal{U}_i} \mathcal{H}\Big(\sum_{y\in\mathcal{Y}_i} p(u|y)p(y)\Big) - H(U_i|Y_i) \\ &\overset{(a)}{\geq} \sum_{u\in\mathcal{U}_i}\sum_{y\in\mathcal{Y}_i} \mathcal{H}\big(p(\mathcal{Y}_i)p(u|y)\big)\frac{p(y)}{p(\mathcal{Y}_i)} - H(U_i|Y_i) \\ &= \sum_{u\in\mathcal{U}_i}\sum_{y\in\mathcal{Y}_i} p(u,y)\log\frac{1}{p(\mathcal{Y}_i)} \\ &\quad + \sum_{u\in\mathcal{U}_i}\sum_{y\in\mathcal{Y}_i} p(u,y)\log\frac{1}{p(u|y)} - H(U_i|Y_i) \\ &= \mathcal{H}(p(\mathcal{Y}_i)) \end{aligned}$$

where (a) follows from the strict concavity of $\mathcal{H}$ and the fact that $\sum_{y\in\mathcal{Y}_i} \frac{p(y)}{p(\mathcal{Y}_i)} = 1$. Since $\frac{p(y)}{p(\mathcal{Y}_i)} > 0$ for all $y \in \mathcal{Y}_i$, (a) holds with equality if and only if $p(u|y) = p(u|y')$ for all $u \in \mathcal{U}_i$ and $y, y' \in \mathcal{Y}_i$, which means that $(\mathcal{U}_i, \mathcal{Y}_i)$ is a ZIC.

Next, consider Part B. Since $H(X_i) = H(X_i|U_i) + I(X_i; U_i)$, we prove Part B by proving that $I(Y_i; U_i) \geq I(X_i; U_i)$ or, equivalently, $H(U_i|X_i) \geq H(U_i|Y_i)$ with equality if and only if $(\mathcal{U}_i, \mathcal{Y}_i)$ is a ZIC. This generalizes the Data Processing Inequality to component random variables. Note that

$$\begin{aligned} H(U_i|X_i) &= \sum_{x\in\mathcal{X}_i}\sum_{u\in\mathcal{U}_i} \mathcal{H}(p(u|x))p(x) \\ &= \sum_{x\in\mathcal{X}_i}\sum_{u\in\mathcal{U}_i} \mathcal{H}\Big(\sum_{y\in\mathcal{Y}_i} p(u|y)p(y|x)\Big)p(x) \\ &\overset{(a)}{\geq} \sum_{x\in\mathcal{X}_i}\sum_{u\in\mathcal{U}_i}\sum_{y\in\mathcal{Y}_i} \mathcal{H}(p(u|y))p(y|x)p(x) \\ &= \sum_{u\in\mathcal{U}_i}\sum_{y\in\mathcal{Y}_i} \mathcal{H}(p(u|y))p(y) \\ &= H(U_i|Y_i) \end{aligned}$$

where (a) follows from the strict concavity of $\mathcal{H}$ and the fact that $\sum_{y\in\mathcal{Y}_i} p(y|x) = 1$. Inequality (a) holds with equality if and only if for each $u \in \mathcal{U}_i$ and $x \in \mathcal{X}_i$, $p(u|y) = p(u|y')$ for all $y, y' \in \mathcal{Y}_i$ satisfying $p(y|x) > 0$ and $p(y'|x) > 0$. We call this condition ($p(u|y) = p(u|y')$ for all $u \in \mathcal{U}_i$ and all $y, y' \in \mathcal{Y}_i$ for which there exists $x \in \mathcal{X}_i$ with $p(y|x) > 0$ and $p(y'|x) > 0$) "Condition A."

All that remains is to show that since $(\mathcal{X}_i, \mathcal{Y}_i)$ is an MDC, Condition A is satisfied if and only if $(\mathcal{U}_i, \mathcal{Y}_i)$ is a ZIC. The forward direction follows immediately from the definition of a ZIC. We next show that since $(\mathcal{X}_i, \mathcal{Y}_i)$ is an MDC, Condition A implies that $(\mathcal{U}_i, \mathcal{Y}_i)$ is a ZIC. The proof is by contradiction. Specifically, we suppose that $(\mathcal{U}_i, \mathcal{Y}_i)$ is not a ZIC and show that this implies that $(\mathcal{X}_i, \mathcal{Y}_i)$ can be broken into smaller DCs, say $(\mathcal{X}_i', \mathcal{Y}_i')$ and $(\mathcal{X}_i'', \mathcal{Y}_i'')$, which gives a contradiction since $(\mathcal{X}_i, \mathcal{Y}_i)$ is an MDC.

The following argument builds up $(\mathcal{X}_i', \mathcal{Y}_i')$ and $(\mathcal{X}_i'', \mathcal{Y}_i'')$ incrementally. If $(\mathcal{U}_i, \mathcal{Y}_i)$ is not a ZIC, then there exist $u \in \mathcal{U}_i$ and $y', y'' \in \mathcal{Y}_i$, for which $p(u|y') \neq p(u|y'')$. Let y' and y'' be the first members of $\mathcal{Y}_i'$ and $\mathcal{Y}_i''$, respectively, and initialize $\mathcal{X}_i'$ and $\mathcal{X}_i''$ as $\mathcal{X}_i' = \{x \in \mathcal{X}_i : p(y'|x) > 0\}$ and $\mathcal{X}_i'' = \{x \in \mathcal{X}_i : p(y''|x) > 0\}$. Notice that $\mathcal{X}_i'$ and $\mathcal{X}_i''$ are disjoint since Condition A allows $p(u|y') \neq p(u|y'')$ only for y' and y'' that are not accessible from the same $x \in \mathcal{X}_i$. Next, we add to $\mathcal{Y}_i'$ all members of the set $\Delta' = \{y \in \mathcal{Y}_i : p(y|x) > 0 \text{ for some } x \in \mathcal{X}_i'\}$ and to $\mathcal{Y}_i''$ all members of the set $\Delta'' = \{y \in \mathcal{Y}_i : p(y|x) > 0 \text{ for some } x \in \mathcal{X}_i''\}$. Notice that $p(u|y) = p(u|y')$ for all $u \in \mathcal{U}_i$ and $y \in \Delta'$ and $p(u|y) = p(u|y'')$ for all $u \in \mathcal{U}_i$ and $y \in \Delta''$. This observation follows from Condition A since for any $y \in \Delta'$ there exists an $x' \in \mathcal{X}_i$ for which $p(y|x') > 0$ and $p(y'|x') > 0$ and for any $y \in \Delta''$ there exists an $x'' \in \mathcal{X}_i$ for which $p(y|x'') > 0$ and $p(y''|x'') > 0$. It further implies that $\Delta' \cap \Delta'' = \varnothing$ since $p(u|y') \neq p(u|y'')$ for some $u \in \mathcal{U}_i$. This process continues, iteratively adding to $\mathcal{X}_i'$ and $\mathcal{X}_i''$ all $x \in \mathcal{X}_i$ that are newly accessible from $\mathcal{Y}_i'$ and $\mathcal{Y}_i''$, respectively, and then adding to $\mathcal{Y}_i'$ and $\mathcal{Y}_i''$ all $y \in \mathcal{Y}_i$ that are newly accessible from $\mathcal{X}_i'$ and $\mathcal{X}_i''$, respectively. At each iteration, Condition A guarantees that $\mathcal{X}_i' \cap \mathcal{X}_i'' = \mathcal{Y}_i' \cap \mathcal{Y}_i'' = \varnothing$. For example, for any y newly added to $\mathcal{Y}_i'$ there exists an $x' \in \mathcal{X}_i'$ and a $\bar{y}'$ already in $\mathcal{Y}_i'$ for which $p(y|x') > 0$ and $p(\bar{y}'|x') > 0$, which implies that $p(u|y)$ is constant for all y in the newly enlarged set $\mathcal{Y}_i'$ and thus that $\mathcal{Y}_i'$ and $\mathcal{Y}_i''$ are disjoint. Similarly, $\mathcal{X}_i'$ and $\mathcal{X}_i''$ are disjoint. Further, since $\mathcal{X}_i$ and $\mathcal{Y}_i$ are finite and the sizes of $\mathcal{X}_i', \mathcal{X}_i'', \mathcal{Y}_i'$, and $\mathcal{Y}_i''$ are nondecreasing from one iteration to the next, the procedure converges. Since the resulting $(\mathcal{X}_i', \mathcal{Y}_i')$ and $(\mathcal{X}_i'', \mathcal{Y}_i'')$ are DCs, we have the desired result. □

Proof of Lemma 8

We show the contrapositive. Namely, we show that if there exists some $u \in \mathcal{U}$ and i such that $p(u|y) \neq p(u|y')$ for some $y, y' \in \mathcal{Y}_i$, then U is not optimal. To show that U is not optimal, we construct an auxiliary random variable V such that $H(X|V) = H(X|U)$ and $I(Y;V) < I(Y;U)$.

Set $\mathcal{U}_i = \{u \in \mathcal{U} : p(u|y) > 0 \text{ for some } y \in \mathcal{Y}_i\}$; that is, $\mathcal{U}_i$ is the set of u's that are connected to $\mathcal{Y}_i$ (notice that $\mathcal{U}_i \cap \mathcal{U}_j$ $(i \neq j)$ need not be empty). By assumption, there exists $u \in \mathcal{U}_i$ such that $p(u|y) \neq p(u|y')$ for some $y, y' \in \mathcal{Y}_i$. For each $u \in \mathcal{U}_i$ define $p_i(u) = \sum_{y\in\mathcal{Y}_i} p(u|y)p(y)$; $p_i(u)$ represents the contribution to the probability of u from the members of $\mathcal{Y}_i$.

We are now ready to define V. The alphabet of V is $\mathcal{V} = \mathcal{V}_o \cup \mathcal{V}_i$. For each $u \in \mathcal{U} \setminus \mathcal{U}_i$, there is a corresponding $v \in \mathcal{V}_o$, and we define $p(v|y) = p(u|y)$ for all $y \in \mathcal{Y}$. For each $u \in \mathcal{U}_i$

there is a corresponding $v \in \mathcal{V}_i$, and we define $p(v|y) = p(u|y)$ for all $y \notin \mathcal{Y}_i$, and $p(v|y) = \frac{p_i(u)}{p(\mathcal{Y}_i)}$ for all $y \in \mathcal{Y}_i$. Notice that V satisfies the property given in the lemma with respect to $\mathcal{Y}_i$.

We begin by proving that V is a valid auxiliary random variable, namely, $\sum_{v\in\mathcal{V}} p(v|y) = 1$ for all y. The result is immediate when $y \notin \mathcal{Y}_i$, since $\sum_{v\in\mathcal{V}} p(v|y) = \sum_{u\in\mathcal{U}} p(u|y) = 1$ for all $y \notin \mathcal{Y}_i$. For $y \in \mathcal{Y}_i$ we have

$$\begin{aligned}\sum_{v\in\mathcal{V}} p(v|y) &\overset{(a)}{=} \sum_{v\in\mathcal{V}_i} p(v|y) = \sum_{u\in\mathcal{U}_i} \frac{p_i(u)}{p(\mathcal{Y}_i)} \\ &= \sum_{u\in\mathcal{U}_i}\sum_{y\in\mathcal{Y}_i} \frac{p(u|y)p(y)}{p(\mathcal{Y}_i)} = \sum_{u\in\mathcal{U}_i} \frac{p(u,\mathcal{Y}_i)}{p(\mathcal{Y}_i)} \\ &= \sum_{u\in\mathcal{U}_i} p(u|\mathcal{Y}_i) = 1\end{aligned}$$

where (a) follows since $p(v|y) = 0$ for all $v \in \mathcal{V}_o$ and $y \in \mathcal{Y}_i$.

Next, we show $H(V) = H(U)$. For any $v \in \mathcal{V}_i$ and its corresponding $u \in \mathcal{U}_i$

$$\begin{aligned}p(v) &= \sum_{y\notin\mathcal{Y}_i} p(v|y)p(y) + \sum_{y\in\mathcal{Y}_i} p(v|y)p(y) \\ &= \sum_{y\notin\mathcal{Y}_i} p(u|y)p(y) + \sum_{y\in\mathcal{Y}_i} \frac{p_i(u)}{p(\mathcal{Y}_i)}p(y) \\ &= \sum_{y\notin\mathcal{Y}_i} p(u|y)p(y) + p_i(u) \\ &= \sum_{y\notin\mathcal{Y}_i} p(u|y)p(y) + \sum_{y\in\mathcal{Y}_i} p(u|y)p(y) \\ &= p(u).\end{aligned}$$

Likewise, for any $v \in \mathcal{V}_o$ and its corresponding $u \in \mathcal{U} \setminus \mathcal{U}_i$, $p(v|y) = p(u|y)$ for all $y \in \mathcal{Y}$, which implies that $p(v) = p(u)$. Thus $H(V) = H(U)$.

Next, we show that $H(X|V) = H(X|U)$ or, equivalently, since $H(V) = H(U)$, we show that $H(V|X) = H(U|X)$. It suffices to show that for any $v \in \mathcal{V}$ and its corresponding $u \in \mathcal{U}$, $p(v|x) = p(u|x)$ for all $x \in \mathcal{X}$. Consider an arbitrary $x \in \mathcal{X}$. If $v \in \mathcal{V}_o$, then

$$\begin{aligned}p(v|x) &= \sum_{y\in\mathcal{Y}} p(v|y)p(y|x) = \sum_{y\notin\mathcal{Y}_i} p(v|y)p(y|x) \\ &= \sum_{y\notin\mathcal{Y}_i} p(u|y)p(y|x) = p(u|x)\end{aligned}$$

as needed. If $v \in \mathcal{V}_i$, then

$$\begin{aligned}p(v|x) &= \sum_{y\notin\mathcal{Y}_i} p(v|y)p(y|x) + \sum_{y\in\mathcal{Y}_i} p(v|y)p(y|x) \\ &= \sum_{y\notin\mathcal{Y}_i} p(u|y)p(y|x) + \sum_{y\in\mathcal{Y}_i} \frac{p_i(u)}{p(\mathcal{Y}_i)}p(y|x).\end{aligned}$$

Therefore, we wish to show that

$$\sum_{y\in\mathcal{Y}_i} \frac{p_i(u)}{p(\mathcal{Y}_i)}p(y|x) = \sum_{y\in\mathcal{Y}_i} p(u|y)p(y|x).$$

If $x \notin \mathcal{X}_i$, then this holds trivially, since $p(y|x) = 0$ for all $y \in \mathcal{Y}_i$. If $x \in \mathcal{X}_i$, then for any $\bar{y} \in \mathcal{Y}_i$

$$\begin{aligned}\sum_{y\in\mathcal{Y}_i} \frac{p_i(u)}{p(\mathcal{Y}_i)}p(y|x) &= \sum_{y\in\mathcal{Y}_i} \frac{p_i(u)}{p(\mathcal{Y}_i)}\frac{p(x|y)p(y)}{p(x)} \\ &\overset{(a)}{=} p_i(u)\frac{p(x|\bar{y})}{p(x)}\sum_{y\in\mathcal{Y}_i}\frac{p(y)}{p(\mathcal{Y}_i)} \\ &= \sum_{y\in\mathcal{Y}_i} \frac{p(u|y)p(y)p(x|\bar{y})}{p(x)} \\ &\overset{(b)}{=} \sum_{y\in\mathcal{Y}_i} p(u|y)\frac{p(y)p(x|y)}{p(x)} \\ &= \sum_{y\in\mathcal{Y}_i} p(u|y)p(y|x)\end{aligned}$$

where (a) and (b) follow since $(\mathcal{X}_i, \mathcal{Y}_i)$ is a ZIC, which implies that $p(x|y) = p(x|y')$ for all $y, y' \in \mathcal{Y}_i$. Hence, $H(V|X) = H(U|X)$ as desired.

It remains to show that $I(Y;V) < I(Y;U)$ or, equivalently, since $H(V) = H(U)$, that $H(V|Y) > H(U|Y)$. Recall that when $v \in \mathcal{V}_o$, its corresponding u is in $\mathcal{U} \setminus \mathcal{U}_i$, and $p(v|y) = p(u|y)$ for all $y \in \mathcal{Y}$. Likewise, when $v \in \mathcal{V}_i$, its corresponding u is in $\mathcal{U}_i$, and $p(v|y) = p(u|y)$ for all $y \notin \mathcal{Y}_i$. Thus

$$\begin{aligned}&H(V|Y) - H(U|Y) \\ &\quad= \sum_{v\in\mathcal{V}_i}\sum_{y\in\mathcal{Y}_i} \mathcal{H}(p(v|y))p(y) - \sum_{u\in\mathcal{U}_i}\sum_{y\in\mathcal{Y}_i} \mathcal{H}(p(u|y))p(y).\end{aligned}$$

Finally

$$\begin{aligned}&\sum_{v\in\mathcal{V}_i}\sum_{y\in\mathcal{Y}_i} \mathcal{H}(p(v|y))p(y) \\ &\quad= \sum_{u\in\mathcal{U}_i}\sum_{y\in\mathcal{Y}_i} \mathcal{H}\Big(\frac{p_i(u)}{p(\mathcal{Y}_i)}\Big)p(y) = \sum_{u\in\mathcal{U}_i} p(\mathcal{Y}_i)\mathcal{H}\Big(\frac{p_i(u)}{p(\mathcal{Y}_i)}\Big) \\ &\quad= \sum_{u\in\mathcal{U}_i} \mathcal{H}(p_i(u)) - \sum_{u\in\mathcal{U}_i} p_i(u)\log\frac{1}{p(\mathcal{Y}_i)} \\ &\quad\overset{(a)}{=} \sum_{u\in\mathcal{U}_i} \mathcal{H}(p_i(u)) - \mathcal{H}(p(\mathcal{Y}_i)) \\ &\quad= \sum_{u\in\mathcal{U}_i} \mathcal{H}\Big(\sum_{y\in\mathcal{Y}_i} p(u|y)p(y)\Big) - \mathcal{H}(p(\mathcal{Y}_i)) \\ &\quad= \sum_{u\in\mathcal{U}_i} \mathcal{H}\Big(\sum_{y\in\mathcal{Y}_i} p(u|y)p(\mathcal{Y}_i)\frac{p(y)}{p(\mathcal{Y}_i)}\Big) - \mathcal{H}(p(\mathcal{Y}_i)) \\ &\quad\overset{(b)}{>} \sum_{u\in\mathcal{U}_i}\sum_{y\in\mathcal{Y}_i} \mathcal{H}\big(p(u|y)p(\mathcal{Y}_i)\big)\frac{p(y)}{p(\mathcal{Y}_i)} - \mathcal{H}(p(\mathcal{Y}_i)) \\ &\quad= \sum_{u\in\mathcal{U}_i}\sum_{y\in\mathcal{Y}_i} \mathcal{H}((p(u|y))p(y) \\ &\qquad+ \sum_{u\in\mathcal{U}_i}\sum_{y\in\mathcal{Y}_i} p(u|y)p(y)\log\frac{1}{p(\mathcal{Y}_i)} - \mathcal{H}(p(\mathcal{Y}_i)) \\ &\quad= \sum_{u\in\mathcal{U}_i}\sum_{y\in\mathcal{Y}_i} \mathcal{H}((p(u|y))p(y) + \sum_{u\in\mathcal{U}_i} p_i(u)\log\frac{1}{p(\mathcal{Y}_i)} \\ &\qquad- \mathcal{H}(p(\mathcal{Y}_i)) \\ &\quad\overset{(c)}{=} \sum_{u\in\mathcal{U}_i}\sum_{y\in\mathcal{Y}_i} \mathcal{H}((p(u|y))p(y)\end{aligned}$$

where (a) and (c) follow because $p_i(u) = \sum_{y \in \mathcal{Y}_i} p(u|y)p(y)$ implies $\sum_{u \in \mathcal{U}_i} p_i(u) = p(\mathcal{Y}_i)$ since $p(u|y) = 0$ for all $u \notin \mathcal{U}_i$ and $y \in \mathcal{Y}_i$; (b) follows from the strict concavity of $\mathcal{H}$ since $\sum_{y \in \mathcal{Y}_i} \frac{p(y)}{p(\mathcal{Y}_i)} = 1$, $\frac{p(y)}{p(\mathcal{Y}_i)} > 0$ for all $y \in \mathcal{Y}_i$, and $p(u|y) \neq p(u|y')$ for some $y, y' \in \mathcal{Y}_i$ and $u \in \mathcal{U}_i$. □

REFERENCES

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

[2] W. Gu, R. Koetter, M. Effros, and T. Ho, "On source coding with coded side information for a binary source with binary side information," in *IEEE Int. Symp. Information Theory (ISIT)*, Nice, France, Jun. 2007, pp. 1456–1460.

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

[4] T. M. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 6, pp. 648–657, Nov. 1980.

[5] T. M. Cover and C. S. K. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 3, pp. 292–298, May 1981.

[6] F. M. J. Willems, "The feedback capacity region of a class of discrete memoryless multiple access channels," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 1, pp. 93–95, Jan. 1982.

[7] R. Ahlswede and T. S. Han, "On source coding with side information via a multiple-access channel and related problems in multi-user information theory," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 396–412, May 1983.

[8] W. Gu and M. Effros, "On approximating the rate region for source coding with coded side information," in *IEEE Information Theory Workshop*, Lake Tahoe, CA., Sep. 2007, pp. 432–435.

[9] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[11] P. Gács and J. Körner, "Common information is far less than mutual information," *Probl. Contr. Inf. Theory*, vol. 2, pp. 149–162, 1973.

[12] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 2, pp. 163–179, Mar. 1975.

[13] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[14] H. Yamamoto, "Coding theorems for Shannon's cipher system with correlated source outputs, and common information," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 85–95, Jan. 1994.

Daniel Marco (S'02–M'04) received the B.Sc. degree in computer engineering from the Technion–Israel Institute of Technology, Haifa, Israel, in 1999 and the M.S. degree in electrical engineering, the M.S. degree in mathematics, and the Ph.D. degree in electrical engineering, all from the University of Michigan, Ann Arbor, in 2001, 2003, and 2004, respectively.

From 2004 to 2006, he was with the California Institute of Technology, Pasadena, as a Postdoctoral Scholar. His research interests lie primarily in information theory, high- and low-resolution quantization theory, rate–distortion theory, coding with side information, and sensor networks.

Michelle Effros (S'93–M'95–SM'03–F'09) received the B.S. degree with distinction in 1989, the M.S. degree in 1990, and the Ph.D. degree in 1994, all in electrical engineering from Stanford University, Stanford, CA.

During the summers of 1988 and 1989, she worked at Hughes Aircraft Company. She joined the faculty at the California Institute of Technology, Pasadena, in 1994 and is currently a Professor of Electrical Engineering. Her research interests include information theory, network coding, data compression, communications, pattern recognition, and image processing.

Prof. Effros received Stanford's Frederick Emmons Terman Engineering Scholastic Award (for excellence in engineering) in 1989, the Hughes Masters Full-Study Fellowship in 1989, the National Science Foundation Graduate Fellowship in 1990, the AT&T Ph.D. Scholarship in 1993, the NSF CAREER Award in 1995, the Charles Lee Powell Foundation Award in 1997, the Richard Feynman-Hughes Fellowship in 1997, an Okawa Research Grant in 2000, and was cited by *Technology Review* as one of the world's top 100 young innovators in 2002. She is a member of Tau Beta Pi, Phi Beta Kappa, Sigma Xi, and the IEEE Information Theory, Signal Processing, and Communications societies. She served as the Editor of the IEEE Information Theory Society NEWSLETTER from 1995 to 1998 and as a Member of the Board of Governors of the IEEE Information Theory Society from 1998 to 2003 and since 2008 to the present. She served on the IEEE Signal Processing Society Image and Multi-Dimensional Signal Processing (IMDSP) Technical Committee since 2001. She was an Associate Editor for the joint special issue on Networking and Information Theory in the IEEE TRANSACTIONS ON INFORMATION THEORY and the IEEE/ACM TRANSACTIONS ON NETWORKING and as Associate Editor for Source Coding for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2004 to 2007.