# A Partial Solution for Lossless Source Coding with Coded Side Information

Daniel Marco
Electrical Engineering Department
California Institute of Technology
Pasadena CA 91125, USA
Email: idaniel@caltech.edu

Michelle Effros
Electrical Engineering Department
California Institute of Technology
Pasadena CA 91125, USA
Email: effros@caltech.edu

***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 the optimal auxiliary random variable, thereby describing part of the rate region explicitly in terms of the distribution of $X$ and $Y$.**

## I. Introduction

In 1975 Ahlswede and Körner [1] introduced the following coding problem. 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 was provided in [1] by means of an auxiliary random variable. Specifically, $X$ can be reconstructed with arbitrarily small probability of error if and only if

$$\begin{aligned} R_X &\geq H(X|U) \\ R_Y &\geq I(Y;U) \ , \end{aligned}$$

for some random variable $U$ such that $X \to Y \to U$ 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 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 straight forward and the bound on the alphabet size of $U$ derives from Carathéodory's theorem.

The above method for describing a rate region in terms of auxiliary random variables is quite common, for example [2] – [6]. The difficulty with such descriptions is that it is non-trivial to actually determine the rate region or even answer basic questions about it: 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 [7] for their famous problem.

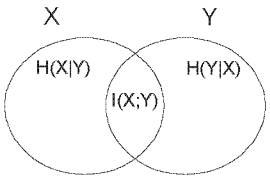


Fig. 1. The relationship between entropies and mutual information of random variables $X$ and $Y$.

In this paper we give a partial solution for the optimal auxiliary random variable of 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, the standard Venn diagram that appears in Figure 1 seems to imply 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 suffice for a complete description of $X$. This, however, turns out not to be the case. In fact, as will be shown, there exist simple examples where we can make $I(X;Y)$ arbitrarily small, and $H(Y)$ arbitrarily large, and yet in order to make full use of the information that $Y$ holds about $X$, one needs to fully describe $Y$. Equivalently, $R_Y \geq H(Y) \gg I(X;Y)$.

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 and considers the alphabet size of these variables. Additionally, it outlines open questions that

[1] This work was supported by the Center for the Mathematics of Information at California Institute of Technology.

need to be resolved in order to obtain a complete explicit solution. Finally, Section V offers concluding remarks.

Due to space limitations, certain proofs are omitted. Some proofs are briefly sketched so as to provide intuition as to how the corresponding results are obtained. The proofs of Theorems 7 and 12 are given in their entirety.

## 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 $\bar{\mathcal{X}} \subseteq \mathcal{X}$, $\bar{\mathcal{Y}} \subseteq \mathcal{Y}$, and $\bar{\mathcal{U}} \subseteq \mathcal{U}$ denote subsets of the possible outcomes of $X$, $Y$, and $U$, respectively. A pair $(\bar{\mathcal{X}}, \bar{\mathcal{Y}})$ or $(\bar{\mathcal{Y}}, \bar{\mathcal{U}})$, and a triplet $(\bar{\mathcal{X}}, \bar{\mathcal{Y}}, \bar{\mathcal{U}})$ 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. Additionally, $p(\bar{\mathcal{X}}) \triangleq \sum_{x \in \bar{\mathcal{X}}} p(x)$, and $p(\bar{\mathcal{Y}})$ and $p(\bar{\mathcal{U}})$ are similarly defined. We let $H(q) = -q \log_2 q$.

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

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

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

*Definition 2:* $(\bar{\mathcal{X}}, \bar{\mathcal{Y}})$ is a *Minimal Disjoint Component* (MDC) if it is a disjoint component that contains no disjoint components other than itself.

*Definition 3:* $(\bar{\mathcal{X}}, \bar{\mathcal{Y}})$ is a *Zero Information Component* (ZIC) if

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

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

The importance of ZICs stems from the fact that knowing that $y \in \bar{\mathcal{Y}}$ occurred gives absolutely no information as to which $x \in \bar{\mathcal{X}}$ occurred. (Note, however, that this does not imply that the conditional distribution of $X$ given $\bar{\mathcal{Y}}$ is uniform, which ordinarily is not the case.) This property will prove very useful. We notice that while disjoint components and MDCs are symmetric in their definitions, ZICs are not. Specifically, if $(\bar{\mathcal{X}}, \bar{\mathcal{Y}})$ is a ZIC, it does not imply that $(\bar{\mathcal{Y}}, \bar{\mathcal{X}})$ is a ZIC. In fact, it is not hard to see that the latter is a ZIC if and only if $(\bar{\mathcal{X}}, \bar{\mathcal{Y}})$ is also an MDC.

Next, we mention two more properties of MDCs and ZICs. First, every $(X, Y)$ imposes a unique decomposition of $(\mathcal{X}, \mathcal{Y})$ into MDCs. Secondly, an MDC $(\bar{\mathcal{X}}, \bar{\mathcal{Y}})$ can be uniquely partitioned into largest ZICs. Specifically, $(\bar{\mathcal{X}}, \bar{\mathcal{Y}}) = \{(\bar{\mathcal{X}}_1, \bar{\mathcal{Y}}_1), \ldots, (\bar{\mathcal{X}}_n, \bar{\mathcal{Y}}_n)\}$, where for each $i \in \{1, \ldots, n\}$, $(\bar{\mathcal{X}}_i, \bar{\mathcal{Y}}_i)$ is a ZIC, and there does not exist $(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}})$ that is a ZIC and strictly contains $(\bar{\mathcal{X}}_i, \bar{\mathcal{Y}}_i)$.

Finally, suppose $(X, Y)$ decomposes $(\mathcal{X}, \mathcal{Y})$ into disjoint components $\{(\bar{\mathcal{X}}_1, \bar{\mathcal{Y}}_1), \ldots, (\bar{\mathcal{X}}_l, \bar{\mathcal{Y}}_l)\}$. And suppose $\mathcal{U} = \{\bar{\mathcal{U}}_1, \ldots, \bar{\mathcal{U}}_l\}$ is such that for all $i \in \{1, \ldots, l\}$ and for all $u \in \bar{\mathcal{U}}_i$, $p(u|y) = 0$ for all $y \notin \bar{\mathcal{Y}}_i$. We define $\bar{Y}_i$ and $\bar{U}_i$ to be the restrictions of $Y$ and $U$ to $\bar{\mathcal{Y}}_i$ and $\bar{\mathcal{U}}_i$, respectively, and call them *component random variables*, (note that $p(\bar{\mathcal{Y}}_i) = p(\bar{\mathcal{U}}_i) < 1$, so these are not really random variables). We then define $R_{\bar{Y}_i} = I(\bar{Y}_i; \bar{U}_i)$, where $I(\bar{Y}_i; \bar{U}_i) = \sum_{y \in \bar{\mathcal{Y}}_i} \sum_{u \in \bar{\mathcal{U}}_i} p(y, u) \log_2 \frac{p(y,u)}{p(y)p(u)}$. It can be shown that $R_Y = \sum_{i=1}^{l} R_{\bar{Y}_i}$.

## III. Results

We focus on identifying key points in the achievable rate region. The point $R_X = H(X)$ and $R_Y = 0$ is trivially in the achievable rate region. Likewise, $R_X = H(X|Y)$ and $R_Y = H(Y)$ is achievable. It is easy to see what auxiliary random variables 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 will be shown, the answer is yes. We define $J(X;Y)$ to be the minimum rate $R_Y$ for which $R_X = H(X|Y)$ is achievable, and note that $I(X;Y) \leq J(X;Y) \leq H(Y)$.

The following three theorems provide a complete answer to the question above. Specifically, Theorem 4 shows necessary and sufficient conditions under which $J(X;Y) = H(Y)$, Theorem 5 provides necessary and sufficient conditions under which $J(X;Y) = I(X;Y)$, and Theorem 6 gives a general characterization of $J(X;Y)$ by providing a formula for computing it.

*Theorem 4:* $J(X;Y) = H(Y)$ if and only if $(X, Y)$ does not contain a ZIC of size greater than one.

*Proof sketch:* One direction is easy. Let $J(X;Y) = H(Y)$. Suppose that there does exist a ZIC $(\bar{\mathcal{X}}, \bar{\mathcal{Y}})$ with $|\bar{\mathcal{Y}}| \geq 2$. We show a contradiction. Let $\mathcal{Y} = \{y_1, \ldots, y_m\}$ and without loss of generality suppose $\bar{\mathcal{Y}} = \{y_1, \ldots, y_l\}$. Construct an auxiliary random variable $U$ as follows. $\mathcal{U} = \{u_1, u_{l+1}, u_{l+2}, \ldots, u_m\}$ such that $p(u_1|y) = 1$ for all $y \in \bar{\mathcal{Y}}$, $p(u_1|y) = 0$ for all $y \notin \bar{\mathcal{Y}}$, and $p(u_j|y_j) = 1$ and $p(u_j|y_r) = 0$ for all $j \in \{l+1, \ldots, m\}$ and $r \in \{1, \ldots, m\}$. It is now easy to see that on the one hand $R_X = H(X|U) = H(X|Y)$, since $U$ distinguishes $Y$ completely unless $Y \in \bar{\mathcal{Y}}$. Distinguishing among members of $\bar{\mathcal{Y}}$ is not necessary since $(\bar{\mathcal{X}}, \bar{\mathcal{Y}})$ is a ZIC. On the other hand, $R_Y = I(Y;U) = H(U) - H(U|Y) = H(U) < H(Y)$, where the inequality derives from the fact that $|\bar{\mathcal{Y}}| \geq 2$. This contradicts the assumption that $J(X;Y) = H(Y)$. The other direction is much longer and its proof is omitted. □

As a corollary to Theorem 4, we have the following example, where $H(Y) \gg I(X;Y)$ and yet $J(X;Y) = H(Y)$. Let $\mathcal{X} = \{x_1, x_2\}$ with both outcomes equally likely, and let $\mathcal{Y} = \{y_1, \ldots, y_m\}$. Let $p(y_1|x_1) = 1-(m-1)q$, $p(y_j|x_1) = q$, $p(y_m|x_2) = 1-(m-1)q$, and $p(y_l|x_2) = q$, where $2 \leq j \leq m$ and $1 \leq l \leq m-1$, and where $q$ is less than, but very close to, $1/m$. It is easy to see that $H(Y) \approx \log_2 m$ and that $I(X;Y) \approx 0$. Since $q < 1/m$, $(X, Y)$ induces no ZICs of size greater than one. Consequently, Theorem 4 implies that $J(X;Y) = H(Y) \gg I(X;Y)$. This example shows that there are cases in which the benefit of knowing $Y$ for the purpose of describing $X$ is minuscule, yet in order to achieve

this minuscule benefit one needs to pay a tremendous price and fully describe $Y$.

*Theorem 5:* $J(X;Y) = I(X;Y)$ if and only if $(X,Y)$ decomposes $(\mathcal{X},\mathcal{Y})$ into $\{(\bar{\mathcal{X}}_1,\bar{\mathcal{Y}}_1),\ldots,(\bar{\mathcal{X}}_k,\bar{\mathcal{Y}}_k)\}$ such that each $(\bar{\mathcal{X}}_i,\bar{\mathcal{Y}}_i)$ is an MDC and a ZIC.

*Proof sketch:* One direction is easy to see. Let each MDC $(\bar{\mathcal{X}}_i,\bar{\mathcal{Y}}_i)$ be also a ZIC. We construct an auxiliary random variable $U$ as follows. Let $\mathcal{U} = \{u_1,\ldots,u_k\}$ and set $p(u_i|y) = 1$ for all $y \in \bar{\mathcal{Y}}_i$ and $p(u_i|y) = 0$ for all $y \notin \bar{\mathcal{Y}}_i$, $i \in \{1,\ldots,k\}$. It is easy to see that $R_X = H(X|U) = H(X|Y)$ since $U$ fully describes $Y$, except for distinguishing between members of a ZIC. Furthermore, as mentioned in Section II, since each $(\bar{\mathcal{X}}_i,\bar{\mathcal{Y}}_i)$ is a ZIC and an MDC, it follows that also $(\bar{\mathcal{Y}}_i,\bar{\mathcal{X}}_i)$ is a ZIC. Thus, $X$ provides no information about $Y$ other than what component occurred in $Y$. This is the same information that $U$ contains about $Y$, and so $R_Y = I(Y;U) = I(X;Y)$ as desired. The other direction is significantly longer and the proof is omitted. □

*Theorem 6:* Let $\{(\bar{\mathcal{X}}_1,\bar{\mathcal{Y}}_1),\ldots,(\bar{\mathcal{X}}_k,\bar{\mathcal{Y}}_k)\}$ be the unique decomposition of $(\mathcal{X},\mathcal{Y})$ into MDCs imposed by $(X,Y)$. Let $(\bar{\mathcal{X}}_i,\bar{\mathcal{Y}}_i) = \{(\bar{\mathcal{X}}_{i1},\bar{\mathcal{Y}}_{i1}),\ldots,(\bar{\mathcal{X}}_{in_i},\bar{\mathcal{Y}}_{in_i})\}$ be the unique partition of $(\bar{\mathcal{X}}_i,\bar{\mathcal{Y}}_i)$ into largest ZICs. Then

$$J(X;Y) = \sum_{i=1}^{k}\sum_{j=1}^{n_i} H(p(\bar{\mathcal{Y}}_{ij})) \ .$$

*Proof sketch:* We obtain $J(X;Y)$ constructively. Let the auxiliary random variable $U$ be as follows. Let $\mathcal{U} = \{u_{11},\ldots,u_{1n_1},u_{21},\ldots,u_{2n_2},\ldots,u_{k1},\ldots,u_{kn_k},\}$ and set $p(u_{ij}|y) = 1$ for all $y \in \bar{\mathcal{Y}}_{ij}$ and $p(u_{ij}|y) = 0$ for all $y \notin \bar{\mathcal{Y}}_{ij}$. It is easy to see that $R_X = H(X|U) = H(X|Y)$. Furthermore, $U$ indeed minimizes $I(Y;U)$ given that $R_X = H(X|Y)$, which can be seen as follows. Any $U$ for which $H(X|U) = H(X|Y)$ must have the property that there is no $u \in \mathcal{U}$ for which $p(u|y) > 0$ and $p(u|y') > 0$ for some $y, y'$ that belong to different ZICs. Given this property, it is not hard to see that the $U$ above minimizes $I(Y;U)$, and hence $J(X;Y) = I(Y;U)$. Therefore,

$$\begin{aligned} J(X;Y) &= H(U) - H(U|Y) = H(U) \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} H(p(u_{ij})) = \sum_{i=1}^{k}\sum_{j=1}^{n_i} H(p(\bar{\mathcal{Y}}_{ij})) \ . \end{aligned}$$

□

Notice that when the decomposition of $(\mathcal{X},\mathcal{Y})$ imposed by $(X,Y)$ does not have ZICs of size greater than one, Theorem 6 implies $J(X;Y) = H(Y)$, which coincides with Theorem 4. Similarly, when all MDCs are ZICs, Theorem 6 implies $J(X;Y) = \sum_{i=1}^{k} H(p(\bar{\mathcal{Y}}_i)) = I(X;Y)$ (the last equality is not hard to verify), which coincides with Theorem 5.

Theorem 6 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 $(J(X;Y),H(X|Y))$[2].

[2]Note that Figure 2 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.

We now proceed by defining another key rate for $R_Y$. Given random variables $(X,Y)$ let $\{(\bar{\mathcal{X}}_1,\bar{\mathcal{Y}}_1),\ldots,(\bar{\mathcal{X}}_k,\bar{\mathcal{Y}}_k)\}$ be the decomposition of $(\mathcal{X},\mathcal{Y})$ into MDCs. Let $\mathcal{U} = \{u_1,\ldots,u_k\}$ be such that $p(u_i|y) = 1$ for all $y \in \bar{\mathcal{Y}}_i$, and $p(u_i|y) = 0$ for all $y \notin \bar{\mathcal{Y}}_i$. We define

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

We observe that $K(X;Y) = 0$ if and only if $k = 1$, i.e., if and only if $(\mathcal{X},\mathcal{Y})$ is an MDC. Additionally, $K(X;Y) = I(X;Y)$ if and only if all MDCs are ZICs, as can easily be seen (see some discussion in the proof sketch of Theorem 5). Lastly, $0 < K(X;Y) < I(X;Y)$ if and only if $k > 1$ and at least one MDC is not a ZIC, as follows from the previous two observations. To summarize, it is always true that $0 \le K(X;Y) \le I(X;Y) \le J(X;Y)$, where the last two inequalities are either both strict or are both equalities, as follows from the second observation above and Theorem 5.

*Theorem 7:* The auxiliary random variable $U$, described above, for which $R_Y = K(X;Y)$ is optimal. Equivalently, $(K(X;Y),H(X|U))$ is a point on the lower convex hull of the achievable rate region.

*Proof:* The proof derives from the following simple observation. If $R_Y + R_X = H(X)$, then $(R_Y,R_X)$ is an optimal rate point. This is easily seen, since it is not possible to obtain $X$ losslessly with a sum rate that is less than the entropy of $X$. It now follows that

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

where the last equality derives from the fact that both $X$ and $Y$ completely determine $U$. □

*Corollary 8:* 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.

Theorem 9 below shows that $K(X;Y)$ is the largest rate $R_Y$ for which the total rate in encoding $X$ and $Y$ separately is no greater than $H(X)$.

*Theorem 9:* If $I(Y;U) > K(X;Y)$, then $I(Y;U) + H(X|U) > H(X)$.

Combining Theorem 7, Corollary 8, and Theorem 9 with the previous results regarding $J(X;Y)$, we obtain a partial description of the lower convex hull of the achievable rate region, which is depicted in Figure 2. The figure shows the upper bound and loose lower bound to the lower convex hull of the achievable rate region for rates $K(X;Y) < R_Y < J(X;Y)$, and the actual lower convex hull of the achievable rate region for rates $0 \le R_Y \le K(X;Y)$ and $R_Y = J(X;Y)$.

## IV. Further Simplifications and Open Questions

The part of the lower convex hull of the achievable rate region that is still not known is that for which $K(X;Y) < R_Y < J(X;Y)$. The following theorem, however, restricts the space of possible optimal auxiliary random variables.

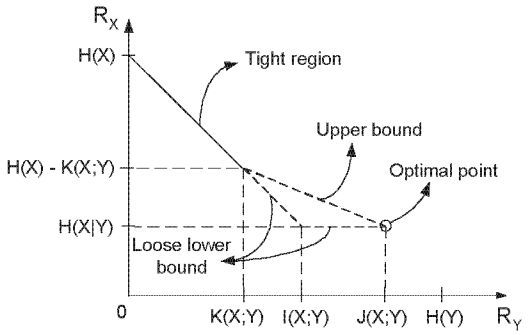


Fig. 2. The achievable rate region as known thus far.

*Theorem 10:* Let $\{(\bar{\mathcal{X}}_1, \bar{\mathcal{Y}}_1), \ldots, (\bar{\mathcal{X}}_k, \bar{\mathcal{Y}}_k)\}$ be the decomposition of $(\mathcal{X}, \mathcal{Y})$ into MDCs induced by $(X, Y)$. Any optimal auxiliary random variable $U$, operating at rate $I(Y;U) > K(X;Y)$, satisfies $\mathcal{U} = \{\bar{\mathcal{U}}_1, \ldots, \bar{\mathcal{U}}_k\}$ such that for all $i \in \{1, \ldots, k\}$ and for all $u \in \bar{\mathcal{U}}_i$, $p(u|y) = 0$ for all $y \notin \bar{\mathcal{Y}}_i$.

This theorem shows that once $R_Y > K(X;Y)$, any optimal $U$ must have a kind of a separation property, namely, it does not connect disjoint components. It follows from this theorem that the only remaining question is how to find an optimal $U$ for a single MDC. If this could be found, then an optimal $U$ for an $(X, Y)$ that imposes more than one MDC could be obtained by separately solving each component. (Of course, one would need to first choose the rate at which each component needs to operate, which would involve a rate allocation type of argument.)

The following theorem simplifies matters further and is useful in helping to focus the effort of finding an optimal auxiliary random variable.

*Theorem 11:* Let $(\widetilde{X}, \widetilde{Y}, \widetilde{U})$ be random variables with joint distribution function $\tilde{p}$ over alphabets $(\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}}, \widetilde{\mathcal{U}})$. Let $(\widetilde{X}, \widetilde{Y})$ induce the decomposition $\{(\bar{\mathcal{X}}_1, \{\tilde{y}_1\}), \ldots, (\bar{\mathcal{X}}_l, \{\tilde{y}_l\})\}$ into ZICs of size one. Suppose that $\widetilde{U}$ is an optimal auxiliary random variable for $(\widetilde{X}, \widetilde{Y})$ at rate $(R_{\widetilde{Y}}, R_{\widetilde{X}})$ with $\widetilde{\mathcal{U}} = \{\tilde{u}_1, \ldots, \tilde{u}_s\}$. Let $(X, Y, U)$ be random variables with joint distribution function $p$ over alphabets $(\mathcal{X}, \mathcal{Y}, \mathcal{U})$. Let $(X, Y)$ induce the decomposition $\{(\bar{\mathcal{X}}_1, \bar{\mathcal{Y}}_1), \ldots, (\bar{\mathcal{X}}_l, \bar{\mathcal{Y}}_l)\}$ into ZICs, where $\bar{\mathcal{X}}_i = \bar{\widetilde{\mathcal{X}}}_i$ for all $i$. If $U$ is such that $\mathcal{U} = \{u_1, \ldots, u_s\}$ and for all $j$, $p(u_j|y) = \tilde{p}(\tilde{u}_j|\tilde{y}_i)$ for all $y \in \bar{\mathcal{Y}}_i$, then $U$ is optimal for $(X, Y)$ at rate $(R_Y, R_X) = (R_{\widetilde{Y}}, R_{\widetilde{X}})$.

Theorem 11 enables us to do away with ZICs of size greater than one when searching for an optimal auxiliary random variable. Specifically, let $\{(\bar{\mathcal{X}}_1, \bar{\mathcal{Y}}_1), \ldots, (\bar{\mathcal{X}}_l, \bar{\mathcal{Y}}_l)\}$ be a decomposition of $(\mathcal{X}, \mathcal{Y})$ into ZICs. Let $(\widetilde{\mathcal{X}}, \widetilde{\mathcal{Y}}) = \{(\bar{\mathcal{X}}_1, \{y_1\}), \ldots, (\bar{\mathcal{X}}_l, \{y_l\})\}$ be such that for all $x$, $p(y_i|x) = \sum_{y \in \bar{\mathcal{Y}}_i} p(y|x)$ for any $i$. In order to find an optimal auxiliary random variable $U$ for $(X, Y)$ at some rate $R_Y$, one can instead find an optimal auxiliary random variable $\widetilde{U}$ for $(\widetilde{X}, \widetilde{Y})$ at rate $R_{\widetilde{Y}} = R_Y$, which is potentially easier. Theorem 11 then shows how to construct an optimal $U$ from $\widetilde{U}$ at the desired rate $R_Y$. In a sense, in generating $(\widetilde{X}, \widetilde{Y})$ we have performed the inverse operation of that performed in Theorem 11. Namely, instead of generating ZICs of size greater than one from $(\bar{\widetilde{\mathcal{X}}}_1, \{\tilde{y}_1\}), \ldots, (\bar{\widetilde{\mathcal{X}}}_l, \{\tilde{y}_l\})$, we collapsed the ZICs of size greater than one $(\bar{\mathcal{X}}_1, \bar{\mathcal{Y}}_1), \ldots, (\bar{\mathcal{X}}_l, \bar{\mathcal{Y}}_l)$ into the former.

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 minimum alphabet size of the auxiliary random variable by the alphabet size of $Y$ plus 2, i.e. $|\mathcal{U}| \leq |\mathcal{Y}| + 2$. We conjecture, however, that this upper bound is loose. Specifically, we conjecture that there always exists an optimal auxiliary random variable $U$ that satisfies $|\mathcal{U}| \leq |\mathcal{Y}|$ and that this bound is sometimes tight. If this conjecture it true, then it reduces further the space of possible optimal auxiliary random variables.

The fact that this bound is sometimes tight derives directly from the construction of an optimal $U$ given in the proof sketch of Theorem 6 (where it is easy to see that the constructed variable has minimal alphabet size among all possible optimal auxiliary random variables), and by selecting $(X, Y)$ that induces no ZICs of size greater than one. The more difficult part is showing that there do not exist situations in which the alphabet size of $U$ needs to exceed that of $Y$. In what follows we show that this is true for a subset of all possible rates, and for the remaining rates we provide a sketch of a proof, which is incomplete and lacks one building block that we have not yet been able to show.

The following theorem provides the subset of rates for which the tighter upper bound to the alphabet size holds.

*Theorem 12:* For any rate $0 \leq R_Y \leq K(X;Y)$ and $R_Y = J(X;Y)$, there exists an optimal auxiliary random variable $U$ such that $|\mathcal{U}| \leq |\mathcal{Y}|$.

*Proof:* If $R_Y = J(X;Y)$, then as follows from the construction in the proof sketch of Theorem 6, $|\mathcal{U}| \leq |\mathcal{Y}|$. Similarly, if $R_Y = K(X;Y)$, then the construction in the proof sketch of Theorem 5 shows that $|\mathcal{U}| \leq |\mathcal{Y}|$. If $R_Y = 0$, then $\mathcal{U} = \{u\}$ such that $p(u|y) = 1$ for all $y \in \mathcal{Y}$ is clearly optimal, and $|\mathcal{U}| \leq |\mathcal{Y}|$. It remains to consider $0 < R_Y < K(X;Y)$. We note that previously we used time sharing to obtain the lower convex hull of the achievable rate region for such rates. However, time sharing might require $|\mathcal{U}| > |\mathcal{Y}|$. Instead, we now give a non time sharing construction for $U$ that attains the lower convex hull for the given rates.

Let $\{(\bar{\mathcal{X}}_1, \bar{\mathcal{Y}}_1), \ldots, (\bar{\mathcal{X}}_K, \bar{\mathcal{Y}}_K)\}$ be the decomposition of $(\mathcal{X}, \mathcal{Y})$ into MDCs. Set $\mathcal{U} = \{u_1, \ldots, u_k\}$ to be the alphabet of $U_q$ such that for all $i$, $p(u_i|y) = 1 - (k-1)q$ for all $y \in \bar{\mathcal{Y}}_i$, and $p(u_i|y) = q$ for all $y \notin \bar{\mathcal{Y}}_i$, where $q \in [0, 1/k]$. First observe that by construction $|\mathcal{U}| \leq |\mathcal{Y}|$ as needed. Next, we show that for any rate $0 < R_Y < K(X;Y)$, there exists a value $q$ for which $I(Y;U_q) = R_Y$, and that $U_q$ is optimal. The former is seen as follows. When $q = 1/k$, all $u$'s are equally likely, from which it follows that $I(Y;U_q) = 0$. Similarly, when $q = 0$, $I(Y;U_q) = K(X;Y)$ as follows from the definition of $K(X;Y)$. Since $I(Y;U_q)$ is a continuous function of $q$, it follows via the mean value theorem that it attains all possible values in $[0, K(X;Y)]$ as $q$ ranges from 0 to $1/k$. The optimality of $U_q$ for any given $q$, is shown by demonstrating that $I(Y;U_q) + H(X|U_q) = H(X)$, which of

course matches the time sharing result. $I(Y;U_q)+H(X|U_q) = H(X) + H(U_q|X) - H(U_q|Y)$. Thus, what is left to show is that $H(U_q|X) = H(U_q|Y)$, which can be shown by explicitly writing the expressions for both conditional entropies. □

Lastly, we consider the rates $K(X;Y) < R_Y < J(X;Y)$. The following lemma asserts that the tighter upper bound $|\mathcal{U}| \leq |\mathcal{Y}|$ holds for these rates if it holds for any $(X,Y)$ that induces a single MDC.

*Lemma 13:* Suppose that for any random variables $(\bar{X},\bar{Y})$ that decompose alphabets $(\bar{\mathcal{X}},\bar{\mathcal{Y}})$ into a single MDC, there exists, for any rate, an optimal auxiliary random variable $\bar{U}$ with $|\bar{\mathcal{U}}| \leq |\bar{\mathcal{Y}}|$. Then for any random variables $(X,Y)$ over alphabets $(\mathcal{X},\mathcal{Y})$ and any rate $R_Y$, $K(X;Y) < R_Y < J(X;Y)$, there exists an optimal $U$ with $|\mathcal{U}| \leq |\mathcal{Y}|$.

*Proof sketch:* Let $(X,Y)$ induce decomposition $\{(\bar{\mathcal{X}}_1,\bar{\mathcal{Y}}_1),\ldots,(\bar{\mathcal{X}}_K,\bar{\mathcal{Y}}_K)\}$ of $(\mathcal{X},\mathcal{Y})$ into MDCs. Let $K(X;Y) < R_Y < J(X;Y)$ be any desired target rate. Theorem 10 shows that any optimal $U$ must have $\mathcal{U} = \{\bar{\mathcal{U}}_1,\ldots,\bar{\mathcal{U}}_k\}$, where for all $i$, $u \in \bar{\mathcal{U}}_i$ connects only to $y \in \bar{\mathcal{Y}}_i$. Thus, in order for $U$ to be optimal, each component random variable $\bar{U}_i$ needs to be optimal for its MDC. Given the desired rate $R_Y = I(Y;U)$, one can appropriately choose rates $R_{\bar{Y}_i}$ for which $R_Y = \sum_{i=1}^{k} R_{\bar{Y}_i}$, such that each $\bar{U}_i$ is optimal at rate $R_{\bar{Y}_i}$. Finally, since by assumption $|\bar{\mathcal{U}}_i| \leq |\bar{\mathcal{Y}}_i|$ for all $i$, it follows that $|\mathcal{U}| \leq |\mathcal{Y}|$. □

It is left to show that the premise of the lemma above is indeed true. The following conjecture makes this claim.

*Conjecture 14:* If random variables $(\bar{X},\bar{Y})$ over alphabets $(\bar{\mathcal{X}},\bar{\mathcal{Y}})$ decompose $(\bar{\mathcal{X}},\bar{\mathcal{Y}})$ into a single MDC, then for any rate there exists an optimal auxiliary random variable $\bar{U}$ such that $|\bar{\mathcal{U}}| \leq |\bar{\mathcal{Y}}|$.

Observe that since $(\bar{X},\bar{Y})$ induces a single MDC, $K(\bar{X};\bar{Y}) = 0$, thus the entire rate region is unknown. We now provide an outline for the proof of this conjecture and focus attention on a certain claim that is needed in order to obtain a complete proof, and which has yet to be proven.

The proof is by contradiction. Suppose there exists some $(\bar{X},\bar{Y})$, for which $(\bar{\mathcal{X}},\bar{\mathcal{Y}})$ is comprised of a single MDC, and some rate $R_{\bar{Y}}$, for which any optimal $\bar{U}$ satisfies $|\bar{\mathcal{U}}| \geq |\bar{\mathcal{Y}}| + 1$. Then replicate $(\bar{\mathcal{X}},\bar{\mathcal{Y}})$ three times, i.e., let $(\mathcal{X},\mathcal{Y}) = \{(\bar{\mathcal{X}}_1,\bar{\mathcal{Y}}_1),(\bar{\mathcal{X}}_2,\bar{\mathcal{Y}}_2),(\bar{\mathcal{X}}_3,\bar{\mathcal{Y}}_3)\}$, where each $(\bar{\mathcal{X}}_i,\bar{\mathcal{Y}}_i)$ is an MDC with identical conditional probabilities to $(\bar{\mathcal{X}},\bar{\mathcal{Y}})$, and where the prior probabilities have been appropriately scaled, i.e., they have all been divided by three. Theorem 10 shows that any optimal $U$ must have $\mathcal{U} = \{\bar{\mathcal{U}}_1,\bar{\mathcal{U}}_2,\bar{\mathcal{U}}_3\}$, where for each $i$, $u \in \bar{\mathcal{U}}_i$ connects only to $y \in \bar{\mathcal{Y}}_i$. Thus, in order for $U$ to be optimal, each component random variable $\bar{U}_i$ needs to be optimal for its MDC. Next, we let each $\bar{U}_i$ be an identical copy of $\bar{U}$. It can be shown that this makes $\bar{U}_i$ optimal for $(\bar{\mathcal{X}}_i,\bar{\mathcal{Y}}_i)$ at rate $R_{\bar{Y}_i} = \frac{1}{3}R_{\bar{Y}} + H(3)$, and that the minimality of the alphabet size of $\bar{U}$ implies that $\bar{U}_i$ has minimal alphabet size as well. Consequently, $U$ is optimal at rate $R_Y = 3(\frac{1}{3}R_{\bar{Y}}+H(3))$ and has minimal alphabet size.

Since $|\bar{\mathcal{U}}_i| \geq |\bar{\mathcal{Y}}_i| + 1$, it follows that $U$, which is optimal for $(X,Y)$ at rate $R_Y$ and has minimal alphabet size, has alphabet size $|\mathcal{U}| \geq |\mathcal{Y}| + 3$, which contradicts the theorem given in [1]. Therefore, it must be that for any rate $R_{\bar{Y}}$, there exists an optimal $\bar{U}$ satisfying $|\bar{\mathcal{U}}| \leq |\bar{\mathcal{Y}}|$.

There is one problem with the above proof. It assumes that it is optimal to let all $\bar{U}_i$ operate at the same rate. While this might appear to be a plausible choice, one cannot rule out the possibility that letting $\bar{U}_i$ operate at different rates that sum to $R_Y$ might yield an optimal $U$ with a smaller alphabet size. A necessary condition for this to be possible is that the lower convex hull of the achievable rate region for $(\bar{X},\bar{Y})$ not be strictly convex (i.e. be linear) over some interval whose interior contains the rate point $R_{\bar{Y}}$. We believe, however, that the lower convex hull of the achievable rate region for a pair of random variable with a single MDC is always strictly convex over the entire rate region, as some numerically evaluated examples seem to support. Therefore, we believe that Conjecture 14 is true, which is the basis for our conjecture that the minimal alphabet size of an optimal auxiliary random variable need not ever exceed the alphabet size of $Y$.

## V. Conclusions

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 the 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 $X$. Some part of the rate region remains unknown explicitly, specifically, the rates $K(X;Y) < R_Y < J(X;Y)$. This part of the region could most likely be explicitly obtained if it were known how to construct an optimal auxiliary random variable for a single MDC that is not a ZIC. Finally, we show that the alphabet size of an optimal auxiliary random variable is bounded from above by the alphabet size of $Y$ for rates $0 \leq R_Y \leq K(X;Y)$ and $R_Y = J(X;Y)$, and we conjecture that this upper bound holds for all rates.

## References


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

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

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

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

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

[6] 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. Info. Thry.*, vol. 29, no. 3, pp. 396–412, May 1983.

[7] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Info. Thry.*, vol. 19, no. 4, pp. 471–480, July 1973.