

# Characterizing the rate region of the coded side-information problem

Ingmar Land<sup>1,2</sup>, Claudio Weidmann<sup>2</sup>, Badri N. Vellambi<sup>1</sup>

<sup>1</sup> Institute for Telecommunications Research  
University of South Australia  
Mawson Lakes Boulevard  
Mawson Lakes SA 5095, Australia  
e-mail: {ingmar.land, badri.vellambi}@unisa.edu.au

<sup>2</sup> ETIS  
CNRS UMR 8051 / ENSEA / Univ Cergy-Pontoise  
6 avenue du Ponceau  
95014 Cergy-Pontoise, France  
e-mail: claudio.weidmann@ensea.fr

**Abstract**—This paper revisits earlier work on the achievable rate-region for the coded side-information problem. For specific source distributions we provide computable extreme rate points. As opposed to previous works, we present short and concise proofs and additional rate points below the time-sharing line of previously known rate points. Our results are based on a formulation as an optimization problem.

## I. INTRODUCTION

The coded side-information problem is depicted in Fig. 1. Two discrete sources,  $X \in \mathcal{X}$  and  $Y \in \mathcal{Y}$ , with finite alphabets  $\mathcal{X}$  and  $\mathcal{Y}$  and a joint distribution  $P_{XY}$ , are separately encoded by encoders  $E_X$  and  $E_Y$  at rates  $R_X$  and  $R_Y$ , respectively. At the destination, the decoder  $D_X$  seeks to *losslessly* reconstruct  $X$  as  $\hat{X}$ ;  $Y$  serves only as side information and is *not* reconstructed. (An example rate region and some properties are provided in Section V.)

Ahlsvede and Körner determined a single-letter characterization of the achievable rate region for this problem [1, 2]. A rate pair  $(R_X, R_Y)$  is achievable if and only if

$$R_X \geq H(X|U) \quad (1)$$

$$R_Y \geq I(Y; U) \quad (2)$$

for an auxiliary random variable  $U \in \mathcal{U}$  satisfying the Markov chain  $X - Y - U$ . The alphabet size  $|\mathcal{U}| \leq |\mathcal{Y}| + 2$  is sufficient, and the rate region is convex.

Although the rate region is completely defined by these conditions, it is not computable from the joint distribution  $P_{XY}$ . Marco and Effros found a way to compute two specific rate pairs on the boundary of the rate region for a given source distribution  $P_{XY}$ , namely  $(R_X = H(X) - K(X; Y), R_Y = K(X; Y))$  and  $(R_X = H(X|Y), R_Y = J(X; Y))$  [3]. In this paper, we will call these the K-point and the J-point, respectively. Before explaining these points, we discuss two properties of the rate region.

There are two obvious necessary conditions for achievable rate pairs:

$$R_X + R_Y \geq H(X) \quad (3)$$

$$R_X \geq H(X|Y). \quad (4)$$

We refer to the boundary line given by equality in the first condition as the X-line, and to the boundary line given by

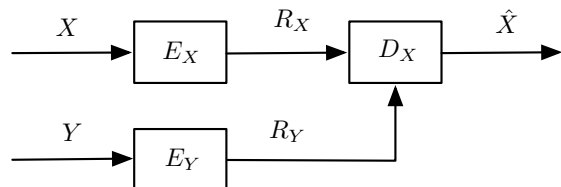


Fig. 1. Lossless reconstruction with coded side-information.

equality in the second condition as the Y-line, cf. Fig. 3. Note that the sum rate is minimal on the X-line, and the rate  $R_X$  is minimal on the Y-line. Two immediate questions arise:

Q1 Starting from the rate point  $(R_X = H(X), R_Y = 0)$ , how much can  $R_X$  be reduced before leaving the X-line?

Q2 Starting from the rate point  $(R_X = H(X|Y), R_Y = H(Y))$ , how much can  $R_Y$  be reduced before leaving the Y-line?

Q1 is equivalent to finding the minimal rate  $R_X$  such that the sum rate is minimal, i.e., meets (3) with equality. Q2 is equivalent to finding the minimal rate  $R_Y$  such that  $R_X$  is minimal, i.e., meets (4) with equality.

The answer to Q1 is the K-point, and the answer to Q2 the J-point, as defined in [3]. Marcos and Effros construct specific auxiliary random variables  $U$ , which are functions of  $Y$ , such that the rate points are achieved by  $R_X = H(X|U)$  and  $R_Y = I(Y; U)$ . As the rate region is convex, the time-sharing line between these two points is, in general, only an inner bound of the rate region.

In the present paper we present new short proofs for the K-point and the J-point. Generalizing the ideas underlying the construction of  $U$  for these two rate points, we derive new rate points on the boundary of the rate region, which are below the time-sharing line between the J-point and the K-point. We derive certain optimality conditions and show that these are fulfilled by a set of candidate rate pairs.

The rest of the paper is organized as follows. Section II and Section III deal with the new proofs for the J-point and the K-point, respectively. Section IV addresses with the optimization problem and the new rate pairs. A numerical example is presented in Section V.

Random variables are denoted by uppercase letters, their realizations by lowercase letters. For the probability of a random variable  $X$  with distribution  $P_X$ , we may simply write  $P_X(x) = p(x)$ ; similarly for joint and conditional probabilities. The support of a probability distribution  $P$  is written as  $\text{supp}(P)$ .

## II. THE K-POINT

The X-line starts at  $(R_X = H(X), R_Y = 0)$  and is formed by rate pairs satisfying

$$H(X|U) + I(Y; U) = H(X)$$

or equivalently

$$H(U|X) = H(U|Y). \quad (5)$$

Consider a graphical representation of  $P_{XY}$  as in Fig. 2, where nodes  $x$  and  $y$  are *connected* by an edge if they have positive joint probability,  $p(x, y) > 0$ . Let the random variable  $A = \phi(Y)$  be a function of  $Y$  that indexes the partition of  $P_{XY}$  into the *largest* number of (graphically) connected components  $\mathcal{X}_a^A \times \mathcal{Y}_a^A$  such that  $x \in \mathcal{X}_a^A$  implies  $\{y : p(x, y) > 0\} \subseteq \mathcal{Y}_a^A$  and  $y \in \mathcal{Y}_a^A$  implies  $\{x : p(x, y) > 0\} \subseteq \mathcal{X}_a^A$ , while  $x \in \mathcal{X}_{a'}^A$ ,  $y \in \mathcal{Y}_{a'}^A$ ,  $a' \neq a$ , implies  $p(x, y) = 0$ .

We use (5) to characterize the X-line:

$$\begin{aligned} H(U|X=x) &= H(p(U|x)) \\ &= - \sum_u \left( \sum_y p(y|x)p(u|y) \right) \log \left( \sum_y p(y|x)p(u|y) \right) \\ &= H \left( \sum_y p(y|x)p(u|y) \right) \\ &\geq \sum_y p(y|x) H(p(U|y)) \\ &= \sum_y p(y|x) H(U|Y=y), \end{aligned}$$

where the  $y$ -summations are over  $y \in \text{supp}(P_{Y|X}(\cdot|x))$ . The inequality follows from the concavity of entropy and Jensen's inequality, and thus equality holds if and only if

$$\begin{aligned} \forall x : p(u|y) = p(u|y, x) &= \sum_{y'} p(y'|x)p(u|y') \\ \iff p(u|y) = \text{const for } y \in \text{supp}(p(\cdot|x)) \end{aligned} \quad (6)$$

Since for each component  $a$ , all  $y \in \mathcal{Y}_a^A$  are connected via some  $x \in \mathcal{X}_a^A$  and  $y' \in \mathcal{Y}_a^A$ , condition (6) must hold transitively for all  $y \in \mathcal{Y}_a^A$ . We conclude that for all points on the X-line,  $p(u|y)$  must be constant over  $\mathcal{Y}_a^A$ , i.e.,  $U$  and  $Y$  are independent given  $A$  and we have the Markov chain  $Y - A - U$ .

The K-point is defined as the point on the X-line with minimal  $R_X$  or, alternatively, with maximal  $R_Y$ . We express the latter as  $I(Y; U) = H(Y) - H(Y|U)$ . Thus the K-point can be reached by minimizing  $H(Y|U)$  on the X-line. Since

on the X-line the Markov chain  $Y - A - U$  holds, we obtain

$$\begin{aligned} H(Y|U) &= H(Y|UA) + I(Y; A|U) \\ &= H(Y|A) + H(A|U) \\ &\geq H(Y|A), \end{aligned}$$

where  $I(Y; A|U) = H(A|U)$  since  $A$  is a function of  $Y$ , and equality holds in the last line if and only if  $A = f(U)$ . Thus  $H(Y|U)$  becomes minimal for  $A = f(U)$  and we will also have  $Y - U - A$ , since  $I(Y; A|U) = 0$ .

*Theorem 1:* Any  $U$  such that  $A = f(U)$  achieves the K-point, maximizing  $R_Y = I(Y; U)$  on the X-line. The choice  $U = A$  is sufficient, and has minimal alphabet size.

*Proof:* By the above derivation. ■

## III. THE J-POINT

The Y-line ends at  $(R_X = H(X|Y), R_Y = H(Y))$  and is formed by points satisfying  $R_X = H(X|U) = H(X|Y)$ . From  $X - Y - U$  we have  $H(X|Y) = H(X|YU)$ , and therefore points on the Y-line must satisfy  $I(X; Y|U) = 0$ . The latter implies  $X - U - Y$ , that is,  $X$  and  $Y$  are independent given  $U$  [2, Sec. 2.8]. Therefore a point lies on the Y-line if and only if  $U$  satisfies  $\forall (u, x, y) \in \text{supp}(P_{U, X, Y})$ :

$$\begin{aligned} p(x, y|u) &= p(y|u)p(x|u, y) = p(y|u)p(x|y) \\ &= p(y|u, x)p(x|u) = p(y|u) \sum_{y'} p(y'|u)p(x|y'), \end{aligned} \quad (7)$$

where the summation is over  $y' \in \text{supp}(P_{Y|U}(\cdot|u))$ ,  $p(x|u, y) = p(x|y)$  follows from  $X - Y - U$ , and  $p(y|u, x) = p(y|u)$  follows from  $X - U - Y$ .

For the following proposition, define

$$\begin{aligned} \mathcal{Y}_u &= \{y : p(u, y) > 0\}, \\ \mathcal{X}_u &= \{x : \exists y \in \mathcal{Y}_u : p(x, y) > 0\} \\ &= \{x : p(u, x) > 0\}, \end{aligned} \quad (8)$$

where the last equality is due to  $X - Y - U$ .

*Proposition 2:* On the Y-line, the conditional distribution  $p(u|y)$  must be such that every  $x \in \mathcal{X}_u$  has a singleton ‘‘reverse fan’’  $\{p(x|y) : y \in \mathcal{Y}_u\}$  towards  $\mathcal{Y}$ . This means that for a given  $x \in \mathcal{X}_u$ , the set  $\{p(x|y) : y \in \mathcal{Y}_u\}$  contains a single element.

*Proof:* By (7),  $p(x|y)$  is the weighted mean of  $|\mathcal{Y}_u|$  probabilities  $p(x|y')$ . Thus we must have  $\min_{y \in \mathcal{Y}_u} p(x|y) = \max_{y \in \mathcal{Y}_u} p(x|y)$ , that is all conditional probabilities  $p(x|y)$  must be equal for given  $x$  and all  $y \in \mathcal{Y}_u$ . ■

The J-point is defined as the point on the Y-line with minimal  $R_Y = I(Y; U) = H(Y) - H(Y|U)$ . Thus the J-point can be reached by maximizing  $H(Y|U)$  while staying on the Y-line.

We define a partition random variable  $B = \psi(Y)$  that is a function of  $Y$  such that  $U = B$  maximizes the sizes of the sets  $\mathcal{Y}_b^B = \{y : p(b, y) > 0\}$  defined in (8). The partition indexed by  $B$  satisfies

$$\mathcal{Y} = \bigcup_b \mathcal{Y}_b^B \quad \text{with } \mathcal{Y}_b^B \cap \mathcal{Y}_{b'}^B = \emptyset \text{ for } b \neq b'$$

and it is not possible to move one or more letters  $y$  from a set  $\mathcal{Y}_b^B$  to another set  $\mathcal{Y}_{b'}^B$  without violating the condition of Proposition 2. It can be seen that the partition  $B$  is a refinement of the partition  $A$ .

*Theorem 3:*  $U = B$  achieves the J-point, minimizing  $R_Y = I(Y; U)$  on the Y-line.

*Proof:* The definition of  $B$  and the conditions in Proposition 2 ensure that no  $u$  may be connected to multiple  $A$ -components, where connected means  $p(u, y) > 0$  and  $p(u, y') > 0$  for  $y \in \mathcal{Y}_a^A$ ,  $y' \in \mathcal{Y}_{a'}^A$  with  $a \neq a'$ . They also ensure that no  $u$  may be connected to multiple  $Y$ -letters within a component  $\mathcal{Y}_a^A$  that are statistically dependent on  $X$ .

We define  $\mathcal{U}_b^B = \{u : p(u, y) > 0, y \in \mathcal{Y}_b^B\}$ , i.e., the set of  $u$  that are connected to  $b$ . By the above, we must have

$$\mathcal{U} = \bigcup_b \mathcal{U}_b^B \quad \text{with } \mathcal{U}_b^B \cap \mathcal{U}_{b'}^B = \emptyset \text{ for } b \neq b',$$

that is, there may be multiple  $u$ 's connected to one  $b$ , but not vice versa.

Now

$$\begin{aligned} H(Y|U) &= \sum_u p(u) H(Y|U = u) \\ &= \sum_b \sum_{u \in \mathcal{U}_b^B} p(u) H(Y|U = u) \\ &\leq \sum_b \left( \sum_{u \in \mathcal{U}_b^B} p(u) \right) H(Y|U \in \mathcal{U}_b^B) \\ &= \sum_b \Pr\{y \in \mathcal{Y}_b^B\} H(Y|Y \in \mathcal{Y}_b^B). \end{aligned}$$

The inequality follows since conditioning reduces entropy, and  $\{U = u\} \subseteq \{u \in \mathcal{U}_b^B\}$ . The final equality follows from the fact that  $\{u \in \mathcal{U}_b^B\}$  and  $\{Y \in \mathcal{Y}_b^B\}$  are the same event. To have equality all along, we must have  $U \perp Y|B$ . This is satisfied by the minimal choice  $U = B = f(Y)$ , which has  $P_{U|Y}(u|y) = 1_{\{u=f(y)\}}$  independently of  $y \in \mathcal{Y}_b^B$ . ■

#### IV. OPTIMIZATION

The goal is to obtain additional achievable rate points that lie below the time-sharing line J–K. Since  $U = A$  achieves the K-point,  $U = B$  achieves the J-point, and  $B$  is a refinement of  $A$ , a straightforward approach is to consider  $\mathcal{U} = \mathcal{B}$  augmented with one letter per  $A$ -component. Then Lagrangian optimization can be used to determine for each component whether it is coded using only  $B$ , only  $A$ , or a mix of both.

To simplify notation, we write  $b_{a,i}$  for the  $b$ 's belonging to the component  $a$ , that is, belonging to the set  $\mathcal{B}_a = \{b : \exists y \text{ s.t. } \phi(y) = a \wedge \psi(y) = b\}$ . To each  $b_{a,i}$ ,  $i = 1, \dots, |\mathcal{B}_a|$ , corresponds an auxiliary symbol  $u_{a,i}$  with conditional probability  $p(u_{a,i}|b_{a,i})$ . The remaining probability is gathered by a ‘‘grouping’’ auxiliary symbol  $u_{a,0}$  for each component  $a$ , with conditional probability  $p(u_{a,0}|b_{a,i})$ . By these definitions we

have

$$\begin{aligned} p(b_{a,i}|u_{a,i}) &= 1, \\ p(u_{a,i}) &= p(b_{a,i})p(u_{a,i}|b_{a,i}) \\ \text{and } p(b_{a,i}|u_{a,0}) &= \frac{p(b_{a,i})p(u_{a,0}|b_{a,i})}{p(u_{a,0})}, \\ \text{with } p(u_{a,0}) &= \sum_{i=1}^{|\mathcal{B}_a|} p(b_{a,i})p(u_{a,0}|b_{a,i}) \end{aligned}$$

Thus the distribution of the auxiliary random variable  $U$  is completely determined by the parameters  $\{p(u_{a,i}|b_{a,i}), p(u_{a,0}|b_{a,i})\}$ . Furthermore, when  $p(u_{a,0}) = p(a) = \sum_i p(b_{a,i})$  for all  $a$ , we obtain  $U = A$ , while for  $p(u_{a,0}) = 0$  for all  $a$ , we obtain  $U = B$ .

The goal is to find a conditional distribution  $P_{U|B}$  that minimizes  $H(X|U)$  subject to the constraints

$$I(Y; U) = H(Y) - H(Y|U) \leq R_Y, \quad (9)$$

$$-p(u_{a,i}|b_{a,i}) \leq 0, \quad (10)$$

$$-p(u_{a,0}|b_{a,i}) \leq 0, \quad (11)$$

$$p(u_{a,i}|b_{a,i}) + p(u_{a,0}|b_{a,i}) = 1. \quad (12)$$

Since  $X - U - A$  and  $Y - U - A$ , we have  $H(X|U) = H(X|A, U)$  and  $H(Y|U) = H(Y|A, U)$ , which helps simplify notation. The Lagrangian function is

$$\begin{aligned} J &= H(X|U) - \lambda H(Y|U) - \sum_{a,i} (\mu_{a,i} - \nu_{a,i}) p(u_{a,i}|b_{a,i}) \\ &\quad - \sum_{a,i} (\mu'_{a,i} - \nu_{a,i}) p(u_{a,0}|b_{a,i}), \end{aligned}$$

with

$$\begin{aligned} H(X|U) &= \sum_{a \in \mathcal{A}} \sum_{i=0}^{|\mathcal{B}_a|} p(u_{a,i}) H(X|U = u_{a,i}) \\ &= - \sum_{a \in \mathcal{A}} \sum_{i=1}^{|\mathcal{B}_a|} p(u_{a,i}) \sum_x p(x|u_{a,i}) \log(p(x|u_{a,i})) \\ &\quad - \sum_{a \in \mathcal{A}} p(u_{a,0}) \sum_x p(x|u_{a,0}) \log(p(x|u_{a,0})), \quad (13) \end{aligned}$$

where

$$p(x|u_{a,i}) = \sum_y p(y|b_{a,i}) p(x|y), \quad i = 1, \dots, |\mathcal{B}_a|, \quad (14)$$

and

$$p(x|u_{a,0}) = \sum_{i=1}^{|\mathcal{B}_a|} p(b_{a,i}|u_{a,0}) p(x|u_{a,i}). \quad (15)$$

$H(Y|U)$  follows analogously, with  $p(y|u_{a,i})$  in lieu of  $p(x|u_{a,i})$ .

Finally, we obtain the derivatives

$$\begin{aligned} \frac{\partial J}{\partial p(u_{a,i}|b_{a,i})} &= p(b_{a,i}) \left[ - \sum_x p(x|u_{a,i}) \log(p(x|u_{a,i})) \right. \\ &\quad \left. + \lambda \sum_y p(y|u_{a,i}) \log(p(y|u_{a,i})) \right] - \mu_{a,i} + \nu_{a,i}, \end{aligned} \quad (16)$$

$$\begin{aligned} \frac{\partial J}{\partial p(u_{a,0}|b_{a,i})} &= p(b_{a,i}) \left[ - \sum_x p(x|u_{a,i}) \log(p(x|u_{a,0})) \right. \\ &\quad \left. + \lambda \sum_y p(y|u_{a,i}) \log(p(y|u_{a,0})) \right] - \mu'_{a,i} + \nu_{a,i}. \end{aligned} \quad (17)$$

For a non-boundary solution (having  $p(u_{a,i}|b_{a,i}) > 0$  and  $p(u_{a,0}|b_{a,i}) > 0$ ), the Karush-Kuhn-Tucker (KKT) necessary conditions require  $\mu_{a,i} = \mu'_{a,i} = 0$ . Then (16) can be used to determine the Lagrange multiplier  $\nu_{a,i}$ , which inserted into (17) yields

$$p(b_{a,i}) [D(P(X|u_{a,i})||P(X|u_{a,0})) - \lambda D(P(Y|u_{a,i})||P(Y|u_{a,0}))] = 0 \quad (18)$$

as a necessary condition for a non-boundary solution. When  $p(u_{a,i}|b_{a,i}) = 0$ , the KKT conditions allow to choose  $\mu_{a,i} > 0$  in order to satisfy (16). Conversely, when  $p(u_{a,0}|b_{a,i}) = 0$ , the KKT conditions allow to choose  $\mu'_{a,i} > 0$  in order to satisfy (17). In both cases, the other necessary condition on  $p(\cdot|b_{a,i})$  can be satisfied. This means that for a given slope  $\lambda$  there will be in general only one  $b_{a,i}$  for which a proper non-boundary optimization takes place, with the exception of multiple identical (sub-)components, when the divergence expression (18) is satisfied for multiple  $b_{a,i}$ . All other  $p(u_{a,i}|b_{a,i})$  will be saturated either at 0 or at 1.

In particular, the limiting cases with  $p(u_{a,0}) = 0$  (i.e.,  $p(u_{a,i}|b_{a,i}) = 1$ ,  $i = 1, \dots, |\mathcal{B}_a|$ ) or  $p(u_{a,0}) = p(a)$  (i.e.,  $p(u_{a,i}|b_{a,i}) = 0$ ,  $i = 1, \dots, |\mathcal{B}_a|$ ) are valid candidate points. These can be used to derive a piecewise linear upper bound on the rate region. Potentially,  $2^{|\mathcal{A}|}$  configurations with either  $p(u_{a,0}) = 0$  or  $p(u_{a,0}) = p(a)$  need to be checked. However, this is not necessary, since we can greedily construct the convex hull of these points, i.e., the piecewise linear upper bound, as follows. Starting from the K-point, in which  $p(u_{a,0}) = p(a)$  for all components  $a \in \mathcal{A}$ , we seek the component  $a_{(1)}$  for which setting  $p(u_{a_{(1)},0}) = 0$  yields the most negative slope

$$s = \frac{H(X|U)|_{p(u_{a,0})=0} - H(X|U)|_{p(u_{a,0})=1}}{H(Y|U)|_{p(u_{a,0})=1} - H(Y|U)|_{p(u_{a,0})=0}}. \quad (19)$$

(It is understood that all other  $p(u_{a',0})$  are maintained at their corresponding values. The sign reversal in the denominator is due to the use of  $H(Y|U)$  instead of  $I(Y;U)$ .) There may be multiple components yielding the minimal slope, in which case they may be selected in any order. This procedure is repeated until all components have  $p(u_{a,0}) = 0$  and we have reached the J-point. At each step, the choice is optimal, since

any choice with less than minimal slope would point to the interior of the convex polytope.

In the example to be presented below, we opted for simple time-sharing between the point with  $p(u_{a,0}) = 0$  and the point with  $p(u_{a,0}) = p(a)$ . Further optimization results are part of ongoing work.

## V. EXAMPLE

The concepts and ideas discussed throughout this paper are illustrated by the following example. Consider  $X$  and  $Y$  with  $|\mathcal{X}| = 6$  and  $|\mathcal{Y}| = 7$  and the joint distribution  $P_{XY}(x, y) = P_{x,y}$ ,

$$P_{x,y} = \begin{bmatrix} \frac{1-q_1}{6} & \frac{q_1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{q_1}{6} & \frac{1-q_1}{6} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1-q_2}{6} & \frac{q_2}{6} & 0 & 0 & 0 \\ 0 & 0 & \frac{q_2}{6} & \frac{1-q_2}{6} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1-q_3}{6} & \frac{q_3}{12} & \frac{q_3}{12} \\ 0 & 0 & 0 & 0 & \frac{q_3}{6} & \frac{1-q_3}{12} & \frac{1-q_3}{12} \end{bmatrix},$$

with  $q_1 = 0.07$ ,  $q_2 = 0.1$ ,  $q_3 = 0.3$ . We have three components, defined by

$$A = \begin{cases} 1 & \text{for } X \in \{1, 2\} \\ 2 & \text{for } X \in \{3, 4\} \\ 3 & \text{for } X \in \{5, 6\} \end{cases},$$

or equivalently by

$$A = \begin{cases} 1 & \text{for } Y \in \{1, 2\} \\ 2 & \text{for } Y \in \{3, 4\} \\ 3 & \text{for } Y \in \{5, 6, 7\} \end{cases}.$$

Furthermore we have

$$B = \begin{cases} Y & \text{for } Y \in \{1, 2, 3, 4, 5\} \\ 6 & \text{for } Y \in \{6, 7\} \end{cases},$$

i.e.,  $Y = 6$  and  $Y = 7$  are mapped to the same  $B = 6$ . Notice that the dependency between  $X$  and  $Y$  is strongest in the first component,  $A = 1$ , and weakest in the third component,  $A = 3$ . The graph of the joint distribution is depicted in Fig. 2: every node corresponds to the realization of a random variable, and two nodes are connected if the corresponding joint probability is positive; remember that we have the Markov chain  $X - Y - B - U$ . The possible  $U$  in the graph describe the candidates considered in the optimization in Section IV.

The rate region for the source  $P_{XY}$  is shown in Fig. 3. Rate points below the X-line and below the Y-line are not achievable, see (4). The point A corresponds to the case without side information, i.e.,  $R_Y = 0$ , and the point D to the case with maximal side information, i.e.,  $R_Y = H(Y)$ . All of the points B, C, J, and K are obtained by

$$p(u_{a0}) \in \{0, p(a)\}.$$

The points J and K lie on the boundary of the rate region.

Starting from point A,  $R_X$  can be reduced while the sum rate remains minimum, i.e., at  $R_X + R_Y = H(X)$ ; the extreme point with this property is K. On the other hand, starting

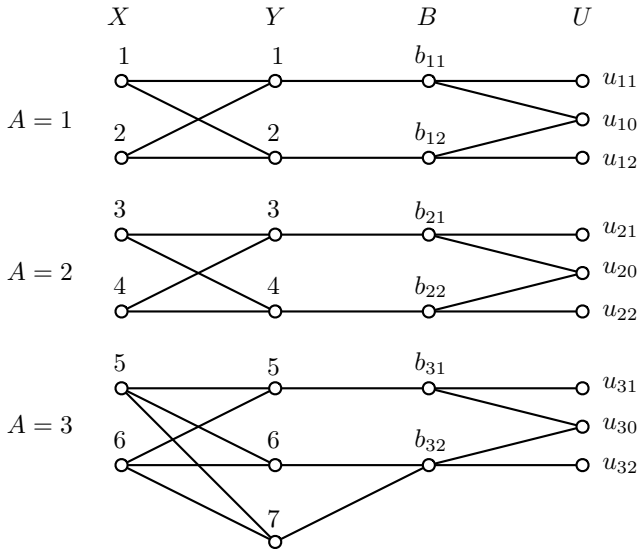


Fig. 2. Graph of the joint distribution  $P_{XYBU}$ .

from D,  $R_Y$  can be reduced while  $R_X$  remains minimum, i.e.,  $R_X = H(X|Y)$ ; the extreme point with this property is J. The time-sharing line between K and J is indicated with a dashed line.

The new rate points B and C are below the time-sharing line between J and K. Point B is obtained by

$$p(u_{10}) = 0, \quad p(u_{20}) = P_A(2), \quad p(u_{30}) = P_A(3),$$

and point C is obtained by

$$p(u_{10}) = 0, \quad p(u_{20}) = 0, \quad p(u_{30}) = P_A(3).$$

(Note that  $P_A(a) \equiv p(a)$ .) These are the optimal choices of the probabilities  $p(u_{a0}) \in \{0, p(a)\}$ . The rate points corresponding to other choices may lie above or below the KJ line.

The convexity of the rate region may be exploited to obtain further computable rate points. For example, denote  $U_C$  and  $U_D$  the auxiliary random variables that lead to the point C and D, respectively. Define now

$$U' = \rho U_C + (1 - \rho) U_D,$$

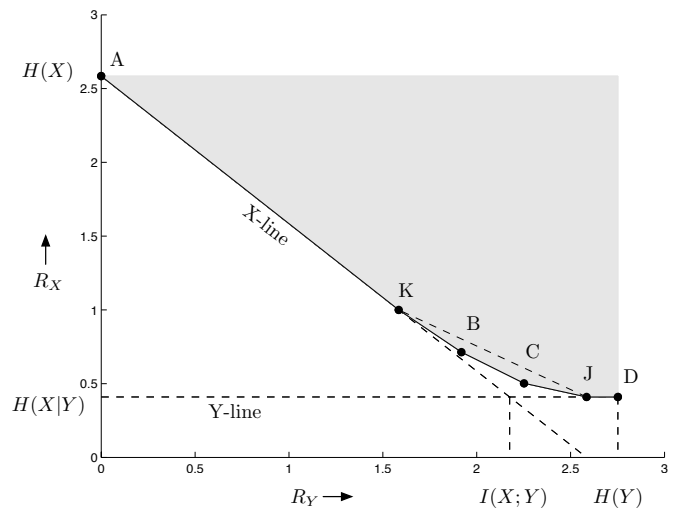


Fig. 3. Rate region and specific rate points for the example.

$\rho \in [0, 1]$ , as a convex combination of  $U_C$  and  $U_D$ . Then the rate point

$$R_X = H(X|U'), \quad R_Y = I(Y; U')$$

can be shown to lie below or on the time-sharing line between C and D for all choices of  $\rho$ . Proceeding in this way with all pairs of neighboring rate points leads to a computable extension of the inner bound given by the polygon defined by the points A, K, B, C, J, D.

#### ACKNOWLEDGMENTS

This work was supported by the Australian Research Council under the ARC Discovery Grant DP120102123.

Ingmar Land's work was supported in part by a visiting professor fellowship from ENSEA, Cergy-Pontoise, France.

#### REFERENCES

- [1] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. 21, no. 6, pp. 629–637, 1975.
- [2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.
- [3] D. Marco and M. Effros, "On lossless coding with coded side information," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3284–3296, 2009.