

Capacity of a Class of Diamond Channels

Wei Kang

Sennur Ulukus

Department of Electrical and Computer Engineering

University of Maryland, College Park, MD 20742

wkang@eng.umd.edu

ulukus@umd.edu

Abstract—We study a special class of diamond channels which was introduced by Schein in 2001. In this special class, each diamond channel consists of a transmitter, a noisy relay, a noiseless relay and a receiver. We prove the capacity of this class of diamond channels by providing an achievable scheme and a converse. The capacity we show is strictly smaller than the cut-set bound. Our result also shows the optimality of a combination of decode-and-forward (DAF) and compress-and-forward (CAF) at the noisy relay node. This is the first example where a combination of DAF and CAF is shown to be capacity achieving. Finally, we note that there exists a duality between this diamond channel coding problem and the Kaspi-Berger source coding problem.

I. PROBLEM STATEMENT AND THE RESULT

The diamond channel was first introduced by Schein in 2001 [1]. The diamond channel consists of one transmitter, two relays and a receiver, where the transmitter and the two relays form a broadcast channel as the first stage and the two relays and the receiver form a multiple access channel as the second stage. The capacity of the diamond channel in its most general form is open. Schein explored several special cases of the diamond channel, one of which [1, Section 3.5] is specified as follows (see Figure 1). The multiple access channel consists of two orthogonal links with rate constraints R_1 and R_2 , respectively. The broadcast channel contains a noisy branch and a noiseless branch, i.e., with input X and two outputs X and Y . We refer to the relay node receiving Y as the noisy relay and the relay node receiving X as the noiseless relay. Schein provided two achievable schemes for this class of diamond channels. In this paper, we will prove the capacity of this special class of diamond channels.

The formal definition of the problem is as follows. Consider a channel with input alphabet \mathcal{X} and output alphabet \mathcal{Y} , which is characterized by the transition probability $p(y|x)$. Assume an n -length block code consisting of (f, g, h, φ) where

$$f : \{1, 2, \dots, M\} \mapsto \mathcal{X}^n \quad (1)$$

$$g : \mathcal{Y}^n \mapsto \{1, 2, \dots, |g|\} \quad (2)$$

$$h : \{1, 2, \dots, M\} \mapsto \{1, 2, \dots, |h|\} \quad (3)$$

$$\varphi : \{1, 2, \dots, |g|\} \times \{1, 2, \dots, |h|\} \mapsto \{1, 2, \dots, M\} \quad (4)$$

Here f denotes the encoding function at the transmitter, g and h denote the processing functions at the noisy and noiseless

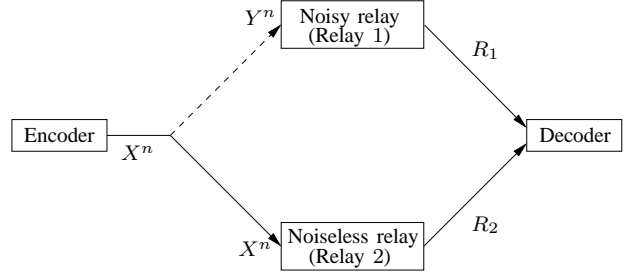


Fig. 1. The diamond channel.

relays, respectively, and φ denotes the decoding function at the receiver.

The encoder sends $x^n = f(m)$ into the channel, where $m \in \{1, 2, \dots, M\}$. The decoder reconstructs $\hat{m} = \varphi(g(Y^n), h(m))$. The average probability of error is defined as

$$P_e \triangleq \frac{1}{M} \sum_{m=1}^M Pr(\hat{m} \neq m | m \text{ is sent}) \quad (5)$$

The rate triple (R, R_1, R_2) is achievable if for every $0 < \epsilon < 1$, $\eta > 0$ and every sufficiently large n , there exists an n -length block code (f, g, h, φ) , such that $P_e \leq \epsilon$ and

$$\frac{1}{n} \ln M \geq R - \eta \quad (6)$$

$$\frac{1}{n} \ln |g| \leq R_1 + \eta \quad (7)$$

$$\frac{1}{n} \ln |h| \leq R_2 + \eta \quad (8)$$

The following theorem characterizes the capacity of the class of diamond channels considered in this paper.

Theorem 1 *The rate triple (R, R_1, R_2) is achievable in the above channel if and only if the following conditions are satisfied*

$$R \leq I(U; Y) + H(X|U) \quad (9)$$

$$R_1 \geq I(Z; Y|U, X) \quad (10)$$

$$R_2 \geq H(X|Z, U) \quad (11)$$

$$R_1 + R_2 \geq R + I(Y; Z|X, U) \quad (12)$$

for some joint distribution

$$p(u, z, x, y) = p(u, x)p(y|x)p(z|u, y) \quad (13)$$

with cardinalities of alphabets satisfying

$$|\mathcal{U}| \leq |\mathcal{X}| + 4 \quad (14)$$

$$|\mathcal{Z}| \leq |\mathcal{U}||\mathcal{Y}| + 3 \leq |\mathcal{X}||\mathcal{Y}| + 4|\mathcal{X}| + 3 \quad (15)$$

II. THE ACHIEVABILITY

Assume a given joint distribution

$$p(u, z, x, y) = p(u, x)p(y|x)p(z|u, y) \quad (16)$$

and consider that the information theoretic quantities on the right hand sides of (9), (10), (11) and (12) are evaluated with this fixed joint probability distribution.

Consider a message W with rate R . If $R \leq H(X|Z, U)$, reliable transmission can be achieved by letting $g(Y^n) = \phi$ (constant) and $h(W) = W$, i.e., by sending the message through the noiseless relay. Thus, we will only consider the case where

$$H(X|Z, U) < R \leq I(U; Y) + H(X|U) \quad (17)$$

We will show that the message can be reliably transmitted with a pair of functions (g, h) such that $(\frac{1}{n} \ln |g|, \frac{1}{n} \ln |h|)$ lies in the inverse pentagon¹ with corners a and b in Figure 2. However, we instead prove reliable transmission with $(\frac{1}{n} \ln |g|, \frac{1}{n} \ln |h|)$ lying in the inverse pentagon with corners a' and b' , which contains the inverse pentagon with corners a and b and thus imposes a stronger condition to prove. It is straightforward to have reliable transmission with the rate pair at point b' by letting $g(Y^n) = \phi$ (constant) and $h(W) = W$. Thus, it remains to prove that reliable transmission is possible with the rate pair at point a' , i.e.,

$$R_1 = I(U; Y) + I(Y; Z|U) \quad (18)$$

$$R_2 = R - I(U; Y) - I(X; Z|U) \quad (19)$$

Let us assume that the message W is decomposed as $W = (W_a, W_b, W_c)$. For a positive number ϵ , let us define

$$M_a \triangleq |W_a| = \exp(n(I(U; Y) - 3\epsilon)) \quad (20)$$

$$M_b \triangleq |W_b| = \frac{M}{M_a M_c} \\ = \exp(\ln M - n(I(U; Y) + I(X; Z|U) + 6\epsilon)) \quad (21)$$

$$M_c \triangleq |W_c| = \exp(n(I(X; Z|U) - 3\epsilon)) \quad (22)$$

Random codebook generation: We use a superposition code structure. The size of the inner code is M_a . For each inner codeword, we independently generate M_b outer codes. The size of each outer code is M_c .

- Independently generate M_a sequences, $u^n(1), u^n(2), \dots, u^n(M_a)$, according to $\prod_{i=1}^n p(u_i)$ where $p(u_i) = p(u)$, for $i = 1, 2, \dots, n$.
- For $u^n(j)$, $j = 1, 2, \dots, M_a$, independently generate M_b codebooks, $\mathcal{C}(j, 1), \mathcal{C}(j, 2), \dots, \mathcal{C}(j, M_b)$.
- In the codebook $\mathcal{C}(j, k)$, $j = 1, 2, \dots, M_a$, $k = 1, 2, \dots, M_b$, independently generate M_c codewords

¹By “inverse pentagon” with corner points a and b , we mean the region in the (R_1, R_2) space that is to the “north-east” of line segment $[a, b]$. More specifically, this is the region described by inequalities in (10), (11) and (12).

$x^n(j, k, 1), x^n(j, k, 2), \dots, x^n(j, k, M_c)$ according to $\prod_{i=1}^n p(x_i|U_i = u_i(j))$, where $p(x_i|U = u_i(j)) = p(x|u)$, for $i = 1, 2, \dots, n$, $j = 1, 2, \dots, M_a$, $k = 1, 2, \dots, M_b$.

There will be no overlapping codebooks with high probability when n is sufficiently large, because

$$\frac{1}{n} \ln M_b M_c < H(X|U) \quad (23)$$

Encoding at the transmitter: Let $W = (W_a, W_b, W_c)$ be the message. We send codeword $X^n = f(W_a, W_b, W_c) \triangleq x^n(W_a, W_b, W_c)$ into the channel.

Processing at the noisy relay: First, after having received Y^n , seek

$$\hat{U}^n = u^n(\hat{W}_a) \in \{u^n(1), u^n(2), \dots, u^n(M_a)\} \quad (24)$$

such that $(\hat{U}^n, Y^n) \in \mathcal{T}_{[UY]}^n$, where the definition of strong typical set can be found in [2, Section 1.2]. If there is not any such \hat{U}^n , then let \hat{U}^n be an arbitrary sequence in $\{u^n(1), u^n(2), \dots, u^n(M_a)\}$. Secondly, construct a conditional rate distortion code according to $\prod_{i=1}^n p(z_i, y_i|\hat{u}_i)$ with encoding function $g'(Y^n, \hat{U}^n)$ and $|g'| = L = \exp(n(I(Y; Z|U) + \tau))$. Finally send \hat{U}^n and $Z^n \triangleq g'(Y^n, \hat{U}^n)$ to the destination, i.e., $g(Y^n) = (\hat{U}^n, Z^n)$, where

$$|g| = M_a \times L \leq \exp(n(I(U; Y) + I(X; Z|U) + \tau - 3\epsilon)) \quad (25)$$

Processing at the noiseless relay: Let $h(f(W_a, W_c, W_b)) = W_b$ where

$$|h| = M_b = \exp(\ln M - n(I(U; Y) + I(X; Z|U) + 6\epsilon)) \quad (26)$$

Decoding: Decoder collects (\hat{U}^n, Z^n) from the noisy relay and W_b from the noiseless relay. The decoder seeks a codeword $x^n(W_a, W_b, i)$ from the codebook $\mathcal{C}(W_a, W_b)$ such that

$$(x^n(\hat{W}_a, W_b, i), Z^n) \in \mathcal{T}_{[XZ|U]}^n(\hat{U}^n) \quad (27)$$

Probability of error: The error occurs when $(\hat{U}, \hat{X}) \neq (U, X)$. The average probability of error

$$Pr(E) \leq Pr(E_1 \cup E_2 \cup E_3) \\ = Pr(E_1) + Pr(E_2 \cap E_1^c) + Pr(E_3 \cap E_1^c \cap E_2^c) \quad (28)$$

where

$$E_1 \triangleq (U^n, X^n, Y^n, Z^n) \notin \mathcal{T}_{[UXYZ]}^n \quad (29)$$

$$E_2 \triangleq \bigcup_{\bar{u}^n \neq U^n, \bar{u}^n \in \{u^n(1), u^n(2), \dots, u^n(M_a)\}} (\bar{u}^n, Y^n) \in \mathcal{T}_{[UY]}^n \quad (30)$$

$$E_3 \triangleq \bigcup_{\bar{x}^n \neq X^n, \bar{x}^n \in \mathcal{C}(W_a, W_b)} (\bar{x}^n, Z^n) \in \mathcal{T}_{[XZ|U]}^n(U^n) \quad (31)$$

We note that

$$Pr(E_1) \leq Pr(U^n \notin \mathcal{T}_{[U]}^n) + Pr((Y^n, Z^n) \notin \mathcal{T}_{[YZ|U]}^n(U^n)) \\ + Pr(X^n \notin \mathcal{T}_{[X|YZU]}^n(Y^n, Z^n, U^n)) \quad (32)$$

where

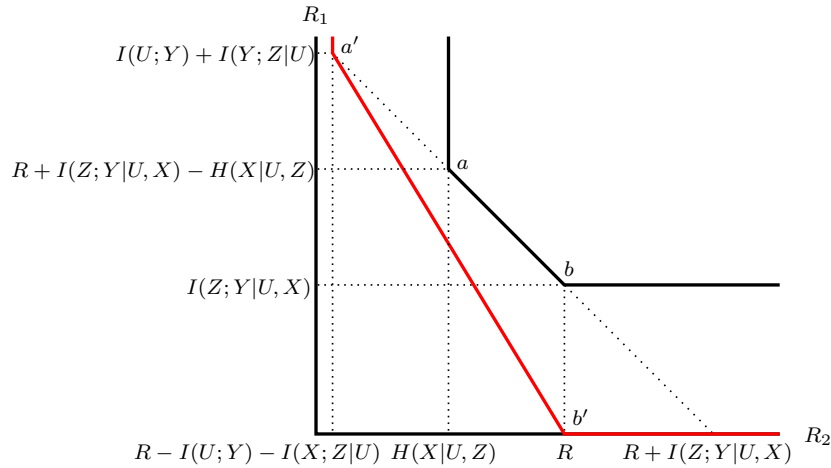


Fig. 2. Rate region of (R_1, R_2) when $H(X|U, Z) \leq R \leq I(U; Y) + I(X; Z|U)$.

- U^n is generated in an i.i.d. fashion with probability $p(u)$. Thus, when n is sufficiently large, we have

$$\Pr(U^n \notin \mathcal{T}_{[U]}^n) \leq \epsilon \quad (33)$$

- Z^n is a conditional rate distortion code for Y^n conditioned on U^n . Thus, when n is sufficiently large, $L = \exp(nI(Y; Z|U) + \tau)$, and $U^n \in \mathcal{T}_{[U]}^n$, we have

$$\Pr((Y^n, Z^n) \notin \mathcal{T}_{[YZ|U]}^n(U^n)) \leq \epsilon \quad (34)$$

- X^n can be viewed as being generated according to an i.i.d. conditional probability $p(x|u, y)$ with respect to (U^n, Y^n) . Thus, when n is sufficiently large and $(Y^n, Z^n, U^n) \in \mathcal{T}_{[YZU]}^n$,

$$\Pr(X^n \notin \mathcal{T}_{[X|YZU]}^n(Y^n, Z^n, U^n)) \leq \epsilon \quad (35)$$

From the above calculation, we have

$$\Pr(E_1) = \Pr((U^n, X^n, Y^n, Z^n) \notin \mathcal{T}_{[UXZ]}^n) \leq 3\epsilon \quad (36)$$

For the second error event, we note that $M_a = \exp(n(I(U; Y) - 3\epsilon))$ and

$$\begin{aligned} & \Pr(E_2 \cap E_1^c) \\ &= \Pr\left(\bigcup_{\bar{u}^n \neq U^n, \bar{u}^n \in \{u^n(1), u^n(2), \dots, u^n(M_a)\}} (\bar{u}^n, Y^n) \in \mathcal{T}_{[UY]}^n | (Y^n) \in \mathcal{T}_{[Y]}^n\right) \\ &\leq \sum_{i=1}^{M_a} \Pr((u^n(i), Y^n) \in \mathcal{T}_{[UY]}^n | Y^n \in \mathcal{T}_{[Y]}^n) \\ &\leq M_a \Pr(u^n(i) \in \mathcal{T}_{[U|Y]}^n(Y^n)) \\ &\leq M_a \exp(-nH(U) + n\epsilon) \exp(nH(U|Y) + n\epsilon) \\ &= \exp(-n\epsilon) \leq \epsilon \end{aligned} \quad (37)$$

for sufficiently large n . We note that $M_c = \exp(n(I(X; Z|U) - 3\epsilon))$, then

$$\Pr(E_3 \cap E_1^c) = \Pr\left(\bigcup_{\bar{x}^n \neq X^n, \bar{x}^n \in \mathcal{C}(W_a, W_b)} (\bar{x}^n, Z^n) \in \mathcal{T}_{[XZ|U]}^n(U^n) | (Z^n, U) \in \mathcal{T}_{[ZU]}^n\right)$$

$$\begin{aligned} & \leq \sum_{i=1}^{M_c} \Pr((x(M_a, M_b, i), Z^n) \in \mathcal{T}_{[XZ|U]}^n(U^n) | \\ & \quad (Z^n, U^n) \in \mathcal{T}_{[ZU]}^n) \\ & \leq M_c \Pr(x(M_a, M_b, i) \in \mathcal{T}_{[X|ZU]}^n(Y^n)) \\ & \leq M_c \exp(-nH(X|U) + n\epsilon) \exp(nH(X|Z, U) + n\epsilon) \\ & = \exp(-n\epsilon) \leq \epsilon \end{aligned} \quad (38)$$

for sufficiently large n . Thus, the average probability error is upper bounded as

$$\Pr(E) \leq 3\epsilon + \epsilon + \epsilon = 5\epsilon \quad (39)$$

which goes to zero when n goes to infinity.

III. THE CONVERSE

Define $Z_i \triangleq g$ and $U_i \triangleq (Y^{i-1}, X_{i+1}^n)$. We note that

$$p(u_i, x_i, y_i, z_i) = p(u_i, x_i)p(y_i|x_i)p(z_i|y_i, u_i) \quad (40)$$

We have

$$\begin{aligned} \ln M &= H(X^n) \\ &= \sum_{i=1}^n H(X_i | X_{i+1}^n) \\ &\leq \sum_{i=1}^n I(Y^{i-1}; Y_i) + H(X_i | X_{i+1}^n) \\ &= \sum_{i=1}^n I(Y^{i-1}, X_{i+1}^n; Y_i) - I(X_{i+1}^n; Y_i | Y^{i-1}) \\ & \quad + H(X_i | Y^{i-1}, X_{i+1}^n) + I(Y^{i-1}; X_i | X_{i+1}^n) \\ &\stackrel{1}{=} \sum_{i=1}^n I(Y^{i-1}, X_{i+1}^n; Y_i) + H(X_i | Y^{i-1}, X_{i+1}^n) \\ &= \sum_{i=1}^n I(U_i; Y_i) + H(X_i | U_i) \end{aligned} \quad (41)$$

where

1) Because of the following equality [3, Lemma 7]

$$\sum_{i=1}^n I(X_{i+1}^n; Y_i | Y^{i-1}) = \sum_{i=1}^n I(Y^{i-1}; X_i | X_{i+1}^n) \quad (42)$$

We have

$$\begin{aligned} \ln |g| &\geq H(g) \\ &\geq H(g|h) \\ &\geq H(g|h) - H(g|h, Y^n) \\ &= I(g; Y^n | h) \\ &= \sum_{i=1}^n I(g; Y_i | h, Y^{i-1}) \\ &= \sum_{i=1}^n I(g, X_{i+1}^n; Y_i | h, Y^{i-1}) - I(X_{i+1}^n; Y_i | g, h, Y^{i-1}) \\ &\stackrel{1}{=} \sum_{i=1}^n I(g, X_{i+1}^n; Y_i | h, Y^{i-1}) - I(Y^{i-1}; X_i | g, h, X_{i+1}^n) \\ &\geq \sum_{i=1}^n I(g, X_{i+1}^n; Y_i | h, Y^{i-1}) - H(X_i | g, h, X_{i+1}^n) \\ &= -H(X^n | g, h) + \sum_{i=1}^n I(g, X_{i+1}^n; Y_i | h, Y^{i-1}) \\ &\stackrel{2}{\geq} \sum_{i=1}^n I(g, X_{i+1}^n; Y_i | h, Y^{i-1}) - \epsilon \\ &\geq \sum_{i=1}^n I(g; Y_i | h, Y^{i-1}, X_{i+1}^n) - \epsilon \\ &\stackrel{3}{\geq} \sum_{i=1}^n I(g; Y_i | h, Y^{i-1}, X_{i+1}^n, X_i) - \epsilon \\ &\stackrel{4}{=} \sum_{i=1}^n I(g; Y_i | Y^{i-1}, X_{i+1}^n, X_i) - \epsilon \\ &= \sum_{i=1}^n I(Z_i; Y_i | U_i, X_i) - \epsilon \end{aligned} \quad (43)$$

where

1) Because of the following equality [3, Lemma 7]

$$\sum_{i=1}^n I(X_{i+1}^n; Y_i | g, h, Y^{i-1}) = \sum_{i=1}^n I(Y^{i-1}; X_i | g, h, X_{i+1}^n) \quad (44)$$

2) Due to Fano's inequality.

3) g is a deterministic function of Y^n . Due to the memoryless property, we have

$$H(g|Y_i, h, Y^{i-1}, X_{i+1}^n, X_i) = H(g|Y_i, h, Y^{i-1}, X_{i+1}^n) \quad (45)$$

4) g is a deterministic function of Y^n and h is a deterministic function of X^n . Due to the memoryless property,

we have

$$H(g|h, Y^{i-1}, X_{i+1}^n, X_i) = H(g|Y^{i-1}, X_{i+1}^n, X_i) \quad (46)$$

$$H(g|h, Y^{i-1}, X_{i+1}^n, X_i, Y_i) = H(g|Y^{i-1}, X_{i+1}^n, X_i, Y_i) \quad (47)$$

We have

$$\begin{aligned} \ln |h| &\geq H(h|g) \\ &\geq I(h; X^n | g) \\ &= H(X^n | g) - H(X^n | g, h) \\ &\stackrel{1}{\geq} H(X^n | g) - n\epsilon \\ &= \sum_{i=1}^n H(X_i | X_{i+1}^n, g) - \epsilon \\ &\geq \sum_{i=1}^n H(X_i | Y^{i-1}, X_{i+1}^n, g) - \epsilon \\ &= \sum_{i=1}^n H(X_i | U_i, Z_i) - \epsilon \end{aligned} \quad (48)$$

where

1) Due to Fano's inequality.

We have

$$\begin{aligned} \ln |g| + \ln |h| &\geq H(g, h) \\ &\geq I(g, h; X^n, Y^n) \\ &\geq I(X^n; g, h) + I(Y^n; g, h | X^n) \\ &= H(X^n) - H(X^n | g, h) + I(Y^n; g, h | X^n) \\ &\stackrel{1}{\geq} \ln M - n\epsilon + I(Y^n; g, h | X^n) \\ &\stackrel{2}{=} \ln M - n\epsilon + I(Y^n; g | X^n) \\ &= \ln M + \sum_{i=1}^n -\epsilon + I(Y_i; g | X^n, Y^{i-1}) \\ &\stackrel{3}{=} \ln M + \sum_{i=1}^n -\epsilon + I(Y_i; g | X_i, Y^{i-1}, X_{i+1}^n) \\ &= \ln M + \sum_{i=1}^n -\epsilon + I(Y_i; Z_i | X_i, U_i) \end{aligned} \quad (49)$$

1) Due to Fano's inequality.

2) h is a deterministic function of X^n

3) g is a deterministic function of Y^n . Due to the memoryless property, we have

$$H(g|X_i, Y^{i-1}, X_{i+1}^n, X^{i-1}) = H(g|X_i, Y^{i-1}, X_{i+1}^n) \quad (50)$$

$$H(g|Y_i, X_i, Y^{i-1}, X_{i+1}^n, X^{i-1}) = H(g|Y_i, X_i, Y^{i-1}, X_{i+1}^n) \quad (51)$$

We note that $\frac{1}{n} \ln M \geq R - \eta$, $\frac{1}{n} \ln |g| \leq R_1 + \eta$ and $\frac{1}{n} \ln |h| \leq R_2 + \eta$, for an arbitrary $\eta > 0$. Assume $\epsilon \rightarrow 0$,

then from (41), (43), (48) and (49), we have

$$R \leq \frac{1}{n} \sum_{i=1}^n I(U_i; Y_i) + H(X_i|U_i) \quad (52)$$

$$R_1 \geq \frac{1}{n} \sum_{i=1}^n I(Z_i; Y_i|U_i, X_i) \quad (53)$$

$$R_2 \geq \frac{1}{n} \sum_{i=1}^n H(X_i|U_i, Z_i) \quad (54)$$

$$R_1 + R_2 \geq R + \frac{1}{n} \sum_{i=1}^n I(Y_i; Z_i|X_i, U_i) \quad (55)$$

Define a time-sharing random variable Q , which is uniformly distributed on $\{1, 2, \dots, n\}$. Also define a set of random variables $(X, Y, \tilde{U}, \tilde{Z})$ such that

$$\begin{aligned} Pr(X = x, Y = y, \tilde{U} = u, \tilde{Z} = z|Q = i) \\ = p(X_i = x, Y_i = y, U_i = u, Z_i = z) \\ i = 1, 2, \dots, n \end{aligned} \quad (56)$$

Define $U = (\tilde{U}, Q)$ and $Z = (\tilde{Z}, Q)$, then

$$\begin{aligned} R &\leq \frac{1}{n} \sum_{i=1}^n I(U_i; Y_i) + H(X_i|U_i) \\ &= I(\tilde{U}; Y|Q) + H(X|\tilde{U}, Q) \\ &\leq I(\tilde{U}, Q; Y) + H(X|\tilde{U}, Q) \\ &= I(U; Y) + H(X|U) \end{aligned} \quad (57)$$

$$\begin{aligned} R_1 &\geq \frac{1}{n} \sum_{i=1}^n I(Z_i; Y_i|U_i, X_i) \\ &= I(\tilde{Z}; Y|\tilde{U}, Q, X) \\ &= I(Z; Y|U, X) \end{aligned} \quad (58)$$

$$\begin{aligned} R_2 &\geq \frac{1}{n} \sum_{i=1}^n H(X_i|U_i, Z_i) \\ &= H(X|\tilde{U}, \tilde{Z}, Q) \\ &= H(X|U, Z) \end{aligned} \quad (59)$$

$$\begin{aligned} R_1 + R_2 &\geq R + \frac{1}{n} \sum_{i=1}^n I(Y_i; Z_i|X_i, U_i) \\ &= R + I(\tilde{Z}; Y|\tilde{U}, X, Q) \\ &= R + I(Z; Y|U, X) \end{aligned} \quad (60)$$

where (57), (58), (59) and (60) are the same as (9), (10), (11) and (12), concluding the proof.

Finally, we note that the bounds on the cardinalities of the alphabets in (14) and (15) can be proven in a way similar to [4, Appendix D].

IV. REMARKS

We have several remarks regarding this result as follows:

1) The capacity is strictly smaller than the cut-set bound

[5], because first

$$R \leq R_1 + R_2 - I(Y; Z|U, X) \quad (61)$$

An operational interpretation is that when the noisy relay cannot fully decode the message, or in other words, when the noisy relay cannot remove the noise completely, the data going through the link from the noisy relay to the receiver contains noise. Thus, the useful information flowing through the multiple access cut will be strictly less than $R_1 + R_2$. Secondly, we note that

$$R \leq I(U; Y) + H(X|U) \leq H(X) \quad (62)$$

An operational interpretation is that when the noisy relay decodes the message with a positive rate, the rate of information flowing through the broadcast cut becomes strictly less than $H(X)$.

Consider the following example. Let X and Y be binary and

$$Y = X \oplus W \quad (63)$$

where the sum is a modulo-2 sum and W has a Bernoulli distribution with entropy 0.5 bits. We assume $R_1 = R_2 = 0.5$ bits. The cut-set bound in this example is 1 bit, which is not achievable. Because if R is equal to 1 bit, we have,

$$R = I(U; Y) + H(X|U) = H(X) = 1 \quad (64)$$

then, U has to be independent of X and Y . Also, we have

$$R = R_1 + R_2 - I(Y; Z|U, X) = R_1 + R_2 = 1 \quad (65)$$

then, Z has to be independent of X and Y if U is independent of X and Y . However, if U and Z are independent of X and Y , we arrive at the following contradiction,

$$0.5 = R_2 \geq H(X|Z, U) = H(X) = 1 \quad (66)$$

which means that the cut-set bound is not achievable in this example. We note that, even in this binary example where $|\mathcal{X}| = |\mathcal{Y}| = 2$, the cardinalities of the auxiliary random variables U and Z are $|\mathcal{U}| \leq 6$ and $|\mathcal{Z}| \leq 15$. These large cardinality bounds make it practically impossible to evaluate the capacity of this diamond channel. However, we note that, even though we were not able to compute the exact value of the capacity in this example, we were able to conclude that the capacity is strictly less than the cut-set bound, which is 1 bit.

We know that the capacity of a diamond channel with four orthogonal links is equal to the cut-set bound in this channel. Our result shows that introducing the broadcast node will reduce the capacity of this all-orthogonal diamond channel. Networks with broadcast nodes have been studied recently from different perspectives, e.g., information theory and network coding

[6]–[8]. We note that our diamond channel model is a simple example of a general network with a broadcast node. Thus, we conclude that the cut-set bound in general is not tight in networks with broadcast nodes.

- 2) The processing at the noisy relay includes two operations: decode the inner code U^n and compress the channel output Y^n to Z^n conditioned on U^n . This processing is essentially the same as Theorem 7 in [9], i.e., combination of DAF and CAF. DAF [9, Theorem 1] has been shown to be optimal in the degraded relay channel [9]. Partial DAF, a special case of [9, Theorem 7] without compression, has been shown to be optimal in semi-deterministic relay channel [10] and the relay channel with orthogonal transmitter-relay link [11]. Recently, CAF [9, Theorem 6] has been shown to be optimal in two special relay channels [12], [13]. To our knowledge, we are the first to show the optimality of the combination of DAF and CAF in some specific channel, even though the channel we consider is not a three-node relay channel in the strict sense, i.e., as in [9].
- 3) If we assume $R = H(X) - R_0$, then the expressions in Theorem 1, i.e.,

$$R \leq I(U; Y) + H(X|U) \quad (67)$$

$$R_1 \geq I(Z; Y|U, X) \quad (68)$$

$$R_2 \geq H(X|Z, U) \quad (69)$$

$$R_1 + R_2 \geq R + I(Y; Z|X, U) \quad (70)$$

for some joint distribution

$$p(u, z, x, y) = p(u, x)p(y|x)p(z|u, y) \quad (71)$$

can be rewritten as follows

$$R_0 \geq I(U; X|Y) \quad (72)$$

$$R_1 \geq I(Z; Y|U, X) \quad (73)$$

$$R_2 \geq I(X; X|Z, U) \quad (74)$$

$$R_0 + R_1 + R_2 \geq I(X, Y; U, X, Z) \quad (75)$$

for some joint distribution

$$p(u, z, x, y) = p(u, x)p(y|x)p(z|u, y) \quad (76)$$

We note that (72), (73), (74) and (75) in addition to the distribution constraint in (76) are the same as the rate region of the rate-distortion problem studied by Kaspi and Berger as shown in Figure 3 [4, Theorem 2.1, Case C].

This duality between our diamond channel coding problem and the Kaspi-Berger source coding problem is similar to the duality between the single-user channel coding problem and the Slepian-Wolf source coding problem [2, Section 3.1] by viewing the codebook information in the channel coding problem as the information sent to all the terminals in the source coding problem, e.g., the information with rate R_0 in Figure 3. Thus, the achievability of our diamond channel coding problem can be obtained from the

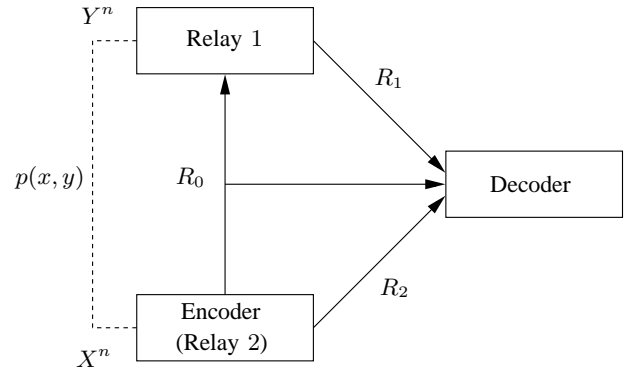


Fig. 3. Kaspi-Berger rate distortion problem.

achievability of Kaspi-Berger source coding problem, in the same way that the achievability of the multiple access channel coding problem can be obtained from the achievability of fork network coding problem [2, Section 3.2].

REFERENCES

- [1] B. E. Schein. *Distributed Coordination in Network Information Theory*. PhD thesis, Massachusetts Institute of Technology, 2001.
- [2] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.
- [3] I. Csiszár and J. Körner. Broadcast channels with confidential messages. *IEEE Trans. Inform. Theory*, 24(3):339–348, 1978.
- [4] A. H. Kaspi and T. Berger. Rate-distortion for correlated sources with partially separated encoders. *IEEE Trans. Inform. Theory*, 28(6):828–840, 1982.
- [5] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons, 1991.
- [6] A. F. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros. Capacity of wireless erasure networks. *IEEE Trans. Inform. Theory*, 52(3):789–804, 2006.
- [7] N. Ratnakar and G. Kramer. The multicast capacity of deterministic relay network with no interference. *IEEE Trans. Inform. Theory*, 52(6):2425–2432, 2006.
- [8] G. Kramer, S. M. S. Tabatabaei Yazdi, and S. A. Savari. Network coding on line networks with broadcast. In *Proc. Conf. Inf. Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2008.
- [9] T. M. Cover and A. El Gamal. Capacity theorems for the relay channel. *IEEE Trans. Inform. Theory*, 25:572–584, Sep. 1979.
- [10] A. El Gamal and M. Aref. The capacity of the semideterministic relay channel. *IEEE Trans. Inform. Theory*, 28(3):536, 1982.
- [11] A. El Gamal and S. Zahedi. Capacity of a class of relay channels with orthogonal components. *IEEE Trans. Inform. Theory*, 51(5):1815–1817, 2005.
- [12] Y. H. Kim. Capacity of a class of deterministic relay channels. *IEEE Trans. Inform. Theory*, 53(3):1328–1329, 2008.
- [13] M. Aleksic, P. Razaghi, and W. Yu. Capacity of a class of modulosum relay channels. *Submitted to IEEE Transactions on Information Theory*, 2007, <http://arxiv.org/pdf/0704.3591>.