

# On the Benefit of Cooperation in Relay Networks

Oliver Kosut, Michelle Effros, Michael Langberg

**Abstract**—This work addresses the cooperation facilitator (CF) model, in which network nodes coordinate through a rate limited communication device. For independent multiple-access channel (MAC) encoders, the CF model is known to show significant rate benefits, even when the rate of cooperation is negligible. Specifically, the benefit in MAC sum-rate, as a function of the cooperation rate  $C_{CF}$ , sometimes has an infinite slope at  $C_{CF} = 0$ . This work studies the question of whether cooperation through a CF can yield similar infinite-slope benefits when applied to internal network encoders in which dependence among MAC transmitters can be established without the help of the CF. Towards this end, this work studies the CF model when applied to relay nodes of a single-source, single-terminal, diamond network consisting of a broadcast channel followed by a MAC. In the relay channel with orthogonal receiver components, careful generalization of the partial-decode-forward/compress-forward lower bound to the CF model yields sufficient conditions for an infinite-slope benefit. Additional results include derivation of a family of diamond networks for which the infinite-slope rate-benefit derives directly from the properties of the corresponding MAC component when studied in isolation.

## I. INTRODUCTION

The information theory and communication literatures approach the goal of improving network communication performance in a variety of ways. While some studies investigate how to get the best possible performance out of existing networks, others seek better designs for future networks. In practice, the way that networks improve over time is somewhere in between — a combination of adding new resources and making better use of what is already there. We here seek new tools for guiding that process, focusing on the questions of whether and where small changes to an existing network can have a big impact on network capacity.

One example of a network in which incremental network modifications can achieve radical network improvement, introduced in [1], employs the multiple-access channel (MAC) and a node called a cooperation facilitator (CF). In practice, the CF is any communicating device that can receive information from multiple transmitters. In any MAC for which dependent channel inputs from the transmitters would yield a higher mutual information between the MAC's inputs and output than is achievable with the independent channel inputs employed in calculating the MAC capacity, adding a small communication link from the CF to either or both of the transmitters yields

a disproportionately large capacity improvement [1]. Specifically, the curve describing the improvement in MAC sum-capacity as a function of the capacity  $C_{CF}$  of the cooperation-enabling CF output link has slope infinity at  $C_{CF} = 0$  [1, Theorem 3]. In some cases, even a single bit — not rate 1, but a single bit no matter what the blocklength, suffices to change the network capacity [2], [3].

Since the infinite-slope improvement in the MAC-capacity results from creating dependence where none could otherwise be observed, it is tempting to believe that the infinite-slope phenomenon cannot occur either in cases where dependence is already attainable or where dependence is not critical to attaining the best possible performance. In this paper, we explore these two intuitions — seeking to understand whether incremental changes can achieve disproportionate channel benefits in these scenarios.

Toward this end, we investigate a single coding framework where both scenarios can arise. We pose this framework as a diamond network in which a single transmitter communicates to a collection of relays, and the relays work independently to transmit information to a shared receiver. Since the communication goal in the diamond network is to transmit information from a single transmitter at the start of the diamond network to a single receiver at its end, dependence at the relays may be available naturally; we investigate whether this availability precludes the possibility of incremental change with disproportionate impact. When the links from relays to the receiver are independent, point-to-point channels, the resulting degenerate MAC fails to meet the prior condition specifying that input dependence should increase sum capacity; we investigate whether this failure precludes the desired small cost, large benefit tradeoff to incremental network modifications.

The rest of this paper is organized as follows. In Section II, we set up the problem of the diamond relay network with  $N$  relay nodes and a cooperation facilitator (Fig. 1), which allows us to pose our main question about the power of cooperation in a relay network. Our results focus on two special cases of this network. The first, covered in Section III, is the relay channel with orthogonal receiver components (Fig. 2). Here, we present an achievability bound for the CF problem, as well as sufficient conditions for the infinite-slope phenomenon to occur. In Section IV, we explore a 3-relay example (Fig. 3) that allows us to exploit the results of [1] on the MAC to demonstrate the infinite-slope phenomenon in a larger network with only one source.

## II. PROBLEM SETUP

*Notation:* For any integer  $k$ ,  $[k]$  denotes the set  $\{1, 2, \dots, k\}$ . Capital letters (e.g.,  $X$ ) denote random variables, lower-case letters (e.g.,  $x$ ) denote realizations of the corresponding

O. Kosut is with the School of Electrical, Computer and Energy Engineering at Arizona State University. Email: okosut@asu.edu

M. Effros is with the Department of Electrical Engineering at the California Institute of Technology. Email: effros@caltech.edu

M. Langberg is with the Department of Electrical Engineering at the University at Buffalo (State University of New York). Email: mikel@buffalo.edu

This work is supported in part by NSF grants CCF-1817241, CCF-1908725, and CCF-1909451.

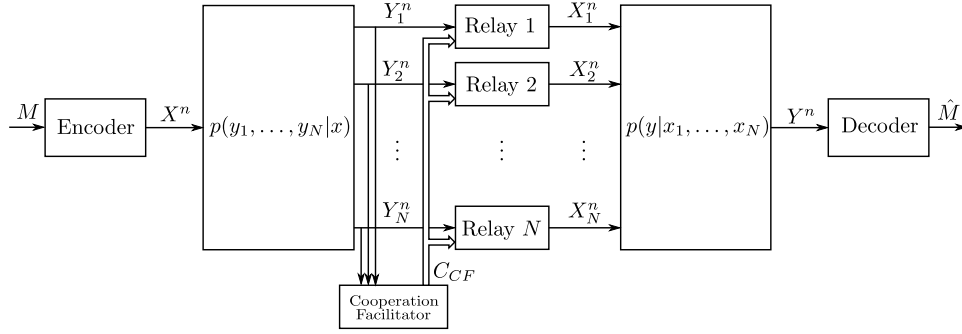


Fig. 1. General diamond relay network with  $N$  nodes and a cooperation facilitator (CF).

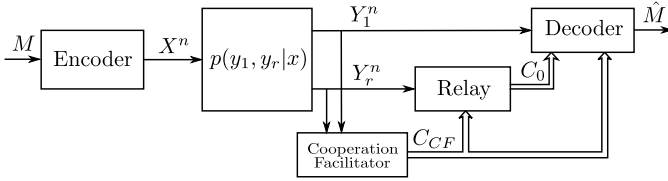


Fig. 2. Relay channel with orthogonal receiver components and a cooperation facilitator, a special case of the general diamond relay network.

variable, and calligraphic letters (e.g.,  $\mathcal{X}$ ) denote the corresponding alphabet. Vectors are denoted with superscript (e.g.,  $x^n = (x_1, \dots, x_n)$ ). We use standard notation for mutual information and entropy.

A diamond relay network with  $N$  relay nodes and a cooperation facilitator (CF)—shown in Fig. 1—is given by a broadcast channel  $p(y_1, \dots, y_N | x)$ , followed by a MAC  $p(y | x_1, \dots, x_N)$ . An  $(n, R)$  code for the diamond relay network is composed of

- an encoder  $f : [2^{nR}] \rightarrow \mathcal{X}^n$ ,
- a CF-encoder  $f_{CF} : \prod_{j \in [N]} \mathcal{Y}_j^n \rightarrow [2^{nC_{CF}}]$ ,
- a relay encoder  $f_j : \mathcal{Y}_j^n \times [2^{nC_{CF}}] \rightarrow \mathcal{X}_j^n$  for each  $j \in [N]$ ,
- a decoder  $g : \mathcal{Y}^n \rightarrow [2^{nR}]$ .

The message  $M$  is assumed to be uniformly drawn from  $[2^{nR}]$ . Encoded message  $X^n = f(M)$  is transmitted by the encoder into the broadcast channel, which outputs  $Y_j^n$  at relay  $j$ ,  $j \in [N]$ . The CF observes all the outputs of the broadcast channel, and encodes  $K = f_{CF}(Y_1^n, \dots, Y_N^n)$ , which is sent to each relay. Relay  $j$  encodes  $X_j^n = f_j(Y_j^n, K)$  and transmits it into the MAC. Finally, the output signal  $Y^n$  is received and decoded to  $\hat{M} = g(Y^n)$ . The overall probability of error is given by  $P_e^{(n)} = P(M \neq \hat{M})$ . We say a rate  $R$  is *achievable* if there exists a sequence of  $(n, R)$  codes with  $P_e^{(n)} \rightarrow 0$ . The *capacity*  $C(C_{CF})$  is the supremum of all achievable rates for a given CF capacity  $C_{CF}$ . This function is non-decreasing in  $C_{CF}$ , and so its derivative  $C'(C_{CF})$  is non-negative.

We are interested in characterizing  $C(C_{CF})$ , but more specifically, we are focused on the following question:

*Main question:* For a given network, is  $C'(0) = \infty$ ?

### III. RELAY CHANNEL WITH ORTHOGONAL RECEIVER COMPONENTS

The first special case of the diamond relay network that we focus on is the relay channel with orthogonal receiver components. Here, we specialize the general model described above in several ways. First, we assume there are only  $N = 2$  relay nodes, and further we assume that the received signal at the decoder is made up of orthogonal components, one from each relay. That is,  $Y = (Y_a, Y_b)$ , where the MAC model factors as

$$p(y_a, y_b | x_1, x_2) = p(y_a | x_1) p(y_b | x_2). \quad (1)$$

Given this factorization, the capacity of the overall network depends on the channels from  $X_1$  to  $Y_a$  and from  $X_2$  to  $Y_b$  only through their capacities [4]. Thus, we can simplify the problem by replacing these noisy channels with rate-limited bit-pipes of capacities  $C_1$  and  $C_0$  from each relay to the decoder. Finally, we assume that  $C_1 = \infty$ ; i.e., we assume that Relay 1 is able to transmit all of its information (consisting of  $Y_1^n$  as well as the CF signal  $K$ ) directly to the decoder. These simplifications yield the network model shown in Fig. 2. Note that this network only has one relay node, so it makes sense to call it a relay channel model rather than a diamond network model. We have also relabelled  $Y_2$  as  $Y_r$  to emphasize that it is the relay's received signal; this also makes the notation consistent with [5], [6].

#### A. Main Achievability Result

*Theorem 1:* Consider a relay channel with orthogonal receiver components with broadcast channel distribution  $p(y_r, y_1 | x)$ , capacity  $C_0$  from relay to destination, and the CF capacity  $C_{CF}$ . Rate  $R$  is achievable if

$$R \leq I(U; Y_r) + \min\{I(X; Y_1, Y_r | U), I(X; Y_1, V | U)\}, \quad (2)$$

$$R \leq \min\{I(U; Y_1), I(U; Y_r)\} + I(X; Y_1 | U) + I(V; X, Y_1 | U) - I(Y_r; V | U) + C_0, \quad (3)$$

$$C_{CF} \geq I(X, Y_1; V | U, Y_r), \quad (4)$$

for some distribution

$$p(u, x) p(y_r, y_1 | x) p(v | x, y_1, y_r, u). \quad (5)$$

*Proof:* See Appendix A. ■

*Remark 1:* Thm. 1 reduces to the well-known combined partial-decode-forward/compress-forward lower bound for the standard problem (without a CF), which originated in [7]. For the relay channel with orthogonal receiver components, [5] showed that this classical bound can be written as follows: rate  $R$  is achievable if

$$R \leq I(U; Y_r) + I(X; Y_1, V|U), \quad (6)$$

$$R \leq \min\{I(U; Y_1), I(U; Y_r)\} + I(X; Y_1|U) + C_0 - I(Y_r; V|U, X, Y_1) \quad (7)$$

for some distribution

$$p(u, x)p(y_r, y_1|x)p(v|y_r, u). \quad (8)$$

In Thm. 1, removing the CF is equivalent to setting  $C_{CF} = 0$ . Thus, (4) implies the Markov chain  $(X, Y_1) - (U, Y_r) - V$ , which implies that the variables have a joint distribution that factors as (8). Thus, by the data processing inequality,  $I(X; Y_1, V|U) \leq I(X; Y_1, Y_r|U)$ , so (2) becomes (6). In addition,

$$I(V; X, Y_1|U) - I(Y_r; V|U) \quad (9)$$

$$= -H(V|U, X, Y_1) + H(V|U, Y_r) \quad (10)$$

$$= -H(V|U, X, Y_1) + H(V|U, Y_r, X, Y_1) \quad (11)$$

$$= -I(Y_r; V|U, X, Y_1), \quad (12)$$

so (3) becomes (7).

### B. Sufficient Conditions for Infinite Slope

The following theorem provides a sufficient condition for which, given a starting achievable point for the partial-decode-forward/compress-forward bound without cooperation (i.e., the bound in (6)–(7)), the achievable rate from Thm. 1 with cooperation improves over the starting point and this improvement has infinite slope as a function of  $C_{CF}$ .

*Theorem 2:* Fix a distribution  $p(u, x)p(v|y_r, u)$ . Let  $R$  be a rate satisfying the no-cooperation achievability conditions in (6)–(7) for this distribution. Suppose  $I(X; Y_1, V|U) < I(X; Y_1, Y_r|U)$ , and there does *not* exist  $\lambda \in [0, 1]$  and  $\gamma(u, x, y_1, y_r) \in \mathbb{R}$  for each  $x, y_1, y_r, u$  such that

$$p(v|u, x, y_1) = \frac{p(v|u, y_1)^\lambda p(v|u, y_r)^{1-\lambda}}{\gamma(u, x, y_1, y_r)} \quad (13)$$

for all  $u, x, y_1, y_r, v$  where  $p(u, x, y_1, y_r) > 0, p(v|y_r, u) > 0$ . Then

$$\lim_{C_{CF} \rightarrow 0} \frac{C(C_{CF}) - R}{C_{CF}} = \infty. \quad (14)$$

*Proof:* See Appendix B. ■

*Remark 2:* We note that there are two ways for (14) to hold: (1)  $C(0) > R$ ; that is, the rate  $R$ , while achievable without cooperation, is smaller than the no-cooperation capacity of the relay channel; (2)  $C(0) = R$ , and  $C'(0) = \infty$ . Here, the rate  $R$  is the no-cooperation capacity, so (14) indicates that the CF really can improve the capacity of the relay network in an infinite-slope manner. Thus, this latter case is the one we are particularly interested in, as it gives an affirmative answer to the Main Question. Unfortunately, for any problem instance for which a matching converse for the no-cooperation setting

is unavailable, even if (14) holds, there is no way to know which situation we are in. Still, if  $R$  represents the best-known achievable rate for a given network, (14) has a non-trivial consequence, showing that the state-of-the-art can be improved disproportionately by a small amount of cooperation.

While the condition in Thm. 2 is sometimes hard to verify, the following corollary provides a simpler sufficient condition for the same conclusion.

*Corollary 3:* Assume that  $p(y_1, y_r|x) > 0$  for all letters  $x, y_1, y_r$ . Consider any distribution  $p(u, x)p(v|u, y_r)$ . Let  $R$  be a rate satisfying (6)–(7) for this distribution. Then at least one of the following possibilities hold:

- 1) there exists a function  $g : \mathcal{U} \times \mathcal{Y}_r \rightarrow \mathcal{V}$  where rate  $R$  satisfies (6)–(7) with  $V = g(U, Y_r)$ .
- 2) (14) holds.

$R_{\text{PD/C}}(0) = p(v|u, y_r)$  is non-deterministic; i.e.,  $0 < p(v|u, y_r) < 1$  for some  $u, v, y_r$ . Then  $R'_{\text{PD/C}}(0) = \infty$ .

*Proof:* See Appendix C. ■

### C. Example Relay Channels

For some relay channels, [8] showed that the compress-forward bound achieves capacity. Thus, it is possible to definitively answer the Main Question for these channels. The following example illustrates one such channel.

*Example 1:* Let  $X \in \{0, 1\}$ ,  $Y_1 = X \oplus Z$ ,  $Y_r = Z \oplus W$ , where  $\oplus$  indicates modulo-2 addition,  $Z \sim \text{Ber}(p)$ ,  $W \sim \text{Ber}(\delta)$ , and  $X, Z, W$  are mutually independent. For this channel, the capacity without cooperation is shown in [8] to be given by

$$C(0) = \max_{p(v|y_r): I(Y_r; V) \leq C_0} 1 - H(Z|V). \quad (15)$$

Moreover, this rate is achieved by compress-forward by choosing  $X \sim \text{Ber}(1/2)$  and setting  $p(v|y_r)$  to be the distribution achieving the maximum in (15). Corollary 3 applies to this channel, since  $p(y_1, y_r|x) > 0$  for all  $(x, y_1, y_r)$  as long as  $p, \delta \in (0, 1)$ . Moreover, the only deterministic distributions from  $Y_r$  to  $V$  are where either  $V$  is a constant, or  $V = Y_r$  (or equivalent). It is easy to see that as long as  $0 < C_0 < H(Y_r) = H(p \oplus \delta)$ , neither of these choices for  $V$  is optimal. Therefore, in all non-trivial cases,  $C'(0) = \infty$  for this channel.

The following relay channel example is one for which the no-cooperation capacity is not known. However, we can verify the sufficient condition from Thm. 2, thus showing that an infinite-slope improvement is possible through cooperation.

*Example 2:* Let  $X \in \{0, 1\}$ , and let  $p(y_1, y_r|x) = p(y_1|x)p(y_r|x)$ , where each of the two component channels is a binary erasure channel (BEC) with erasure probability  $p$ . An achievable rate for the no-cooperation case from (6)–(7) is given by taking  $U = \emptyset$ ,  $X$  to be uniform on  $\{0, 1\}$ , and  $p(v|y_r)$  to be a channel that further erases any un-erased bit with probability  $q$ ; that is,

$$p(v|y_r) = \begin{cases} 1 - q & v = y_r \in \{0, 1\} \\ q & v = e, y_r \in \{0, 1\}, \\ 1 & v = y_r = e, \\ 0 & \text{otherwise.} \end{cases} \quad (16)$$

This leads to the achievable rate

$$R = \max_{q \in [0,1]} \min\{(1-p)(1+p(1-q)), 1-p-H((1-p)(1-q)) + (1-p)H(q) + C_0\} \quad (17)$$

where  $H(\cdot)$  is the binary entropy function.

Note that this channel does not satisfy the conditions of Corollary 3, since  $p(y_1, y_r|x)$  is not always positive. Instead, we verify the sufficient condition of Thm. 2 directly. Suppose that there exists a  $\lambda$  and  $\gamma$  satisfying (13). Note that  $Y_1 - X - Y_r - V$  is a Markov chain, so  $p(v|x, y_1) = p(v|x)$ . For  $x \in \{0, 1\}$  and any  $y_1 \in \{x, e\}$

$$\frac{p_{V|X}(x|x)}{p_{V|X}(e|x)} = \frac{(1-p)(1-q)}{1-(1-p)(1-q)} \quad (18)$$

$$= \frac{p_{V|Y_1}(x|y_1)^\lambda (1-q)^{1-\lambda}}{\gamma'(x, y_1, x)} \cdot \frac{\gamma'(x, y_1, x)}{p_{V|Y_1}(e|y_1)^\lambda q^{1-\lambda}} \quad (19)$$

$$= \left(\frac{1-q}{q}\right)^{1-\lambda} \left(\frac{p_{V|Y_1}(x|y_1)}{p_{V|Y_1}(e|y_1)}\right)^\lambda \quad (20)$$

$$= \left(\frac{1-q}{q}\right)^{1-\lambda} \begin{cases} \left(\frac{(1-p)(1-q)}{1-(1-p)(1-q)}\right)^\lambda, & y_1 = x \\ \left(\frac{\frac{1}{2}(1-p)(1-q)}{1-(1-p)(1-q)}\right)^\lambda, & y_1 = e. \end{cases} \quad (21)$$

This cannot hold with equality for both  $y_1 = x$  and  $y_1 = e$  unless  $p = 0$ ,  $p = 1$ , or  $q = 1$ . Therefore, except in these trivial cases, infinite slope improvement occurs.

#### IV. 3-RELAY NETWORK EXAMPLE

In the analysis of Theorem 1 and  $C'(0)$  for the orthogonal-receiver setting given in Sections III-A and III-B, relay-cooperation is governed by the statistics of the broadcast channel  $p(y_1, y_r|x)$  and, roughly speaking, is designed to “remove” from  $Y_r^n$  message information that can be obtained at the receiver from  $Y_1^n$ . In this aspect, we say that the design of cooperation information looks *backwards* and is governed by the broadcast channel of the diamond network.

In this section, we study a *forward* form of cooperation, that takes into account the MAC appearing in the second stage of the diamond network. For forward-looking cooperation, it is tempting to treat the MAC “in isolation”, rather than as part of a larger network, and to design cooperation solely based on the MAC noise statistics, as done in [1], [3]. In general, designing cooperation by treating the MAC as an isolated component may not suffice to improve communication of the diamond network, because MAC encoders in the diamond network potentially hold dependent information resulting from the broadcast stage of communication. Nevertheless, in what follows, we present a family of 3-relay diamond networks for which cooperation-gain in the network as a whole is derived directly from the MAC cooperation-gain when studied in isolation; the latter is well understood and given in [1]. Our network family is described below and depicted in Figure 3.

Consider the *diamond network* defined by broadcast channel  $(\mathcal{X}, p(y_0, y_1, y_2|x), \mathcal{Y}_0, \mathcal{Y}_1, \mathcal{Y}_2)$ , and MAC  $(\mathcal{X}_0, \mathcal{X}_1, \mathcal{X}_2, q(y|x_0, x_1, x_2), \mathcal{Y})$ . More specifically, as depicted in Figure 3, consider the case in which  $\mathcal{X} = \mathcal{Y}_0 = \mathcal{Y}_1 = \mathcal{X}_0 = \mathcal{X}_1 = \{0, 1\}$ ;  $\mathcal{Y}_2 = \mathcal{X}_2 = \{0, 1\}^2$ ;

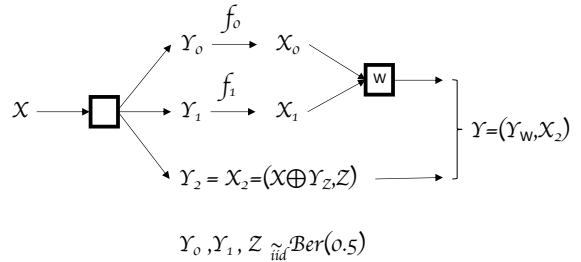


Fig. 3. The 3-relay diamond network of Claim 4.

$\mathcal{Y} = \mathcal{Y}_W \times \mathcal{X}_2$  for a given memoryless 2-user binary MAC  $W$ :  $(\mathcal{X}_0, \mathcal{X}_1, p_W(y_W|x_0, x_1), \mathcal{Y}_W)$ ; for any  $x$ ,  $p(Y_0, Y_1, Y_2|x)$  induces  $(Y_0, Y_1, Y_2)$  where  $Y_0, Y_1, Z$  are independent Bernoulli(0.5) random variables and  $Y_2 = (x \oplus Y_Z, Z)$ ;  $X_0^n = f_0(Y_0^n)$ ,  $X_1^n = f_1(Y_1^n)$ ,  $X_2^n = f_2(Y_2^n)$  for relay encoders  $f_0, f_1$ , and  $f_2$ ; and  $q(Y|x_0, x_1, x_2)$  for which  $Y = (Y_W, x_2)$ . As  $Y^n$  holds the value of  $X_2^n \in \{0, 1\}^n$ , which in turn depends on  $Y_2^n \in \{0, 1\}^n$  through  $f_2$ , we assume without loss of generality that  $X_2^n = Y_2^n$ .

Using the independent nature of relays  $Y_0$  and  $Y_1$ , in Claim 4 below we tie the cooperation gain of the 2-transmitter MAC  $W$  with the cooperation gain of the diamond network.

**Claim 4:** Let  $C_{\text{sum}}(C_{CF})$  be the sum-capacity of the 2-transmitter MAC  $W$  with user cooperation of rate  $C_{CF}$ . Then the capacity  $C(C_{CF})$  of the diamond network satisfies  $C'(0) = \infty$  if  $C'_{\text{sum}}(0) = \infty$ .

*Proof:* See Appendix D. ■

#### REFERENCES

- [1] P. Noorzad, M. Effros, and M. Langberg, “The unbounded benefit of encoder cooperation for the k-user MAC,” *IEEE Transactions on Information Theory*, vol. 64, no. 5, pp. 3655–3678, 2018.
- [2] M. Langberg and M. Effros, “On the capacity advantage of a single bit,” in *IEEE Workshop on Network Coding and Applications (NetCod)*, 2016, pp. 1–6.
- [3] O. Kosut, M. Effros, and M. Langberg, “Every bit counts: Second-order analysis of cooperation in the multiple-access channel,” in *IEEE International Symposium on Information Theory (ISIT)*, 2021, pp. 2214–2219.
- [4] R. Koetter, M. Effros, and M. Médard, “A theory of network equivalence – Part I: Point-to-point channels,” *IEEE Trans. Inf. Theory*, vol. 57, pp. 972–995, Feb. 2011.
- [5] A. El Gamal, A. Gohari, and C. Nair, “Achievable rates for the relay channel with orthogonal receiver components,” in *2021 IEEE Information Theory Workshop (ITW)*, 2021, pp. 1–6.
- [6] —, “Strengthened cutset upper bound on the capacity of the relay channel and applications,” in *2021 IEEE International Symposium on Information Theory (ISIT)*, 2021, pp. 1344–1349.
- [7] T. Cover and A. El Gamal, “Capacity theorems for the relay channel,” *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, 1979.
- [8] M. Aleksic, P. Razaghi, and W. Yu, “Capacity of a class of modulo-sum relay channels,” *IEEE Transactions on Information Theory*, vol. 55, no. 3, pp. 921–930, 2009.
- [9] A. El Gamal and Y. Kim, *Network Information Theory*. Cambridge University Press, 2011.

APPENDIX A  
PROOF OF THEOREM 1

We denote  $T_\epsilon^{(n)}$  as the robustly typical set. See [9] for the definition, as well as the formal statement of the packing lemma, which will be used in the proof.

We employ the following lemma, which is a slight variation on the covering lemma from [9].

*Lemma 5:* Let  $(U, X, \hat{X}) \sim p(u, x, \hat{x})$  and  $\epsilon' < \epsilon$ . Let  $(u^n, x^n) \in T_{\epsilon'}(U, X)$  be a pair of fixed sequences, and let  $\hat{X}^n(m), m \in \mathcal{A}$ , where  $|\mathcal{A}| \geq 2^{nR}$ , be random sequences, conditionally independent of each other, each uniformly distributed on  $T_\epsilon^{(n)}(\hat{X}|u^n)$ . Let  $M^*$  be the smallest  $m$  for which

$$\hat{X}^n(m) \in T_\epsilon^{(n)}(\hat{X}|u^n, x^n). \quad (22)$$

If there is no such  $m$ , we say  $M^*$  is undefined. Then,

- 1) there exists  $\delta(\epsilon)$  that tends to zero as  $\epsilon \rightarrow 0$  such that  $\lim_{n \rightarrow \infty} P(M^* \text{ is undefined}) = 0$ , if  $R > I(X; \hat{X}|U) + \delta(\epsilon)$ ,
- 2) conditioning on the event that  $M^*$  is defined,  $\hat{X}^n(M^*)$  is uniformly distributed in  $T_\epsilon^{(n)}(\hat{X}|u^n, x^n)$ .

*Proof:* For any  $m \in \mathcal{A}$ , we have

$$P((u^n, x^n, \hat{X}^n(m)) \in T_\epsilon^{(n)}|U^n = u^n, X^n = x^n) \quad (23)$$

$$= P((u^n, x^n, \hat{X}^n(m)) \in T_\epsilon^{(n)}|U^n = u^n) \quad (24)$$

$$= \sum_{\hat{x}^n \in T_\epsilon^{(n)}(\hat{X}|u^n, x^n)} \frac{1}{|T_\epsilon^{(n)}(\hat{X}|u^n)|} \quad (25)$$

$$= \frac{|T_\epsilon^{(n)}(\hat{X}|u^n, x^n)|}{|T_\epsilon^{(n)}(\hat{X}|u^n)|} \quad (26)$$

$$\geq 2^{-n(I(X; \hat{X}|U) + \delta(\epsilon))}. \quad (27)$$

The remainder of the proof of statement 1 follows from an identical argument to that of the standard covering lemma, i.e., [9, Lemma 3.3].

To prove statement 2, note first that, if  $M^*$  is defined, then by definition  $\hat{X}^n(M^*) \in T_\epsilon^{(n)}(\hat{X}|u^n, x^n)$ . Now, for any  $\hat{x}^n \in T_\epsilon^{(n)}(\hat{X}|u^n, x^n)$ ,

$$P(\hat{X}^n(M^*) = \hat{x}^n | M^* \text{ is defined}) \quad (28)$$

$$= \sum_{m \in \mathcal{A}} P(M^* = m | M^* \text{ is defined}) P(\hat{X}^n(m) = \hat{x}^n |$$

$$\hat{X}^n(m') \notin T_\epsilon^{(n)}(\hat{X}|u^n, x^n) \text{ for all } m' < m,$$

$$\hat{X}^n(m) \in T_\epsilon^{(n)}(\hat{X}|u^n, x^n)) \quad (29)$$

$$= \sum_{m \in \mathcal{A}} P(M^* = m | M^* \text{ is defined}) P(\hat{X}^n(m) = \hat{x}^n |$$

$$\hat{X}^n(m) \in T_\epsilon^{(n)}(\hat{X}|u^n, x^n)) \quad (30)$$

$$= \sum_{m \in \mathcal{A}} P(M^* = m | M^* \text{ is defined}) \frac{1}{|T_\epsilon^{(n)}(\hat{X}|u^n, x^n)|} \quad (31)$$

$$= \frac{1}{|T_\epsilon^{(n)}(\hat{X}|u^n, x^n)|} \quad (32)$$

where (30) holds since  $\hat{X}^n(m)$  are independent for different  $m$ , and (31) since  $\hat{X}^n(m)$  is uniform on  $T_\epsilon^{(n)}(\hat{X}|u^n)$ , and  $T_\epsilon^{(n)}(\hat{X}|u^n) \subset T_\epsilon^{(n)}(\hat{X}|u^n, x^n)$ . This proves statement 2. ■

We now proceed to the main proof of the theorem. The following argument combines partial-decode-forward/compress-forward strategies, as discussed in Remark 1. Fix rates  $R_a, R_b, S$  to be determined, where  $R_a + R_b = R$ . Also fix small constants  $0 < \epsilon' < \epsilon$ . We construct a code as follows.

*Codebook generation:*

- For each  $m_a \in [2^{nR_a}]$ , generate  $u^n(m_a) \sim \prod_{i=1}^n p_U(u_i)$ .
- For each  $m_a \in [2^{nR_a}], m_b \in [2^{nR_b}]$ , generate  $x^n(m_a, m_b) \sim \prod_{i=1}^n p_{X|U}(x_i|u_i(m_a))$ .
- For each  $m_a \in [2^{nR_a}], \ell \in [2^{nS}], k \in [2^{nC_{CF}}]$ , generate  $v^n(m_a, \ell, k) \sim \text{Unif}[T_\epsilon^{(n)}(V|u^n(m_a))]$ , and corresponding source coding bins  $k \in [2^{nC_{CF}}]$ , generate  $m_0(m_a, \ell, k) \sim \text{Unif}[2^{nC_0}]$ .
- For each  $m_a \in [2^{nR_a}], y_r^n \in \mathcal{Y}_r^n$  and  $k \in [2^{nC_{CF}}]$ , let  $\ell(m_a, y_r^n, k)$  be the smallest  $\ell \in [2^{nS}]$  such that

$$(u^n(m_a), y_r^n, v^n(m_a, \ell, k)) \in T_\epsilon^{(n)}(U, Y_r, V). \quad (33)$$

If there is no such  $\ell$ , we say  $\ell(m_a, y_r^n, k)$  is undefined.

*Encoding:* At the transmitter, given message  $m = (m_a, m_b)$ , send  $x^n(m_a, m_b)$ .

*CF coding:* At the CF, given  $y_1^n$  and  $y_r^n$ , first find the unique pair  $\hat{m}_a, \hat{m}_b$  such that

$$(u^n(\hat{m}_a), x^n(\hat{m}_a, \hat{m}_b), y_1^n, y_r^n) \in T_\epsilon^{(n)}(U, X, Y_1, Y_r). \quad (34)$$

Next, find the smallest  $k \in [2^{nC_{CF}}]$  such that  $\ell(m_a, y_r^n, k)$  is defined, and

$$(u^n(\hat{m}_a), x^n(\hat{m}_a, \hat{m}_b), y_r^n, y_1^n, v^n(\hat{m}_a, \ell(\hat{m}_a, y_r^n, k), k)) \in T_\epsilon^{(n)}(U, X, Y_r, Y_1, V). \quad (35)$$

Send this  $k$ . If there is no such  $k$ , declare an error.

*Relay coding:* At the relay, given  $y_r^n$  and  $k$ , first find  $\hat{m}_a$  such that

$$(u^n(\hat{m}_a), y_r^n) \in T_\epsilon^{(n)}(U, Y_r). \quad (36)$$

Then let  $\ell = \ell(\hat{m}_a, y_r^n, k)$ , and send  $m_0(\hat{m}_a, \ell, k)$ . If  $\ell(\hat{m}_a, y_r^n, k)$  is undefined, declare an error.

*Decoding:* At the decoder, given  $y_1^n, m_0$ , and  $k$ , find  $\hat{m}_a, \hat{m}_b, \hat{\ell}$  such that

$$(u^n(\hat{m}_a), x^n(\hat{m}_a, \hat{m}_b), y_1^n, v^n(\hat{m}_a, \hat{\ell}, k)) \in T_\epsilon^{(n)}, \quad (37)$$

$$m_0(\hat{m}_a, \hat{\ell}, k) = m_0. \quad (38)$$

*Error analysis:* Throughout the error analysis, we assume without loss of generality that  $m_a = m_b = 1$ . To prove that  $m_a$  is decoded correctly at both the CF and the relay, and that  $m_b$  is decoded correctly at the CF, it suffices to consider the following error events:

$$\mathcal{E}_1 = \{(u^n(1), x^n(1, 1), y_r^n, y_1^n) \notin T_{\epsilon'}^{(n)}(U, X, Y_r, Y_1)\}, \quad (39)$$

$$\mathcal{E}_2 = \{(u^n(m_a), y_r^n) \in T_\epsilon^{(n)}(U, Y_r) \text{ for some } m_a \neq 1\} \quad (40)$$

$$\mathcal{E}_3 = \{(u^n(1), x^n(1, m_b), y_r^n, y_1^n) \in T_\epsilon^{(n)}(U, X, Y_r, Y_1) \text{ for some } m_b \neq 1\}. \quad (41)$$

By the law of large numbers  $P(\mathcal{E}_1) \rightarrow 0$ . By the packing lemma,  $P(\mathcal{E}_2) \rightarrow 0$  and  $P(\mathcal{E}_3) \rightarrow 0$  if

$$R_a < I(U; Y_r), \quad (42)$$

$$R_b < I(X; Y_1, Y_r | U). \quad (43)$$

We now show that  $\ell(1, y_r^n, k)$  is defined for most values of  $k$ . For each  $k \in [2^{nC_{CF}}]$ , define the event

$$\mathcal{E}_4(k) = \{(u^n(1), y_r^n, v^n(1, \ell, k)) \notin T_\epsilon^{(n)}(U, Y_r, V) \text{ for all } \ell \in [2^{nS}]\}. \quad (44)$$

Since we have already established that  $P(\mathcal{E}_1) \rightarrow 0$ , by Lemma 5,  $P(\mathcal{E}_4(k)) \rightarrow 0$  if

$$S > I(Y_r; V | U). \quad (45)$$

Moreover, Lemma 5 asserts that, given  $\ell(1, y_r^n, k)$  is defined,  $v^n(1, \ell(1, y_r^n, k), k)$  is uniformly distributed on  $T_\epsilon^{(n)}(V | u^n(1), y_r^n)$ . Now consider the event

$$\mathcal{E}_5 = \{|\{k : \ell(1, y_r^n, k) \text{ is defined}\}| < (1 - \epsilon)2^{nC_{CF}}\}. \quad (46)$$

Since  $\ell(1, y_r, k)$  is defined if and only if  $\mathcal{E}_4(k)$  does not occur, it is straightforward to show that  $P(\mathcal{E}_5) \rightarrow 0$  as  $n \rightarrow \infty$ . Now consider the error event in which the CF cannot find a value of  $k$  to transmit, i.e.,

$$\mathcal{E}_6 = \{(u^n(1), x^n(1, 1), y_r^n, y_1^n, v^n(1, \ell(1, y_r^n, k), k)) \notin T_\epsilon^{(n)} \text{ for all } k \in [2^{nC_{CF}}]\}. \quad (47)$$

We may now apply Lemma 5 a second time to find that  $P(\mathcal{E}_6) \rightarrow 0$  if

$$C_{CF} > I(X, Y_1; V | U, Y_r). \quad (48)$$

Let us further assume without loss of generality that  $k = 1$ ,  $\ell(1, y_r^n, 1) = 1$ , and  $m_0(1, 1, 1) = 1$ . Assuming that error events  $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_6$  do not occur, the relay selects  $\hat{m}_a = \ell = m_0 = 1$ , and

$$(u^n(1), x^n(1, 1), y_r^n, y_1^n, v^n(1, 1, 1)) \in T_\epsilon^{(n)}. \quad (49)$$

Now consider the following decoding error events:

$$\mathcal{E}_7 = \{(u^n(1), x^n(1, m_b), y_r^n, v^n(1, 1, 1)) \in T_\epsilon^{(n)} \text{ for some } m_b \neq 1\}, \quad (50)$$

$$\mathcal{E}_8 = \{(u^n(1), x^n(1, m_b), y_r^n, v^n(1, \ell, 1)) \in T_\epsilon^{(n)}, m_0(1, \ell, 1) = 1 \text{ for some } m_b \neq 1, \ell \neq 1\}, \quad (51)$$

$$\mathcal{E}_9 = \{(u^n(m_a), x^n(m_a, m_b), y_r^n, v^n(m_a, \ell, 1)) \in T_\epsilon^{(n)}, m_0(m_a, \ell, 1) = 1 \text{ for some } m_a \neq 1, m_b, \ell\}. \quad (52)$$

Applying the packing lemma several times,  $P(\mathcal{E}_7) \rightarrow 0$  if

$$R_b < I(X; Y_1, V | U), \quad (53)$$

$P(\mathcal{E}_8) \rightarrow 0$  if

$$R_b + S < I(X; Y_1 | U) + I(V; X, Y_1 | U) + C_0, \quad (54)$$

and  $P(\mathcal{E}_9) \rightarrow 0$  if

$$R_a + R_b + S < I(U, X; Y_1) + I(V; X, Y_1 | U) + C_0. \quad (55)$$

We now collect the various rate conditions required for all of the error event probabilities to vanish. It is advantageous if  $S$  is as small as possible; from the lower limit in (45), we may assume that  $S$  is slightly larger than  $I(Y_r; V | U)$ . We have three conditions on  $R_b$ , namely (43), (53), and (54). Combining each

of these with the condition on  $R_a$  in (42), and recalling that  $R = R_a + R_b$ , we need

$$R < I(U; Y_r) + I(X; Y_1, Y_r | U), \quad (56)$$

$$R < I(U; Y_r) + I(X; Y_1, V | U), \quad (57)$$

$$R < I(U; Y_r) + I(X; Y_1 | U) + I(V; X, Y_1 | U) + C_0 - I(Y_r; V | U). \quad (58)$$

Furthermore, from (55) we need

$$R < I(U, X; Y_1) + I(V; X, Y_1 | U) - I(Y_r; V | U) + C_0 \quad (59)$$

$$= I(U; Y_1) + I(X; Y_1 | U) + I(V; X, Y_1 | U) - I(Y_r; V | U) + C_0. \quad (60)$$

Therefore, the conditions in the statement of the theorem imply that  $R_a, R_b, S$  can be found such that all of the above conditions hold.

## APPENDIX B PROOF OF THEOREM 2

Under the starting distribution

$$p(u, x)p(y_1, y_r | x)p(v | u, y_r), \quad (61)$$

$I(X, Y_1; V | U, Y_r) = 0$ . To show (14), we modify this distribution slightly, in a way that gives  $I(X, Y_1; V | U, Y_r) > 0$ , which corresponds to positive  $C_{CF}$ , while increasing the achieved rate. In particular, we leave  $p(u, x)$  fixed, but change the conditional distribution for  $v$  to

$$q(v | u, x, y_1, y_r) = p(v | u, y_r) + \alpha r(v | u, x, y_1, y_r) \quad (62)$$

where  $\alpha \approx 0$ . For a variable  $A \subset \{U, X, Y_1, Y_r\}$ , we further define  $r(v | a)$ , for example by

$$r(v | u, y_r) = \sum_{x, y_1} p(x, y_1 | u, y_r) r(v | u, x, y_1, y_r). \quad (63)$$

Thus  $q(v | a) = p(v | a) + \alpha r(v | a)$ . In order for  $q$  to be a valid distribution, we need

$$\sum_v r(v | u, x, y_1, y_r) = 0 \text{ for all } u, x, y_1, y_r. \quad (64)$$

Thus, these  $r$  functions are not really distributions; instead they satisfy  $\sum_v r(v | a) = 0$  for any variable  $A$ . Moreover, if  $p(v | u, y_r) = 0$ , then in order for  $q$  to be valid, we need  $r(v | u, x, y_1, y_r) \geq 0$ ; we here make the simplifying assumption that  $r(v | u, x, y_1, y_r) = 0$  for any  $u, y_r, v$  where  $p(v | u, y_r) = 0$ . This assumption has the following consequence. Suppose for some  $(u, x, y_1, y_r, v)$ ,  $p(u, x, y_1, y_r, v) = 0$ . Recalling

$$p(u, x, y_1, y_r, v) = p(u, x, y_1, y_r)p(v | u, y_r) \quad (65)$$

it must be true that either  $p(u, x, y_1, y_r)$  or  $p(v | u, y_r)$  is zero. That is, if  $p(u, x, y_1, y_r) > 0$ , then  $r(v | u, x, y_1, y_r) = 0$ , so  $q(v | u, y_1, y_r, u) = 0$ . In particular

$$p(u, x, y_1, y_r)q(v | u, x, y_1, y_r) \quad (66)$$

vanishes if and only if  $p(u, x, y_1, y_r, v)$  vanishes.

We are changing the distribution of  $V$ , but not of  $U$ , so many terms in the lower bounds in Thm. 1 do not change with  $\alpha$ . Define the following functions:

$$\begin{aligned} f_1(\alpha) &= I_q(X; V|U, Y_1), \\ f_2(\alpha) &= I_q(V; X, Y_1|U) - I_q(Y_r; V|U), \\ C_{CF}(\alpha) &= I_q(X, Y_1; V|U, Y_r). \end{aligned}$$

Note that  $f_1$  and  $f_2$  include all the terms that change with  $V$  in (2) and (3) respectively. Since by assumption,  $I(X; Y_1, V|U) < I(X; Y_1, Y_r|U)$ , if  $f_1$  increases with  $\alpha$  then so does the right-hand side of (2). Thus, to prove (14), it is enough for

$$C'_{CF}(0) = 0, \quad f'_1(0) > 0, \quad f'_2(0) > 0. \quad (67)$$

We first show that, given the assumptions we have already made,  $C'_{CF}(0) = 0$ . We have

$$C_{CF}(\alpha) = I_q(X, Y_1; V|U, Y_r) \quad (68)$$

$$= D(q(u, x, y_1, y_r, v) \| q(u, x, y_1, y_r) q(v|u, y_r)) \quad (69)$$

$$= \sum_{u, x, y_1, y_r} p(u, x, y_1, y_r) D(q(v|u, x, y_1, y_r) \| q(v|u, y_r)). \quad (70)$$

Note that  $p(v|x, y_1, y_r, u) = p(v|u, y_r)$ , so  $C_{CF}(0) = 0$ . To find  $C'_{CF}(\alpha)$ , consider an arbitrary function of the form

$$f(\alpha) = D(p(x) + \alpha r_1(x) \| p(x) + \alpha r_2(x)) \quad (71)$$

$$= \sum_x (p(x) + \alpha r_1(x)) \log \frac{p(x) + \alpha r_1(x)}{p(x) + \alpha r_2(x)} \quad (72)$$

where  $\sum_x r_i(x) = 0$  for  $i = 1, 2$ , and  $r_1(x) = r_2(x) = 0$  whenever  $p(x) = 0$ . Thus, the only relevant terms in the summation are where  $p(x) > 0$ , so

$$f'(0) = \lim_{\alpha \rightarrow 0} f'(\alpha) \quad (73)$$

$$= \lim_{\alpha \rightarrow 0} \sum_{x:p(x)>0} \left[ r_1(x) \log \frac{p(x) + \alpha r_1(x)}{p(x) + \alpha r_2(x)} + r_1(x) \right. \quad (74)$$

$$\left. - r_2(x) \frac{p(x) + \alpha r_1(x)}{p(x) + \alpha r_2(x)} \right] \quad (75)$$

$$= \sum_{x:p(x)>0} (r_1(x) - r_2(x)) \quad (76)$$

$$= 0. \quad (77)$$

To apply this analysis to the function  $C_{CF}(\alpha)$  from (70), given any  $u, x, y_1, y_r$ , consider

$$D(q(v|u, x, y_1, y_r) \| q(v|u, y_r)). \quad (78)$$

Recall that

$$q(v|u, x, y_1, y_r) = p(v|u, y_r) + \alpha r(v|u, x, y_1, y_r), \quad (79)$$

$$q(v|u, y_r) = p(v|u, y_r) + \alpha r(v|u, y_r). \quad (80)$$

Moreover, we have made the assumption that  $r(v|u, x, y_1, y_r) = 0$  whenever  $p(v|u, y_r) = 0$ , so we have a scenario matching the above assumptions on  $f(\alpha)$ . Thus,  $C'_{CF}(0) = 0$ .

We now consider the conditions when  $f'_1(0), f'_2(0) > 0$ . Consider a variable  $A \subset \{X, Y_1, Y_r, U\}$ . Recalling the fact that if  $p(a) > 0$  and  $p(v|a) = 0$ , then  $q(v|a) = 0$ , we may write

$$H_q(V|A) = - \sum_{a,v:q(v|a)>0} p(a)q(v|a) \log q(v|a) \quad (81)$$

$$= - \sum_{\substack{u,x,y_1,y_r,v: \\ p(u,x,y_1,y_r,v)>0, \\ q(v|u,x,y_1,y_r)>0}} p(u, x, y_1, y_r) q(v|u, x, y_1, y_r) \log q(v|a) \quad (82)$$

$$= - \sum_{\substack{u,x,y_1,y_r,v: \\ p(u,x,y_1,y_r,v)>0}} p(u, x, y_1, y_r) q(v|u, x, y_1, y_r) \log q(v|a) \quad (83)$$

where we have used the fact that  $p$  and  $q$  have precisely the same support. Thus

$$\frac{d}{d\alpha} H_q(V|A) \quad (84)$$

$$= - \sum_{\substack{u,x,y_1,y_r,v: \\ p(u,x,y_1,y_r,v)>0}} p(u, x, y_1, y_r) \frac{d}{d\alpha} q(v|u, x, y_1, y_r) \log q(v|a) \quad (85)$$

$$= - \sum_{\substack{u,x,y_1,y_r,v: \\ p(u,x,y_1,y_r,v)>0}} p(u, x, y_1, y_r) \left[ r(v|u, x, y_1, y_r) \log q(v|a) + \frac{q(v|u, x, y_1, y_r)r(v|a)}{q(v|a)} \right] \quad (86)$$

$$= - \sum_{\substack{u,x,y_1,y_r,v: \\ p(u,x,y_1,y_r,v)>0}} p(u, x, y_1, y_r) r(v|u, x, y_1, y_r) \log q(v|a) + \sum_{a,v} p(a)r(v|a) \quad (87)$$

$$= - \sum_{\substack{u,x,y_1,y_r,v: \\ p(u,x,y_1,y_r,v)>0}} p(u, x, y_1, y_r) r(v|u, x, y_1, y_r) \log q(v|a). \quad (88)$$

In particular,

$$\begin{aligned} \frac{d}{d\alpha} H_q(V|A) \Big|_{\alpha=0} \\ = - \sum_{\substack{u,x,y_1,y_r,v: \\ p(u,x,y_1,y_r,v)>0}} p(u, x, y_1, y_r) r(v|u, x, y_1, y_r) \log p(v|a). \end{aligned} \quad (89)$$

Now we may easily write

$$\begin{aligned} f'_1(0) &= \sum_{\substack{u,x,y_1,y_r,v: \\ p(u,x,y_1,y_r,v)>0}} p(u, x, y_1, y_r) r(v|u, x, y_1, y_r) \\ &\quad \cdot \log \frac{p(v|u, x, y_1)}{p(v|u, y_1)}, \end{aligned} \quad (90)$$

$$\begin{aligned} f'_2(0) &= \sum_{\substack{u,x,y_1,y_r,v: \\ p(u,x,y_1,y_r,v)>0}} p(u, x, y_1, y_r) r(v|u, x, y_1, y_r) \\ &\quad \cdot \log \frac{p(v|u, x, y_1)}{p(v|u, y_r)}. \end{aligned} \quad (91)$$

Recall that we are interested in showing that  $f'_1(0), f'_2(0) > 0$ . Since each of these is a linear function of  $r$ , we consider a generic linear set up. In particular, we are interested in whether there exists a vector  $z$  such that  $a^T z > 0$ ,  $b^T z > 0$ , and  $Az = 0$ . That is, we are interested in

$$\max_{z: Az=0} \min\{a^T z, b^T z\} \quad (92)$$

$$= \max_{z: Az=0} \min_{\lambda \in [0,1]} \lambda a^T z + (1-\lambda)b^T z \quad (93)$$

$$= \max_z \min_{\lambda \in [0,1], \gamma} \lambda a^T z + (1-\lambda)b^T z + \gamma^T Az \quad (94)$$

$$= \max_z \min_{\lambda \in [0,1], \gamma} (\lambda a + (1-\lambda)b + A^T \gamma)^T z \quad (95)$$

$$= \min_{\lambda \in [0,1], \gamma} \max_z (\lambda a + (1-\lambda)b + A^T \gamma)^T z \quad (96)$$

$$= \min_{\lambda \in [0,1], \gamma} \begin{cases} 0 & \lambda a + (1-\lambda)b + A^T \gamma = 0 \\ \infty & \text{otherwise.} \end{cases} \quad (97)$$

That is, there exists no  $z$  of interest if and only if there exists  $\lambda \in [0, 1]$  and  $\gamma$  where

$$\lambda a + (1-\lambda)b + A^T \gamma = 0. \quad (98)$$

Applying this principle to our situation, there does *not* exist such an  $r$  function if and only if there exists  $\lambda \in [0, 1]$ ,  $\gamma(u, x, y_1, y_r)$  where

$$\begin{aligned} & p(u, x, y_1, y_r) \left[ \lambda \log \frac{p(v|u, x, y_1)}{p(v|y_1)} \right. \\ & \left. + (1-\lambda) \log \frac{p(v|u, x, y_1)}{p(v|u, y_r)} \right] + \gamma(u, x, y_1, y_r) = 0, \\ & \text{for all } x, y_1, y_r, v : p(u, x, y_1, y_r, v) > 0. \quad (99) \end{aligned}$$

Dividing by  $p(u, x, y_1, y_r)$  and rearranging gives

$$\begin{aligned} p(v|u, x, y_1) &= \frac{p(v|u, y_1)^\lambda p(v|u, y_r)^{1-\lambda}}{\gamma(u, x, y_1, y_r)} \\ & \text{for all } x, y_1, y_r, v : p(u, x, y_1, y_r, v) > 0. \quad (100) \end{aligned}$$

#### APPENDIX C PROOF OF COROLLARY 3

It is enough to show that either Possibility 1 in the corollary statement holds, or the sufficient condition in Thm. 2 holds. Thus, it is enough to prove that, if the sufficient condition in Thm. 2 does *not* hold, then Possibility 1 must hold. That is, we assume (13) holds for all  $u, x, y_1, y_r, v$  such that  $p(u, x, y_1, y_r) > 0, p(v|y_r, u) > 0$ , and we prove the existence of a function  $g$ . We may assume without loss of generality that  $U$  has full support, since if not we may simply delete any zero-probability letters. Fix any  $u \in \mathcal{U}$ , and let  $x \in \mathcal{X}_u$ , where

$$\mathcal{X}_u = \{x \in \mathcal{X} : p(x|u) > 0\}. \quad (101)$$

Thus, by the assumption of the corollary, for any  $y_1, y_r$ ,

$$p(u, x, y_1, y_r) = p(u, x)p(y_1, y_r|x) > 0. \quad (102)$$

In the following portion of the proof, we will only focus on this  $u$  value, and so for convenience we will drop the dependence

on  $u$  in the conditional distributions. That is, we re-write (13) as

$$p(v|x, y_1) = \frac{p(v|y_1)^\lambda p(v|y_r)^{1-\lambda}}{\gamma(x, y_1, y_r)} \quad (103)$$

which must hold for all  $x, y_1, y_r, v$  where  $x \in \mathcal{X}_u$  and  $p(v|y_r) > 0$ .

Consider the graph on vertex set  $\mathcal{V}$  with edge set given by

$$\mathcal{E} = \{(v_a, v_b) : \text{there exists } y_r \in \mathcal{Y}_r \text{ such that } p(v_a|y_r), p(v_b|y_r) > 0\}. \quad (104)$$

Consider any pair  $(v_a, v_b) \in \mathcal{E}$ . By definition there exists a  $y_r$  such that  $p(v_a|y_r), p(v_b|y_r) > 0$ . For any  $x \in \mathcal{X}_u, y_1 \in \mathcal{Y}_1$ , from (103) we have

$$\frac{p(v_a|x, y_1)}{p(v_b|x, y_1)} = \frac{p(v_a|y_1)^\lambda p(v_a|y_r)^{1-\lambda}}{\gamma(x, y_1, y_r)} \quad (105)$$

$$\cdot \frac{\gamma(x, y_1, y_r)}{p(v_b|y_1)^\lambda p(v_b|y_r)^{1-\lambda}} \quad (106)$$

$$= \left( \frac{p(v_a|y_1)}{p(v_b|y_1)} \right)^\lambda \left( \frac{p(v_a|y_r)}{p(v_b|y_r)} \right)^{1-\lambda}. \quad (107)$$

We define the quantity in (107) as  $\beta(v_a, v_b, y_1)$ . If there is more than one valid  $y_r$ , then by (107) they each much produce the same value. Thus for all  $x \in \mathcal{X}_u$  and  $y_1$ ,

$$\frac{p(v_a|x, y_1)}{p(v_b|x, y_1)} = \beta(v_a, v_b, y_1). \quad (108)$$

Suppose  $v_a$  and  $v_b$  are connected in the graph  $(\mathcal{V}, \mathcal{E})$ . That is, there exists a sequence of letters  $v_a = v_1, \dots, v_k = v_b$  where

$$(v_1, v_2), \dots, (v_{k-1}, v_k) \in \mathcal{E}. \quad (109)$$

Thus, for all  $x, y_1$ ,

$$\frac{p(v_a|x, y_1)}{p(v_b|x, y_1)} = \prod_{\ell=1}^{k-1} \frac{p(v_\ell|x, y_1)}{p(v_{\ell+1}|x, y_1)} = \prod_{\ell=1}^{k-1} \beta(v_\ell, v_{\ell+1}, y_1). \quad (110)$$

We may define the latter as  $\beta(v_a, v_b, y_1)$  for any connected  $v_a, v_b$ .

The graph  $(\mathcal{V}, \mathcal{E})$  splits into connected sub-graphs with vertex sets  $\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_m$ , where these vertex sets represent a partition of  $\mathcal{V}$ . Define a random variable  $W$ , with alphabet  $\{1, \dots, m\}$ , where  $W = w$  whenever  $V \in \mathcal{V}_w$ . Thus  $W$  is a deterministic function of  $V$ . Moreover,  $W$  is a deterministic function of  $Y_r$ , since for any  $y_r$ , all letters  $v$  where  $p(v|y_r) > 0$  must be in the same sub-graph  $\mathcal{V}_w$  for some  $w$ . Consider any  $w \in \{1, \dots, m\}$ . Let  $v_w$  be a designated element of  $\mathcal{V}_w$ . For any  $v \in \mathcal{V}_w$ , we have

$$\frac{p(v|x, y_1)}{p(v_w|x, y_1)} = \beta(v, v_w, y_1) \text{ for all } x. \quad (111)$$

Thus, for any  $v \in \mathcal{V}_w$

$$p(v|x, y_1, w) = \frac{p(v, w|x, y_1)}{p(w|x, y_1)} \quad (112)$$

$$= \frac{p(v|x, y_1)}{p(w|x, y_1)} \quad (113)$$

$$= \frac{p(v|x, y_1)}{\sum_{v' \in \mathcal{V}_w} p(v'|x, y_1)} \quad (114)$$



$$= \frac{p(v|x, y_1)/p(v_w|x, y_1)}{\sum_{v' \in \mathcal{V}_w} p(v'|x, y_1)/p(v_w|x, y_1)} \quad (115)$$

$$= \frac{\beta(v, v_w, y_1)}{\sum_{v' \in \mathcal{V}_w} \beta(v', v_w, y_1)}. \quad (116)$$

If  $v \notin \mathcal{V}_w$ , then obviously  $p(v|x, y_1, w) = 0$ . Thus

$$p(v|x, y_1, w) = 1(v \in \mathcal{V}_w) \frac{\beta(v, v_w, y_1)}{\sum_{v' \in \mathcal{V}_w} \beta(v', v_w, y_1)}. \quad (117)$$

Since the RHS does not depend on  $x$ , we must have

$$p(v|x, y_1, w) = p(v|y_1, w). \quad (118)$$

We now reintroduce the dependence on  $u$ . Since the above analysis holds for any  $u \in \mathcal{U}$ , it must be that

$$p(v|u, x, y_1, w) = p(v|u, y_1, w) \quad (119)$$

That is,  $X - (U, Y_1, W) - V$  is a Markov chain. Recall that by assumption, rate  $R$  satisfies (6)–(7). We show that replacing  $V$  by  $W$  does not reduce the achieved rate in this bound. For (6), note that

$$I(X; Y_1, V|U) = I(X; Y_1, W, V|U) \quad (120)$$

$$= I(X; Y_1, W|U) + I(X; V|U, Y_1, W) \quad (121)$$

$$= I(X; Y_1, W|U). \quad (122)$$

For (7), note that

$$I(Y_r; V|U, X, Y_1) = I(Y_r; W, V|U, X, Y_1) \quad (123)$$

$$\geq I(Y_r; W|U, X, Y_1). \quad (124)$$

Thus,  $R$  must satisfy (6)–(7) with  $V$  replaced by  $W$ . The proof is completed by recalling that  $W$  is a deterministic function of  $U$  and  $Y_r$ .

#### APPENDIX D PROOF OF CLAIM 4

We first show that  $C(0) \leq C_{sum}(0)/2$ . We then show that  $C(C_{CF}) \geq C_{sum}(C_{CF})/2$  for any  $C_{CF} > 0$ . Together, these imply our assertion.

Let  $\varepsilon > 0$  be an arbitrarily small parameter, and let  $n$  be sufficiently large. To show that  $C(0) \leq C_{sum}(0)/2$ , consider any functions  $f_0^n(Y_0^n) = X_0^n$  and  $f_1^n(Y_1^n) = X_1^n$ . For  $n$  sufficiently large,

$$\begin{aligned} C(0)n - \varepsilon n &\leq I(X^n; Y^n) \\ &= I(X^n; Y_2^n) + I(X^n; Y_W^n | Y_2^n) \\ &= I(X^n; Y_W^n | Y_2^n) \\ &= I(X^n; Y_W^n | X^n \oplus Y_{Z^n}^n, Z^n) \\ &\leq I(X^n, X^n \oplus Y_{Z^n}^n; Y_W^n | Z^n) \\ &= I(X^n, Y_{Z^n}^n; Y_W^n | Z^n) \\ &= I(Y_{Z^n}^n; Y_W^n | Z^n) + I(X^n; Y_W^n | Y_{Z^n}^n, Z^n) \\ &= I(Y_{Z^n}^n; Y_W^n | Z^n) \\ &\stackrel{(a)}{=} \frac{1}{2} (I(Y_{Z^n}^n; Y_W^n | Z^n) + I(Y_{(Z^n)^c}^n; Y_W^n | Z^n)) \\ &\leq \frac{1}{2} (I(Y_{Z^n}^n; Y_W^n | Z^n) + I(Y_{(Z^n)^c}^n; Y_W^n, Y_{Z^n}^n | Z^n)) \end{aligned}$$

$$= \frac{1}{2} (I(Y_{Z^n}^n; Y_W^n | Z^n) + I(Y_{(Z^n)^c}^n; Y_{Z^n}^n | Z^n) + I(Y_{(Z^n)^c}^n; Y_W^n | Y_{Z^n}^n, Z^n))$$

$$= \frac{1}{2} (I(Y_{Z^n}^n; Y_W^n | Z^n) + I(Y_{(Z^n)^c}^n; Y_W^n | Y_{Z^n}^n, Z^n))$$

$$= \frac{1}{2} I(Y_{Z^n}^n, Y_{(Z^n)^c}^n; Y_W^n | Z^n)$$

$$= \frac{1}{2} I(Y_0^n, Y_1^n; Y_W^n) \leq \frac{C_{sum}(0)}{2} n.$$

In (a) above we use the equality  $I(Y_{Z^n}^n; Y_W^n | Z^n) = I(Y_{(Z^n)^c}^n; Y_W^n | Z^n)$  which follows from the independent and symmetric nature of  $Z^n$ . Thus, the capacity of the diamond network above is at most half the sum-capacity of  $W$ .

To show that  $C(C_{CF}) \geq C_{sum}(C_{CF})/2$ , consider any rate vector  $(r_0, r_1)$  achievable on the 2-transmitter MAC  $W$  with a cooperation facilitator of rate  $C_{CF}$ . Assume that  $r_0 \geq r_1$  (a symmetric argument is used otherwise). We construct a rate  $r = (r_0 + r_1)/2$  scheme for the diamond network in which the first two relays cooperate at rate  $C_{CF}$ . Consider any  $rn$  bit message  $M$  for the diamond network. We treat  $M$  as two messages, an  $r_1 n$ -bit message  $M_1$  and an  $(rn - r_1 n)$ -bit message  $M_2$  with  $rn - r_1 n = (r_0 - r_1)n/2$ . Encode  $M$  to a binary word  $X^n$  using a three-part code: the first  $r_1 n$  bits of  $X^n$  equal  $M_1$ , the next  $(r_0 - r_1)n$  bits of  $X^n$  are a rate- $(1/2)$  erasure encoding of  $M_2$ , the remaining bits of  $X^n$  are all set to zero. Let message  $m_0$  for  $W$  be the first  $r_0 n$  bits of  $Y_0^n$ . Let message  $m_1$  for  $W$  be the first  $r_1 n$  bits of  $Y_1^n$ . Let  $X_0^n = f_0^n(Y_0^{r_0 n})$  and  $X_1^n = f_1^n(Y_1^{r_1 n})$ , where  $f_1^n$  and  $f_2^n$  are the encoding functions of  $W$  that achieve rate vector  $(r_0, r_1)$ . Now, from the outcome  $Y_W^n$  of  $W$ , the first  $r_0 n$  bits of  $Y_0^n$  and the first  $r_1 n$  bits of  $Y_1^n$  can be decoded (using the decoder of  $W$ ). This implies, using  $Y_2^n$ , that the first  $r_1 n$  bits of  $X^n$  (and thus  $M_1$ ) can be decoded. Moreover, roughly speaking, out of the next  $(r_0 - r_1)n$  bits of  $Y_0^n$  (approximately)  $(r_0 - r_1)n/2$  bits of  $X^n$  (at random locations according to  $Z^n$ ) can be decoded, implying that  $M_2$  can also be decoded using the capacity  $1/2$  erasure decoder. All in all,  $M$  is decoded successfully. To be more precise, for any  $\varepsilon > 0$ , in the arguments above one defines  $r$  to be  $\frac{r_0 + r_1}{2} - \varepsilon$ , uses a rate  $\frac{1}{2} - \varepsilon$  erasure code, and, through standard concentration, shows that indeed a rate- $r$  message  $M$  is decoded successfully with probability that depends on  $\varepsilon$  and tends to one when  $\varepsilon$  tends to zero. As  $\varepsilon$  is arbitrary, we conclude that  $C(C_{CF}) \geq C_{sum}(C_{CF})/2$ .