On the Benefit of Cooperation in Relay Networks

Oliver Kosut, Michelle Effros, Michael Langberg

2022 IEEE International Symposium on Information Theory (ISIT) | 978-1-6654-2159-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISIT50566.2022.9834763

***Abstract*—This work addresses the cooperation facilitator (CF) model, in which network nodes coordinate through a rate limited communication device. For 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 $C_{CF} = 0$ when the CF enables transmitter dependence where none was possible otherwise. This work asks whether cooperation through a CF can yield similar infinite-slope benefits when dependence among MAC transmitters has no benefit or when it can be established without the help of the CF. Specifically, this work studies the CF model when applied to relay nodes of a single-source, single-terminal, diamond network comprising 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 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 [2], 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 [2]. 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 $C_{CF} = 0$ [2, 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 [3], [4].

Since the infinite-slope improvement in the MAC-capacity results from creating dependence where none is possible otherwise, 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). When viewed in isolation, the MAC employed derives no benefit from the addition of a CF. We investigate whether an infinite slope is possible employing this MAC in a larger network. 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 [2] on the MAC to demonstrate the infinite-slope phenomenon in certain diamond networks.

II. Problem Setup

Notation: For any integer k, $[k]$ denotes the set $\{1, 2, \ldots, k\}$. 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. The full version of this work appears in [1].

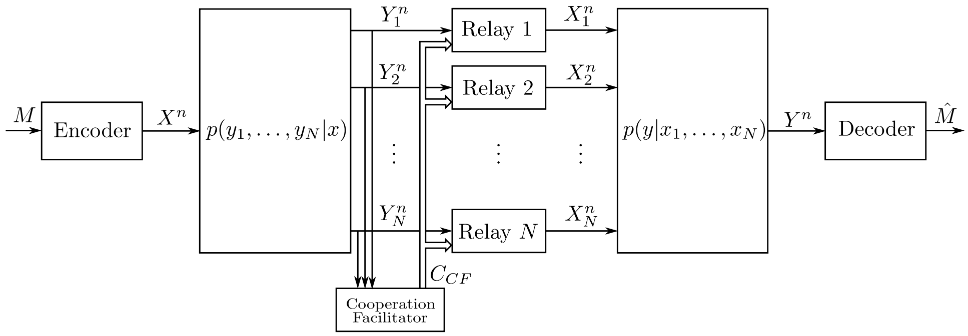

Fig. 1. General diamond relay network with N nodes and a cooperation facilitator (CF).

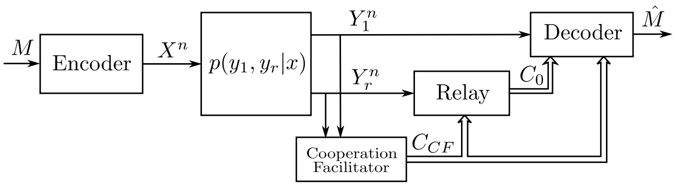

Fig. 2. Relay channel with orthogonal receiver components and a cooperation facilitator, a special case of the general diamond relay network.

variable, and calligraphic letters (e.g., $\mathcal{X}$) denote the corresponding alphabet. Vectors are denoted with superscript (e.g., $x^n = (x_1, \ldots, x_n)$). 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 $p(y_1, \ldots, y_N|x)$, followed by a MAC $p(y|x_1, \ldots, x_N)$. An (n, R) code for the diamond relay network is composed of

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

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 $K = f_{CF}(Y_1^n, \ldots, Y_N^n)$, 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 $P_e^{(n)} \to 0$. 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

Consider a diamond relay network in which the relay nodes communicate to the receiver through independent channels. This example specializes 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, $Y = (Y_a, Y_b)$, and the MAC model factors as

$$p(y_a, y_b|x_1, x_2) = p(y_a|x_1)p(y_b|x_2). \tag{1}$$

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 [5]. 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 can 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 [6], [7].

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 CF capacity C_{CF}. Rate R is achievable if

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

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

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

for some distribution $p(u,x)p(y_r,y_1|x)p(v|x,y_1,y_r,u)$.

Proof: We describe the codebook generation and encoding/decoding steps but leave the error analysis to the full version [1]. We use $T_\epsilon^{(n)}$ for the robustly typical set (see [8] for definition). Fix rates R_a, R_b, S to be determined, where $R_a + R_b = R$.

Codebook generation:

- For each $m_a \in [2^{nR_a}]$, generate cloud center

$$u^n(m_a) \sim \textstyle\prod_{i=1}^n p_U(u_i). \tag{5}$$

- For each $m_a \in [2^{nR_a}]$ and $m_b \in [2^{nR_b}]$, generate transmit sequence
$$x^n(m_a, m_b) \sim \prod_{i=1}^n p_{X|U}(x_i|u_i(m_a)). \tag{6}$$
- For each $m_a \in [2^{nR_a}]$, $\ell \in [2^{nS}]$, and $k \in [2^{nC_{CF}}]$, generate rate-distortion codeword
$$v^n(m_a, \ell, k) \sim \text{Unif}\left[T_\epsilon^{(n)}(V|u^n(m_a))\right]. \tag{7}$$
and the corresponding source coding bins
$$m_0(m_a, \ell, k) \sim \text{Unif}[2^{nC_0}]. \tag{8}$$
- 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
$$(u^n(m_a), y_r^n, v^n(m_a, \ell, k)) \in T_\epsilon^{(n)}(U, Y_r, V). \tag{9}$$
If there is no such ℓ, 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
$$(u^n(\hat{m}_a), x^n(\hat{m}_a, \hat{m}_b), y_1^n, y_r^n) \in T_\epsilon^{(n)}(U, X, Y_1, Y_r). \tag{10}$$
Next, find the smallest $k \in [2^{nC_{CF}}]$ such that $\ell(m_a, y_r^n, k)$ is defined, and
$$(u^n(\hat{m}_a), x^n(\hat{m}_a, \hat{m}_b), y_r^n, y_1^n, v^n(\hat{m}_a, \ell(\hat{m}_a, y_r^n, k), k)) \in T_\epsilon^{(n)}(U, X, Y_r, Y_1, V). \tag{11}$$
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
$$(u^n(\hat{m}_a), y_r^n) \in T_\epsilon^{(n)}(U, Y_r). \tag{12}$$
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
$$(u^n(\hat{m}_a), x^n(\hat{m}_a, \hat{m}_b), y_1^n, v^n(\hat{m}_a, \hat{\ell}, k)) \in T_\epsilon^{(n)}(U, X, Y_1, V), \tag{13}$$
$$m_0(\hat{m}_a, \hat{\ell}, k) = m_0. \tag{14}$$
■

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 [9]. For the relay channel with orthogonal receiver components, [6] showed that this classical bound can be written as follows: rate R is achievable if
$$R \le I(U; Y_r) + I(X; Y_1, V|U), \tag{15}$$
$$R \le \min\{I(U; Y_1), I(U; Y_r)\} + I(X; Y_1|U) + C_0 - I(Y_r; V|U, X, Y_1) \tag{16}$$
for some distribution
$$p(u, x)p(y_r, y_1|x)p(v|y_r, u). \tag{17}$$

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 joint distribution factors as in (17). Thus, by the data processing inequality, $I(X; Y_1, V|U) \le I(X; Y_1, Y_r|U)$, so (2) becomes (15). In addition,
$$I(V; X, Y_1|U) - I(Y_r; V|U) \tag{18}$$
$$= -H(V|U, X, Y_1) + H(V|U, Y_r) \tag{19}$$
$$= -H(V|U, X, Y_1) + H(V|U, Y_r, X, Y_1) \tag{20}$$
$$= -I(Y_r; V|U, X, Y_1), \tag{21}$$
so (3) becomes (16).

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 (15)–(16)), 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 (15)–(16) 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
$$p(v|u, x, y_1) = \frac{p(v|u, y_1)^\lambda p(v|u, y_r)^{1-\lambda}}{\gamma(u, x, y_1, y_r)} \tag{22}$$
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
$$\lim_{C_{CF} \to 0} \frac{C(C_{CF}) - R}{C_{CF}} = \infty. \tag{23}$$

Proof: Under the starting distribution
$$p(u, x)p(y_1, y_r|x)p(v|u, y_r), \tag{24}$$
$I(X, Y_1; V|U, Y_r) = 0$. To show (23), we modify this distribution slightly, in a way that $I(X, Y_1; V|U, Y_r) > 0$, 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
$$q(v|u, x, y_1, y_r) = p(v|u, y_r) + \alpha\, r(v|u, x, y_1, y_r) \tag{25}$$
where $\alpha \approx 0$. For a variable $A \subset \{U, X, Y_1, Y_r\}$, we further define $r(v|a)$, for example by
$$r(v|u, y_r) = \sum_{x, y_1} p(x, y_1|u, y_r) r(v|u, x, y_1, y_r). \tag{26}$$
Thus $q(v|a) = p(v|a) + \alpha\, r(v|a)$. In order for q to be a valid distribution, we need
$$\sum_v r(v|u, x, y_1, y_r) = 0 \text{ for all } u, x, y_1, y_r. \tag{27}$$
Thus, these r functions are not really distributions; instead they satisfy $\sum_v r(v|a) = 0$ for any variable A. Moreover, we assume $r(v|u, x, y_1, y_r) = 0$ for any u, y_r, v where $p(v|u, y_r) = 0$.

We are only changing the distribution of V, not U, so many terms in Thm. 1 do not change with α. The only terms that do change with α are given by the following functions:
$$f_1(\alpha) = I_q(X; V|U, Y_1),$$

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

To prove (23), it is enough for

$$C'_{CF}(0) = 0, \quad f'_1(0) > 0, \quad f'_2(0) > 0. \tag{28}$$

In fact, the assumption that $r(v|u, x, y_1, y_r) = 0$ whenever $p(v|u, y_r) = 0$ implies that $C'_{CF}(0) = 0$ (see [1] for details). Moreover, one can evaluate the other derivatives as

$$f'_1(0) = \sum_{\substack{u,x,y_1,y_r,v:\\ p(u,x,y_1,y_r,v)>0}} p(u, x, y_1, y_r) r(v|u, x, y_1, y_r) \cdot \log \frac{p(v|u, x, y_1)}{p(v|u, y_1)}, \tag{29}$$

$$f'_2(0) = \sum_{\substack{u,x,y_1,y_r,v:\\ p(u,x,y_1,y_r,v)>0}} p(u, x, y_1, y_r) r(v|u, x, y_1, y_r) \cdot \log \frac{p(v|u, x, y_1)}{p(v|u, y_r)}. \tag{30}$$

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

$$\max_{z:Az=0} \min\{a^T z, b^T z\} \tag{31}$$
$$= \max_{z:Az=0} \min_{\lambda\in[0,1]} \lambda a^T z + (1-\lambda) b^T z \tag{32}$$
$$= \max_{z} \min_{\lambda\in[0,1],\gamma} \lambda a^T z + (1-\lambda) b^T z + \gamma^T A z \tag{33}$$
$$= \min_{\lambda\in[0,1],\gamma} \max_{z} (\lambda a + (1-\lambda) b + A^T \gamma)^T z \tag{34}$$
$$= \min_{\lambda\in[0,1],\gamma} \begin{cases} 0 & \lambda a + (1-\lambda) b + A^T\gamma = 0 \\ \infty & \text{otherwise.} \end{cases} \tag{35}$$

That is, there exists no z of interest if and only if there exists $\lambda \in [0, 1]$ and γ where $\lambda a + (1-\lambda)b + A^T \gamma = 0$. 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

$$p(u, x, y_1, y_r) \left[\lambda \log \frac{p(v|u, x, y_1)}{p(v|y_1)} + (1-\lambda) \log \frac{p(v|u, x, y_1)}{p(v|u, y_r)}\right] + \gamma(u, x, y_1, y_r) = 0,$$
$$\text{for all } x, y_1, y_r, v : p(u, x, y_1, y_r, v) > 0. \tag{36}$$

Dividing by $p(u, x, y_1, y_r)$ and rearranging gives (22). ■

Remark 2: We note that there are two ways for (23) 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 (23) 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 (23) 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, (23) 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 (with proof in [1]) 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 (15)–(16) for this distribution. Then at least one of the following possibilities hold:

1) there exists a function $g : \mathcal{U} \times \mathcal{Y}_r \to \mathcal{V}$ where rate R satisfies (15)–(16) with $V = g(U, Y_r)$.
2) (23) holds.

C. Example Relay Channels

For some relay channels, [10] shows 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 [10] to be given by

$$C(0) = \max_{p(v|y_r):I(Y_r;V)\le C_0} 1 - H(Z|V). \tag{37}$$

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 (37). 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 (15)–(16) 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, if $y_r \in \{0, 1\}$, then $p(v|y_r) = 1-q$ for $v = y_r$ and $p(v|y_r) = q$ for $v = e$; if $y_r = e$ then $p(v|y_r) = 1$ if $v = e$. This leads to the achievable rate

$$R = \max_{q\in[0,1]} \min\{(1-p)(1 + p(1-q)), 1 - p - H((1-p)(1-q)) + (1-p)H(q) + C_0\} \tag{38}$$

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 λ and γ satisfying (22). 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 $y_1 \in \{x, e\}$

$$\frac{p_{V|X}(x|x)}{p_{V|X}(e|x)} = \frac{(1-p)(1-q)}{1-(1-p)(1-q)} \tag{39}$$

$$= \frac{p_{V|Y_1}(x|y_1)^{\lambda}(1-q)^{1-\lambda}}{\gamma'(x, y_1, x)} \cdot \frac{\gamma'(x, y_1, x)}{p_{V|Y_1}(e|y_1)^{\lambda} q^{1-\lambda}} \tag{40}$$

$$= \left(\frac{1-q}{q}\right)^{1-\lambda} \left(\frac{p_{V|Y_1}(x|y_1)}{p_{V|Y_1}(e|y_1)}\right)^{\lambda} \tag{41}$$

$$= \left(\frac{1-q}{q}\right)^{1-\lambda} \begin{cases} \left(\frac{(1-p)(1-q)}{1-(1-p)(1-q)}\right)^{\lambda}, & y_1 = x \\ \left(\frac{\frac{1}{2}(1-p)(1-q)}{1-(1-p)(1-q)}\right)^{\lambda}, & y_1 = e. \end{cases} \tag{42}$$

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 [2], [4]. In general, designing cooperation by treating the MAC as an isolated component may not suffice to improve communication of the diamond network. This is 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 [2]. 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 $(\mathcal{X}_0, \mathcal{X}_1, \mathcal{X}_2, q(y|x_0, x_1, x_2), \mathcal{Y})$. As depicted in Figure 3, consider the case in which $\mathcal{X} = \mathcal{Y}_0 = \mathcal{Y}_1 = \mathcal{X}_0 = \mathcal{X}_1 = \{0, 1\}$; $\mathcal{Y}_2 = \mathcal{X}_2 = \{0, 1\}^2$; $\mathcal{Y} = \mathcal{Y}_{\mathtt{W}} \times \mathcal{X}_2$ for a given memoryless 2-user binary MAC $\mathtt{W}$: $(\mathcal{X}_0, \mathcal{X}_1, p_{\mathtt{W}}(y_{\mathtt{W}}|x_0, x_1), \mathcal{Y}_{\mathtt{W}})$; 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 $Y = (Y_{\mathtt{W}}, x_2)$. As Y^n holds the value of $X_2^n \in \{0, 1\}^n$, which in turn depends on

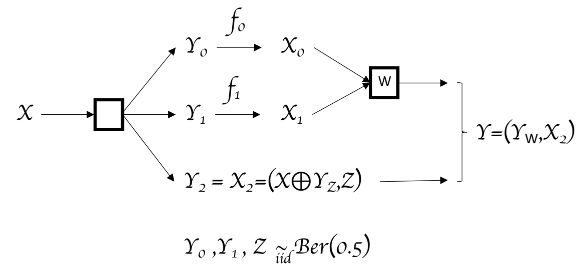

Fig. 3. The 3-relay diamond network of Claim 4.

$Y_2^n \in \{0, 1\}^n$ 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 $\mathtt{W}$ 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 $\mathtt{W}$ with user cooperation of rate C_{CF}. Then the capacity $C(C_{CF})$ of the diamond network satisfies $C'(0) = \infty$ if $C'_{\text{sum}}(0) = \infty$.

Claim 4 is proven in detail in the full version of this work [1]. We here present a rough outline. To prove the assertion, it suffices to show that $C(0) \le C_{\text{sum}}(0)/2$ for $C_{CF} = 0$ and $C(C_{CF}) \ge C_{\text{sum}}(C_{CF})/2$ for any $C_{CF} > 0$. The lower bound of $C(C_{CF}) \ge C_{\text{sum}}(C_{CF})/2$ is obtained using the encoders for $\mathtt{W}$ that achieve sum-rate $C_{\text{sum}}(C_{CF})$ as the relay encoders f_0 and f_1 of the diamond network applied to (a subset of) the independent bits Y_0^n and Y_1^n. The decoded information from $Y_{\mathtt{W}}^n$ combined with the knowledge of Z^n can then be used to distill parts of X^n which in turn (using a variant of rate-(1/2) erasure encoding for X^n) suffice to reveal the message m of rate $C_{\text{sum}}(C_{CF})/2$. For the upper bound of $C(0) \le C_{\text{sum}}(0)/2$ we argue that any decoder for the diamond network achieving rate $C(0)$ must be able to recover at least $C(0)n$ entries of Y_{Z^n}, which in turn, using the nature of Z^n, implies that $C_{\text{sum}}(0) \ge 2C(0)$.

V. Conclusion

This paper has explored whether it is possible for a small amount of cooperation between nodes in a relay network to have a disproportionate impact on capacity. The 3-relay network example of Fig. 3 illustrates that it is possible for cooperation to have a disproportionate effect in essentially the same manner as the MAC on its own. Example 1 also illustrates this phenomenon, but for a different reason: the proof is based on relay-channel coding rather than MAC-coding. In Example 1, the two channel outputs of the broadcast channel are independent. Thus, we have not definitively demonstrated whether the disproportionate benefit phenomenon can occur when the information available at the relays is dependent. We *do* show that there can be disproportionate benefit for dependent relays *compared to the best-known achievability bound without cooperation*, but we have only been able to show this (e.g., in Example 2) for channels with no matching converse bound.

References

[1] O. Kosut, M. Effros, and M. Langberg, "On the benefit of cooperation in relay networks," *arXiv preprint arXiv:2202.01878*, 2022.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] ——, "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.

[8] A. El Gamal and Y. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[9] 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.

[10] 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.