Capacity of a Class of State-Dependent Orthogonal Relay Channels

Iñaki Estella Aguerri and Deniz Gündüz

***Abstract*—The class of orthogonal relay channels in which the orthogonal channels connecting the source terminal to the relay and the destination, and the relay to the destination, depend on a state sequence, is considered. It is assumed that the state sequence is fully known at the destination while it is not known at the source or the relay. The capacity of this class of relay channels is characterized, and shown to be achieved by the partial decode-compress-and-forward (pDCF) scheme. Then the capacity of certain binary and Gaussian state-dependent orthogonal relay channels are studied in detail, and it is shown that the compress-and-forward (CF) and partial-decode-and-forward (pDF) schemes are suboptimal in general. To the best of our knowledge, this is the first single relay channel model for which the capacity is achieved by pDCF, while pDF and CF schemes are both suboptimal. Furthermore, it is shown that the capacity of the considered class of state-dependent orthogonal relay channels is in general below the cut-set bound. The conditions under which pDF or CF suffices to meet the cut-set bound, and hence, achieve the capacity, are also derived.**

***Index Terms*—Capacity, channels with state, relay channel, decode-and-forward, compress-and-forward, partial decode-compress-and forward.**

I. Introduction

We consider a state-dependent orthogonal relay channel, in which the channels connecting the source to the relay, and the source and the relay to the destination are orthogonal, and are governed by a state sequence, which is assumed to be known only at the destination. We call this model the *state-dependent orthogonal relay channel with state information available at the destination*, and refer to it as the ORC-D model. See Figure 1 for an illustration of the ORC-D channel model.

Many practical communication scenarios can be modelled by the ORC-D model. For example, consider a cognitive network with a relay, in which the transmit signal of the secondary user interferes simultaneously with the received primary user signals at both the relay and the destination. After decoding the secondary user's message, the destination obtains information about the interference affecting the source-relay channel, which can be exploited to decode the primary transmitter's message. Note that the relay may be oblivious to the presence of the secondary user, and hence, may not have access to the side information. Similarly, consider a mobile network with a relay (e.g., a femtostation), in which the base station (BS) operates in the full-duplex mode, and transmits on the downlink channel to a user, in parallel to the uplink transmission of a femtocell user, causing interference for the uplink transmission at the femtostation. While the femtostation, i.e., the relay, has no prior information about this interfering signal, the BS knows it perfectly and can exploit this knowledge to decode the uplink user's message forwarded by the femtostation.

The best known transmission strategies for the three terminal relay channel are the decode-and-forward (DF), compress-and-forward (CF) and partial decode-compress-and-forward (pDCF) schemes, which were all introduced by Cover and El Gamal in [2]. In DF, the relay decodes the source message and forwards it to the destination together with the source terminal. DF is generalized by the partial decode-and-forward (pDF) scheme in which the relay decodes and forwards only a part of the message. In the ORC-D model, pDF would be optimal when the channel state information is not available at the destination [3]; however, when the state information is known at the destination, fully decoding and re-encoding the message transmitted on the source-relay link renders the channel state information at the destination useless. Hence, we expect that pDF is suboptimal for ORC-D in general.

In CF, the relay does not decode any part of the message, and simply compresses the received signal and forwards the compressed bits to the destination using Wyner-Ziv coding followed by separate channel coding. Using CF in the ORC-D model allows the destination to exploit its knowledge of the state sequence; and hence, it can decode messages that may not be decodable by the relay. However, CF also forwards some noise to the destination, and therefore, may be suboptimal in certain scenarios. For example, as the dependence of the source-relay channel on the state sequence weakens, i.e., when the state information becomes less informative, CF performance is expected to degrade.

pDCF combines both schemes: part of the source message is decoded by the relay, and forwarded, while the remaining signal is compressed and forwarded to the destination. Hence, pDCF can optimally adapt its transmission to the dependence of the orthogonal channels on the state sequence. Indeed, we show that pDCF achieves the capacity in the ORC-D channel model, while pure DF and CF are in general suboptimal. The main results of the paper are summarized as follows:

- We derive an upper bound on the capacity of the ORC-D model, and show that it is achievable by the pDCF

This paper was presented in part at the IEEE Information Theory Workshop (ITW), Laussane, Switzerland, Sept. 2012 [1]. I. Estella Aguerri is with the Mathematical and Algorithmic Sciences Laboratory, France Research Center, Huawei Technologies Company, Ltd., Boulogne-Billancourt 92100, France E-mail: inaki.estella@huawei.com. Deniz Gündüz is with the Department of Electrical and Electronic Engineer, Imperial College London, London SW7 2BT, U.K. E-mail: d.gunduz@imperial.ac.uk. The research presented here was done during the Ph.D. studies of Iñaki Estella Aguerri at Imperial College London.

scheme. This characterizes the capacity of this class of relay channels.
- Focusing on the multi-hop binary and Gaussian models, we show that applying either only the CF or only the DF scheme is in general suboptimal.
- We show that the capacity of the ORC-D model is in general below the cut-set bound. We identify the conditions under which pure DF or pure CF meet the cut-set bound. Under these conditions the cut-set bounds is tight, and either DF or CF scheme is sufficient to achieve the capacity.

While the capacity of the general relay channel is still an open problem, there have been significant achievements within the last decade in understanding the capabilities of various transmission schemes, and the capacity of some classes of relay channels has been characterized. For example, DF is shown to be optimal for physically degraded relay channels and inversely degraded relay channels in [2]. In [3], the capacity of the orthogonal relay channel is characterized, and shown to be achieved by the pDF scheme. It is shown in [4] that pDF achieves the capacity of semi-deterministic relay channels as well. CF is shown to achieve the capacity in deterministic primitive relay channels in [5]. While all of these capacity results are obtained by using the cut-set bound for the converse proof [6], the capacity of a class of modulo-sum relay channels is characterized in [7], and it is shown that the capacity, achievable by the CF scheme, can be below the cut-set bound. The pDCF scheme is shown to achieve the capacity of a class of diamond relay channels in [8].

The state-dependent relay channel has also attracted considerable attention in the literature. Key to the investigation of the state-dependent relay channel model is whether the state sequence controlling the channel is known at the nodes of the network, the source, relay or the destination in a causal or non-causal manner. The relay channel in which the state information is non-causally available only at the source is considered in [9], [10], and both causally and non-causally available state information is considered in [11]. The model in which the state is non-causally known only at the relay is studied in [12] while causal and non-causal knowledge is considered in [13]. Similarly, the relay channel with state causally known at the source and relay is considered in [14], and state non-causally known at the source, relay and destination in [15]. Recently a generalization of pDF, called the cooperative-bin-forward scheme, has been shown to achieve the capacity of state-dependent semi-deterministic relay channels with causal state information at the source and destination [16]. The compound relay channel with informed relay and destination are discussed in [17] and [18]. The state-dependent relay channel with structured state has been considered in [19] and [20]. To the best of our knowledge, this is the first work that focuses on the state-dependent relay channel in which the state information is available only at the destination.

The rest of the paper is organized as follows. In Section II we provide the system model and our main result. Section III is devoted to the proofs of the achievability and converse for the main result. In Section IV, we provide two examples

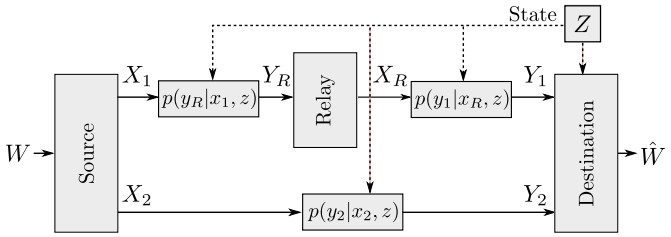

Fig. 1. Orthogonal state-dependent relay channel with channel state information available at the destination, called the ORC-D model.

demonstrating the suboptimality of pDF and CF schemes on their own, and in Section V we show that the capacity is in general below the cut-set bound, and we provide conditions under which pure DF and CF schemes meet the cut-set bound. Finally, Section VII concludes the paper.

We use the following notation in the rest of the paper: $X_i^j \triangleq (X_i, X_{i+1}, ..., X_j)$ for $i < j$, $X^n \triangleq (X_1, ..., X_n)$ for the complete sequence, $X_{n+1}^n \triangleq \emptyset$, and $Z^{n\backslash i} \triangleq (Z_1, ..., Z_{i-1}, Z_{i+1}, ..., Z_n)$.

II. System Model and Main Result

We consider the class of orthogonal relay channels depicted in Figure 1. The source and the relay are connected through a memoryless channel characterized by $p(y_R|x_1, z)$, the source and the destination are connected through an orthogonal memoryless channel characterized by $p(y_2|x_2, z)$, while the relay and the destination are connected by a memoryless channel $p(y_1|x_R, z)$. The three memoryless channels depend on an independent and identically distributed (i.i.d.) state sequence $\{Z\}_{i=1}^n$, which is available at the destination. The input and output alphabets are denoted by $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{X}_R$, $\mathcal{Y}_1$, $\mathcal{Y}_2$ and $\mathcal{Y}_R$, and the state alphabet is denoted by $\mathcal{Z}$.

Let W be the message to be transmitted to the destination with the assistance of the relay. The message W is assumed to be uniformly distributed over the set $\mathcal{W} = \{1, ..., M\}$. An (M, n, ν_n) code for this channel consists of an encoding function at the source:

$$f : \{1, ..., M\} \to \mathcal{X}_1^n \times \mathcal{X}_2^n, \tag{1}$$

a set of encoding functions $\{f_{r,i}\}_{i=1}^n$ at the relay, whose output at time i depends on the symbols it has received up to time $i-1$:

$$X_{Ri} = f_{r,i}(Y_{R1}, ..., Y_{R(i-1)}), \quad i = 1, ..., n, \tag{2}$$

and a decoding function at the destination

$$g : \mathcal{Y}_1^n \times \mathcal{Y}_2^n \times \mathcal{Z}^n \to \{1, ..., M\}. \tag{3}$$

The probability of error, ν_n, is defined as

$$\nu_n \triangleq \frac{1}{M} \sum_{w=1}^{M} \Pr\{g(Y_1^n, Y_2^n, Z^n) \neq w | W = w\}. \tag{4}$$

The joint probability mass function (pmf) of the involved random variables over the set $\mathcal{W} \times \mathcal{Z}^n \times \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{X}_R^n \times \mathcal{Y}_R^n \times \mathcal{Y}_1^n \times \mathcal{Y}_2^n$ is given by

$$p(w, z^n, x_1^n, x_2^n, x_R^n, y_R^n, y_1^n, y_2^n) = p(w) \prod_{i=1}^{n} p(z_i) p(x_{1i}, x_{2i}|w) \cdot p(y_{Ri}|z_i, x_{1i}) p(x_{Ri}|y_R^{i-1}) p(y_{1i}|x_{Ri}, z_i) p(y_{2i}|x_{2i}, z_i).$$

A rate R is said to be *achievable* if there exists a sequence of $(2^{nR}, n, \nu_n)$ codes such that $\lim_{n\to\infty} \nu_n = 0$. The *capacity*, $\mathcal{C}$, of this class of state-dependent orthogonal relay channels, denoted as ORC-D, is defined as the supremum of the set of all achievable rates.

We define R_1 and R_1 as follows, which can be thought as the capacities of the individual links from the relay to the destination, and from the source to the destination, respectively, when the channel state sequence is available at the destination:

$$R_1 \triangleq \max_{p(x_R)} I(X_R; Y_1|Z), \quad R_2 \triangleq \max_{p(x_2)} I(X_2; Y_2|Z). \quad (5)$$

Let $p^*(x_R)$ and $p^*(x_2)$ be the channel input distributions achieving R_1 and R_2, respectively.

Let us define $\mathcal{P}$ as the set of all joint pmf's given by

$$\mathcal{P} \triangleq \{p(u, x_1, z, y_R, \hat{y}_R) : \quad (6)$$
$$p(u, x_1, z, y_R, \hat{y}_R) = p(u, x_1)p(z)p(y_R|x_1, z)p(\hat{y}_R|y_R, u)\},$$

where U and $\hat{Y}_R$ are auxiliary random variables defined over the alphabets $\mathcal{U}$ and $\hat{\mathcal{Y}}_R$, respectively.

The main result of this work, provided in the next theorem, is the capacity of the class of relay channels described above.

Theorem 1. *The capacity of the ORC-D relay channel is given by*

$$\mathcal{C} = \sup_{\mathcal{P}} R_2 + I(U; Y_R) + I(X_1; \hat{Y}_R|UZ),$$
$$\text{s.t. } R_1 \geq I(U; Y_R) + I(Y_R; \hat{Y}_R|UZ), \quad (7)$$

where $|\mathcal{U}| \leq |\mathcal{X}_1| + 3$ *and* $|\hat{\mathcal{Y}}_R| \leq |\mathcal{U}||\mathcal{Y}_R| + 1$.

Proof. The achievability part of the theorem is proven in Section III-A, while the converse proof can be found in Section III-B. □

In the next section, we show that the capacity of this class of state-dependent relay channels is achieved by the pDCF scheme. To the best of our knowledge, this is the first single-relay channel model for which the capacity is achieved by pDCF, while the pDF and CF schemes are both suboptimal in general. In addition, the capacity of this relay channel is in general below the cut-set bound [6]. These issues are discussed in more detail in Sections IV and V.

It follows from Theorem 1 that the transmission over the relay-destination and source-destination links can be independently optimized to operate at the corresponding capacities, and these links in principle act as error-free channels of capacity R_1 and R_2, respectively. We also note that the relay can acquire some knowledge about the channel state sequence Z^n from its channel output Y_R^n, and could use it in the transmission over the relay-destination link, which depends on the same state information sequence. In general, non-causal state information available at the relay can be exploited to increase the achievable throughput in multi-user setups [21], [22]. However, it follows from Theorem 1 that this knowledge is useless. This is because the channel state information acquired from Y_R^n can be seen as delayed feedback to the relay, which does not increase the capacity in point-to-point channels.

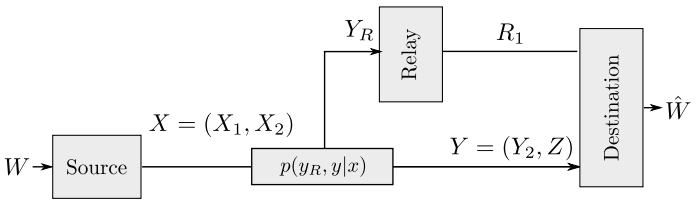

Fig. 2. The ORC-D as a class of primitive relay channel.

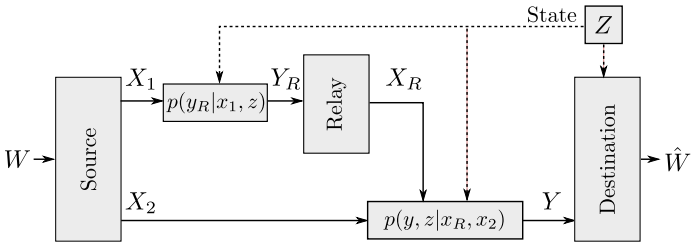

Fig. 3. The ORC-D is a particular case of the state dependent orthogonal relay channel with orthogonal components.

A. Comparison with previous relay channel models

Here, we compare ORC-D with other relay channel models in the literature, and discuss the differences and similarities. The discrete memoryless relay channel consists of four finite sets $\mathcal{X}$, $\mathcal{X}_R$, $\mathcal{Y}$ and $\mathcal{Y}_R$, and a probability distribution $p(y, y_R|x, x_R)$. In this setup, X corresponds to the source input to the channel, Y to the channel output available at the destination, while Y_R is the channel output available at the relay, and X_R is the channel input symbol chosen by the relay. We note that the three-terminal relay channel model in [2] reduces to ORC-D by setting $X^n = (X_1^n, X_2^n)$, $Y^n = (Y_1^n, Y_2^n, Z^n)$, and $p(y, y_R|x_1x_R) = p(y_1, y_2, y_R, z|x_1, x_2, x_R) = p(z)p(y_R|x_1, z)p(y_1|x_R, z)p(y_2|x_2, z)$.

By considering the channel from the relay to the destination as an error-free link with finite capacity, the ORC-D is included in the class of primitive relay channels proposed in [5] and [23] as seen in Figure 2, for which the channel distribution satisfies $p(y, y_R|x, x_R) = p(y, y_R|x)$. Although the capacity of this channel remains unknown in general, it has been characterized for certain special cases. CF has been shown to achieve the cut-set bound, i.e., to be optimal, in [5], if the relay output, Y_R, is a deterministic function of the source input and output at the destination, i.e., $Y_R = f(X, Y)$. The capacity of a class of primitive relay channels under a particular modulo sum structure is shown to be achievable by CF in [7], and to be below the cut-set bound. Theorem 1 provides the optimality of pDCF for a class of primitive relay channels, not included in any of the previous relay models for which the capacity is known. It is discussed in [23] that for the primitive relay channel, CF and DF do not outperform one another in general. It is also noted that their combination in the form of pDCF might not be sufficient to achieve the capacity in general. We will see in Section IV that both DF and CF are in general suboptimal, and that pDCF is necessary and sufficient to achieve the capacity for the class of primitive

relay channels considered in this paper.

It is also interesting to compare the ORC-D model with the orthogonal relay channel proposed in [3], in which the source-relay link is orthogonal to the multiple-access channel from the source and relay to the destination, i.e., $p(y, y_R|x, x_R) = p(y|x_2, x_R)p(y_R|x_1, x_R)$. The capacity for this model is shown to be achievable by pDF, and coincides with the cut-set bound. For the ORC-D, we have $p(y, y_R|x, x_R) = p(z)p(y_2|x_2, z)p(y_1|x_R, z)p(y_R|x_1, x_R, z)$, i.e., given the channel inputs, the orthogonal channel outputs at the relay and the destination are still dependent due to Z. Therefore, the ORC-D does not fall within the class of orthogonal channels considered in [3]. We can consider the class of state dependent relay channel with orthogonal components satisfying $p(y, z, y_R|x, x_R) = p(z)p(y|x_2, x_R, z)p(y_R|x_1, x_R, z)$ as shown in Figure 3. This class includes the orthogonal relay channel in [3] and the ORC-D as a particular cases. However, the capacity for this class of state dependent relay channel remains open in general.

III. Proof of Theorem 1

We first show in Section III-A that the capacity claimed in Theorem 1 is achievable by pDCF. Then, we derive the converse result for Theorem 1 in Section III-B.

A. Achievability

We derive the rate achievable by the pDCF scheme for ORC-D using the achievable rate expression for the pDCF scheme proposed in [2] for the general relay channel. Note that the three-terminal relay channel in [2] reduces to ORC-D by setting $X^n = (X_1^n, X_2^n)$ and $Y^n = (Y_1^n, Y_2^n, Z^n)$, as discussed in Section II-A.

In pDCF for the general relay channel, the source applies message splitting, and the relay decodes only a part of the message. The part to be decoded by the relay is transmitted through the auxiliary random variable U^n, while the rest of the message is superposed onto this through channel input X^n. Block Markov encoding is used for transmission. The relay receives Y_R^n and decodes only the part of the message that is conveyed by U^n. The remaining signal Y_R^n is compressed into $\hat{Y}_R^n$. The decoded message is forwarded through V^n, which is correlated with U^n, and the compressed signal is superposed onto V^n through the relay channel input X_R^n. At the destination the received signal Y^n is used to recover the message. See [2] for details. The achievable rate of the pDCF scheme is given below.

Theorem 2. (Theorem 7, [2]) *The capacity of a relay channel $p(y, y_R|x, x_R)$ is lower bounded by the following rate:*

$$\begin{aligned} R_{\mathrm{pDCF}} = \sup \min \{ & I(X; Y, \hat{Y}_R|X_R, U) + I(U; Y_R|X_R, V), \\ & I(X, X_R; Y) - I(\hat{Y}_R; Y_R|X, X_R, U, Y)\}, \\ \text{s.t. } & I(\hat{Y}_R; Y_R|Y, X_R, U) \le I(X_R; Y|V), \end{aligned} \quad (8)$$

where the supremum is taken over all joint pmf's of the form

$$p(v)p(u|v)p(x|u)p(x_R|v)p(y, y_R|x, x_R)p(\hat{y}_R|x_R, y_R, u).$$

Since ORC-D is a special case of the general relay channel model, the rate R_{pDCF} is achievable in an ORC-D as well. The capacity achieving pDCF scheme for ORC-D is obtained from (8) by setting $V = \emptyset$, and generating X_R^n and X_1^n independent of the rest of the variables with distribution $p^*(x_R)$ and $p^*(x_1)$, respectively, as given in the next lemma.

Lemma 1. *For the class of relay channels characterized by the ORC-D model, the capacity expression $\mathcal{C}$ defined in (7) is achievable by the pDCF scheme.*

Proof. See Appendix A. □

The optimal pDCF scheme for ORC-D applies independent coding over the source-destination and the source-relay-destination branches. The source applies message splitting. Part of the message is transmitted over the source-destination branch and decoded at the destination using Y_2^n and Z^n. In the relay branch, the part of the message to be decoded at the relay is transmitted through U^n, while the rest of the message is superposed onto this through the channel input X_1^n. At the relay the part conveyed by U^n is decoded from Y_R^n, and the remaining signal Y_R^n is compressed into $\hat{Y}_R^n$ using binning and assuming that Z^n is available at the decoder. Both U^n and the bin index corresponding to $\hat{Y}_R^n$ are transmitted over the relay-destination channel using X_R^n. At the destination, X_R^n is decoded from Y_1^n, and U^n and the bin index are recovered. Then, the decoder looks for the part of message transmitted over the relay branch jointly typical with $\hat{Y}_R^n$ within the corresponding bin and Z^n.

B. Converse

The proof of the converse consists of two parts. First we derive a single-letter upper bound on the capacity, and then, we provide an alternative expression for this bound, which coincides with the rate achievable by pDCF.

Lemma 2. *The capacity of the class of relay channels characterized by the ORC-D model is upper bounded by*

$$R_{up} = \sup_{\mathcal{P}} \min\{R_2 + I(U; Y_R) + I(X_1; \hat{Y}_R|UZ), \quad (9)$$

$$R_1 + R_2 - I(\hat{Y}_R; Y_R|X_1UZ)\}. \quad (10)$$

Proof. See Appendix B. □

As stated in the next lemma, the upper bound R_{up}, given in Lemma 2, is equivalent to the capacity expression $\mathcal{C}$ given in Theorem 1. Since the achievable rate meets the upper bound, this concludes the proof of Theorem 1.

Lemma 3. *The upper bound on the achievable rate R_{up} given in Lemma 2 is equivalent to the capacity expression $\mathcal{C}$ in Theorem 1.*

Proof. See Appendix C. □

IV. The Multihop Relay Channel with State: Suboptimality of Pure pDF and CF schemes

We have seen in Section III that the pDCF scheme is capacity-achieving for the class of relay channels characterized

by the ORC-D model. In order to prove the suboptimality of the pure DF and CF schemes for this class of relay channels, we consider a simplified system model, called the *multihop relay channel with state information available at the destination* (MRC-D), which is obtained by simply removing the direct channel from the source to the destination, i.e., $R_2 = 0$.

The capacity of this multihop relay channel model and the optimality of pDCF follows directly from Theorem 1. However, the single-letter capacity expression depends on the joint pmf of X_1, Y_R, X_R and Y_1 together with the auxiliary random variables U and $\hat{Y}_R$. Unfortunately, the numerical characterization of the optimal joint pmf of these random variables is very complicated for most channels. A simple and computable upper bound on the capacity can be obtained from the cut-set bound [24]. For MRC-D, the cut-set bound is given by

$$R_{\mathrm{CS}} = \min\{R_1, \max_{p(x_1)} I(X_1; Y_R|Z)\}. \tag{11}$$

Next, we characterize the rates achievable by the DF and CF schemes for MRC-D. Since they are special cases of the pDCF scheme, their achievable rates can be obtained by particularizing the achievable rate of pDCF for this setup.

1) DF Scheme: If we consider a pDCF scheme that does not perform any compression at the relay, i.e., $\hat{Y_R} = \emptyset$, we obtain the rate achievable by the pDF scheme. Note that the optimal distribution of X_R is given by $p^*(x_r)$. Then, we have

$$R_{\mathrm{pDF}} = \min\{R_1, \sup_{p(x_1,u)} I(U; Y_R)\}. \tag{12}$$

From the Markov chain $U - X_1 - Y_R$, we have $I(U; Y_R) \leq I(X_1; Y_R)$, where the equality is achieved by $U = X_1$. That is, the performance of pDF is maximized by letting the relay decode the whole message. Therefore, the maximum rate achievable by pDF and DF for MRC-D coincide, and is given by

$$R_{\mathrm{DF}} = R_{\mathrm{pDF}} = \min\{R_1, \max_{p(x_1)} I(X_1; Y_R)\}. \tag{13}$$

We note that considering more advanced DF strategies based on list decoding as in [23] does not increase the achievable rate in the MRC-D, since there is no direct link.

2) CF Scheme: If the pDCF scheme does not perform any decoding at the relay, i.e., $U = V = \emptyset$, pDCF reduces to CF. Then, the achievable rate for the CF scheme in MRC-D is given by

$$\begin{aligned} R_{\mathrm{CF}} = \sup\, & I(X_1; \hat{Y}_R|Z) \\ & \text{s.t.} \;\; R_1 \geq I(\hat{Y}_R; Y_R|Z), \\ & \text{over } p(x_1)p(z)p(y_R|x_1, z)p(\hat{y}_R|y_R). \end{aligned} \tag{14}$$

A. Multihop Parallel Binary Symmetric Channel

In this section we consider a special MRC-D as shown in Figure 4, which we call the *parallel binary symmetric MRC-D*. For this setup, we characterize the optimal performance of the DF and CF schemes, and show that in general pDCF outperforms both, and that in some cases the cut-set bound is tight and coincides with the channel capacity. This example proves the suboptimality of both DF and CF on their own for ORC-D.

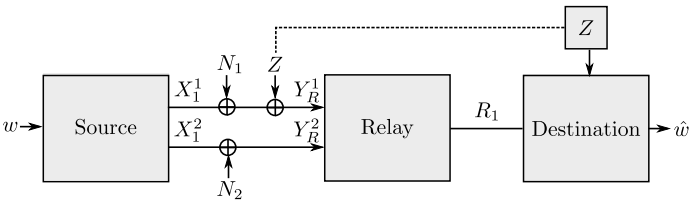

Fig. 4. The parallel binary symmetric MRC-D with parallel source-relay links. The destination has side information about only one of the source-relay links.

In this scenario, the source-relay channel consists of two parallel binary symmetric channels. We have $X_1 = (X_1^1, X_1^2)$, $Y_R = (Y_R^1, Y_R^2)$ and $p(y_R|x_R, z) = p(y_R^1|x_1^1, z)p(y_R^2|x_1^2)$ characterized by

$$Y_R^1 = X_1^1 \oplus N_1 \oplus Z, \quad \text{and} \quad Y_R^2 = X_1^2 \oplus N_2,$$

where N_1 and N_2 are i.i.d. Bernoulli random variables with $\Pr\{N_1 = 1\} = \Pr\{N_2 = 1\} = \delta$, i.e., $N_1 \sim \mathrm{Ber}(\delta)$ and $N_2 \sim \mathrm{Ber}(\delta)$. We consider a Bernoulli distributed state Z, $Z \sim \mathrm{Ber}(p_z)$, which affects one of the two parallel channels, and is available at the destination. We have $\mathcal{X}_1^1 = \mathcal{X}_1^2 = \mathcal{Y}_R^1 = \mathcal{Y}_R^1 = \mathcal{N}_1 = \mathcal{N}_2 = \mathcal{Z} = \{0, 1\}$.

From (11), the cut-set bound is given by

$$\begin{aligned} R_{\mathrm{CS}} &= \min\{R_1, \max_{p(x_1^1 x_1^2)} I(X_1^1 X_1^2; Y_R^1 Y_R^2|Z)\} \\ &= \min\{R_1, 2(1 - h_2(\delta))\}, \end{aligned} \tag{15}$$

where $h_2(\cdot)$ is the binary entropy function defined as $h_2(p) \triangleq -p\log p - (1-p)\log(1-p)$.

The maximum DF rate is achieved by $X_1^1 \sim \mathrm{Ber}(1/2)$ and $X_1^2 \sim \mathrm{Ber}(1/2)$, and is found to be

$$\begin{aligned} R_{\mathrm{DF}} &= \min\{R_1, \max_{p(x_1^1 x_1^2)} I(X_1^1 X_1^2; Y_R^1 Y_R^2)\} \\ &= \min\{R_1, 2 - h_2(\delta \star p_z) - h_2(\delta)\}, \end{aligned} \tag{16}$$

where $\alpha \star \beta \triangleq \alpha(1-\beta) + (1-\alpha)\beta$.

Following (14), the rate achievable by the CF scheme in the parallel binary symmetric MRC-D is given by

$$\begin{aligned} R_{\mathrm{CF}} = \max\, & I(X_1^1 X_1^2, \hat{Y}_R|Z), \\ & \text{s.t.} \;\; R_1 \geq I(Y_R^1 Y_R^2; \hat{Y}_R|Z) \\ & \text{over } p(z)p(x_1^1 x_1^2)p(y_R^1|z, x_1^1)p(y_R^2|x_2)p(\hat{y}_R|y_R^1 y_R^2). \end{aligned} \tag{17}$$

Let us define $h_2^{-1}(q)$ as the inverse of the entropy function $h_2(p)$ for $q \geq 0$. For $q < 0$, we define $h_2^{-1}(q) = 0$.

As we show in the next lemma, the achievable CF rate in (17) is maximized by transmitting independent channel inputs over the two parallel links to the relay by setting $X_1^1 \sim \mathrm{Ber}(1/2)$, $X_1^2 \sim \mathrm{Ber}(1/2)$, and by independently compressing each of the channel outputs Y_R^1 and Y_R^2 as $\hat{Y}_R^1 = Y_R^1 \oplus Q_1$ and $\hat{Y}_R^2 = Y_R^2 \oplus Q_2$, respectively, where $Q_1 \sim \mathrm{Ber}(h_2^{-1}(1 - R_1/2))$ and $Q_2 \sim \mathrm{Ber}(h_2^{-1}(1 - R_1/2))$. Note that for $R_1 \geq 2$, the channel outputs can be compressed errorlessly. The maximum achievable CF rate is given in the following lemma.

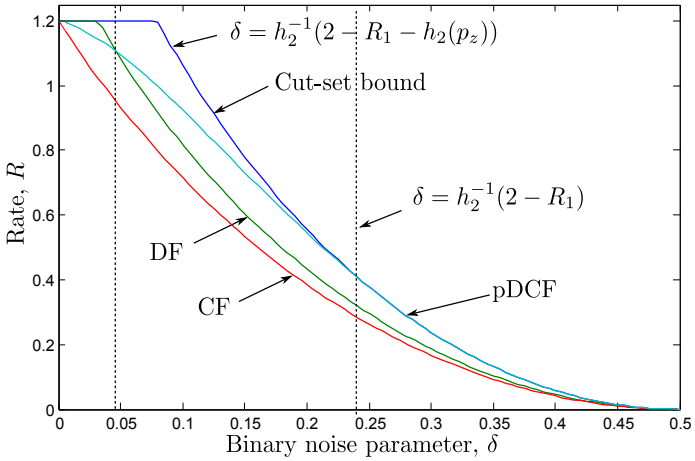

Fig. 5. Achievable rates and the cut-set upper bound for the parallel binary symmetric MRC-D with respect to the binary noise parameter δ, for $R_1 = 1.2$ and $p_z = 0.15$.

Lemma 4. *The maximum rate achievable by CF over the parallel binary symmetric MRC-D is given by*

$$R_{\mathrm{CF}} = 2\left(1 - h_2\left(\delta \star h_2^{-1}\left(1 - \frac{R_1}{2}\right)\right)\right). \tag{18}$$

Proof. See Appendix D. □

Now, we consider the pDCF scheme for the parallel binary symmetric MRC-D. Although we have not been able to characterize the optimal choice of $(U, \hat{Y}_R, X_1^1, X_1^2)$ in general, we provide an achievable scheme that outperforms both DF and CF schemes and meets the cut-set bound in some regimes. Let $X_1^1 \sim \mathrm{Ber}(1/2)$ and $X_1^2 \sim \mathrm{Ber}(1/2)$ and $U = X_1^2$, i.e., the relay decodes the channel input X_1^2, while Y_R^1 is compressed using $\hat{Y}_R = Y_R^1 + Q$, where $Q \sim \mathrm{Ber}(h_2^{-1}(2 - h_2(\delta) - R_1))$. The rate achievable by this scheme is given in the following lemma.

Lemma 5. *A lower bound on the achievable pDCF rate over the parallel binary symmetric MRC-D is given by*

$$R_{\mathrm{pDCF}} \geq \min\{R_1, 2 - h_2(\delta) - h_2\left(\delta \star h_2^{-1}\left(2 - h_2(\delta) - R_1\right)\right)\}.$$

Proof. See Appendix E. □

We notice that for $p_z \leq h_2^{-1}\left(2 - h_2(\delta) - R_1\right)$, or equivalently, $\delta \leq h_2^{-1}\left(2 - h_2(p_z) - R_1\right)$, the proposed pDCF is outperformed by DF. In this regime, pDCF can achieve the same performance by decoding both channel inputs, reducing to DF.

Comparing the cut-set bound expression in (15) with R_{DF} in (16) and R_{CF} in (18), we observe that DF achieves the cut-set bound if $R_1 \leq 2 - h(\delta \star p_z) - h(\delta)$ while R_{CF} coincides with the cut-set bound if $R_1 \geq 2$. On the other hand, the proposed suboptimal pDCF scheme achieves the cut-set bound if $R_1 \geq 2 - h_2(\delta)$, i.e., for $\delta \geq h_2^{-1}(2 - R_1)$. Hence, the capacity of the parallel binary symmetric MRC-D in this regime is achieved by pDCF, while both DF and CF are suboptimal, as stated in the next lemma.

Lemma 6. *If $R_1 < 2$ and $\delta \geq h_2^{-1}(2 - R_1)$, pDCF achieves the capacity of the parallel binary symmetric MRC-D, while pure CF and DF are both suboptimal under these constraints. For $R_1 \geq 2$, both CF and pDCF achieve the capacity.*

The achievable rates of DF, CF and pDCF, together with the cut-set bound are shown in Figure 5 with respect to δ for $R_1 = 1.2$ and $p_z = 0.15$. We observe that in this setup, DF outperforms CF in general, while for $\delta \leq h_2^{-1}(2 - R_1 - h_2(p_z)) = 0.0463$, DF outperforms the proposed suboptimal pDCF scheme as well. We also observe that pDCF meets the cut-set bound for $\delta \geq h_2^{-1}(2 - R_1) = 0.2430$, characterizing the capacity in this regime, and proving the suboptimality of both the DF and CF schemes when they are used on their own.

B. Multihop Binary Symmetric Channel

In order to gain further insights into the proposed pDCF scheme, we look into the *binary symmetric* MRC-D, in which, there is only a single channel connecting the source to the relay, given by

$$Y_R = X_1 \oplus N \oplus Z, \tag{19}$$

where $N \sim \mathrm{Ber}(\delta)$ and $Z \sim \mathrm{Ber}(p_z)$.

Similarly to Section IV-A, the cut-set bound and the maximum achievable rates for DF and CF are found as

$$R_{\mathrm{CS}} = \min\{R_1, 1 - h_2(\delta)\}, \tag{20}$$
$$R_{\mathrm{DF}} = \min\{R_1, 1 - h_2(\delta \star p_z)\}, \tag{21}$$
$$R_{\mathrm{CF}} = 1 - h_2(\delta \star h_2^{-1}(1 - R_1))), \tag{22}$$

where R_{DF} is achieved by $X_1 \sim \mathrm{Ber}(1/2)$, and R_{CF} can be shown to be maximized by $X_1 \sim \mathrm{Ber}(1/2)$ and $\hat{Y}_R = Y_R \oplus Q$, where $Q \sim \mathrm{Ber}(h_2^{-1}(1 - R_1))$ similarly to Lemma 4. Note that, for Y_R independent of Z, i.e., $p_z = 0$, DF achieves the cut-set bound while CF is suboptimal. However, CF outperforms DF whenever $p_z \geq h_2^{-1}(1 - R_1)$.

For the pDCF scheme, we consider binary $(U, X_1, \hat{Y}_R)$, with $U \sim \mathrm{Ber}(p)$, a superposition codebook $X_1 = U \oplus W$, where $W \sim \mathrm{Ber}(q)$, and $\hat{Y}_R = Y_R \oplus Q$ with $Q \sim \mathrm{Ber}(\alpha)$. As stated in the next lemma, the maximum achievable rate of this pDCF scheme is obtained by reducing it to either DF or CF, depending on the values of p_z and R_1.

Lemma 7. *For the binary symmetric MRC-D, pDCF with binary $(U, X_1, \hat{Y}_R)$ achieves the following rate.*

$$R_{\mathrm{pDCF}} = \max\{R_{\mathrm{DF}}, R_{\mathrm{CF}}\} \tag{23}$$
$$= \begin{cases} \min\{R_1, 1 - h_2(\delta \star p_z)\} & \text{if } p_z < h_2^{-1}(1 - R_1), \\ 1 - h_2(\delta \star h_2^{-1}(1 - R_1)) & \text{if } p_z \geq h_2^{-1}(1 - R_1). \end{cases}$$

This result justifies the pDCF scheme proposed in Section IV-A for the parallel binary symmetric MRC-D. Since the channel $p(y_1^2|x_2)$ is independent of the channel state Z, the largest rate is are achieved if the relay decodes X_1^2 from Y_R^2. However, for channel $p(y_1^1|x_1, z)$, which depends on Z, the relay either decodes X_1^1, or compress Y_R^1, depending on p_z.

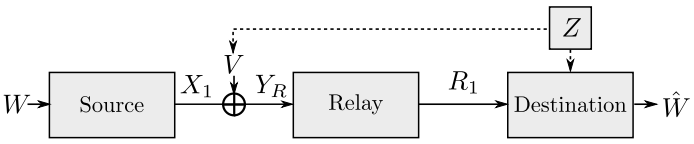

Fig. 6. The multihop Gaussian relay channel with source-relay channel state information available at the destination.

C. Multihop Gaussian Channel with State

Next, we consider an AWGN multihop channel, called *Gaussian* MRC-D, in which the source-relay link is characterized by $Y_R = X_1 + V$, while the destination has access to correlated state information Z. We assume that V and Z are zero-mean jointly Gaussian random variables with a covariance matrix

$$\mathbf{C}_{ZV} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \tag{24}$$

The channel input at the source has to satisfy the power constraint $E[|X_1^n|^2] \le nP$. Finally, the relay and the destination are connected by a noiseless link of rate R_1 (see Figure 6 for the channel model).

In this case, the cut-set bound is given by

$$R_{\mathrm{CS}} = \min\left\{R_1, \frac{1}{2}\log\left(1 + \frac{P}{1-\rho^2}\right)\right\}. \tag{25}$$

It easy to characterize the optimal DF rate, achieved by a Gaussian input, as follows:

$$R_{\mathrm{DF}} = \min\left\{R_1, \frac{1}{2}\log(1+P)\right\}. \tag{26}$$

For CF and pDCF, we consider the achievable rate when the random variables $(X_1, U, \hat{Y}_R)$ are constrained to be jointly Gaussian, which is a common assumption in evaluating achievable rates, yet potentially suboptimal. For CF, we generate the compression codebook using $\hat{Y}_R = Y_R + Q$, where $Q \sim \mathcal{N}(0, \sigma_q^2)$. Optimizing over σ_q^2, the maximum achievable rate is given by

$$R_{\mathrm{CF}} = R_1 - \frac{1}{2}\log\left(\frac{P + 2^{2R_1}(1-\rho^2)}{P+1-\rho^2}\right). \tag{27}$$

For pDCF, we let $U \sim \mathcal{N}(0, \alpha P_1)$, and $X_1 = U + T$ to be a superposition codebook where T is independent of U and distributed as $T \sim \mathcal{N}(0, \bar{\alpha}P_1)$, where $\bar{\alpha} \triangleq 1 - \alpha$. We generate a quantization codebook using the test channel $\hat{Y}_R = Y_R + Q$ as in CF. Next lemma shows that with this choice of random variables, pDCF reduces either to pure DF or pure CF, similarly to the multihop binary model in Section IV-B.

Lemma 8. *The optimal achievable rate for pDCF with jointly Gaussian $(X_1, U, \hat{Y}_R)$ is given by*

$$\begin{aligned} R_{\mathrm{pDCF}} &= \max\{R_{\mathrm{DF}}, R_{\mathrm{CF}}\} \qquad (28)\\ &= \begin{cases} \min\{R_1, 1/2\log(1+P)\} & \text{if } \rho^2 \le 2^{-2R_1}(1+P), \\ R_1 - \frac{1}{2}\log\left(\frac{P+2^{2R_1}(1-\rho^2)}{P+1-\rho^2}\right) & \text{if } \rho^2 > 2^{-2R_1}(1+P). \end{cases} \end{aligned}$$

Proof. See Appendix F. □

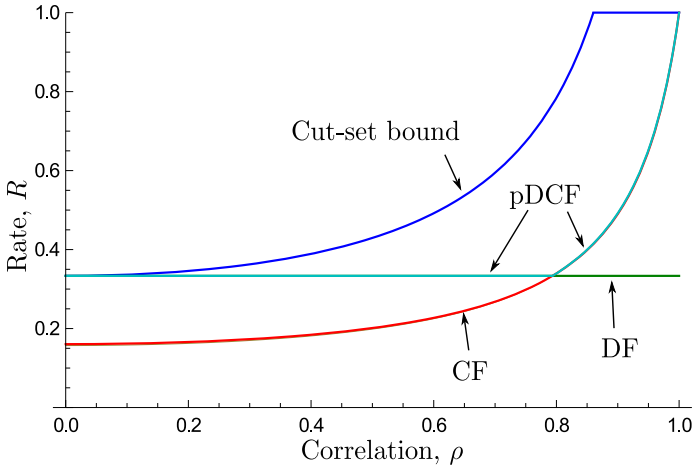

Fig. 7. Achievable rates and the cut-set upper bound for the multihop AWGN relay channel with source-relay channel state information at the destination for $R_1 = 1$ and $P = 0.3$.

In Figure 7 the achievable rates are compared with the cut-set bound. It is shown that DF achieves the best rate when the correlation coefficient ρ is low, i.e., when the destination has low quality channel state information, while CF achieves higher rates for higher values of ρ. It is seen that pDCF achieves the best of the two transmission schemes. Note also that for $\rho = 0$ DF meets the cut-set bound, while for $\rho = 1$ CF meets the cut-set bound.

Although this example proves the suboptimality of the DF scheme for the channel model under consideration, it does not necessarily lead to the suboptimality of the CF scheme as we have constrained the auxiliary random variables to be Gaussian.

V. Comparison with the Cut-Set Bound

In the examples considered in Section IV, we have seen that for certain conditions, the choice of certain random variables allows us to show that the cut-set bound and the capacity coincide. For example, we have seen that for the parallel binary symmetric MRC-D the proposed pDCF scheme achieves the cut-set bound for $\delta \ge h_2^{-1}(2 - R_1)$, or Gaussian random variables meet the cut-set bound for $\rho = 0$ or $\rho = 1$ in the Gaussian MRC-D. An interesting question is whether the capacity expression in Theorem 1 always coincides with the cut-set bound or not; that is, whether the cut-set bound is tight for the relay channel model under consideration.

To address this question, we consider the multihop binary channel in (19) for $Z \sim \mathrm{Ber}(1/2)$. The capacity $\mathcal{C}$ of this channel is given in the following lemma.

Lemma 9. *The capacity of the binary symmetric MRC-D with $Y_R = X_1 \oplus N \oplus Z$, where $N \sim Ber(\delta)$ and $Z \sim Ber(1/2)$, is achieved by CF and pDCF, and is given by*

$$\mathcal{C} = 1 - h_2(\delta \star h_2^{-1}(1 - R_1)). \tag{29}$$

Proof. See Appendix G. □

From (20), the cut-set bound is given by $R_{\mathrm{CS}} = 1 - h_2(\delta)$. It then follows that in general the capacity is below the cut-set

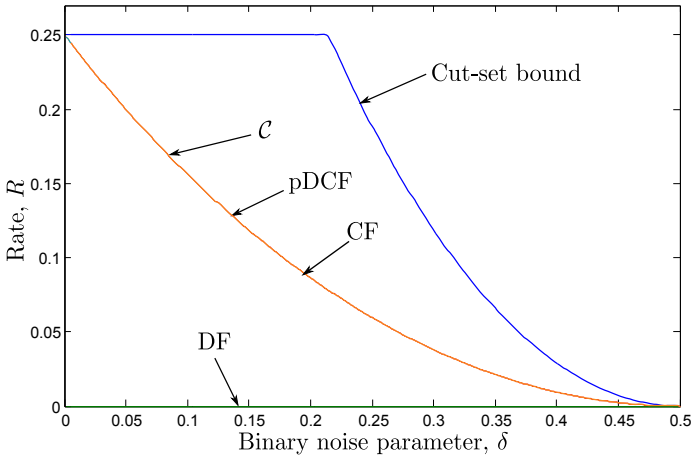

Fig. 8. Achievable rates, capacity and cut-set upper bound for the multihop binary relay channel with respect to δ for $R_1 = 0.25$ and $p_z = 0.5$.

bound. Note that for this setup, $R_{\rm DF} = 0$ and pDCF reduces to CF, i.e., $R_{\rm pDCF} = R_{\rm CF}$. See Figure 8 for comparison of the capacity with the cut-set bound for varying δ values.

CF suffices to achieve the capacity of the binary symmetric MRC-D for $Z \sim \text{Ber}(1/2)$. While in general pDCF outperforms DF and CF, in certain cases these two schemes are sufficient to achieve the cut-set bound, and hence, the capacity. For the ORC-D model introduced in Section II, the cut-set bound is given by

$$R_{\rm CS} = R_2 + \min\{R_1, \max_{p(x_1)} I(X_1; Y_R|Z)\}. \tag{30}$$

Next, we present four cases for which the cut-set bound is achievable, and hence, is the capacity:

Case 1) If $I(Z;Y_R) = 0$, the setup reduces to the class of orthogonal relay channels studied in [3], for which the capacity is known to be achieved by pDF.

Case 2) If $H(Y_R|X_1Z) = 0$, i.e., Y_R is a deterministic function of X_1 and Z, the capacity, given by

$$R_2 + \min\{R_1, \max_{p(x_1)} I(X_1; Y_R|Z)\},$$

is achievable by CF.

Case 3) If $\max_{p(x_1)} I(X_1;Y_R) \geq R_1$, the capacity, given by $\mathcal{C} = R_2 + R_1$, is achievable by pDF.

Case 4) Let $\arg\max_{p(x_1)} I(X_1;Y_R|Z) = \bar{p}(x_1)$. If $R_1 > H(\bar{Y}_R|Z)$ for $\bar{Y}_R$ induced by $\bar{p}(x_1)$, the capacity, given by $R_2 + I(\bar{X}_1; \bar{Y}_R|Z)$, is achievable by CF.

Proof. See Appendix H. □

These cases can be observed in the examples from Section IV. For example, in the Gaussian MRC-D with $\rho = 0$, Y_R is independent of Z, and thus, DF meets the cut-set bound as stated in Case 1. Similarly, for $\rho = 1$ CF meets the cut-set bound since Y_R is a deterministic function of X_R and Z, which corresponds to Case 2.

For the parallel binary symmetric MRC-D in Section IV-A, pDCF achieves the cut-set bound if $\delta \geq h_2^{-1}(2 - R_1)$ due to the following reasoning. Since Y_R^1 is independent of X_1^1, from Case 1, DF should achieve the cut-set bound. Once X_1^1 is decoded, the available rate to compress Y_2 is given by $R_1 - I(X_1;Y_1) = R_1 - 1 + h_2(\delta)$, and the entropy of Y_2 conditioned on the channel state at the destination is given by $H(Y_2|Z) = 1 - h_2(\delta)$. For $\delta \geq h_2^{-1}(2 - R_1)$ we have $R_1 - I(X_1;Y_1) \geq H(Y_2|Z)$. Therefore the relay can compress Y_2 losslessly, and transmit to the destination. This corresponds to Case 4. Thus, the capacity characterization in the parallel binary symmetric MRC-D is due to a combination of Case 1 and Case 4.

VI. Conclusion

We have considered a class of orthogonal relay channels, in which the channels connecting the source to the relay and the destination, and the relay to the destination, depend on a state sequence, known at the destination. We have characterized the capacity of this class of relay channels, and shown that it is achieved by the partial decode-compress-and-forward (pDCF) scheme. This is the first three-terminal relay channel model for which the pDCF is shown to be capacity achieving while partial decode-and-forward (pDF) and compress-and-forward (CF) schemes are both suboptimal in general. We have also shown that, in general, the capacity of this channel is below the cut-set bound.

Appendix A
Proof of Lemma 1

In the rate expression and joint pmf in Theorem 2, we set $X^n = (X_1^n, X_2^n)$, $Y^n = (Y_1^n, Y_2^n, Z^n)$, $V = \emptyset$, and generate X_R^n and X_2^n independent of the rest of the random variables with distributions $p^*(x_R)$ and $p^*(x_2)$, which maximize the mutual information terms in (5), respectively. Under this set of distributions we have

$$\begin{aligned}
I(X; Y\hat{Y}_R|X_R U) &= I(X_1X_2; Y_1Y_2\hat{Y}_R Z|X_R, U)\\
&\overset{(a)}{=} I(X_1X_2; Y_2\hat{Y}_R|X_R U Z)\\
&\overset{(b)}{=} I(X_2; Y_2|Z) + I(X_1; \hat{Y}_R|UZ)\\
&= R_2 + I(X_1; \hat{Y}_R|UZ),\\
I(U; Y_R|X_R V) &= I(U; Y_R|X_R) \overset{(c)}{=} I(U; Y_R),\\
I(XX_R; Y) &= I(X_1X_2X_R; Y_1Y_2Z)\\
&\overset{(d)}{=} I(X_2X_R; Y_1Y_2|Z)\\
&\overset{(e)}{=} I(X_R; Y_1) + I(X_2; Y_2|Z) = R_1 + R_2,\\
I(\hat{Y}_R; Y_R|XX_R UY) &= I(\hat{Y}_R; Y_R|X_R X_1X_2 UY_1Y_2Z)\\
&\overset{(f)}{=} I(\hat{Y}_R; Y_R|X_R X_1X_2 UY_2Z)\\
&\overset{(g)}{=} I(\hat{Y}_R; Y_R|X_1 UZ),\\
I(\hat{Y}_R; Y_R|YX_R U) &= I(\hat{Y}_R; Y_R|Y_1Y_2ZX_R U)\\
&\overset{(h)}{=} I(\hat{Y}_R; Y_R|UZ),\\
I(X_R; Y|V) &= I(X_R; Y_1Y_2Z) = I(X_R; Y_1) = R_1,
\end{aligned}$$

where (a) is due to the Markov chain $(X_1X_2) - X_R - Y_1$; $(b), (c), (e), (f), (g), (h)$ are due to the independence of (U, X_1) and X_R, and (d) is due to the Markov chain $(Y_1Y_2) - (X_2X_RZ) - X_1$.

Then, (8) reduces to the following rate

$$R = \sup_{\mathcal{P}} \min\{I(U;Y_R) + R_2 + I(X_1;\hat{Y}_R|UZ), \quad (31)$$
$$R_2 + R_1 - I(\hat{Y}_R;Y_R|X_1UZ)\},$$
$$\text{s.t. } R_1 \geq I(\hat{Y}_R;Y_R|UZ). \quad (32)$$

Focusing on the joint distributions in $\mathcal{P}$ such that the minimum in R is achieved for the first argument, i.e.,

$$R_1 - I(\hat{Y}_R;Y_R|X_1UZ) \geq I(U;Y_R) + I(X_1;\hat{Y}_R|UZ),$$

and using the chain rule for the mutual information, the rate achievable by pDCF is lower bounded by

$$R \geq \sup_{\mathcal{P}} R_2 + I(U;Y_R) + I(X_1;\hat{Y}_R|UZ)$$
$$\text{s.t. } R_1 \geq I(U;Y_R) + I(X_1Y_R;\hat{Y}_R|UZ), \quad (33)$$
$$R_1 \geq I(\hat{Y}_R;Y_R|UZ). \quad (34)$$

From (33), we have

$$\begin{aligned} R_1 &\geq I(U;Y_R) + I(X_1Y_R;\hat{Y}_R|UZ) \\ &\overset{(a)}{=} I(U;Y_R) + I(\hat{Y}_R;Y_R|UZ) \\ &\geq I(\hat{Y}_R;Y_R|UZ), \end{aligned} \quad (35)$$

where (a) is due to the Markov chain $\hat{Y}_R - (UY_R) - (X_1Z)$. Hence, (33) implies (34), i.e., the latter condition is redundant, and $R \geq \mathcal{C}$. Therefore the capacity expression $\mathcal{C}$ in (7) is achievable by pDCF. This concludes the proof.

Appendix B
Proof of Lemma 2

Consider any sequence of $(2^{nR}, n, \nu_n)$ codes such that $\lim_{n\to\infty} \nu_n \to 0$. We need to show that $R \leq R_{up}$.

Let us define $U_i \triangleq (Y_{R1}^{i-1} X_{1i+1}^n Z^{n\setminus i})$ and $\hat{Y}_{Ri} \triangleq (Y_{1i+1}^n)$. For such $\hat{Y}_{Ri}$ and U_i, the following Markov chain holds

$$\hat{Y}_{Ri} - (U_i, Y_{Ri}) - (X_{1i}, X_{2i}, Z_i, Y_{1i}, Y_{2i}, X_{Ri}). \quad (36)$$

From Fano's inequality, we have

$$H(W|Y_1^nY_2^nZ^n) \leq n\epsilon_n, \quad (37)$$

such that $\epsilon_n \to 0$ as $n \to \infty$.

First, we derive the following set of inequalities related to the capacity of the source-destination channel.

$$\begin{aligned} nR &= H(W) \\ &\overset{(a)}{=} I(W;Y_1^nY_2^n|Z^n) + H(W|Y_1^nY_2^nZ^n) \\ &\overset{(b)}{\leq} I(X_1^nX_2^n;Y_1^nY_2^n|Z^n) + n\epsilon_n, \end{aligned} \quad (38)$$

where (a) follows from the independence of Z^n and W; and (b) follows from Fano's inequality in (37).

We also have the following inequalities:

$$\begin{aligned} I(X_2^n;Y_2^n|Z^n) &= \sum_{i=1}^n H(Y_{2i}|Z^nY_{21}^{i-1}) - H(Y_{2i}|Z^nY_{21}^{i-1}X_2^n) \\ &\overset{(a)}{\leq} \sum_{i=1}^n H(Y_{2i}|Z_i) - H(Y_{2i}|Z_iX_{2i}) \\ &= \sum_{i=1}^n I(X_{2i};Y_{2i}|Z_i) \\ &\overset{(b)}{=} nI(X_{2Q'};Y_{2Q'}|Z_{Q'}Q') \\ &\overset{(c)}{\leq} nI(X_{2Q'};Y_{2Q'}|Z_{Q'}) \\ &\overset{(d)}{\leq} nR_2, \end{aligned} \quad (39)$$

where (a) follows since conditioning reduces entropy, (b) follows by defining Q' as a uniformly distributed random variable over $\{1,...,n\}$ and $(X_{2Q'}, Y_{2Q'}, Z_{Q'})$ as a pair of random variables satisfying $\Pr\{X_{2i} = x_2, Y_{2i} = y_2, Z_i = z\} = \Pr\{X_{2Q'} = x_2, Y_{2Q'} = y_2, , Z_{Q'} = z|Q = i\}$ for $i = 1,...,n$, (c) follows from the Markov chain relation $Q' - X_{2Q'} - Y_{2Q'}$ and (d) follows from the definition of R_2 in (5). Following the same steps, we obtain

$$I(X_R^n;Y_1^n|Z^n) \leq nR_1. \quad (40)$$

Then, we can bound the achievable rate as,

$$\begin{aligned} nR &= I(W;Y_1^nY_2^nZ^n) + H(W|Y_1^nY_2^nZ^n) \\ &\overset{(a)}{\leq} I(W;Y_1^nY_2^nZ^n) + n\epsilon_n \\ &\overset{(b)}{=} I(W;Y_2^n|Z^n) + I(W;Y_1^n|Y_2^nZ^n) + n\epsilon_n \\ &\overset{(c)}{\leq} I(X_2^n;Y_2^n|Z^n) + I(W;Y_1^n|Y_2^nZ^n) + n\epsilon_n \\ &\overset{(d)}{\leq} nR_2 + H(Y_1^n|Y_2^nZ^n) - H(Y_1^n|WZ^n) + n\epsilon_n \\ &\overset{(e)}{\leq} nR_2 + H(Y_1^n|Z^n) - H(Y_1^n|WX_1^nZ^n) + n\epsilon_n \\ &\overset{(f)}{=} nR_2 + I(X_1^n;Y_1^n|Z^n) + n\epsilon_n \\ &\overset{(g)}{\leq} nR_2 + H(X_1^n) - H(X_1^n|Y_1^nZ^n) + n\epsilon_n \\ &= nR_2 + \sum_{i=1}^n H(X_{1i}|X_{1i+1}^n) - H(X_1^n|Y_1^nZ^n) + n\epsilon_n \\ &\overset{(h)}{\leq} nR_2 + \sum_{i=1}^n \Big[I(Y_{R1}^{i-1}Z^{n\setminus i};Y_{Ri}) \\ &\quad + H(X_{1i}|X_{1i+1}^n) \Big] - H(X_1^n|Y_1^nZ^n) + n\epsilon_n \\ &= nR_2 + \sum_{i=1}^n \Big[I(Y_{R1}^{i-1}Z^{n\setminus i}X_{1i+1}^n;Y_{Ri}) \\ &\quad - I(X_{1i+1}^n;Y_{Ri}|Y_{R1}^{i-1}Z^{n\setminus i}) + H(X_{1i}|Y_{R1}^{i-1}Z^{n\setminus i}X_{1i+1}^n) \\ &\quad + I(X_{1i};Y_{R1}^{i-1}Z^{n\setminus i}|X_{1i+1}^n) \Big] - H(X_1^n|Y_1^nZ^n) + n\epsilon_n \\ &\overset{(i)}{=} nR_2 + \sum_{i=1}^n \Big[I(Y_{R1}^{i-1}Z^{n\setminus i}X_{1i+1}^n;Y_{Ri}) \\ &\quad + H(X_{1i}|Y_{R1}^{i-1}Z^{n\setminus i}X_{1i+1}^n) \Big] - H(X_1^n|Y_1^nZ^n) + n\epsilon_n \end{aligned}$$

$$
\begin{aligned}
&= nR_2 + \sum_{i=1}^{n}\Big[I(U_i;Y_{Ri}) + H(X_{1i}|U_i)\Big] \\
&\quad -H(X_1^n|Y_1^nZ^n) + n\epsilon_n \\
&\overset{(j)}{\leq} nR_2 + \sum_{i=1}^{n}\Big[I(U_i;Y_{Ri}) + H(X_{1i}|U_i) \\
&\quad -H(X_{1i}|U_iZ_i\hat{Y}_{Ri})\Big] + n\epsilon_n \\
&\overset{(k)}{=} nR_2 + \sum_{i=1}^{n}\Big[I(U_i;Y_{Ri}) + I(X_{1i};\hat{Y}_{Ri}|U_iZ_i)\Big] + n\epsilon_n,
\end{aligned}
$$

where (a) is due to Fano's inequality; (b) is due to the chain rule and the independence of Z^n from W; (c) is due to the data processing inequality; (d) is due to the Markov chain relation $Y_2^n - (W, Z^n) - Y_1^n$ and (39); (e) is due to the fact that conditioning reduces entropy, and that X_1^n is a deterministic function of W; (f) is due to the Markov chain relation $Y_1^n - X_1^n - W$; (g) is due to the independence of Z^n and X_1^n; (i) follows because

$$
\begin{aligned}
\sum_{i=1}^{n} I(X_{1i};Y_{R1}^{i-1}Z^{n\setminus i}|X_{1i+1}^n) &\overset{(l)}{=} \sum_{i=1}^{n} I(X_{1i};Y_{R1}^{i-1}|X_{1i+1}^nZ^{n\setminus i}) \\
&\overset{(m)}{=} \sum_{i=1}^{n} I(X_{1i+1}^n;Y_{Ri}|Y_{R1}^{i-1}Z^{n\setminus i}),
\end{aligned}
$$

where (l) is due to the independence of Z^n and X_1^n; and (m) is the conditional version of Csiszár's equality [24]. Inequality (j) is due to the following bound,

$$
\begin{aligned}
H(X_1^n|Y_1^nZ^n) &= \sum_{i=1}^{n} H(X_{1i}|X_{1i+1}^nZ^nY_1^n) \\
&\geq \sum_{i=1}^{n} H(X_{1i}|Y_{R1}^{i-1}X_{1i+1}^nZ^nY_1^n) \\
&\overset{(n)}{=} \sum_{i=1}^{n} H(X_{1i}|Y_{R1}^{i-1}X_{1i+1}^nZ^nY_{1i+1}^n) \\
&= \sum_{i=1}^{n} H(X_{1i}|U_iZ_i\hat{Y}_{Ri}),
\end{aligned} \tag{41}
$$

where (n) is follows from the Markov chain relation $X_{1i} - (Y_{R1}^{i-1}X_{1i+1}^nZ^nY_{1i+1}^n) - Y_{11}^i$, and noticing that $X_{Ri} = f_{r,i}(Y_{R1}^{i-1})$. Finally, (k) is due to the fact that Z_i independent of (X_{1i}, U_i).

We can also obtain the following sequence of inequalities

$$
\begin{aligned}
&nR_1+nR_2 \\
&\overset{(a)}{\geq} I(X_R^n;Y_1^n|Z^n) + I(X_2^n;Y_2^n|Z^n) \\
&\overset{(b)}{\geq} H(Y_2^n|Z^n) - H(Y_2^n|X_2^nZ^n) \\
&\quad +H(Y_1^n|Y_2^nZ^n) - H(Y_1^n|X_R^nZ^n) \\
&= H(Y_1^nY_2^n|Z^n) - H(Y_2^n|X_2^nZ^n) - H(Y_1^n|X_R^nZ^n) \\
&\overset{(c)}{=} H(Y_1^nY_2^n|Z^n) - H(Y_2^n|X_1^nX_2^nY_R^nZ^n) \\
&\quad -H(Y_1^n|X_R^nX_1^nX_2^nY_R^nY_2^nZ^n)
\end{aligned} \tag{42}
$$

$$
\begin{aligned}
&\overset{(d)}{\geq} H(Y_1^nY_2^n|Z^n) - H(Y_1^nY_2^n|X_1^nX_2^nY_R^nZ^n) \\
&= I(Y_1^nY_2^n;X_1^nX_2^nY_R^n|Z^n) \\
&= I(X_1^nX_2^n;Y_1^nY_2^n|Z^n) + I(Y_1^nY_2^n;Y_R^n|X_1^nX_2^nZ^n) \\
&\overset{(e)}{\geq} nR + I(Y_1^nY_2^n;Y_R^n|X_1^nX_2^nZ^n) - n\epsilon_n \\
&\overset{(f)}{=} nR + I(Y_1^n;Y_R^n|X_1^nZ^n) - n\epsilon_n \\
&= nR + \sum_{i=1}^{n} I(Y_1^n;Y_{Ri}|X_1^nY_{R1}^{i-1}Z^n) - n\epsilon_n \\
&\geq nR + \sum_{i=1}^{n} I(Y_{1i+1}^n;Y_{Ri}|X_1^nY_{R1}^{i-1}Z^n) - n\epsilon_n \\
&\overset{(g)}{\geq} nR + \sum_{i=1}^{n} I(Y_{1i+1}^n;Y_{Ri}|X_{1i}^nY_{R1}^{i-1}Z^n) - n\epsilon_n \\
&= nR + \sum_{i=1}^{n} I(\hat{Y}_{Ri};Y_{Ri}|X_{1i}U_iZ_i) - n\epsilon_n,
\end{aligned}
$$

where (a) follows from (39) and (40); (b) is due to the fact that conditioning reduces entropy; (c) is due to the Markov chains $Y_2^n - (X_2^nZ^n) - (X_1^nY_R^n)$ and $Y_1^n - (X_R^nZ^n) - (X_1^nX_2^nY_R^nY_2^n)$; (d) follows since conditioning reduces entropy; (e) is due to the expression in (38); (f) is due to the Markov chain $(Y_R^nY_1^n) - (X_1^nZ^n) - (X_2^nY_2^n)$ and; (g) is due to the Markov chain $(Y_{1i+1}^n) - (X_{1i}^nY_{R1}^{i-1}Z^n) - X_{11}^{i-1}$.

A single letter expression can be obtained by using the usual time-sharing random variable arguments. Let Q be a time sharing random variable uniformly distributed over $\{1, ..., n\}$, independent of all the other random variables. Also, define a set of random variables $(X_{1Q}, Y_{RQ}, U_Q, \hat{Y}_{RQ}, Z_Q)$ satisfying

$$
\begin{aligned}
&\Pr\{X_{1Q} = x_1, Y_{RQ} = y_R, U_Q = u, \hat{Y}_{RQ} = \hat{y}_R, Z_Q = z|Q = i\} \\
&= \Pr\{X_{1i} = x_1, Y_{Ri} = y_R, U_i = u, \hat{Y}_{Ri} = \hat{y}_D, Z_i = z\} \\
&\qquad \text{for } i = 1, ..., n.
\end{aligned} \tag{43}
$$

Define $U \triangleq (U_Q, Q)$, $\hat{Y}_R \triangleq \hat{Y}_{RQ}$, $X_1 \triangleq X_{1Q}$, $Y_{RQ} \triangleq Y_R$ and $Z \triangleq Z_Q$. We note that the pmf of the tuple $(X_1, Y_R, U, \hat{Y}_R, Z)$ belongs to $\mathcal{P}$ in (6) as follows:

$$
\begin{aligned}
&p(u, x_1, y_R, z, \hat{y}_R) \\
&= p(q, u_Q, x_{1Q}, y_{RQ}, z_Q, \hat{y}_{RQ}) \\
&= p(q, u_Q, x_{1Q})p(z_Qy_{RQ}\hat{y}_{RQ}|q, u_Qx_{1Q}) \\
&= p(q, u_Q, x_{1Q})p(z_Q|q, u_Q, x_{1Q})p(y_{RQ}, \hat{y}_{RQ}|q, u_Q, x_{1Q}, z_Q) \\
&\overset{(a)}{=} p(q, u_Q, x_{1Q})p(z)p(y_{RQ}|q, u_Q, x_{1Q}, z_Q) \\
&\quad \cdot p(\hat{y}_{RQ}|q, u_Q, x_{1Q}, z_Q, y_{RQ}) \\
&\overset{(b)}{=} p(q, u_Q, x_{1Q})p(z)p(y_R|x_1, z)p(\hat{y}_{RQ}|q, u_Q, x_{1Q}, z_Q, y_{RQ}) \\
&\overset{(c)}{=} p(q, u_Q, x_{1Q})p(z)p(y_R|x_1, z)p(\hat{y}_{RQ}|q, u_Q, y_{RQ}) \\
&= p(u, x_1)p(z)p(y_R|x_1, z)p(\hat{y}_R|u, y_R),
\end{aligned}
$$

where (a) follows since the channel state Z^n is i.i.d; and thus $p(z_Q|q, u_Q, x_{1Q}) = p(z_Q|q) = p(z)$; (b) follows since $p(y_{RQ}|q, u_Q, x_{1Q}, z_Q) = p(y_{RQ}|q, x_{1Q}, z_Q) = p(y_R|x_1, z)$; (c) follows from the Markov chain in (36).

Then, we get the single letter expression,

$$\begin{aligned} R &\leq R_2 + \frac{1}{n}\sum_{i=1}^{n}[I(U_i;Y_{Ri}) + I(X_{1i};\hat{Y}_{Ri}|U_iZ_i)] + \epsilon_n \\ &= R_2 + I(U_Q;Y_{RQ}|Q) + I(X_{1Q};\hat{Y}_{RQ}|U_QZ_QQ) + \epsilon_n \\ &\leq R_2 + I(U_QQ;Y_{RQ}) + I(X_{1Q};\hat{Y}_{RQ}Q|U_QZ_Q) + \epsilon_n \\ &= R_2 + I(U;Y_R) + I(X_1;\hat{Y}_R|UZ) + \epsilon_n, \end{aligned}$$

and

$$\begin{aligned} R_1 + R_2 &\geq R + \frac{1}{n}\sum_{i=1}^{n} I(\hat{Y}_{Ri};Y_{Ri}|X_{1i}U_iZ_i) - n\epsilon_n \\ &= R + I(\hat{Y}_{RQ};Y_{RQ}|X_{1Q}U_QZ_QQ) - n\epsilon_n \\ &= R + I(\hat{Y}_R;Y_R|X_1UZ) - n\epsilon_n. \end{aligned}$$

The cardinality of the bounds on the alphabets of U and $\hat{Y}_R$ can be found using the usual techniques [24]. This completes the proof.

Appendix C
Proof of Lemma 3

Now, we will show that the expression of R_{up} in (9) is equivalent to the expression $\mathcal{C}$ in (7). First we will show that $\mathcal{C} \leq R_{up}$. Consider the subset of pmf's in $\mathcal{P}$ such that

$$\begin{aligned} &R_1 + R_2 - I(\hat{Y}_R;Y_R|X_1UZ) && (44)\\ &\quad \geq R_2 + I(U;Y_R) + I(X_1;\hat{Y}_R|UZ), && (45) \end{aligned}$$

holds. Then, similarly to (35) in Appendix A this condition is equivalent to

$$R_1 \geq I(U;Y_R) + I(Y_R;\hat{Y}_R|UZ). \tag{46}$$

Hence, we have $\mathcal{C} \leq R_{up}$.

Then, it remains to show that $\mathcal{C} \geq R_{up}$. As R_2 can be extracted from the supremum, it is enough to show that, for each $(X_1, U, Z, Y_R, \hat{Y}_R)$ tuple with a joint pmf $p_e \in \mathcal{P}$ satisfying

$$\begin{aligned} &R(p_e) \leq I(U;Y_R) + I(X_1;\hat{Y}_R|UZ), \\ &\text{where } R(p_e) \triangleq R_1 - I(\hat{Y}_R;Y_R|X_1UZ), \end{aligned} \tag{47}$$

there exist random variables $(X_1^*, U^*, Z, Y_R^*, \hat{Y}_R^*)$ with joint pmf $p_e^* \in \mathcal{P}$ that satisfy

$$\begin{aligned} &R(p_e) = I(U^*;Y_R) + I(X_1^*;\hat{Y}_R^*|U^*Z) \text{ and} \\ &R(p_e) \leq R_1 - I(\hat{Y}_R^*;Y_R|X_1^*U^*Z). \end{aligned} \tag{48}$$

This argument is proven next.

Let B denote a Bernoulli random variable with parameter $\lambda \in [0,1]$, i.e., $B = 1$ with probability λ, and $B = 0$ with probability $1-\lambda$. We define the triplets of random variables:

$$(U', X_1', \hat{Y}_R') = \begin{cases} (U, X_1, \hat{Y}_R) & \text{if } B = 1, \\ (X_1, X_1, \emptyset) & \text{if } B = 0, \end{cases} \tag{49}$$

and

$$(U'', X_1'', \hat{Y}_R'') = \begin{cases} (X_1, X_1, \emptyset) & \text{if } B = 1, \\ (\emptyset, \emptyset, \emptyset) & \text{if } B = 0. \end{cases} \tag{50}$$

We first consider the case $R(p_e) > I(X_1;Y_R)$. Let $U^* = (U', B)$, $X_1^* = X_1'$, $\hat{Y}_R^* = (\hat{Y}_R', B)$. For $\lambda = 1$,

$$\begin{aligned} I(U^*;Y_R) + I(X_1^*;\hat{Y}_R^*|U^*Z) &= I(U;Y_R) + I(X_1;\hat{Y}_R|UZ) \\ &> R(p_e), \end{aligned} \tag{51}$$

and for $\lambda = 0$,

$$\begin{aligned} I(U^*;Y_R) + I(X_1^*;\hat{Y}_R^*|U^*Z) &= I(X_1;Y_R) \\ &< R(p_e). \end{aligned} \tag{52}$$

As $I(U^*;Y_R) + I(X_1^*;\hat{Y}_R^*|U^*Z)$ is a continuous function of λ, by the intermediate value theorem, there exists a $\lambda \in [0,1]$ such that $I(U^*;Y_R) + I(X_1^*;\hat{Y}_R^*|U^*Z) = R(p_e)$. We denote the corresponding joint distribution by p_e^*.

We have

$$\begin{aligned} I(\hat{Y}_R^*;Y_R|X_1^*U^*Z) &= I(\hat{Y}_R';Y_R|X_1'U'ZB) \\ &= \lambda I(\hat{Y}_R;Y_R|X_1UZ) \\ &\leq I(\hat{Y}_R;Y_R|X_1UZ), \end{aligned} \tag{53}$$

which implies that p_e^* satisfies (48) since

$$\begin{aligned} R(p_e) &= R_1 - I(\hat{Y}_R;Y_R|X_1UZ) \\ &\leq R_1 - I(\hat{Y}_R^*;Y_R|X_1^*U^*Z). \end{aligned} \tag{54}$$

Next we consider the case $R(p_e) \leq I(X_R;Y_1)$. We define $U^* = (U'', B)$, $X_1^* = X_1''$ and $\hat{Y}_R^* = (\hat{Y}_R'', B)$. Then, for $\lambda = 1$,

$$\begin{aligned} I(U^*;Y_R) + I(X_1^*;\hat{Y}_R^*|U^*Z) &= I(X_1;Y_R) \\ &\geq R(p_e), \end{aligned}$$

and for $\lambda = 0$,

$$\begin{aligned} I(U^*;Y_R) + I(X_1^*;\hat{Y}_R^*|U^*Z) &= 0 \\ &< R(p_e). \end{aligned} \tag{55}$$

Once again, as $I(U^*;Y_R) + I(X_1^*;\hat{Y}_R^*|U^*Z)$ is a continuous function of λ, by the intermediate value theorem, there exists a $\lambda \in [0,1]$ such that $I(U^*;Y_R) + I(X_1^*;\hat{Y}_R^*|U^*Z) = R(p_e)$. Again, we denote this joint distribution by p_e^*. On the other hand, we have $I(\hat{Y}_R^*;Y_R|X_1^*U^*Z) = 0$, which implies that

$$\begin{aligned} R(p_e) &= R_1 - I(\hat{Y}_R;Y_R|X_1UZ) \\ &\leq R_1 \\ &= R_1 - I(\hat{Y}_R^*;Y_R|X_1^*U^*Z). \end{aligned} \tag{56}$$

That is, p_e^* also satisfies (48).

We have shown that for any joint pmf $p_e \in \mathcal{P}$ satisfying (47), there exist another joint pmf, p_e^*, that satisfies (48). For a distribution satisfying (48) we can write

$$\begin{aligned} R_1 &> I(U^*;Y_R) + I(X_1^*;\hat{Y}_R^*|U^*Z) + I(\hat{Y}_R^*;Y_R|X_1^*U^*Z) \\ &= I(U^*;Y_R) + I(Y_RX_1^*;\hat{Y}_R^*|U^*Z) \\ &\overset{(a)}{=} I(U^*;Y_R) + I(\hat{Y}_R^*;Y_R|U^*Z) \end{aligned}$$

where (a) is due to Markov chain $X_1^* - (Y_RZU^*) - \hat{Y}_R^*$. This concludes the proof.

APPENDIX D
PROOF OF LEMMA 4

Before deriving the maximum achievable rate by CF in Lemma 4, we provide some definitions that will be used in the proof.

Let X and Y be a pair of discrete random variables, where $\mathcal{X} = \{1, 2, ..., n\}$ and $\mathcal{Y} = \{1, 2, ..., m\}$, for $n, m < \infty$. Let $\mathbf{p}_Y \in \Delta_m$ denote the distribution of Y, where Δ_k denotes the $(k-1)$-dimensional simplex of probability k-vectors. We define T_{XY} as the $n \times m$ stochastic matrix with entries $T_{XY}(j,i) = \Pr\{X = j | Y = i\}$. Note that the joint distribution $p(x, y)$ is characterized by T_{XY} and $\mathbf{p}_Y$.

Next, we define the conditional entropy bound from [25], which lower bounds the conditional entropy between two variables. Note the relabeling of the variables in [25] to fit our model.

Definition 1 (Conditional Entropy Bound)**.** *Let $\mathbf{p}_Y \in \Delta_m$ be the distribution of Y and T_{XY} denote the channel matrix relating X and Y. Then, for $\mathbf{q} \in \Delta_m$ and $0 \le s \le H(Y)$, define the function*

$$F_{T_{XY}}(\mathbf{q}, s) \triangleq \inf_{\substack{p(w|y):\ X-Y-W,\\ H(Y|W)=s,\ \mathbf{p}_Y=\mathbf{q}.}} H(X|W). \tag{57}$$

That is, $F_{T_{XY}}(\mathbf{q}, s)$ is the infimum of $H(X|W)$ given a specified distribution $\mathbf{q}$ and the value of $H(Y|W)$. Many properties of $F_{T_{XY}}(\mathbf{q}, s)$ are derived in [25], such as its convexity on $(\mathbf{q}, s)$ [25, Theorem 2.3] and its non-decreasing monotonicity in s [25, Theorem 2.5].

Consider a sequence of N random variables $\mathbf{Y} = (Y_1, ..., Y_N)$ and denote by $\mathbf{q}_i$ the distribution of Y_i, for $i = 1, ..., N$, by $\mathbf{q}^{(N)}$ the joint distribution of $\mathbf{Y}$ and by $\mathbf{q} \triangleq \frac{1}{N}\sum_{i=1}^{N} \mathbf{q}_i$ the average distribution. Note that $Y_1, ..., Y_N$ can have arbitrary correlation. Define the sequence $\mathbf{X} = (X_1, ..., X_N)$, in which X_i, $i = 1, ..., N$, is jointly distributed with each Y_i through the stochastic matrix T_{XY} and denote by $T_{XY}^{(N)}$ the Kronecker product of N copies of the stochastic matrix T_{XY}.

Then, the theorem given in [25, Theorem 2.4] can be straightforwardly generalized to non i.i.d. sequences as given in the following lemma.

Lemma 10. *For $N \in \mathbb{Z}^+$, and $0 \le Ns \le H(\mathbf{Y})$, we have*

$$F_{T_{XY}^{(N)}}(\mathbf{q}^{(N)}, Ns) \ge N F_{T_{XY}}(\mathbf{q}, s), \tag{58}$$

where equality holds for i.i.d. Y_i components following $\mathbf{q}$.

Proof. Let $W, \mathbf{X}, \mathbf{Y}$ be a Markov chain, such that $H(\mathbf{Y}|W) = Ns$. Then, using the standard identity we have

$$H(\mathbf{Y}|W) = \sum_{k=1}^{N} H(Y_k | \mathbf{Y}_1^{k-1}, W), \tag{59}$$

$$H(\mathbf{X}|W) = \sum_{k=1}^{N} H(X_k | \mathbf{X}_1^{k-1}, W). \tag{60}$$

Letting $s_k = H(Y_k | \mathbf{Y}_1^{k-1}, W)$, we have

$$\frac{1}{N}\sum_{k=1}^{N} s_k = s. \tag{61}$$

Also, from the Markov chain $X_k - (\mathbf{Y}_1^{k-1}, W) - \mathbf{X}_1^{k-1}$, we have

$$H(X_k | \mathbf{X}_1^{k-1}, W) \ge H(X_k | \mathbf{Y}_1^{k-1}, \mathbf{X}_1^{k-1}, W) \tag{62}$$

$$= H(X_k | \mathbf{Y}_1^{k-1}, W). \tag{63}$$

Applying the conditional entropy bound in (57) we have

$$H(X_k | \mathbf{Y}_1^{k-1}, W) \ge F_{T_{XY}}(\mathbf{q}_k, s_k). \tag{64}$$

Combining (60), (62) and (64) we have

$$H(\mathbf{X}|W) \ge \sum_{k=1}^{N} F_{T_{XY}}(\mathbf{q}_k, s_k) \ge N F_{T_{XY}}(\mathbf{q}, s),$$

where the last inequality follows from the convexity of $F_T(\mathbf{q}, s)$ in $\mathbf{q}$ and s and (61).

If we let $W, \mathbf{X}, \mathbf{Y}$ be N independent copies of the random variables W, X, Y, that achieve $F_{T_{XY}}(\mathbf{q}, s)$, we have $H(\mathbf{Y}|W) = Ns$ and $H(\mathbf{X}|W) = F_{T_{XY}^{(N)}}(\mathbf{q}^N) = N F_{T_{XY}}(\mathbf{q}, s)$. Therefore, $F_{T_{XY}^{(N)}}(\mathbf{q}^N) \le N F_{T_{XY}}(\mathbf{q}, s)$, and the equality holds for i.i.d. components of $\mathbf{Y}$. □

Now, we look into the binary symmetric channel $Y = X \oplus N$ where $N \sim \text{Ber}(\delta)$. Due to the binary modulo-sum operation, we have $X = Y \oplus N$, and we can characterize the channel T_{XY} of this model as

$$T_{XY} = \begin{bmatrix} 1-\delta & \delta \\ \delta & 1-\delta \end{bmatrix}. \tag{65}$$

When Y and X are related through channel T_{XY} in (65), $F_{T_{XY}}(\mathbf{q}, s)$ is characterized as follows [25].

Lemma 11. *Let $Y \sim Ber(q)$, i.e., $\mathbf{q} = [q, 1-q]$, and T_{XY} be given as in (65). Then the conditional entropy bound is*

$$F_{T_{XY}}(q, s) = h_2(\delta \star h_2^{-1}(s)), \quad \textit{for } 0 \le s \le h_2(q).$$

In the following, we use the properties of $F_{T_{XY}}(\mathbf{q}, s)$ to derive the maximum rate achievable by CF in the parallel binary symmetric MRC-D. From (17), we have

$$I(Y_R^1, Y_R^2; \hat{Y}_R | Z) = I(X_1^1 \oplus N_1 \oplus Z, X_1^2 \oplus N_2; \hat{Y}_R | Z)$$
$$= I(X_1^1 \oplus N_1, X_1^2 \oplus N_2; \hat{Y}_R | Z).$$

Let us define $\bar{Y}_R^1 \triangleq X_1^1 \oplus N_1$ and $\bar{\mathbf{Y}}_R \triangleq (\bar{Y}_R^1, Y_R^2)$, and the channel input $\mathbf{X} \triangleq (X_1^1, X_1^2)$. Note that the distribution of $\bar{\mathbf{Y}}_R$, given by $\mathbf{q}^{(2)}$, determines the distribution of $\mathbf{X}$ via $T_{XY}^{(2)}$, the Kronecker product of T_{XY} in (65). Then, we can rewrite the achievable rate for CF in (17) as follows

$$R_{\text{CF}} = \max_{p(\mathbf{x})p(z)p(\bar{\mathbf{y}}_R|\mathbf{x})p(\hat{y}_R|\bar{\mathbf{y}}_R, z)} I(\mathbf{X}, \hat{Y}_R | Z)$$
$$\text{s.t. } R_1 \ge I(\bar{\mathbf{Y}}_R; \hat{Y}_R | Z). \tag{66}$$

Next, we derive a closed form expression for R_{CF}. First, we note that if $R_1 \ge 2$, we have $H(\bar{\mathbf{Y}}_R) \le R_1$ and $R_{\text{CF}} = 2(1 - h(\delta))$, i.e., CF meets the cut-set bound.

For fixed $\mathbf{q}^{(2)}$, if $H(\bar{\mathbf{Y}}_R) \le R_1 \le 2$, the constraint in (66) is satisfied by any $\hat{Y}_R$, and can be ignored. Then, due to the Markov chain $\mathbf{X} - \bar{\mathbf{Y}}_R - \hat{Y}_R Z$, and the data processing inequality, the achievable rate is upper bounded by

$$R_{\text{CF}} \le I(\mathbf{X}, \bar{\mathbf{Y}}_R) = H(\bar{\mathbf{Y}}_R) - 2h(\delta) \le R_1 - 2h(\delta). \tag{67}$$

For $R_1 \leq H(\bar{\mathbf{Y}}_R) \leq 2$, the achievable rate by CF is upper bounded as follows.

$$\begin{aligned}
R_{\mathrm{CF}} &\overset{(a)}{=} \max_{p(\mathbf{x})p(z)p(\bar{\mathbf{y}}_R|\mathbf{x})p(\hat{y}_R|\bar{\mathbf{y}}_R,z)} H(\mathbf{X}) - H(\mathbf{X}|Z\hat{Y}_R) \\
&\quad \text{s.t. } H(\bar{\mathbf{Y}}_R|Z\hat{Y}_R) \geq H(\bar{\mathbf{Y}}_R) - R_1 \\
&\overset{(b)}{\leq} \max_{p(\mathbf{x})p(\bar{\mathbf{y}}_R|\mathbf{x})p(w|\bar{\mathbf{y}}_R)} H(\mathbf{X}) - H(\mathbf{X}|W) \\
&\quad \text{s.t. } H(\bar{\mathbf{Y}}_R|W) \geq H(\bar{\mathbf{Y}}_R) - R_1 \\
&= \max_{p(\mathbf{x})p(\bar{\mathbf{y}}_R|\mathbf{x})} [H(\mathbf{X}) - \min_{p(w|\bar{\mathbf{y}}_R)} H(\mathbf{X}|W)] \\
&\quad \text{s.t. } H(\bar{\mathbf{Y}}_R|W) \geq H(\bar{\mathbf{Y}}_R) - R_1 \\
&\overset{(c)}{=} \max_{p(\mathbf{x})p(\bar{\mathbf{y}}_R|\mathbf{x}),0\leq s\leq H(\bar{\mathbf{Y}}_R)} [H(\mathbf{X}) - F_{T_{XY}^{(2)}}(\mathbf{q}^{(2)}, s)] \\
&\quad \text{s.t. } s \geq H(\bar{\mathbf{Y}}_R) - R_1 \\
&\overset{(d)}{=} \max_{p(\mathbf{x})p(\bar{\mathbf{y}}_R|\mathbf{x})} [H(\mathbf{X}) - F_{T_{XY}^{(2)}}(\mathbf{q}^{(2)}, H(\bar{\mathbf{Y}}_R) - R_1)] \\
&\overset{(e)}{\leq} \max_{p(\mathbf{x})p(\bar{\mathbf{y}}_R|\mathbf{x})} [H(\mathbf{X}) - 2F_{T_{XY}}(\mathbf{q}, (H(\bar{\mathbf{Y}}_R) - R_1)/2)],
\end{aligned}$$

where (a) follows from the independence of Z from $\mathbf{X}$ and $\bar{\mathbf{Y}}_R$; (b) follows since optimizing over W can only increase the value compared to optimizing over $(Z, \hat{Y}_R)$; (c) follows from the definition of the conditional entropy bound in (57); (d) follows from the nondecreasing monotonicity of $F_{T_{XY}^{(2)}}(\mathbf{q}^{(2)}, s)$ in s; and (e) follows from Lemma 10, and $\mathbf{q} \triangleq [q, 1-q] = \frac{1}{2}(\mathbf{q}_1 + \mathbf{q}_2)$ is the average distribution of $\mathbf{Y}$.

Now, we lower bound $H(\bar{\mathbf{Y}}_R)$. Since conditioning reduces entropy, we have $H(\bar{\mathbf{Y}}_R) \geq H(\bar{\mathbf{Y}}_R|N_1N_2) = H(\mathbf{X})$, and then we can lower bound $H(\bar{\mathbf{Y}}_R)$ as follows:

$$\max\{H(\mathbf{X}), R_1\} \leq H(\bar{\mathbf{Y}}_R) \leq 2. \tag{68}$$

Let $\nu \triangleq (\max\{H(\mathbf{X}), R_1\} - R_1)/2$. Then, we have

$$\begin{aligned}
&R_{\mathrm{CF}} \qquad (69)\\
&\overset{(a)}{\leq} \max_{p(\mathbf{x})p(\bar{\mathbf{y}}_R|\mathbf{x})} [H(\mathbf{X}) - 2F_{T_{XY}}(\mathbf{q}, \nu)], \\
&\overset{(b)}{=} \max_{p(\mathbf{x})p(\bar{\mathbf{y}}_R|\mathbf{x})} [H(\mathbf{X}) - 2h_2(\delta \star h_2^{-1}(\nu))] \\
&\quad \text{s.t. } 0 \leq \nu \leq h_2(q) \\
&\overset{(c)}{\leq} \max_{p(\mathbf{x})} [H(\mathbf{X}) - 2h_2(\delta \star h_2^{-1}(\nu))] \\
&\quad \text{s.t. } R_1 \leq \max\{H(\mathbf{X}), R_1\} \leq 2 + R_1 \\
&\overset{(d)}{=} \max_{p(\mathbf{x})} [H(\mathbf{X}) - 2h_2(\delta \star h_2^{-1}(\max\{H(\mathbf{X}), R_1\} - R_1)/2))] \\
&\quad \text{s.t. } \max\{H(\mathbf{X}), R_1\} \leq 2 \\
&\overset{(e)}{=} \max_{0\leq\alpha\leq 1} [2\alpha - 2h_2(\delta \star h_2^{-1}((\max\{2\alpha, R_1\} - R_1)/2))] \\
&\quad \text{s.t. } \max\{R_1, 2\alpha\} \leq 2,
\end{aligned}$$

where (a) follows from (68) and $F_{T_{XY}}(\mathbf{q}, s)$ being nondecreasing in s; equality (b) follows from the definition of $F_{T_{XY}}(q, s)$ for the binary symmetric channel; (c) follows since $h_2(q) \leq 1$, and we are enlarging the optimization domain; (d) follows since there is no loss in generality by reducing the optimization set, since $\max\{H(\mathbf{X}), R_1\} \geq R_1$ and from (68), any $(\mathbf{X}, \bar{\mathbf{Y}}_R)$ following $p(\mathbf{x}, \bar{\mathbf{y}}_R)$ satisfy $\max\{H(\mathbf{X}), R_1\} \leq 2$; and (e) follows from defining $H(\mathbf{X}) \triangleq 2\alpha$, for $0 \leq \alpha \leq 1$.

Then, for $2\alpha \leq R_1$, we have

$$R_{\mathrm{CF}} \leq \max_{0\leq\alpha\leq R_1/2} [2\alpha - 2h_2(\delta)] = R_1 - 2h_2(\delta), \tag{70}$$

and for $2\alpha > R_1$, we have

$$R_{\mathrm{CF}} \leq \max_{R_1/2<\alpha\leq 1} [2\alpha - 2h_2(\delta \star h_2^{-1}(\alpha - R_1/2))]. \tag{71}$$

Now, we solve (71). Let us define $f(u) \triangleq h_2(\delta \star h_2^{-1}(u))$ for $0 \leq u \leq 1$. Then, we have the following lemma from [26].

Lemma 12 ([26]). *Function $f(u)$ is convex for $0 \leq u \leq 1$.*

We define $g(\alpha) \triangleq \alpha - h_2(\delta \star h_2^{-1}(\alpha - R_1/2))$, such that $R_{\mathrm{CF}} \leq \max_{R_1/2<\alpha\leq 1} 2g(\alpha)$. We have that $g(\alpha)$ is concave in α, since is a shifted version by α, which is linear, of the composition of the concave function $-f(u)$ and the affine function $\alpha - R_1/2$.

Proposition 1. *$g(\alpha)$ is monotonically increasing for $R_1/2 \leq \alpha \leq 1 + R_1/2$.*

Proof. Using the chain rule for composite functions, we have

$$\frac{d^2 g(\alpha)}{d\alpha^2} = -f''(\alpha - R_1/2), \tag{72}$$

where $f''(u) \triangleq d^2 f/du^2(u)$.

Since $g(\alpha)$ is convex, and is defined over a convex region, it follows that its unique maximum is achieved either for $f''(\alpha - R_1/2) = 0$, or at the boundaries of the region. It is shown in [26, Lemma 2] that $f''(u) > 0$ for $0 < u < 1$. Therefore, the maximum is achieved either at $u = 0$ or at $u = 1$, or equivalently, for $\alpha = R_1/2$ or $\alpha = 1 + R_1/2$. Since $g(R_1/2) = R_1/2 - h_2(\delta)$ and $g(1 + R_1/2) = R_1/2$, i.e., $g(R_1/2) < g(1 + R_1/2)$, it follows that $g(\alpha)$ is monotonically increasing in α for $R_1/2 \leq \alpha \leq 1 + R_1/2$. □

From Proposition 1 if follows that for $R_1/2 \leq \alpha \leq 1$, $g(\alpha)$ achieves its maximum at $\alpha = 1$. Then, for $2\alpha > R_1$, we have

$$R_{\mathrm{CF}} \leq 2(1 - h_2(\delta \star h_2^{-1}(1 - R_1/2))). \tag{73}$$

Thus, from (70) and (73), for $R_1 \leq H(\bar{\mathbf{Y}}_R)$ we have

$$\begin{aligned}
R_{\mathrm{CF}} &\leq 2\max\{R_1/2 - h_2(\delta), 1 - h_2(\delta \star h_2^{-1}(1 - R_1/2))\} \\
&= 2(1 - h_2(\delta \star h_2^{-1}(1 - R_1/2))),
\end{aligned} \tag{74}$$

where the equality follows from Proposition 1 by noting that the first element in the maximum coincides with $g(R_1/2) = R_1/2 - h_2(\delta)$, and the second one coincides with $g(1)$.

Finally, R_{CF} is upper bounded by the maximum over the joint distributions satisfying $H(\bar{\mathbf{Y}}_R) \leq R_1$ given in (67), and the upper bound for the joint distributions satisfying $R_1 \leq H(\bar{\mathbf{Y}}_R)$ given in (74). Since (67) coincides with $g(R_1/2)$, (74) serves as an upper bound on R_{CF} when $R_1 \leq H(\bar{\mathbf{Y}}_R)$.

Next, we show that the upper bound in (74) is achievable by considering the following variables

$$\begin{aligned}
&X_1^1 \sim \mathrm{Ber}(1/2), \quad X_1^2 \sim \mathrm{Ber}(1/2), \quad \hat{Y}_R = (\hat{Y}_R^1, \hat{Y}_R^2) \\
&\hat{Y}_R^1 = Y_R^1 \oplus Q_1,\ Q_1 \sim \mathrm{Ber}(h_2^{-1}(1 - R_1/2)). \\
&\hat{Y}_R^2 = Y_R^2 \oplus Q_2,\ Q_2 \sim \mathrm{Ber}(h_2^{-1}(1 - R_1/2)).
\end{aligned}$$

Let $Q_i \sim \text{Ber}(\nu)$ for $i = 1, 2$. Then from the constraint in (17) we have

$$\begin{aligned} & I(Y_R^1, Y_R^2; \hat{Y}_R|Z) \\ &= H(\hat{Y}_R|Z) - H(\hat{Y}_R|Y_R^1 Y_R^2 Z) \\ &= H(X_1^1 \oplus N_1 \oplus Q_1, X_1^2 \oplus N_2 \oplus Q_2) - H(Q_1, Q_2) \\ &\overset{(a)}{=} 2 - 2h_2(\nu), \end{aligned}$$

where (a) follows since $X_1^i \sim \text{Ber}(1/2)$, $i = 1, 2$ and from the independence of Q_1 and Q_2. We have $2h_2(\nu) \geq 2 - R_1$, and thus, $\nu \geq h_2^{-1}(1 - R_1/2)$.

Then, the achievable rate in (17) is given by

$$\begin{aligned} I(\mathbf{X}; \hat{Y}_R|Z) &= H(\hat{Y}_R|Z) - H(\hat{Y}_R|\mathbf{X}Z) \\ &= H(X_1^1 \oplus N_1 \oplus Q_1, X_1^2 \oplus N_2 \oplus Q_2) \\ &\quad - H(N_1 \oplus Q_1, N_2 \oplus Q_2) \\ &= 2 - 2h(\delta \star \nu) \\ &\leq 2 - 2h_2(\delta \star h_2^{-1}(1 - R_1/2)), \end{aligned}$$

where the last inequality follows from the bound on ν. This completes the proof.

Appendix E
Proof of Lemma 5

From (7), the achievable rate for the proposed pDCF scheme is given by

$$\begin{aligned} R_{\text{pDCF}} &= I(X_1^1; Y_R^1) + I(X_1^2; \hat{Y}_R|Z) \\ &\text{s.t. } R_1 \geq I(X_1^1; Y_R^1) + I(Y_R^2; \hat{Y}_R|Z). \end{aligned}$$

First, we note that the constraint is always satisfied for the choice of variables:

$$\begin{aligned} & I(X_1^1; Y_R^1) + I(Y_R^2; \hat{Y}_R|Z) \\ &= H(Y_R^1) - H(N_1) + H(X_1^2 \oplus N_2 \oplus Q) - H(Q) \\ &= 1 - h_2(\delta) + 1 - h_2(h_2^{-1}(2 - h(\delta) - R_1)) \\ &= R_1, \end{aligned} \tag{75}$$

where $H(Y_R^1) = 1$ since $X_1^1 \sim \text{Ber}(1/2)$ and $H(X_1^2 \oplus N_2 \oplus Q) = 1$ since $X_1^2 \sim \text{Ber}(1/2)$. Then, similarly the achievable rate is given by

$$\begin{aligned} R_{\text{pDCF}} &= I(X_1^1; Y_R^1) + I(X_1^2; \hat{Y}_R|Z) \\ &= H(Y_R^1) - H(N_1) + H(X_1^2 \oplus N_2 \oplus Q) - H(V \oplus Q) \\ &= 1 - h_2(\delta) + 1 - h_2(\delta \star h_2^{-1}(2 - h(\delta) - R_1)), \end{aligned}$$

which completes the proof.

Appendix F
Proof of Lemma 8

By evaluating (7) with the considered Gaussian random variables, we get

$$\begin{aligned} R &= \frac{1}{2} \log \left(1 + \frac{\alpha P}{\bar{\alpha} P + 1}\right) \left(1 + \frac{\bar{\alpha} P}{(1 - \rho^2) + \sigma_q^2}\right) \\ &\text{s.t. } R_1 \geq \frac{1}{2} \log \left(1 + \frac{\alpha P}{\bar{\alpha} P + 1}\right) \left(1 + \frac{\bar{\alpha} P + (1 - \rho^2)}{\sigma_q^2}\right). \end{aligned}$$

We can rewrite the constraint on R_1 as,

$$\sigma_q^2 \geq f(\alpha) \triangleq \frac{(P + 1)(\bar{\alpha} P + 1 - \rho^2)}{2^{2R_1}(\bar{\alpha} P + 1) - (P + 1)}. \tag{76}$$

Since R is increasing in σ_q^2, it is clear that the optimal σ_q^2 is obtained by $\sigma_q^2 = f(\alpha)$, where α is chosen such that $f(\alpha) \geq 0$. It is easy to check that $f(\alpha) \geq 0$ for

$$\alpha \in \left[0, \min\left\{(1 - 2^{-2R_1})\left(1 + \frac{1}{P}\right), 1\right\}\right]. \tag{77}$$

Now, we substitute $\sigma_q^2 = f(\alpha)$ in (76), and write the achievable rate as a function of α as

$$R(\alpha) = \frac{1}{2} \log G(\alpha), \tag{78}$$

where

$$\begin{aligned} G(\alpha) &\triangleq \left(1 + \frac{\alpha P}{\bar{\alpha} P + 1}\right) \left(1 + \frac{\bar{\alpha} P}{(1 - \rho^2) + f(\alpha)}\right) \\ &= \frac{2^{2R_1}(1 + P)(1 - \rho^2 + \bar{\alpha} P)}{(1 - \rho^2) 2^{2R_1}(1 + \bar{\alpha} P) + \bar{\alpha} P(1 + P)}. \end{aligned} \tag{79}$$

We take the derivative of $G(\alpha)$ with respect to α:

$$G'(\alpha) \triangleq \frac{2^{2R_1} P(1 + P)\left(1 - \rho^2\right)\left(P + 1 - 2^{2R_1}\rho^2\right)}{\left[P(1 + P)\bar{\alpha} + 2^{2R_1}(1 + \bar{\alpha} P)\left(1 - \rho^2\right)\right]^2}.$$

We note that if $\rho^2 \geq 2^{-R_1}(P+1)$, then $G'(\alpha) < 0$, and hence, $G(\alpha)$ is monotonically decreasing. Achievable rate R is maximized by setting $\alpha^* = 0$. When $\rho^2 < 2^{-R_1}(P + 1)$, we have $G'(\alpha) > 0$, and hence $\alpha^* = \min\left\{(1 - 2^{-R_1})\left(1 + \frac{1}{P}\right), 1\right\} = 1$, since we have $(1 - 2^{-R_1})\left(1 + \frac{1}{P}\right) \geq (1 + \frac{1 - \rho^2}{P}) > 1$.

Appendix G
Proof of Lemma 9

In order to characterize the capacity of the binary symmetric MRC-D, we find the optimal distribution of $(U, X_1, \hat{Y}_R)$ in Theorem 1 for $Z \sim \text{Ber}(1/2)$. First, we note that U is independent of Y_R since

$$I(U; Y_R) \leq I(X_1; Y_R) = 0, \tag{80}$$

where the inequality follows from the Markov chain $U - X_1 - Y_R$, and the equality follows since for $Z \sim \text{Ber}(1/2)$ the channel output of the binary channel $Y_R = X_1 \oplus N \oplus Z$ is independent of the channel input X_1 [6]. Then, the capacity region in (7) is given by

$$\mathcal{C} = \sup\{I(X_1; \hat{Y}_R|UZ) : R_1 \geq I(Y_R; \hat{Y}_R|UZ)\},$$

where the supremum is taken over the set of pmf's in the form

$$p(u, x_1)p(z)p(y_R|x_1, z)p(\hat{y}_R|y_R, u).$$

Let us define $\bar{Y} \triangleq X_1 \oplus N$. The capacity is equivalent to

$$\mathcal{C} = \sup\{I(X_1; \hat{Y}_R|UZ) : H(\bar{Y}|\hat{Y}UZ) \geq H(\bar{Y}|U) - R_1\},$$

over the joint pmf's of the form

$$p(u, x_1)p(z)p(\bar{y}|x_1)p(\hat{y}_R|\bar{y}, u, z), \tag{81}$$

where we have used the fact that $\bar{Y}$ is independent from Z.

For any joint distribution for which $0 \leq H(\bar{Y}|U) \leq R_1$, the constraint in (81) is also satisfied. It follows from the Markov chain $X_1 - \bar{Y} - \hat{Y}_R$ given U, Z, and the data processing inequality, that

$$\begin{aligned}\mathcal{C} &\leq \max_{p(u,x_1)} \{I(X_1;\bar{Y}|ZU) : H(\bar{Y}|U) \leq R_1\} \qquad (82)\\ &= \max_{p(u,x_1)} \{H(\bar{Y}|U) - h_2(\delta) : H(\bar{Y}|U) \leq R_1\}\\ &\leq R_1 - h_2(\delta).\end{aligned}$$

We next consider the joint distributions for which $R_1 \leq H(\bar{Y}|U)$. Let $p(u) = \Pr[U = u]$ for $u = 1, .., |\mathcal{U}|$, and we can write

$$I(X_1;\hat{Y}_R|UZ) = H(X_1|U) - \sum_u p(u) H(X_1|\hat{Y}_R Z u), \qquad (83)$$

and

$$\begin{aligned}I(Y_R;\hat{Y}_R|UZ) &\overset{(a)}{=} I(\bar{Y};\hat{Y}_R|UZ)\\ &\overset{(b)}{=} H(\bar{Y}|U) - \sum_u p(u) H(\bar{Y}|\hat{Y}_R Z u), \qquad (84)\end{aligned}$$

where (a) follows from the definition of $\bar{Y}$, and (b) follows from the independence of Z from $\bar{Y}$ and U.

For each u, the channel input X_1 corresponds to a binary random variable $X_u \sim \text{Ber}(\nu_u)$, where $\nu_u \triangleq \Pr[X_1 = 1|U = u] = p(1|u)$ for $u = 1, ..., |\mathcal{U}|$. The channel output for each X_u is given by $\bar{Y}_u = X_u \oplus N$. We denote by $q_u \triangleq \Pr[Y_u = 1] = \Pr[Y_R = 1|U = u]$. Similarly, we define $\hat{Y}_u$ as $\hat{Y}_R$ for each u value. Note that for each u, $X_u - \bar{Y}_u - \hat{Y}_u$ form a Markov chain.

Then, we have $H(X_1|u) = h_2(\nu_u)$ and $H(\bar{Y}|u) = h_2(\delta \star \nu_u)$. We define $s_u \triangleq H(\bar{Y}|\hat{Y}_R Z u)$, such that $0 \leq s_u \leq H(\bar{Y}_u)$. Substituting (83) and (84) in (81) we have

$$\begin{aligned}\mathcal{C} = &\max_{p(u,x_1)p(\hat{y}_R|y_R,u)} [H(X_1|U) - \sum_u p(u) H(X_1|\hat{Y}_R Z u)]\\ &\text{s.t. } R_1 \geq H(\bar{Y}|U) - \sum_u p(u) H(\bar{Y}|\hat{Y}_R Z u)\\ \overset{(a)}{=} &\max_{p(u,x_1),} [H(X_1|U) - \sum_u p(u) F_{T_{XY}}(q_u, s_u)]\\ &\text{s.t. } R_1 \geq H(\bar{Y}|U) - \sum_u p(u) s_u,\ 0 \leq s_u \leq H(\bar{Y}_u)\\ \overset{(b)}{=} &\max_{p(u,x_1)} [H(X_1|U) - \sum_u p(u) h_2(\delta \star h_2^{-1}(s_u))]\\ &\text{s.t. } R_1 \geq H(\bar{Y}|U) - \sum_u p(u) s_u,\ 0 \leq s_u \leq H(\bar{Y}_u),\\ \overset{(c)}{\leq} &\max_{p(u,x_1)} H(X_1|U) - h_2\left(\delta \star h_2^{-1}\left(\sum_u p(u) s_u\right)\right)\\ &\text{s.t. } \sum_u p(u) s_u \geq H(\bar{Y}|U) - R_1,\end{aligned}$$

where (a) follows from the definition of $F_{T_{XY}}(q,s)$ for channel $\bar{Y}_u = X_u \oplus N$, which for each u has a matrix T_{XY} as in (65), (b) follows from the expression of $F_{T_{XY}}(q,s)$ for the binary channel T_{XY} in Lemma 11, (c) follows from noting that $-h_2(\delta \star h_2^{-1}(s_u))$ is concave on s_u from Lemma 12 and applying Jensen's inequality. We also drop the conditions on s_u, which can only increase $\mathcal{C}$.

Then, similarly to the proof of Lemma 4, we have $H(\bar{Y}|U) \geq H(\bar{Y}|UV) = H(X_1|U)$, and we can upper bound the capacity as follows

$$\begin{aligned}\mathcal{C} \leq &\max_{p(x_1,u)} \left[H(X_1|U) - h_2\left(\delta \star h_2^{-1}\left(\sum_u p(u) s_u\right)\right)\right]\\ &\text{s.t. } \sum_u p(u) s_u \geq \max\{H(X_1|U), R_1\} - R_1\\ \leq &\max_{0\leq\alpha\leq 1} \alpha - h_2(\delta \star h_2^{-1}(\max\{\alpha, R_1\} - R_1)), \qquad (85)\end{aligned}$$

where we have defined $\alpha \triangleq H(X_1|U)$.

The optimization problem can be solved similarly to the proof in Appendix D as follows. If $0 \leq \alpha \leq R_1$, we have $\bar{s} \geq 0$ and

$$\mathcal{C} \leq \max_{0\leq\alpha\leq R_1} \alpha - h_2(\delta) = R_1 - h_2(\delta). \qquad (86)$$

For $R_1 \leq \alpha \leq 1$, we have

$$\mathcal{C} \leq \max_{R_1\leq\alpha\leq 1} \alpha - h_2(\delta \star h_2^{-1}(\alpha - R_1)). \qquad (87)$$

Then, it follows from a scaled version of Proposition 1 that the upper bound is maximized for $\alpha = 1$. Then, by noticing that (86) corresponds to the value of the bound in (87) for $\alpha = R_1$, it follows that

$$\mathcal{C} \leq 1 - h_2(\delta \star h_2^{-1}(1 - R_1)). \qquad (88)$$

This bound is achievable by CF. This completes the proof.

Appendix H
Proof of the Cut-Set Bound Optimality Conditions

Cases 1 and 2 are straightforward since under these assumptions, the ORC-D studied here becomes a particular case of the channel models in [27] and [5], respectively.

To prove Case 3 we use the following arguments. For any channel input distribution to the ORC-D, we have

$$\begin{aligned}I(X_1;Y_R|Z) &= H(X_1|Z) - H(X_1|Y_R, Z)\\ &\geq H(X_1) - H(X_1|Y_R) \qquad (89)\\ &= I(X_1;Y_R),\end{aligned}$$

where we have used the independence of X_1 and Z, and the fact that conditioning reduces entropy. Then, the condition $\max_{p(x_1)} I(X_1;Y_R) \geq R_1$, implies $\max_{p(x_1)} I(X_1;Y_R|Z) \geq R_1$; and hence, the cut-set bound is given by $R_{\text{CS}} = R_2 + R_1$, which is achievable by DF scheme.

In Case 4, the cut-set bound is given by $R_2 + \min\{R_1, I(\bar{X}_1;\bar{Y}_R|Z)\} = R_2 + I(\bar{X}_1;\bar{Y}_R|Z)$ since $R_1 \geq H(\bar{Y}_R|Z)$. CF achieves the capacity by letting X_1 be distributed with $\bar{p}(x_1)$, and choosing $\hat{Y}_R = \bar{Y}_R$. This choice is always possible as the CF constraint

$$R_1 \geq I(\hat{Y}_R;\bar{Y}_R|Z) = H(\bar{Y}_R|Z) - H(\bar{Y}_R|Z,\hat{Y}_R) = H(\bar{Y}_R|Z),$$

always holds. Then, the achievable rate for CF is $R_{\text{CF}} = R_2 + I(\bar{X}_1;\hat{Y}_R|Z) = R_2 + I(\bar{X}_1;\bar{Y}_R|Z)$, which is the capacity.

REFERENCES

[1] I. Estella-Aguerri and D. Gunduz, "Capacity of a class of relay channels with state," in *Proc. of 2012 IEEE Information Theory Workshop (ITW)*, Sep. 2012, pp. 277–281.

[2] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. on Information Theory*, vol. 25, no. 5, pp. 572 – 584, Sep. 1979.

[3] A. El Gamal and S. Zahedi, "Capacity of a class of relay channels with orthogonal components," *IEEE Trans. on Information Theory*, vol. 51, no. 5, pp. 1815 –1817, May 2005.

[4] A. Gamal and M. Aref, "The capacity of the semideterministic relay channel (corresp.)," *IEEE Trans. on Information Theory*, vol. 28, no. 3, p. 536, May 1982.

[5] T. M. Cover and Y.-H. Kim, "Capacity of a class of deterministic relay channels," in *Proc. of 2007 IEEE International Symposium on Information Theory (ISIT)*, Jun. 2007, pp. 591 –595.

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 1991.

[7] M. Aleksic, P. Razaghi, and W. Yu, "Capacity of a class of modulo-sum relay channels," *IEEE Trans. on Information Theory*, vol. 55, no. 3, pp. 921 –930, Mar. 2009.

[8] W. Kang and S. Ulukus, "Capacity of a class of diamond channels," *IEEE Trans. on Information Theory*, vol. 57, no. 8, pp. 4955 –4960, Aug. 2011.

[9] M. Khormuji and M. Skoglund, "On cooperative downlink transmission with frequency reuse," in *Proc. of 2009 IEEE International Symposium on Information Theory Proceedings (ISIT)*, Jun. 2009, pp. 849–853.

[10] A. Zaidi, S. Shamai, P. Piantanida, and L. Vandendorpe, "Bounds on the capacity of the relay channel with noncausal state information at source," in *Proc. of 2010 IEEE International Symposium on Information Theory Proceedings (ISIT)*, Jun. 2010, pp. 639 –643.

[11] B. Akhbari, M. Mirmohseni, and M.-R. Aref, "Achievable rate regions for relay channel with private messages with states known at the source," in *Proc. of 2010 IEEE International Conference on Information Theory and Information Security (ICITIS),*, Dec. 2010, pp. 1080–1087.

[12] A. Zaidi, S. Shamai, P. Piantanida, and L. Vandendorpe, "Bounds on the capacity of the relay channel with noncausal state at the source," *IEEE Trans. on Information Theory*, vol. 59, no. 5, pp. 2639–2672, May 2013.

[13] B. Akhbari, M. Mirmohseni, and M. Aref, "Compress-and-forward strategy for relay channel with causal and non-causal channel state information," *IET Communications*, vol. 4, no. 10, pp. 1174–1186, Jul. 2010.

[14] M. Mirmohseni, B. Akhbari, and M.-R. Aref, "Compress-and-forward strategy for the relay channel with causal state information," in *Proc. of 2009 IEEE Information Theory Workshop (ITW)*, Oct. 2009, pp. 426–430.

[15] M. Khormuji, A. El Gamal, and M. Skoglund, "State-dependent relay channel: Achievable rate and capacity of a semideterministic class," *IEEE Trans. on Information Theory*, vol. 59, no. 5, pp. 2629–2638, May 2013.

[16] R. Kolte, A. Özgür, and H. H. Permuter, "Cooperative binning for semideterministic channels," *CoRR*, vol. abs/1508.05149, 2015. [Online]. Available: http://arxiv.org/abs/1508.05149

[17] O. Simeone, D. Gunduz, and S. Shamai, "Compound relay channel with informed relay and destination," in *Proc. of 47th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2009*, Oct. 2009, pp. 692 –699.

[18] A. Behboodi and P. Piantanida, "On the simultaneous relay channel with informed receivers," in *Proc. of 2009 IEEE International Symposium on Information Theory (ISIT)*, Jul. 2009, pp. 1179 –1183.

[19] K. Bakanoglu, E. Erkip, O. Simeone, and S. Shitz, "Relay channel with orthogonal components and structured interference known at the source," *IEEE Transactions on Comm.*, vol. 61, no. 4, pp. 1277–1289, Apr. 2013.

[20] K. Bakanoglu, E. Erkip, and O. Simeone, "Half-duplex Gaussian diamond relay channel with interference known at one relay," in *Proc. 2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Nov. 2011, pp. 1353 –1357.

[21] M. Li, O. Simeone, and A. Yener, "Message and state cooperation in a relay channel when only the relay knows the state," *CoRR*, vol. abs/1102.0768, 2011. [Online]. Available: http://arxiv.org/abs/1102.0768

[22] A. Zaidi, S. Kotagiri, J. Laneman, and L. Vandendorpe, "Cooperative relaying with state available noncausally at the relay," *IEEE Trans. on Information Theory*, vol. 56, no. 5, pp. 2272–2298, May 2010.

[23] Y.-H. Kim, "Coding techniques for primitive relay channels," in *Proc. of Allerton Conf. Commun., Control, Comput., Monticello, IL*, Sep. 2007.

[24] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[25] H. S. Witsenhausen and A. D. Wyner, "A conditional entropy bound for a pair of discrete random variables," *IEEE Trans. on Information Theory*, vol. 21, no. 5, pp. 493–501, Sep. 1975.

[26] A. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications–i," *IEEE Trans. on Information Theory*, vol. 19, no. 6, pp. 769–772, Nov. 1973.

[27] A. El Gamal and S. Zahedi, "Capacity of a class of relay channels with orthogonal components," *IEEE Trans. on Information Theory*, vol. 51, no. 5, pp. 1815–1817, May 2005.

Iñaki Estella Aguerri Iñaki Estella Aguerri received his B.Sc. and M.Sc. degrees in Telecommunication engineering from Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 2008 and 2011, respectively. He received his Ph.D. degree from Imperial College London, London, UK, in January 2015. In November 2014, he joined the Mathematical and Algorithmic Sciences Lab, France Research Center, Huawei Technologies Co. Ltd., Boulogne-Billancourt, France. From 2008 to 2013, he was research assistant at Centre Tecnolgic de Telecomunicacions de Catalunya (CTTC) in Barcelona. His primary research interests are in the areas of information theory and communication theory, with special emphasis on joint source-channel coding and cooperative communications for wireless networks.

Deniz Gunduz Deniz Gunduz [S'03-M'08-SM'13] received the B.S. degree in electrical and electronics engineering from METU, Turkey in 2002, and the M.S. and Ph.D. degrees in electrical engineering from NYU Polytechnic School of Engineering in 2004 and 2007, respectively. After his PhD he served as a postdoctoral research associate at the Department of Electrical Engineering, Princeton University, and as a consulting assistant professor at the Department of Electrical Engineering, Stanford University. He also served as a research associate at CTTC in Barcelona, Spain. Since September 2012 he is a Lecturer in the Electrical and Electronic Engineering Department of Imperial College London, UK. He also held a visiting researcher position at Princeton University from November 2009 until November 2011.

Dr. Gunduz is an Associate Editor of the IEEE Transactions on Communications, and an Editor of the IEEE Journal on Selected Areas in Communications (JSAC) Series on Green Communications and Networking. He is the recipient of a Starting Grant of the European Research Council, the 2014 IEEE Communications Society Best Young Researcher Award for the Europe, Middle East, and Africa Region, and the Best Student Paper Award at the 2007 IEEE International Symposium on Information Theory (ISIT). He is serving as a co-chair of the IEEE Information Theory Society Student Committee, and the co-director of the Imperial College Probability Center. He is the General Co-chair of the 2016 IEEE Information Theory Workshop. He served as a Technical Program Chair of the Network Theory Symposium at the 2013 and 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP), and General Co-chair of the 2012 IEEE European School of Information Theory (ESIT). His research interests lie in the areas of communication theory and information theory with special emphasis on joint source-channel coding, multi-user networks, energy efficient communications and privacy in cyber-physical systems.