Research Article

Lower Bounds on the Capacity of the Relay Channel with States at the Source

Abdellatif Zaidi and Luc Vandendorpe

École Polytechnique de Louvain, Université Catholique de Louvain, Louvain La Neuve 1348, Belgium

Correspondence should be addressed to Abdellatif Zaidi, abdellatif.zaidi@uclouvain.be

We consider a state-dependent three-terminal full-duplex relay channel with the channel states noncausally available at only the source, that is, neither at the relay nor at the destination. This model has application to cooperation over certain wireless channels with asymmetric cognition capabilities and cognitive interference relay channels. We establish lower bounds on the channel capacity for both discrete memoryless (DM) and Gaussian cases. For the DM case, the coding scheme for the lower bound uses techniques of rate-splitting at the source, decode-and-forward (DF) relaying, and a Gel'fand-Pinsker-like binning scheme. In this coding scheme, the relay decodes only partially the information sent by the source. Due to the rate-splitting, this lower bound is better than the one obtained by assuming that the relay decodes all the information from the source, that is, full-DF. For the Gaussian case, we consider channel models in which each of the relay node and the destination node experiences on its link an additive Gaussian outside interference. We first focus on the case in which the links to the relay and to the destination are corrupted by the *same* interference; and then we focus on the case of *independent* interferences. We also discuss a model with *correlated* interferences. For each of the first two models, we establish a lower bound on the channel capacity. The coding schemes for the lower bounds use techniques of dirty paper coding or carbon copying onto dirty paper, interference reduction at the source and decode-and-forward relaying. The results reveal that, by opposition to carbon copying onto dirty paper and its root Costa's initial dirty paper coding (DPC), it may be beneficial in our setup that the informed source uses a part of its power to partially cancel the effect of the interference so that the uninformed relay benefits from this cancellation, and so the source benefits in turn.

1. Introduction

In this work, we consider a state-dependent three-terminal full-duplex relay channel (RC) in which the outputs Y_2 at the relay and Y_3 at the destination are controlled by the channel inputs X_1 from the source and X_2 from the relay, along with a random parameter S that represents the channel state, through a given conditional probability $W_{Y_2,Y_3|X_1,X_2,S}$. The channel state S is generated according to a given memoryless probability law Q_S, and is known, in a noncausal manner, to only the source—for instance, the relay and the destination do not know the channel states. The considered channel model is shown in Figure 1. In this model, the source wishes to transmit a message W to the destination through the state-dependent RC in n channel uses, with the help of the relay. The destination estimates the message sent by the source from the received channel output. In this work, we study

the capacity of this communication model. We refer to this model as *RC with informed source*.

1.1. Background and Related Work. Channels whose probabilistic input-output relation depends on random parameters, or channel states, have spurred much interest and can model a large variety of problems, each related to some physical situation of interest. The random states sequence may be known in a *causal* or *noncausal* manner. For single user models, the concept of channel states available at only the transmitter dates back to Shannon [1] for the causal channel state case, and to Gel'fand and Pinsker [2] for the noncausal channel state case. In [3], Heegard and El Gamal study a model in which the state sequence is known noncausally to only the encoder or to only the decoder. They also derive achievable rates for the case in which partial channel state information (CSI) is given at varying rates
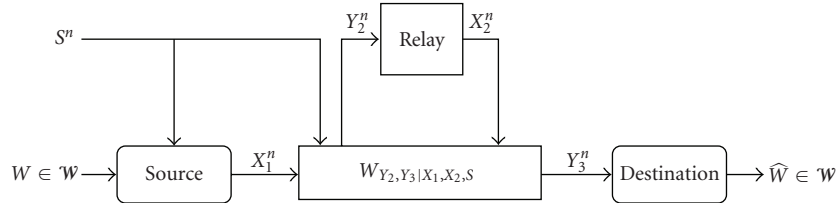
FIGURE 1: Relay channel with state information S^n available noncausally at only the source.

to both the encoder and the decoder. In [4], Costa studies an additive Gaussian channel with additive Gaussian state known at only the encoder, and shows that Gel'fand-Pinsker coding with a specific auxiliary random variable, widely known as *dirty paper coding* (DPC), achieves the channel capacity. Interestingly, in this case, the DPC removes the effect of the additive channel state on the capacity as if there were no channel state present in the model or the channel state were known to the decoder as well. For a comprehensive review of state-dependent channels and related work, the reader may refer to [5].

For multiuser channels, different state-dependent channel models under different setups are investigated in the literature. One key element in the study of state-dependent multiuser channels is whether the parameters controlling the channel are known to *all* or *only some* of the users in the communication model. For example, the broadcast channel (BC) with states available at the transmitter is studied in [6, 7], and the multiaccess channel (MAC) with partial state information at all encoders and full state information at the decoder is studied in [8].

In the Gaussian setup, the result on the property that a known additive state does not affect capacity as long as a full knowledge of this state is available at the transmitter, which was originally shown by Costa for a single-user channel, is shown to continue to hold for a number of multiuser channels, including the Gaussian BC [6], the Gaussian MAC [6], the physically degraded relay channel (RC) [9] and the physically degraded relay broadcast channel (RBC) [10]. For these channels, the key feature in enabling complete mitigation of the interference is the *symmetric* availability of the state at *all* the encoders. If the state is available at only some encoders (i.e., in the *asymmetric* case), complete mitigation of the interference is difficult to hope for and, in general, one has to expect some rate penalty due to the lack of knowledge of the state at the noninformed encoders. For example, the state-dependent multiaccess channel (MAC) with only one informed encoder is considered in [11–17] and the state-dependent relay channel (RC) with only informed relay is considered in [18, 19]. For all these models, in the Gaussian case, the informed encoder applies a slightly generalized DPC (GDPC) [11, 13] in which the channel input random variable and the channel state random variable are negatively correlated. Also, in these models, the uninformed encoders benefit from the GDPC applied by the informed encoders because the negative correlation between the codewords at the informed encoders and the channel state can be interpreted as partial state cancellation. For the state-dependent MAC with one informed encoder and one

noninformed encoder, and with the former sending both a common message and a private message and the latter sending only the common message, the capacity region for the Gaussian case is obtained by deriving a nontrivial outer bound that permits to characterize the rate loss due to not knowing the state at the noninformed encoder [15, 16]. Two upper bounds that use similar bounding techniques, but with different proofs, are derived in [17] for another case of degraded message sets for the state-dependent MAC with one informed encoder which is obtained by swapping the roles of the encoders in [15, 16], and in [18, 19] for the case of state-dependent RC with informed relay. The state-dependent MAC studied in [17] has some connection with the model for the RC with informed relay studied in [18, 19], when considering transmission from the source and the relay to the destination, that is, the multiaccess part of the relay channel.

1.2. Motivation. The model shown in Figure 1 may find application in cooperative information embedding and data hiding [20–24], fading in time-varying cooperative wireless channels and interfering signals in interference environments. The channel state may also model the information gained by some specific terminals by means of cognition in certain cognitive systems [25–27]. In this application, one common assumption, made for example, in [25, 26], is that some transmitters can know the signals that are sent by some other transmitters noncausally; and, so, they can then use that knowledge to increase the system spectral efficiency. The problem of collaborative signal transmission in presence of interference and with some cognitive terminals is studied, for example, in [28] for an interference channel with a cognitive transmitter or degraded message sets, and in [29] for a state-dependent cognitive interference channel. Interference relay channels and relay strategies for relaying in presence of an outside interference are also studied, for example, in [30–32]. In the model we study, the source may model a cognitive radio that cooperates with the relay, as shown in Figure 2.

1.3. Main Contributions. For the discrete-memoryless (DM) case, we derive a lower bound on the capacity of the general state-dependent RC with informed source. This lower bound is obtained by a coding scheme that uses techniques of rate-splitting at the source, regular encoding sliding-window decoding [33] for decode-and-forward (DF) relaying [34, Theorem 4] and a Gel'fand-Pinsker like binning scheme. In this coding scheme, the relay decodes only partially the information sent by the source, that is, *Partial*-DF [35]. By
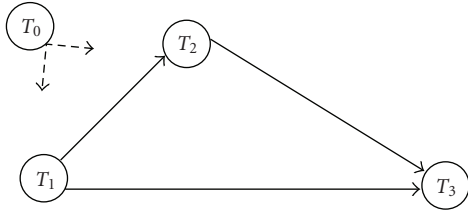
FIGURE 2: Example relay channel with asymmetric cognition capabilities. Only the source T_0 knows the interference from the competing source T_0. The source helps the relay T_2 to cancel the effect of the interference.

specializing this achievability result to the case in which the relay decodes all the information sent by the source, we readily obtain a lower bound for the case in which the relay operates in *Full*-DF [34].

Furthermore, we also consider memoryless Gaussian models in which each of the relay node and the destination node experiences on its link an additive Gaussian outside interference in addition to additive Gaussian noise. The interferences are known noncausally to only the source, and play the role of additive channel states. We first focus on the model in which the links to the relay and to the destination are corrupted by the *same* interference. For this model we derive a lower bound on the channel capacity, based on a coding scheme that combines techniques of generalized DPC [11, 13] and partial decode-and-forward relaying. Then, we focus on the case of *independent* interferences. For this case, we obtain a lower bound on the channel capacity by using a coding scheme that combines carefully the techniques of carbon copying onto dirty paper (CC) [36], interference reduction at the source, and decode-and-forward relaying. We also discuss a case of *correlated* interferences.

For the Gaussian model with independent interferences studied in this paper, we mention that a major difference from the Gaussian multicast problem studied in [36] is that, here, one of the two receivers (the relay) also helps the other receiver (the destination) by relaying the information intended to it from the source. Also, the carbon copying onto dirty paper technique that we employ as part of our coding scheme differs from the CC in [36] in that (i) the codeword sent by the informed transmitter is (negatively) *correlated* with the channel state, and (ii) the noise terms at the two receivers have different variances. We refer to the employed CC with negative correlation between the codewords at the source and the channel state as a *generalized* carbon copying onto dirty paper (GCC). We show that our lower bound is larger than the one that is obtained by a similar combination of regular CC and decode-and-forward relaying. The allowed improvement is due to that the proposed GCC improves upon regular CC through the allowed negative correlation, just in the same way GDPC improves upon regular DPC.

Finally, we mention that in this work the established lower bounds are compared to the well-known *max-flow min-cut bound* or *cut-set bound*. In Section 3.2, we will comment onto why deriving nontrivial upper bounds on the channel capacity for the present scenario is a challenging task, and is more involved than for the seemingly similar

informed relay scenario [18, 19] or the related MAC with asymmetric CSI and degraded message sets [12, 16, 17].

1.4. Outline and Notation. An outline of the remainder of this paper is as follows. Section 2 describes the communication model that we consider in this work. Section 3 provides lower bounds on the capacity of the general DM RC with informed source. Section 4 provides lower bounds on the capacity of the Gaussian RC with informed source. This section also contains some discussions as well as some numerical results for illustration purposes. Finally, Section 5 concludes the paper.

We use the following notations throughout the paper. Upper case letters, for example, X, are used to denote random variables; lower case letters, that is, x, are used to denote realizations of random variables; and calligraphic letters, that is, $\mathcal{X}$, designate alphabet sets. The probability distribution of a random variable X is denoted by $P_X(x)$. Sometimes, for convenience, we write it as P_X. The shorthand notation X_i^j indicates a sequence of random variables $(X_i, X_{i+1}, \ldots, X_j)$ and x_i^j denotes a particular realization of a random sequence X_i^j. For convenience the n-vector x^n will occasionally be denoted by the boldface notation $\mathbf{x}$ as well. We use the notation $\mathbb{E}_X[\cdot]$ to denote the expectation of random variable X. The set of probability distributions defined on an alphabet $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$. A probability distribution of a random variable Y given X is denoted by $P_{Y|X}$. The Gaussian distribution with mean μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$. Finally, throughout the paper, the logarithm function is to base 2, and the complement to unity of a scalar $u \in [0, 1]$ is denoted by $\bar{u}$, that is, $\bar{u} = 1 - u$.

2. System Model and Definitions

In this section, we formally present our communication model and the definitions related to it. As shown in Figure 1, we consider a state-dependent relay channel denoted by $W_{Y_2, Y_3 | X_1, X_2, S}$ whose outputs $Y_2 \in \mathcal{Y}_2$ and $Y_3 \in \mathcal{Y}_3$ for the relay and the destination, respectively, are controlled by the channel inputs $X_1 \in \mathcal{X}_1$ from the source and $X_2 \in \mathcal{X}_2$ from the relay, along with a random state parameter $S \in \mathcal{S}$. It is assumed that the channel state S_i at time instant i is independently drawn from a given distribution Q_S and the channel states S^n are noncausally known at the source.

The source wants to transmit a message W to the destination with the help of the relay, in n channel uses. The message W is assumed to be uniformly distributed over the set $\mathcal{W} = \{1, \ldots, M\}$. The information rate R is defined as $\log M/n$ bits per transmission.

An (M, n) code for the state-dependent relay channel with informed source consists of an encoding function at the source

$$\phi_1^n : \{1, \ldots, M\} \times \mathcal{S}^n \longrightarrow \mathcal{X}_1^n, \qquad (1)$$

a sequence of encoding functions at the relay

$$\phi_{2,i} : \mathcal{Y}_{2,1}^{i-1} \longrightarrow \mathcal{X}_2, \qquad (2)$$

for $i = 1, 2, \ldots, n$, and a decoding function at the destination

$$\psi^n : \mathcal{Y}_3^n \longrightarrow \{1, \ldots, M\}. \tag{3}$$

Let an (M, n) code be given. The sequences X_1^n and X_2^n from the source and the relay, respectively, are transmitted across a state-dependent relay channel modeled as a *memoryless* and *time invariant* channel in the sense that

$$
p\left(y_{2,i}, y_{3,i} \mid x_1^i, x_2^i, s^n, y_2^{i-1}, y_3^{i-1}\right)
$$
$$
= W_{Y_2,Y_3|X_1,X_2,S}\left(y_{2,i}, y_{3,i} \mid x_{1,i}, x_{2,i}, s_i\right) \tag{4}
$$

for all $i = 1, \ldots, n$.

The destination estimates the message sent by the source from the channel output Y_3^n. The average probability of error is defined as $P_e^n = \Pr[\psi^n(Y_3^n) \neq W]$.

An (ϵ, n, R) code for the state-dependent RC with informed source is an $(2^{nR}, n)$—code $(\phi_1^n, \phi_2^n, \psi^n)$ having average probability of error P_e^n not exceeding ϵ.

A rate R is said to be achievable if there exists a sequence of (ϵ_n, n, R)—codes with $\lim_{n \to \infty} \epsilon_n = 0$. The capacity $\mathcal{C}$ of the state-dependent RC with informed source is defined as the supremum of the set of achievable rates.

The channel is said to be physically degraded if the conditional distribution $W_{Y_2,Y_3|X_1,X_2,S}$ factorizes as

$$W_{Y_2,Y_3|X_1,X_2,S} = W_{Y_2|X_1,X_2,S} W_{Y_3|Y_2,X_2,S}. \tag{5}$$

3. The Discrete Memoryless RC with Informed Source

In this section, we assume that the alphabets $\mathcal{S}$, $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{Y}_2$, $\mathcal{Y}_3$ in the model are all discrete and finite.

3.1. Lower Bound on Channel Capacity

Definition 1. For given distributions Q_S and $W_{Y_2,Y_3|X_1,X_2,S}$, let $\mathcal{P}$ be the set of joint distributions $P_{S,U,U_1,X_1,X_2,Y_2,Y_3}$ of the form

$$P_{S,U,U_1,X_1,X_2,Y_2,Y_3} = Q_S P_{X_2} P_{U|S,X_2} P_{U_1,X_1|U,S,X_2} W_{Y_2,Y_3|X_1,X_2,S} \tag{6}$$

that satisfy

$$0 \le I(U; Y_2 \mid X_2) - I(U; S \mid X_2), \tag{7a}$$

$$0 \le I(U_1; Y_3 \mid U, X_2) - I(U_1; S \mid U, X_2), \tag{7b}$$

$$0 \le I(U, U_1; Y_3 \mid X_2) - I(U, U_1; S \mid X_2), \tag{7c}$$

where $U \in \mathcal{U}$ and $U_1 \in \mathcal{U}_1$ are auxiliary random variables with

$$|\mathcal{U}| \le |\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2| + 2, \tag{8a}$$

$$|\mathcal{U}_1| \le (|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2| + 2)|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2| + 1, \tag{8b}$$

respectively.

The following theorem provides a lower bound on the capacity of the state-dependent DM RC with informed source.

Theorem 1. *The capacity of the state-dependent DM RC with informed source is lower-bounded by*

$$
R = \max \min \{ I(U; Y_2 \mid X_2) + I(U_1; Y_3 \mid U, X_2)
$$
$$
- I(U, U_1; S \mid X_2),
$$
$$
I(U; Y_2 \mid X_2) - I(U; S \mid X_2) \tag{9}
$$
$$
+ I(U, U_1; Y_3 \mid X_2) - I(U, U_1; S \mid X_2),
$$
$$
I(U, U_1, X_2; Y_3) - I(U, U_1; S \mid X_2) \},
$$

where the maximization is over all probability distributions $P_{S,U,U_1,X_1,X_2,Y_2,Y_3} \in \mathcal{P}$.

Proof of Theorem 1. A formal proof of Theorem 1 with complete error analysis is given in Appendix A. We now give a description of a random coding scheme which we use to obtain the lower bound given in Theorem 1. This coding scheme is based on a combination of the techniques of rate-splitting [37], regular-encoding sliding-window decoding for DF [33], and a variation of Gel'fand-Pinsker binning.

We split the message W to be transmitted into two independent parts, $W = (W_r, W_d)$, where W_r is sent through the relay at rate R_r and W_d is sent directly to the destination at rate R_d. The total rate is then $R = R_r + R_d$. Transmission is performed over $B + 1$ blocks, each of length n. During each of the first B blocks, the source encodes a message $w_i = (w_{r,i}, w_{d,i})$ and sends it over the channel, where $w_{r,i} \in [1, 2^{nR_r}]$, $w_{d,i} \in [1, 2^{nR_d}]$ and $i = 1, \ldots, B$ denotes the index of the block. During the last block, the source sends $w_{r,B+1} = 1$ and some $w_{d,B+1} \in [1, 2^{nR_d}]$. For fixed n, the average rate $R_d + R_r(B/(B + 1))$ over $B + 1$ blocks approaches R as $B \to +\infty$. $\square$

3.1.1. Codebook Generation. Fix a measure $P_{S,U,U_1,X_1,X_2,Y_2,Y_3}$ satisfying (6). Fix $\epsilon > 0$ and let

$$
J = 2^{n[I(U;S|X_2)+2\epsilon]} \qquad J_1 = 2^{n[I(U_1;S|U,X_2)+2\epsilon]}
$$
$$
M = 2^{n[R_r-4\epsilon]} \qquad M_1 = 2^{n[R_d-6\epsilon]}. \tag{10}
$$

We generate two statistically independent codebooks (codebooks 1 and 2) by following the steps outlined below twice. We will use these codebooks for blocks with odd and even indices, respectively.

(1) At the relay, we generate M independent and identically distributed (i.i.d.) codewords $\{\mathbf{x}_2(w_r')\}$ indexed by $w_r' = 1, \ldots, M$, each with i.i.d. components drawn according to P_{X_2}.

(2) For each $\mathbf{x}_2(w_r')$, we generate a collection of JM i.i.d. auxiliary codewords $\{\mathbf{u}(w_r', w_r, j)\}$ at the source, indexed by $w_r = 1, \ldots, M$, $j = 1, \ldots, J$. Each codeword $\mathbf{u}(w_r', w_r, j)$ is with i.i.d. components given $\mathbf{x}_2(w_r')$ drawn according to $P_{U|X_2}$. For notational convenience, we sometimes denote $\mathbf{u}(w_r', w_r, j)$ as $\mathbf{u}_{w_r', w_r, j}$. For the collection $\{\mathbf{u}_{w_r', w_r, j}\}$ of auxiliary codewords for given $\mathbf{x}_2(w_r')$, w_r indexes bins and j indexes sequences within a particular bin.

TABLE 1: Regular encoding for the state-dependent RC with informed source with rate-splitting. At the beginning of block $\mathbf{i}$, the relay sends $\mathbf{x}_2(\mathbf{w_{r,i-1}})$ and the source sends a vector $\mathbf{x}_1(\mathbf{w_{r,i-1}},\mathbf{w_{r,i}},\mathbf{w_{d,i}})$ with i.i.d. components given $\mathbf{x}_2(\mathbf{w_{r,i-1}})$, $\mathbf{u}(\mathbf{w_{r,i-1}},\mathbf{w_{r,i}},j^\star[\mathbf{i}])$, $\mathbf{u}_{1\mathbf{w_{d,i}},j_1^\star[\mathbf{i}]}(\mathbf{u}(\mathbf{w_{r,i-1}},\mathbf{w_{r,i}},j^\star[\mathbf{i}]))$ and $\mathbf{s}[\mathbf{i}]$, drawn according to the marginal $\mathbf{P}_{X_1|X_2,U,U_1,S}$.

	Block 1	Block 2	Block 3	Block 4
Relay codewords	$\mathbf{x}_2(1)$	$\mathbf{x}_2(w_{r,1})$	$\mathbf{x}_2(w_{r,2})$	$\mathbf{x}_2(w_{r,3})$
	$\mathbf{u}(1,w_{r,1},j^\star[1])$	$\mathbf{u}(w_{r,1},w_{r,2},j^\star[2])$	$\mathbf{u}(w_{r,2},w_{r,3},j^\star[3])$	$\mathbf{u}(w_{r,3},1,j^\star[4])$
Source codewords	$\mathbf{u}_{1w_{d,1},j_1^\star[1]}(\mathbf{u}(1,w_{r,1},j^\star[1]))$	$\mathbf{u}_{1w_{d,2},j_1^\star[2]}(\mathbf{u}(w_{r,1},w_{r,2},j^\star[2]))$	$\mathbf{u}_{1w_{d,3},j_1^\star[3]}(\mathbf{u}(w_{r,2},w_{r,3},j^\star[3]))$	$\mathbf{u}_{1w_{d,4},j_1^\star[4]}(\mathbf{u}(w_{r,3},1,j^\star[4]))$
	$\mathbf{x}_1(1,w_{r,1},w_{d,1})$	$\mathbf{x}_1(w_{r,1},w_{r,2},w_{d,2})$	$\mathbf{x}_1(w_{r,2},w_{r,3},w_{d,3})$	$\mathbf{x}_1(w_{r,3},1,w_{d,4})$

(3) For each $\mathbf{x}_2(w_r')$, for each $\mathbf{u}_{w_r',w_r,j}$, we generate a collection of J_1M_1 i.i.d. codewords $\{\mathbf{u}_{1w_d,j_1}(\mathbf{u}_{w_r',w_r,j})\}$ at the source, indexed by $w_d = 1,\ldots,M_1$, $j_1 = 1,\ldots,J_1$. Each codeword $\mathbf{u}_{1w_d,j_1}(\mathbf{u}_{w_r',w_r,j})$ is with i.i.d. components given $(\mathbf{x}_2(w_r'),\mathbf{u}_{w_r',w_r,j})$ drawn according to $P_{U_1|U,X_2}$. For the collection of auxiliary codewords $\{\mathbf{u}_{1w_d,j_1}(\mathbf{u}_{w_r',w_r,j})\}$ for given $\mathbf{u}_{w_r',w_r,j}$, w_d indexes bins and j_1 indexes sequences within a particular bin.

3.1.2. Encoding. The encoders at the source and the relay encode messages using codebook 1 for blocks with odd indices, and codebook 2 for blocks with even indices. This is done because some of the decoding steps are performed jointly over two adjacent blocks, and so having independent codebooks makes the error events corresponding to these blocks independent and their probabilities easier to evaluate.

We denote by $(w_{r,k},w_{d,k})$ the message pair to be sent from the source node at the beginning of block $k = 1,\ldots,B+1$. We pick up the story in block $i-1$, $i \in \{0,\ldots,B\}$. First, let us assume that the relay has decoded correctly message $w_{r,i-1}$ and the destination has decoded correctly both messages $w_{r,i-2}$ and $w_{d,i-2}$, at the end of block $i-1$. We will show that our code construction allows the relay to decode correctly message $w_{r,i}$ and the destination to decode correctly both messages $w_{r,i-1}$ and $w_{d,i-1}$ at the end of block i (with a probability of error $\leq \epsilon$). Thus, the information state $(w_{r,i-1},w_{d,i-1})$ propagates forward and a recursive calculation of the probability of error can be made, yielding a probability of error $\leq (B+1)\epsilon$.

Continuing with the strategy, let $\mathbf{s}[i]$ and $(w_{r,i},w_{d,i})$ be the state sequence in block i and the new message pair to be sent at the beginning block i, respectively. At the beginning of block i the relay knows $w_{r,i-1}$ and sends $\mathbf{x}_2(w_{r,i-1})$. The source looks in the bin indexed by $w_{r,i}$ for the smallest $j \in \{1,\ldots,J\}$ such that $\mathbf{u}_{w_{r,i-1},w_{r,i},j}$ and $\mathbf{s}[i]$ are jointly typical given $\mathbf{x}_2(w_{r,i-1})$. Denote this j by $j^\star[i] = j(\mathbf{s}[i],w_{r,i-1},w_{r,i})$. If such j does not exist, or if the observed state is not typical, an error is declared and $j^\star[i]$ is set to J. Next, the source looks for the smallest $j_1 \in \{1,\ldots,J_1\}$ such that $\mathbf{u}_{1w_{d,i},j_1}(\mathbf{u}_{w_{r,i-1},w_{r,i},j^\star[i]})$ and $\mathbf{s}[i]$ are jointly typical given $\mathbf{x}_2(w_{r,i-1})$ and $\mathbf{u}_{w_{r,i-1},w_{r,i},j^\star[i]}$. Denote this j_1 by $j_1^\star[i] = j_1(\mathbf{s}[i],w_{r,i-1},w_{r,i},w_{d,i})$. If such j_1 does not exist, an error is declared and $j_1^\star[i]$ is set to J_1. The source then sends a vector $\mathbf{x}_1(w_{r,i-1},w_{r,i},w_{d,i})$ which has i.i.d. components conditionally given $\mathbf{x}_2(w_{r,i-1})$, $\mathbf{u}_{w_{r,i-1},w_{r,i},j^\star[i]}$, $\mathbf{u}_{1w_{d,i},j_1^\star[i]}(\mathbf{u}_{w_{r,i-1},w_{r,i},j^\star[i]})$, $\mathbf{s}[i]$, drawn according to the marginal $P_{X_1|X_2,U,U_1,S}$ induced by the distribution (6).

For convenience we list the codewords at the source and the relay that are used for transmission in the first four blocks in Table 1 .

3.1.3. Decoding. Decoding is based on a combination of joint typicality and sliding-window. The decoding procedures at the end of block i are as follows.

(1) The relay knows $w_{r,i-1}$ and declares that $\hat{w}_{r,i}$ is sent if there is a unique $\hat{w}_{r,i} \in \{1,\ldots,M\}$ such that $\mathbf{u}_{w_{r,i-1},\hat{w}_{r,i},j}$ is jointly typical with $\mathbf{y}_2[i]$ given $\mathbf{x}_2(w_{r,i-1})$ for some $j \in \{1,\ldots,J\}$, where $\mathbf{y}_2[i]$ denotes the output sequence at the relay in block i. One can show that the decoding error in this step is small for sufficiently large n if

$$R_r \leq I(U;Y_2 \mid X_2) - I(U;S \mid X_2). \tag{11}$$

(2) The destination knows $w_{r,i-2}$ and decodes $w_{r,i-1}$ and $w_{d,i-1}$ based on the information received in block $i-1$ and block i. It declares that the pair $(\hat{w}_{r,i-1},\hat{w}_{d,i-1})$ is sent if there is a unique triple $(\hat{w}_{r,i-1},\hat{j},\hat{w}_{d,i-1})$ such that $\mathbf{x}_2(w_{r,i-2})$, $\mathbf{u}_{w_{r,i-2},\hat{w}_{r,i-1},\hat{j}}$, $\mathbf{u}_{1\hat{w}_{d,i-1},j_1}(\mathbf{u}_{w_{r,i-2},\hat{w}_{r,i-1},\hat{j}})$, $\mathbf{y}_3[i-1]$ are jointly typical *and* $\mathbf{x}_2(\hat{w}_{r,i-1})$ is jointly typical with $\mathbf{y}_3[i]$. One can show that the decoding error in this step is small for sufficiently large n if

$$R_d \leq I(U_1;Y_3 \mid U,X_2) - I(U_1;S \mid U,X_2),$$
$$R_d \leq I(U,U_1;Y_3 \mid X_2) - I(U,U_1;S \mid X_2), \tag{12}$$
$$R_r + R_d \leq I(U,U_1,X_2;Y_3) - I(U,U_1;S \mid X_2).$$

The analysis of the probability of error of this scheme, the details about how (9) is obtained from (11) and (12) as well as the proof that the rate (9) is not altered if the sizes of the alphabets of the auxiliary random variables U and U_1 are restricted as in (8a), and ,(8b), are given in Appendix A.

The rate in Theorem 1 requires the relay to decode only one part of the message sent by the source. It readily specializes to the case in which the relay decodes the message sent by the source *fully*, that is, full-DF. This can be obtained by sending only the message W_r in the above coding scheme, that is, $R_d = 0$. The result is stated in the following corollary.

Corollary 2. *The capacity of the state-dependent DM RC with informed source is lower-bounded by*

$$R = \max \min\{I(U;Y_2 \mid X_2) - I(U;S \mid X_2),$$
$$I(U,X_2;Y_3) - I(U;S \mid X_2)\}, \tag{13}$$

where the maximization is over all probability distributions of the form

$$P_{S,U,X_1,X_2,Y_2,Y_3} = Q_S P_{X_2} P_{U,X_1|S,X_2} W_{Y_2,Y_3|X_1,X_2,S}, \quad (14)$$

and $U \in \mathcal{U}$ is an auxiliary random variable with

$$|\mathcal{U}| \leq |\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2| + 1. \quad (15)$$

The formal proof of Corollary 2 is similar to that of Theorem 1 and, hence, it is omitted for brevity.

We close this section by noting that, here, we focused on the case in which the relay decodes and forwards (either fully or partially) the source message. The relay can employ other relaying schemes to assist the source, such as estimate-and-forward [34], amplify-and-forward [38–40] or combinations of these schemes. However, in general, none of these schemes truly extracts the potential benefits of cooperation even in the standard case in which the channel is state-independent, as no one of these schemes outperforms all the others (in terms of achievable rate) for all possible choices of channel parameters. The analysis of these alternative schemes in the context of the considered RC with informed source is beyond the scope of this work.

3.2. Comments on Upper Bounding Techniques. As we indicated in the introduction section, the model for the RC with informed source that we study in this paper seemingly exhibits some similarities with the RC with informed relay considered in [18, 19], and it also connects with the MAC with asymmetric CSI and degraded message sets [12, 16, 17]. In [18, 19], the authors derive a nontrivial upper bound on the capacity of the RC with informed relay, that is, one which is strictly tighter than the cut-set bound. A similar upper bounding technique is developed for multiaccess channels with asymmetric CSI and degraded message sets in [15–17].

However, establishing a nontrivial upper bound for the present scenario is more involved, comparatively. Partly, this is due to the following reason. In the two models mentioned above, the uninformed encoder transmits a codeword which is function of *only* the message to transmit. For the present scenario however, it is potentially possible for the uninformed encoder (the relay) to get some information about the channel states from the past received sequence from the informed encoder (the source). That is, at time i, the input of the relay $X_{2,i}$ can potentially depend on the channel states through $Y_2^{i-1} = (Y_{2,1}, \ldots, Y_{2,i-1})$. Further, since, for $j = 1, \ldots, i - 1$, Y_2^j depends on the channel states in a noncausal manner (through the source codeword $X_{1,j}(W, S^n)$), and not only through the current state S_i, so does the input of the relay, potentially.

A trivial upper bound on the capacity of the general DM RC with informed source is obtained by assuming that the channel states are also available at the source and the destination, that is, the *max-flow min-cut bound* or *cut-set bound*,

$$R_{\text{triv}}^{\text{up}} = \max \min\{I(X_1; Y_2, Y_3 \mid S, X_2), I(X_1, X_2; Y_3 \mid S)\}, \quad (16)$$

where the maximization is over all distributions of the form

$$P_{S,X_1,X_2,Y_2,Y_3} = Q_S P_{X_1,X_2|S} W_{Y_2,Y_3|X_1,X_2,S}. \quad (17)$$

4. The Gaussian RC with Informed Source

In this section, we consider a state-dependent three-terminal full-duplex RC in which the channel states and the noise are additive and Gaussian. Furthermore, we extend the discrete memoryless model considered in Section 3 to accommodate two state sequences; one state sequence affects the transmission to the relay and the other state sequence affects the transmission to the destination. In this model, the channel states model additive Gaussian interferences which are assumed to be known (noncausally) to only the source. We first consider a scenario in which transmission to the relay and to the destination are corrupted by the *same* interference. We derive a lower bound on the channel capacity for this scenario in Section 4.2. Then, we consider the case of two independent interferences in Section 4.3. In Section 4.4 we discuss a case of correlated interferences.

4.1. Channel Model. For the state-dependent Gaussian RC, the channel outputs $Y_{2,i}$ and $Y_{3,i}$ at time instant i for the relay and the destination, respectively, are related to the channel input $X_{1,i}$ from the source and $X_{2,i}$ from the relay, and the channel states $S_{2,i}$ and $S_{3,i}$ by

$$Y_{2,i} = X_{1,i} + S_{2,i} + Z_{2,i},$$
$$Y_{3,i} = X_{1,i} + X_{2,i} + S_{3,i} + Z_{3,i}, \quad (18)$$

where $S_{2,i}$ models the interference on the link to the relay and $S_{3,i}$ models the interference on the link to the destination. The channel states $S_{2,i}$ and $S_{3,i}$ are zero mean Gaussian random variables with variance Q, and only the source knows the states sequences S_2^n and S_3^n (noncausally). The noises $Z_{2,i}$ and $Z_{3,i}$ are zero mean Gaussian random variables with variances N_2 and N_3, respectively, and are mutually independent and independent from the states sequences (S_2^n, S_3^n) and the channel inputs (X_1^n, X_2^n).

We consider the following individual power constraints on the average transmitted power at the source and the relay

$$\sum_{i=1}^{n} X_{1,i}^2 \leq nP_1, \qquad \sum_{i=1}^{n} X_{2,i}^2 \leq nP_2. \quad (19)$$

The definition of a code for this channel is the same as that given in Section 2, with the additional constraint that the channel inputs should satisfy the power constraints (19).

For convenience we define the following functions $R_1(\cdot)$, $R_2(\cdot)$ and $\varsigma(\cdot)$ which we will use throughout the remaining sections.

Definition 2. Let

$$R_1\left(P_1, Q, N_2, \zeta, \varrho, \alpha\right)$$

$$= \frac{1}{2} \log\left(\frac{P_1 \zeta \left[P_1 \zeta + \left(\varrho\sqrt{P_1} + \sqrt{Q} \right)^2 + N_2 \right]}{P_1 Q \zeta (1-\alpha)^2 + N_2 \left[P_1 \zeta + \left(\varrho\sqrt{P_1} + \alpha\sqrt{Q} \right)^2 \right]} \right),$$

$$R_2\left(P_1, P_2, Q, N_3, \zeta, \varrho, \alpha\right)$$

$$= \frac{1}{2} \log\left(\frac{P_1 \zeta \left[P_1 \zeta + P_c + \left(\varrho\sqrt{P_1} + \sqrt{Q} \right)^2 + N_3 \right]}{P_1 Q \zeta (1-\alpha)^2 + N_3 \left[P_1 \zeta + \left(\varrho\sqrt{P_1} + \alpha\sqrt{Q} \right)^2 \right]} \right),$$

$$\varsigma\left(P_1, Q, \zeta, \varrho, \alpha\right)$$

$$= \frac{\left[(1-\alpha)^2 + 1 \right] P_1 \zeta + \left(\varrho\sqrt{P_1} + \alpha\sqrt{Q} \right)^2}{P_1 \zeta + \left(\varrho\sqrt{P_1} + \alpha\sqrt{Q} \right)^2}, \tag{20}$$

where $P_c := \left(\sqrt{1 - \zeta - \varrho^2}\sqrt{P_1} + \sqrt{P_2} \right)^2$, for nonnegative P_1, $P_2, Q, N_2, N_3, \zeta \in [0,1], \varrho \in [-1,0]$ and $\alpha \in \mathcal{A}(\zeta, \varrho) = \{x \in \mathbb{R} : R_1(P_1, Q, N_2, \zeta, \varrho, x) \geq 0, \ R_2(P_1, P_2, Q, N_3, \zeta, \varrho, x) \geq 0\}$.

4.2. Case of One Interference. In this section, we let $S_{2,i} = S_{3,i} = S_i$ in (18), that is, we consider the case in which the relay and the destination are corrupted by the *same* interference:

$$Y_{2,i} = X_{1,i} + S_i + Z_{2,i},$$
$$Y_{3,i} = X_{1,i} + X_{2,i} + S_i + Z_{3,i}. \tag{21}$$

The results obtained in Section 4 for the DM case can be applied to memoryless channels with discrete time and continuous alphabets using standard techniques [41]. We use the lower bound established in Theorem 1 to compute a lower bound on the channel capacity of the Gaussian model (21).

The following theorem provides a lower bound on the channel capacity of the model (21).

Theorem 3. *The capacity of the state-dependent Gaussian RC with informed source is lower-bounded by*

$$R_G = \max \min\{R_1\left(P_{1r}, Q, N_2 + P_{1d}, \xi, \rho, \alpha\right),$$
$$R_2\left(P_{1r}, P_2, Q, N_3 + P_{1d}, \xi, \rho, \alpha\right)\} \tag{22}$$
$$+ \frac{1}{2}\log\left(1 + \frac{P_{1d}}{N_3}\right),$$

where the functions $R_1(\cdot)$ and $R_2(\cdot)$ are defined as in Definition 2, and the maximization is over nonnegative P_{1r} and P_{1d} such that $P_{1r} + P_{1d} \leq P_1, \xi \in [0,1], \rho \in [-\sqrt{1-\xi}, 0]$, and, for fixed $(P_{1r}, P_{1d}, \xi, \rho)$, $\alpha \in \mathcal{B}_1(P_{1r}, P_{1d}, \xi, \rho) := \{x \in \mathbb{R} : R_1(P_{1r}, Q, N_2 + P_{1d}, \xi, \rho, x) \geq 0, \ R_1(P_{1r}, Q, N_3 + P_{1d}, \xi, \rho, x) \geq 0\}$.

Proof. A formal proof of Theorem 3 is given in Appendix B. □

An informal proof is as follows. As we outlined after Theorem 1, we decompose the message W to be sent from the source into two parts W_r and W_d. The input X_1^n from the source is divided accordingly into two independent parts, that is, $X_1^n = X_{1r}^n + X_{1d}^n$, where X_{1r}^n carries message W_r and has power constraint nP_{1r} and X_{1d}^n carries message W_d and has power constraint nP_{1d}, with $P_1 = P_{1r} + P_{1d}$. The message W_r is sent through the relay at rate R_r and the message W_d is sent directly to the destination at rate R_d. The total rate is $R_G = R_r + R_d$. Since message W_d is to be decoded by only the destination, it is precoded against the interference on the link to the destination, using a standard DPC. The message W_r, however, experiences the same interference S^n but different noise terms on its way to the relay and to the destination, and it is precoded against the interference S^n through a GDPC. The GDPC can be interpreted as a partial cancellation of the interference, by the source for the relay, combined with standard DPC [11, 13]. The relay benefits from this cancellation and can then transmit more reliably to the destination, and so the source benefits in turn.

More formally, we decompose the source input random variable X_1 as

$$X_1 = X_{1r} + X_{1d}, \tag{23}$$

where X_{1r} is zero mean Gaussian with variance P_{1r}, is independent from X_{1d} and is correlated with both the relay input X_2 and the state S, with $\mathbb{E}[X_{1r}X_2] = \rho_{12}\sqrt{P_{1r}P_2}$ and $\mathbb{E}[X_{1r}S] = \rho_{1s}\sqrt{P_{1r}Q}$, for some $\rho_{12} \in [-1,1], \rho_{1s} \in [-1,1]$; and X_{1d} is zero mean Gaussian with variance P_{1d}, and is independent from both the relay input X_2 and the state S. (Note that $X_{1d,i}$ also can be chosen to be negatively correlated with the state S_i. The rate achievable in this case can be obtained in a straightforward manner from the analysis in Section 4.2.) For the GDPC, we choose the random variable U as

$$U = X_{1r} + \alpha S, \tag{24}$$

for some $\alpha \in \mathbb{R}$. For the standard DPC, we choose the random variable U_1 as

$$U_1 = X_{1d} + \frac{P_{1d}}{P_{1d} + N_3}(1 - \alpha)S. \tag{25}$$

Let $\xi := 1 - \rho_{12}^2 - \rho_{1s}^2$ and $\rho := \rho_{1s}$. As it will become clear from the proof in Appendix B, allowable values of the correlation coefficients ρ_{12} and ρ_{1s} are such that the vector $(S, X_{1r}, X_{1d}, X_2, Z_2, Z_3)$ has a nonnegative discriminant, that is, for nonzero P_{1r}, P_{1d}, Q,

$$\xi = 1 - \rho_{12}^2 - \rho_{1s}^2 \geq 0, \tag{26}$$

and all values of α such that $R_1(P_{1r}, Q, N_2 + P_{1d}, \xi, \rho, \alpha)$ and $R_1(P_{1r}, Q, N_3 + P_{1d}, \xi, \rho, \alpha)$ are nonnegative real.

The result in Theorem 3 readily specializes to the case of full DF at the relay.

Corollary 4. *The capacity of the model (21) for the state-dependent Gaussian RC with informed source is lower-bounded by*

$$R_{\text{G-DF}} = \max \min \{ R_1(P_1, Q, N_2, \xi, \rho, \alpha), \tag{27}$$
$$R_2(P_1, P_2, Q, N_3, \xi, \rho, \alpha) \},$$

where the functions $R_1(\cdot)$ and $R_2(\cdot)$ are defined as in Definition 2 and the maximization is over parameters $\xi \in [0,1]$, $\rho \in [-\sqrt{1-\xi}, 0]$, and $\alpha \in \mathcal{B}_2(\xi, \rho) = \{ x \in \mathbb{R} : R_1(P_1, Q, N_2, \xi, \rho, x) \geq 0 \text{ and } R_2(P_1, P_2, Q, N_3, \xi, \rho, x) \geq 0 \}$.

4.3. Case of Two Independent Interferences. In this section we consider a model in which the links source-to-relay and source-to-destination are corrupted by *independent* interferences, that is, the model (18) of Section 4.1,

$$Y_{2,i} = X_{1,i} + S_{2,i} + Z_{2,i},$$
$$Y_{3,i} = X_{1,i} + X_{2,i} + S_{3,i} + Z_{3,i}, \tag{28}$$

with $S_{2,i}$ and $S_{3,i}$ being mutually independent zero mean Gaussian random variables with variance Q.

It is interesting to observe that if the relay operates in a decode-and-forward scheme the channel model (28) has a (high-level) connection with the Gaussian multicast model studied in [36], in that the signal sent by the informed transmitter (i.e., the source) experiences different interferences on its way to the two receivers. However, a major difference from [36] is that, here, one receiver (the relay) also helps the transmitter by relaying the information to the other receiver, that is, a relay setup. Also, unlike the model in [36], the noise terms at the receivers have different variances.

The following theorem provides a lower bound on the channel capacity of the model (28). The coding scheme that we use to establish this theorem builds upon the scheme of [36] (named carbon copying onto dirty paper (CC) therein), and it also generalizes it as will become clear from the proof.

Theorem 5. *The capacity of the channel model (28) is lower-bounded by*

$$R_G^{\text{ind}} = \max T_1 + T_2 \tag{29}$$

with

$$T_1$$
$$= \nu \min \left\{ R_1 \left(P_{1a}, \frac{Q}{2}, \tilde{N}_2, \xi, \rho, \alpha_a \right), \right.$$
$$\left. R_2 \left(P_{1a}, P_{2a}, \frac{Q}{2}, \tilde{N}_3 + P_{2d}, \xi, \rho, \alpha_a \right) \right\}$$
$$+ \nu \min \left\{ \frac{1}{2} \log \left(1 + \frac{P_{1d}}{N_2} \right), \right.$$
$$\left. \frac{1}{2} \log \left(1 + \frac{P_{2d}}{N_3 + P_{1d} + (Q/2)\varsigma(P_{1a}, Q/2, \xi, \rho, \alpha_a)} \right) \right\},$$

$$T_2 = \bar{\nu} \min \left\{ R_1 \left(P_{1a}, \frac{Q}{2}, \tilde{N}_2, \xi, \rho, \alpha_a \right), \right.$$
$$\left. R_2 \left(P_{1a}, P_2, \frac{Q}{2}, \tilde{N}_3, \xi, \rho, \alpha_a \right) \right\} + \frac{\bar{\nu}}{2} \log \left(1 + \frac{P_{1d}}{N_3} \right), \tag{30}$$

where $\tilde{N}_2 = N_2 + P_{1d} + Q/2$, $\tilde{N}_3 = N_3 + P_{1d} + Q/2$ and the maximization is over parameters: $\nu \in [0,1]$; nonnegative P_{1a}, P_{1d}, P_{2a}, P_{2d} such that $P_{1a} + P_{1d} \leq P_1$ and $P_{2a} + P_{2d} \leq P_2$; $\xi \in [0,1]$, $\rho \in [-\sqrt{1-\xi}, 0]$; and, for fixed P_{1a}, P_{1d}, P_{2a}, P_{2d}, $\alpha_a \in \mathcal{B}_3(\xi, \rho) := \{ x \in \mathbb{R} : R_1(P_{1a}, Q/2, \tilde{N}_2, \xi, \rho, x) > 0, R_1(P_{1a}, Q/2, \tilde{N}_3, \xi, \rho, x) > 0, R_1(P_{1a}, Q/2, \tilde{N}_3 + P_{2d}, \xi, \rho, x) > 0 \}$.

Proof. A formal proof of Theorem 5 is given in Appendix C. In the proof, we develop a coding scheme that uses decode-and-forward relaying and a carbon copying like scheme. As it can be seen from the proof, the applied CC scheme has two differences from [36]. First, the signal sent by the informed transmitter is *correlated* with the channel state. Second, the noise terms at the two receivers are with different variances. One direct consequence of the latter dissimilarity is that, in our case, for the transmission of the message that is sent to the two receivers at the same time, one cannot derive an optimal choice for Costa's parameter (i.e., one that permits to remove the effect of the interference simultaneously for the two links via one single DPC as in [36]). (This explains why Costa's parameter α_a is left to be optimized over in the encoding step (C.1) in Appendix C, as opposed to in [36] where the choice of Costa's parameter is optimal for both receivers.) We outline the coding scheme in the following.   □

4.3.1. Outline of the Coding Scheme. Let $S_a^n := (S_2^n + S_3^n)/2$ and $S_d^n := (S_2^n - S_3^n)/2$. The states S_a^n and S_d^n are mutually independent and they can be used to represent the interferences on the link to the relay and the link to the destination, as

$$S_2^n = S_a^n + S_d^n,$$
$$S_3^n = S_a^n - S_d^n. \tag{31}$$

We denote by T the total transmission time. We divide this time into two transmission periods of duration νT and $\bar{\nu} T$, respectively, with $0 \leq \nu \leq 1$. Also, we decompose the massage W to be transmitted from the source into two submessages W_a and W_d. At time i, the input $X_{1,i}$ from the source is divided accordingly into two independent parts, that is, $X_{1,i} = X_{1a,i} + X_{1d,i}$, where $X_{1a,i}$ carries message W_a and has power constraint P_{1a} and $X_{1d,i}$ carries message W_d and has power constraint P_{1d}, with $P_1 = P_{1a} + P_{1d}$. The message W_a will be sent to both the relay and the destination at the same time, during the two periods νT and $\bar{\nu} T$; and it is precoded against the interference $S_{a,i}$. The precoding for message W_a is performed through a GDPC which is similar to that in Section 4.2. The message W_d will be sent to only the relay during the period νT and to only the destination during

the period $\bar{\nu}T$. Hence, it is precoded against the interference on the link to the relay during the period νT and against the interference on the link to the destination during the period $\bar{\nu}T$. The precoding for message W_d is performed through standard DPC. The relay decodes-and-forwards the message W_a during both transmission periods and also decodes-and-forwards message W_d during the period νT.

During the period νT, the relay sends a superposition of two independent Gaussian codewords, $X_{2,i} = X_{2a,i} + X_{2d,i}$, where codeword $X_{2a,i}$ enhances the transmission of message W_a by the source to the destination and codeword $X_{2d,i}$ carries message W_d to the destination. (Note that during this period the destination decodes message W_d from only the transmission by the relay. For instance, the decoder at the destination treats the codeword $X_{1d,i}$ from the source as part of the noise, as it can be seen from the decoding procedure in Appendix C.) Thus, the relay shares power between $X_{2a,i}$ and $X_{2d,i}$, that is, $X_{2a,i}$ is sent with power constraint P_{2a} and $X_{2d,i}$ is sent with power constraint P_{2d}, with $P_{2a} + P_{2d} \leq P_2$. During the period $\bar{\nu}T$, the relay sends only $X_{2a,i}$, with power constraint P_2. In this case, the destination obtains message W_d from the direct transmission from the source. A block diagram of the communication protocol is shown in Figure 3.

Remark 1. The rate of Theorem 5 includes the rates of the following schemes.

(i) full-DF with precoding against the interference S_a^n using a GDPC; This is obtained by putting $\nu = 0$, $P_{1d} = 0$ in (29).

(ii) partial-DF with precoding for the message sent through the relay against the interference S_a^n using a GDPC, and precoding for the additional message to be decoded only by the destination against the interference on the link to the destination using a standard DPC; This is obtained by putting $\nu = 0$ in (29).

(iii) time-sharing DPC between the relay and the destination; This is obtained by setting $P_{1a} = P_{2a} = 0$ during both transmission periods.

Remark 2. So far, we have assumed that the parameters P_{1a}, P_{1d} and $(\rho_{12} = \sqrt{1 - \xi - \rho^2}, \rho_{1s} = \rho, \alpha_a)$ are identical for the two periods νT and $\bar{\nu}T$. One can obtain larger achievable rates if one allows the source to send with different powers and use different parameters $(\rho_{12}, \rho_{1s}, \alpha_a)$ for the two periods. However, the numerical computation of the obtained rates becomes tedious in this case.

4.3.2. Extreme Cases.
We now focus on the behavior of the above bounds in some trivial extreme cases.

(1) In the limit of strong interference (i.e., $Q \to \infty$), the lower bound (29) reduces to

$$R_G = \frac{1}{2} \log\left(1 + \frac{P_1}{N_3}\right), \tag{32}$$

which can also be achieved using a standard DPC at the source and keeping the relay off.

Period νT:
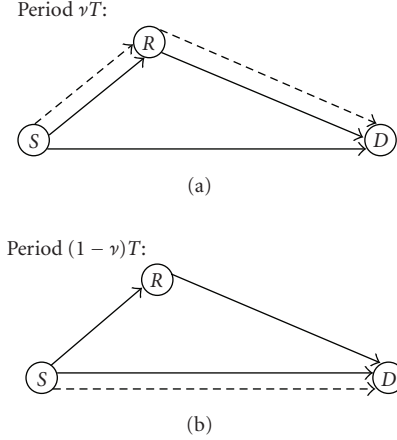
(a)

Period $(1 - \nu)T$:

(b)

FIGURE 3: Block diagram for the communication protocol for the RC with independent interferences. Solid lines correspond to transmission of message W_a and dashed lines correspond to transmission of message W_d. Message W_a is precoded against the interference $S_{a,i}$ using a GDPC, during both periods. Message W_d is precoded using standard DPC, against the interference $(1 - \alpha_a)S_{a,i} + S_{d,i}$ during the period νT and against the interference $(1 - \alpha_a)S_{a,i} - S_{d,i}$ during the period $(1 - \nu)T$.

(2) For $Q = 0$, the lower bound (29) reduces to the one achievable with partial DF in the standard interference-free Gaussian RC. Also, in this case, putting $P_{1d} = 0$ in (29) we get that the resulting lower bound meets with the cut-set bound for the degraded Gaussian case, and yields

$$\mathcal{C}_{DG} = \max_{0 \leq \beta \leq 1} \min\left\{\frac{1}{2} \log\left(1 + \frac{P_1(1 - \beta^2)}{N_2}\right), \right.$$
$$\left. \frac{1}{2} \log\left(1 + \frac{P_1 + P_2 + 2\beta\sqrt{P_1 P_2}}{N_3}\right)\right\}, \tag{33}$$

which is the capacity of the standard degraded Gaussian RC [34, Theorem 5].

(3) If $P_2 = 0$, the capacity of the model (28) is given by

$$\mathcal{C}_G = \frac{1}{2} \log\left(1 + \frac{P_1}{N_3}\right). \tag{34}$$

In this case, the effect of the interference on the link to the destination is removed by a regular DPC at the source.

4.4. Discussion: A Case of Correlated Interferences.
Consider a variation of the model (18) in Section 4.1 with $S_{2,i} = h_2 S_i$ and $S_{3,i} = h_3 S_i$, where the CSI S_i is a zero-mean Gaussian random variable with variance Q, and the sequence S^n is known noncausally to only the source. For purposes of exposition we assume that the coefficients h_2 and h_3 are real-valued and nonnegative, and are fixed and known to all the terminals, with $h_2 \neq h_3$. This model can be seen as an extension of the model (21) to the case in which the interference corrupting

the transmission is scaled differently on the links to the relay and to the destination, as

$$Y_{2,i} = X_{1,i} + h_2 S_i + Z_{2,i},$$
$$Y_{3,i} = X_{1,i} + X_{2,i} + h_3 S_i + Z_{3,i}, \tag{35}$$

where h_2 and h_3 may model known channel coefficients in fading environments. Proceeding similarly to the model (28) in Section 4.3, let $S_a^{(k)} = h_a S^{(k)}$ and $S_d^{(k)} = h_d S^{(k)}$, with $h_a = (h_2 + h_3)/2$, $h_d = (h_2 - h_3)/2$ and $k = 1, 2$.

It is interesting to note that the coding scheme that we used to establish **Theorem 5** does *not* apply directly to the model (35) because the noise will be correlated with both the channel input and the CSI in the steps in which $S_d^{(k)}$, $k = 1, 2$, is treated as part of the noise and $(S_a^{(k)})^n$ as CSI at the encoder. For example, observe that, for the model (35), treating $S_d^{(1)}$ as part of the noise as we did in (C.7), (C.8) during the period νT, the resulting noise term $(X_{1d}^{(1)} + S_d^{(1)} + Z_2)$ will be correlated with both the input $X_{1a}^{(1)}$ and the CSI $S_a^{(1)}$, with $\mathbb{E}[(X_{1d}^{(1)} + S_d^{(1)} + Z_2) X_{1a}^{(1)}] = \rho_{1s} h_d \sqrt{P_{1a} Q}$ and $\mathbb{E}[(X_{1d}^{(1)} + S_d^{(1)} + Z_2) S_a^{(1)}] = h_a h_d Q$. The same applies for treating $S_d^{(1)}$ as part of the noise as in (C.12), (C.13) during the period νT, and for treating $S_d^{(2)}$ as part of the noise as in (C.21), (C.22) during the period $(1 - \nu) T$.

In [36] the authors develop a coding scheme for a Gaussian multicast problem with independent interferences known noncausally to the transmitter; they also mention that with some modifications (especially, the common randomness that is mentioned below) their coding scheme also applies for a model in which the interferences are scaled differently on the links to the two receivers (one which is similar to (35), but for the multicast problem). In [36] the direct application of the coding scheme developed for the independent interferences case to the model with scaled interferences incurs (only) a correlation between the noise and the CSI, and the authors mention that such a correlation does not reduce the rates relative to the case in which they are independent if the encoder and the decoders have access to a source of common randomness. The underlying code construction is based on the lattice strategies of [42], and for instance the *Inflated Lattice Lemma* [42, Lemma 6]. In fact the code construction for the Gaussian multicast problem with independent interferences in [36] can be seen as being basically a careful superimposition of two DPCs and, thus, an appropriate superimposition of two lattice codes with good quantization properties, each designed as in [42], clearly achieves the same rates asymptotically with the dimension of the employed lattices.

In our case, however, as we already mentioned, there is also the correlation between the noise and the channel input. This correlation is due to that the source input is chosen to be (negatively) correlated with the CSI, for the sake of reducing its effect on the link relay-to-destination. While it is possible to get rid of this correlation by transforming the channel into an equivalent channel in which the (equivalent) channel input is independent from the (equivalent) CSI, it is still to be proved that the rate bounds on the rates $R_a^{(k)}$ and $R_d^{(k)}$,
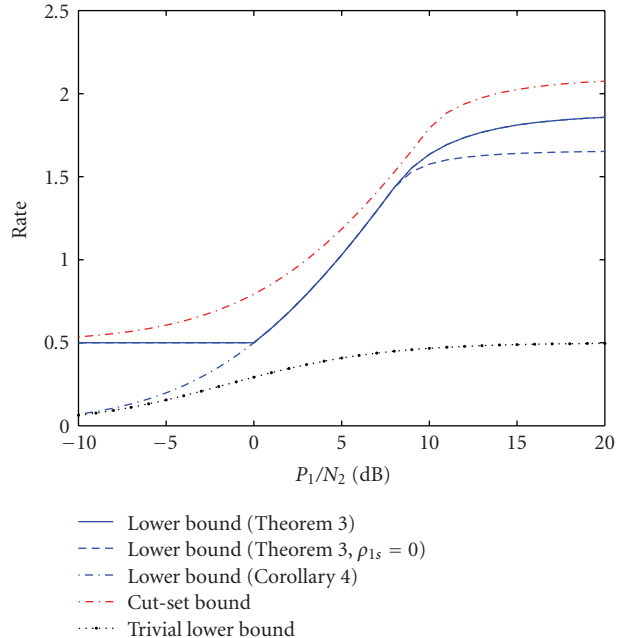

FIGURE 4: Illustration of the lower bounds in **Theorem 3** and **Corollary 4**. The bounds are plotted versus the SNR in the link source-to-relay. Numerical values are: $P_1 = 10$ dB, $P_2 = 20$ dB, $Q = 10$ dB, $N_3 = 10$ dB.

$k = 1, 2$, in the proof of **Theorem 5** which are established using random binning arguments can be achieved using a proper choice of linear lattice strategies at the source and codes at the relay. (This can be obtained by dividing the source input into two independent parts: one part which is *proportional* to the known CSI $S_a^{(k)}$, $k = 1, 2$, and thus is considered as part of the equivalent CSI, and another part which is independent from it and is considered as channel input for the equivalent channel.)

4.5. Numerical Examples and Discussion. In this section we discuss some numerical examples for the general Gaussian case.

4.5.1. Generalized DPC versus Standard DPC. First, we illustrate the rates given in **Theorem 3** and **Corollary 4**, and the efficiency of the coding ideas used therein, through an numerical example. Figure 4 depicts the evolution of the lower bound (22) for a numerical example for the Gaussian RC model (21), as function of the signal-to-noise-ratio (SNR) at the relay, SNR $= P_1/N_2$ (in decibels). Also shown for comparison are: the rate of **Corollary 4**, the cut-set bound (16) computed here for the Gaussian channel model (21) and the trivial lower bound obtained by treating the interference at the relay and the destination as unknown noises. Investigating the two curves depicting the rates of **Theorem 3** and **Corollary 4**, shows that, as expected, splitting the message to be transmitted into two parts, and so having the relay decode only one part of it, is beneficial at small SNR. However, this improvement vanishes at large SNR, as the relay can decode *all* the information transmitted by the
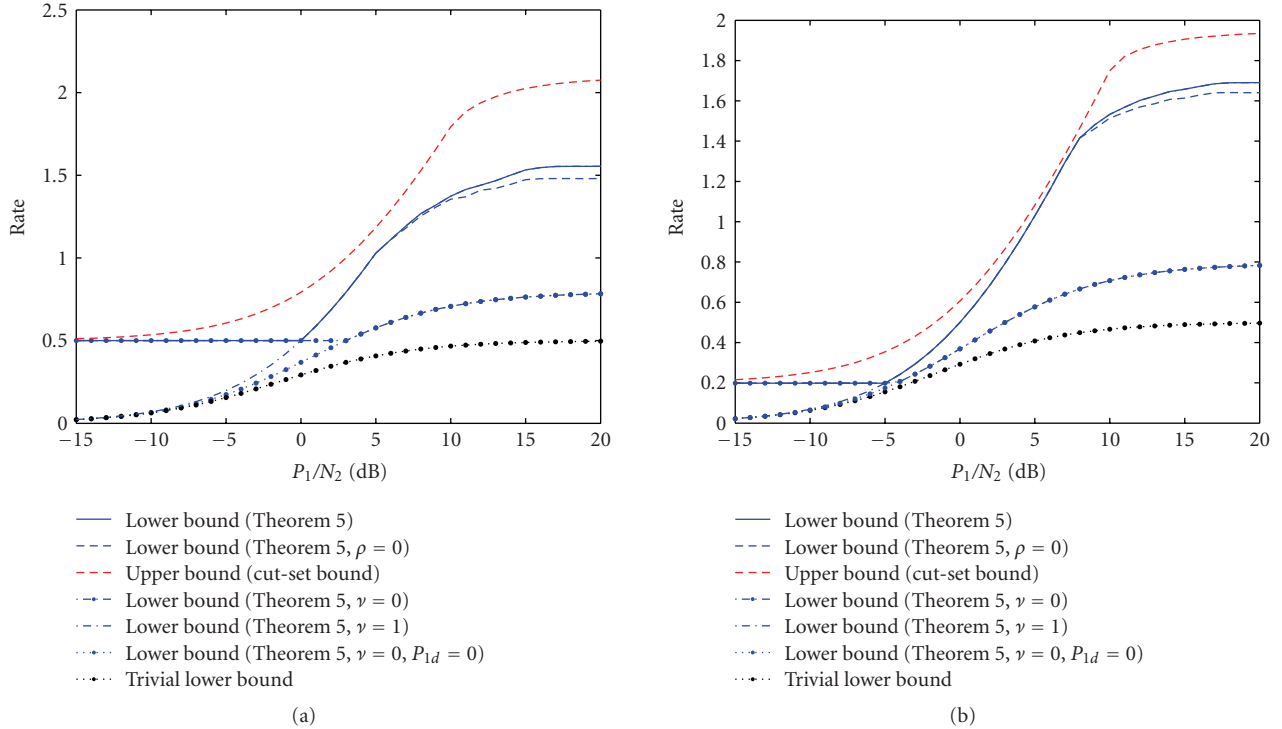
FIGURE 5: Illustration of the lower bound in Theorem 5 and some special cases of it. The bounds are plotted versus the SNR in the link source-to-relay. Numerical values are: (a) $P_1 = 10$ dB, $P_2 = 20$ dB, $Q = 10$ dB, $N_3 = 10$ dB and (b) $P_1 = 5$ dB, $P_2 = 20$ dB, $Q = 5$ dB, $N_3 = 10$ dB.

source, and therefore there is no benefit from splitting the message at such a range of SNR.

Furthermore, Figure 4 also shows the rate obtained with rate-splitting and standard DPC at the source, that is, the special case of (22) obtained by setting $\rho := \rho_{1s} = 0$. Comparing this rate to the one of Theorem 3 (which is based on a coding scheme that employs GDPC or, equivalently, partial cancellation of the interference combined with standard DPC) it can be seen that, GDPC always improves upon standard DPC. This means that, even if only the source knows the interference, both the source and the relay benefit from this knowledge. This is made possible by having the source partially cancel the interference for the relay. For instance, the relay benefits since its transmit signal faces less interference on its way to the destination, and the source benefits in turn since the advantage taken from being assisted by a relay which is actually more efficient favourably counterbalances the loss incurred by spending some power in partially cleaning the channel for the relay.

It is worth mentioning that the improvement brought by GDPC (over standard DPC) is mainly visible at large SNR. This is because, as a prerequisite for the DF relaying strategy (or its variants), the relay can assist the source efficiently only if it decodes the source transmit symbol reliably (at least partially); and so, if it does not so, that is, at low SNR, it is not worth that the source spends power in facilitating relay transmission by (partially) cleaning the channel for it, as this would be accomplished at the cost of some power that could be allocated to strengthen the source transmit signal so that the relay can decode it more reliably, instead.

4.5.2. Generalized CC versus Standard CC. Figure 5 depicts the evolution of the lower bound (29) for two numerical examples for the Gaussian RC model (28), as function of the signal-to-noise-ratio (SNR) at the relay, SNR = P_1/N_2 (in decibels). The figure also shows the cut-set bound (16) computed here for the Gaussian channel model (28) and the curves corresponding to some other achievable rates obtained as special cases of the rate in Theorem 5 as we mentioned in Remark 1—all shown for comparison purposes. Furthermore, in order to show the improvement brought up by GCC over standard CC, Figure 5 also shows the rate obtained by the latter scheme, that is, the special case of (29) obtained by setting $\rho = 0$. We mention that the latter scheme is the one that can be obtained by a *natural*, but careful, extension of the initial CC [36], which was developed for a multicast setup as we already mentioned, to the relay setup.

It is interesting to observe that, just as the scheme that uses GDPC improves upon the one that uses regular DPC in Theorem 3 as we explained previously, here also the coding scheme that employs GCC improves upon the one that employs regular CC. Moreover, the range of SNR for which the improvement is visible corresponds to when the multiaccess part of the bound in (29) is operative, that is, when the obtained rate is given by the information that the source and the relay together can transfer to the destination.

The coding scheme that we developed for the case of one interference (21) in Section 4.2 also applies for the model (28); the allowed rate is obtained by setting $\nu = 0$ in (29) as we indicated in Remark 1. This rate is shown by the
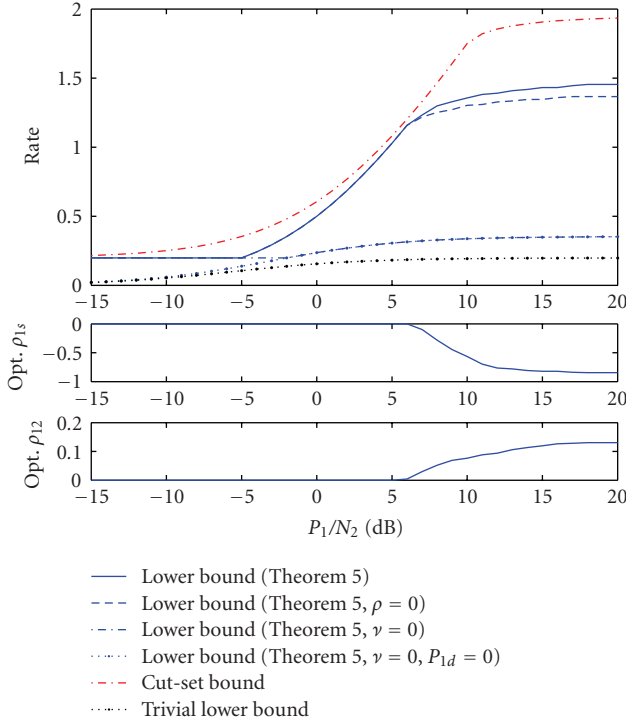
FIGURE 6: Lower bounds on the capacity of the state-dependent general Gaussian RC with informed source (28), together with the maximizing $\rho_{12} = \sqrt{1 - \xi - \rho^2}$ and $\rho_{1s} := \rho$ for the lower bound (29). Numerical values are: $P_1 = 5\,\text{dB}$, $P_2 = 20\,\text{dB}$, $Q = 10\,\text{dB}$, $N_3 = 10\,\text{dB}$.

dash-dotted curve in Figure 5. Comparing this rate with (29), it is insightful to observe that because the model (28) comprises two interferences, time sharing the superimposed GDPC and regular DPC that are applied during the period νT with those that are applied during the period $\bar{\nu} T$ (i.e., varying $\nu \in [0, 1]$) is advantageous. While the achieved improvement is well expected since optimizing over ν as in Theorem 5 can only increase the rate (relative to the one obtained by fixing $\nu = 0$ in (29), that is, by applying the coding scheme of Theorem 3 to the model (28)), it is insightful to comment on this improvement. Investigating the effect of fixing $\nu = 0$, it can be seen that this causes the information sent through the relay to suffer from the interference $S_{d,i} = (S_{2,i} - S_{3,i})/2$ during all the transmission time (recall that this interference is considered as an unknown noise with power $Q/2$ at the decoder). In fact, with the coding scheme used to establish Theorem 5 also, the message W_a sent through the relay suffers from the same interference during all the transmission time. However, in this latter case, the relay also helps transmitting message W_d. For small SNR, however, there is no benefit from relaying message W_d as well; and this explains why the two schemes give the same rate for such a range of SNR, that is, the optimal choice of ν in Theorem 5 is zero for small SNR.

Another numerical example is shown in Figure 6. For this numerical example, the figure also shows the variations of the maximizing $\rho_{1s} := \rho$ and $\rho_{12} = \sqrt{1 - \xi - \rho^2}$ in (29), as function of the SNR. This shows how the informed source

allocates its power among combating the interference for the relay (related to the value of ρ_{1s}), sending signals that are coherent with the transmission from the relay (related to the values of ρ_{12}) and sending additional information (related to the values of $1 - \rho_{12}^2 - \rho_{1s}^2$).

5. Conclusion

In this paper, we consider a state-dependent three-terminal full-duplex relay channel (RC) with the states of the channel known noncausally at only the source, that is, neither at the relay nor at the destination. We refer to this communication model as *state-dependent RC with informed source*. This setup may model the basic scenario of cooperation over a wireless network in which only the sources are cognitive of the states of the channel. We study this problem in the discrete memoryless (DM) setup and in the Gaussian setup. For the Gaussian setup, the channel states model additive Gaussian outside interferences.

For the DM case, we establish a lower bound on the channel capacity. This lower bound is obtained by a coding scheme that uses techniques of rate-splitting at the source, regular encoding sliding-window decoding [33] for decode-and-forward (DF) relaying [34, Theorem 4] and a Gel'fand-Pinsker like binning scheme. Due to the rate-splitting at the source, this lower bound is better than the one obtained by assuming that the he relay decodes the source message fully, that is, full-DF combined with a Gel'fand-Pinsker like binning scheme.

For the Gaussian setup, we consider channel models in which each of the relay node and the destination node experiences on its link an additive Gaussian outside interference in addition to additive Gaussian noise. The interferences are known noncausally to only the source, and play the role of additive channel states. We focus on the case of *one* interference corrupting both transmissions to the relay and to the destination, and also the case of two *independent* interferences each corrupting one link. We establish lower bounds on channel capacity for each of these two models. Furthermore, we also discuss a case of *correlated* interferences. For the case of *one* interference, the applied coding scheme combines techniques of rate-splitting and generalized dirty paper coding [11, 13] at the source, and decode-and-forward relaying. For the case of two independent interferences, as part of our coding scheme we employ a carbon copying onto dirty paper (CC) that builds carefully upon the initial CC [36] that was initially developed for a multicast setup; it also generalizes it by allowing negative correlation between the codewords at the source at the known channel states. For both studied Gaussian models, the uniformed relay benefits from the allowed negative correlation and, so, the source benefits in turn.

Appendices

Throughout this section we denote the set of strongly jointly ϵ-typical sequences [43, Chapter 14.2] with respect to the distribution $P_{X,Y}$ as $T_{\epsilon}^n(P_{X,Y})$.

A. Proof of Theorem 1

Consider the random coding scheme that we described in Section 3.1. We first show that the average probability of error goes to zero as $n \to \infty$.

A.1. Analysis of Probability of Error. Fix a probability distribution $P_{S,U,U_1,X_1,X_2,Y_2,Y_3}$ satisfying (6). Let $(w_{r,i-1}, w_{d,i-1})$ and $(w_{r,i}, w_{d,i})$ be the message pair sent in block $i-1$ and the message pair sent in block i, respectively. Let $\mathbf{s}[i]$ denote any random state sequence observed in block i. As we already outlined after Theorem 1, at the beginning of block i the relay has decoded $w_{r,i-1}$ and transmits $\mathbf{x}_2(w_{r,i-1})$, and the source transmits a vector $\mathbf{x}_1(w_{r,i-1}, w_{r,i}, w_{d,i})$ with i.i.d. components conditionally given $\mathbf{s}[i]$, $\mathbf{x}_2(w_{r,i-1})$, $\mathbf{u}(w_{r,i-1}, w_{r,i}, j^\star[i])$ and $\mathbf{u}_{1w_{d,i}, j_1^\star[i]}(\mathbf{u}(w_{r,i-1}, w_{r,i}, j^\star[i]))$ drawn according to the marginal $P_{X_1|X_2,U,U_1,S}$, with $j^\star[i] = j(\mathbf{s}[i], w_{r,i-1}, w_{r,i})$ and $j_1^\star[i] = j_1(\mathbf{s}[i], w_{r,i-1}, w_{r,i}, w_{d,i})$.

The average probability of error is given by

$$\Pr(\text{Error}) = \sum_{\mathbf{s} \in \mathcal{S}^n} \Pr(\mathbf{s}) \Pr(\text{error} \mid \mathbf{s})$$

$$\leq \sum_{\mathbf{s} \notin T_\epsilon^n(Q_S)} \Pr(\mathbf{s}) + \sum_{\mathbf{s} \in T_\epsilon^n(Q_S)} \Pr(\mathbf{s}) \Pr(\text{error} \mid \mathbf{s}). \tag{A.1}$$

The first term, $\Pr(\mathbf{s} \notin T_\epsilon^n(Q_S))$, on the RHS of (A.1) goes to zero as $n \to \infty$, by the strong asymptotic equipartition property (AEP) [43]. Thus, it is sufficient to upper bound the second term on the RHS of (A.1).

We now examine the probabilities of the error events associated with the encoding and decoding procedures. The error event is contained in the union of the following error events; where the events E_{1i} and E_{2i} correspond to encoding errors at block i; the events E_{3i} and E_{4i} correspond to decoding errors at the relay at block i; and the events E_{5i}, E_{6i}, E_{7i}, and E_{8i} correspond to decoding errors at the destination at block i.

Let E_{1i} be the event that there is no sequence $\mathbf{u}(w_{r,i-1}, w_{r,i}, j)$ jointly typical with $\mathbf{s}[i]$ given $\mathbf{x}_2(w_{r,i-1})$, that is,

$$E_{1i} = \{\nexists\, j \in \{1, \ldots, J\}$$

$$\text{s.t.} (\mathbf{x}_2(w_{r,i-1}), \mathbf{u}(w_{r,i-1}, w_{r,i}, j), \mathbf{s}[i]) \in T_\epsilon^n(P_{X_2,U,S})\}. \tag{A.2}$$

To bound the probability of the event E_{1i}, we use a standard argument [2]. More specifically, for $\mathbf{u}(w_{r,i-1}, w_{r,i}, j)$ and $\mathbf{s}[i]$ generated independently given $\mathbf{x}_2(w_{r,i-1})$ with i.i.d. components drawn according to $P_{U|X_2}$ and Q_S, respectively, the probability that $\mathbf{u}(w_{r,i-1}, w_{r,i}, j)$ is jointly typical with $\mathbf{s}[i]$ given $\mathbf{x}_2(w_{r,i-1})$ is greater than $(1-\epsilon)2^{-n(I(U;S|X_2)+\epsilon)}$ for sufficiently large n. There is a total of J such $\mathbf{u}$'s in each bin. The probability of the event E_{1i}, the probability that there is no such $\mathbf{u}$, is therefore bounded as

$$\Pr(E_{1i}) \leq \left[1 - (1-\epsilon)2^{-n(I(U;S|X_2)+\epsilon)}\right]^J. \tag{A.3}$$

Taking the logarithm on both sides of (A.3) and substituting J using (10) we obtain that $\ln(\Pr(E_{1i})) \leq -(1-\epsilon)2^{n\epsilon}$. Thus, $\Pr(E_{1i}) \to 0$ as $n \to \infty$.

Let E_{2i} be the event that there is no sequence $\mathbf{u}_{1w_{d,i}, j_1}(\mathbf{u}(w_{r,i-1}, w_{r,i}, j^\star))$ jointly typical with $\mathbf{s}[i]$ given $\mathbf{x}_2(w_{r,i-1})$ and $\mathbf{u}(w_{r,i-1}, w_{r,i}, j^\star)$, that is,

$$E_{2i} = \{\nexists\, j_1 \in \{1, \ldots, J_1\}$$

$$\text{s.t.} (\mathbf{x}_2(w_{r,i-1}), \mathbf{u}(w_{r,i-1}, w_{r,i}, j^\star[i]),$$

$$\mathbf{u}_{1w_{d,i}, j_1}(\mathbf{u}(w_{r,i-1}, w_{r,i}, j^\star[i])), \mathbf{s}[i]$$

$$\in T_\epsilon^n(P_{X_2,U,U_1,S})\}. \tag{A.4}$$

Proceeding similarly to for the event E_{1i} above, it can be shown that, conditioned on E_{1i}^c, the complement event of E_{1i}, we have $\Pr(E_{2i} \mid E_{1i}^c) \to 0$ as $n \to \infty$.

Let E_{3i} be the event that $\mathbf{u}(w_{r,i-1}, w_{r,i}, j^\star[i])$ and $\mathbf{y}_2[i]$ are not jointly typical given $\mathbf{x}_2(w_{r,i-1})$, that is,

$$E_{3i} = \{(\mathbf{x}_2(w_{r,i-1}), \mathbf{u}(w_{r,i-1}, w_{r,i}, j^\star[i]), \mathbf{y}_2[i]) \notin T_\epsilon^n(P_{X_2,U,Y_2})\}. \tag{A.5}$$

Conditioned on E_{1i}^c, E_{2i}^c, we have that $\mathbf{u}(w_{r,i-1}, w_{r,i}, j^\star[i])$ and $\mathbf{u}_{1w_{d,i}, j_1^\star[i]}(\mathbf{u}(w_{r,i-1}, w_{r,i}, j^\star[i]))$ are jointly typical with $(\mathbf{s}[i], \mathbf{x}_2(w_{r,i-1}))$, and with the source input $\mathbf{x}_1(w_{r,i-1}, w_{r,i}, w_{d,i})$. Thus, $\Pr(E_{3i} \mid E_{1i}^c, E_{2i}^c) \to 0$ as $n \to \infty$, by the Markov Lemma [43].

Let E_{4i} be the event that $\mathbf{u}(w_{r,i-1}, w'_{r,i}, j)$ and $\mathbf{y}_2[i]$ are jointly typical given $\mathbf{x}_2(w_{r,i-1})$ for some $w'_{r,i} \in \{1, \ldots, M\}$, $j \in \{1, \ldots, J\}$ with $w'_{r,i} \neq w_{r,i}$, that is,

$$E_{4i} = \Big\{\exists\, w'_{r,i} \in \{1, \ldots, M\}, j \in \{1, \ldots, J\} \text{ s.t. } w'_{r,i} \neq w_{r,i},$$

$$\left(\mathbf{x}_2(w_{r,i-1}), \mathbf{u}(w_{r,i-1}, w'_{r,i}, j), \mathbf{y}_2[i]\right) \in T_\epsilon^n(P_{X_2,U,Y_2})\Big\}. \tag{A.6}$$

Using the union bound and standard arguments on jointly typical sequences, the probability of the event E_{4i} conditioned on $E_{1i}^c, E_{2i}^c, E_{3i}^c$ can be easily bounded as

$$\Pr(E_{4i} \mid E_{1i}^c, E_{2i}^c, E_{3i}^c) \leq MJ2^{-n(I(U;Y_2|X_2)-\epsilon)}$$

$$= 2^{-n(I(U;Y_2|X_2)-I(U;S|X_2)-R_r+\epsilon)}. \tag{A.7}$$

Thus, $\Pr(E_{4i} \mid E_{1i}^c, E_{2i}^c, E_{3i}^c) \to 0$ as $n \to \infty$ if $R_r \leq I(U; Y_2 \mid X_2) - I(U; S \mid X_2)$.

For decoding the triple $(w_{r,i-1}, j^\star[i-1], w_{d,i-1})$ at the destination, let E_{5i} be the union of the following two events,

$$E_{5i}^{(1)} = \{(\mathbf{x}_2(w_{r,i-2}), \mathbf{u}(w_{r,i-2}, w_{r,i-1}, j^\star[i-1]),$$

$$\mathbf{u}_{1w_{d,i-1}, j_1^\star[i-1]}(\mathbf{u}(w_{r,i-2}, w_{r,i-1}, j^\star[i-1])),$$

$$\mathbf{y}_3[i-1]) \notin T_\epsilon^n(P_{X_2,U,U_1,Y_3})\} \tag{A.8}$$

$$E_{5i}^{(2)} = \{(\mathbf{x}_2(w_{r,i-1}), \mathbf{y}_3[i]) \notin T_\epsilon^n(P_{X_2,Y_3})\}.$$

For $\mathbf{s}[i-1], \mathbf{x}_2(w_{r,i-2}), \mathbf{u}(w_{r,i-2}, w_{r,i-1}, j^\star[i-1]), \mathbf{u}_{1w_{d,i-1}, j_1^\star[i-1]}$
$\times (\mathbf{u}(w_{r,i-2}, w_{r,i-1}, j^\star[i-1]))$ and $\mathbf{x}_1(w_{r,i-2}, w_{r,i-1}, w_{d,i-1})$

jointly typical, $\Pr(E_{5i}^{(1)} \mid \cap_{k=1}^{4} E_{ki}^{c}) \to 0$ as $n \to \infty$ by the Markov Lemma. Similarly, $\Pr(E_{5i}^{(2)} \mid \cap_{k=1}^{4} E_{ki}^{c}) \to 0$ as $n \to \infty$. Thus, $\Pr(E_{5i} \mid \cap_{k=1}^{4} E_{ki}^{c}) \to 0$ as $n \to \infty$.

For decoding the triple $(w_{r,i-1}, j^{\star}[i-1], w_{d,i-1})$ at the destination, let E_{6i} be the event

$$
\begin{aligned}
E_{6i} = \Big\{ & \exists\ w'_{r,i-1} \in \{1, \ldots, M\}, j \in \{1, \ldots, J\}, w_{d,i-1} \\
& \in \{1, \ldots, M\}, j_1 \in \{1, \ldots, J_1\} \text{ s.t. } w'_{r,i-1} \neq w_{r,i-1}, \\
& \Big\{ \big(\mathbf{x}_2(w_{r,i-2}), \mathbf{u}\big(w_{r,i-2}, w'_{r,i-1}, j\big), \\
& \quad \mathbf{u}_{1 w_{d,i-1}, j_1}\big(\mathbf{u}\big(w_{r,i-2}, w'_{r,i-1}, j\big)\big), \\
& \quad \mathbf{y}_3[i-1]\big) \in T_{\epsilon}^{n}\big(P_{X_2, U, U_1, Y_3}\big) \Big\}, \\
& \Big\{ \big(\mathbf{x}_2\big(w'_{r,i-1}\big), \mathbf{y}_3[i] \big) \in T_{\epsilon}^{n}\big(P_{X_2, Y_3}\big) \Big\} \Big\}.
\end{aligned}
$$
(A.9)

Conditioned on the events $E_{1i}^{c}, E_{2i}^{c}, E_{3i}^{c}, E_{4i}^{c}, E_{5i}^{c}$, the probability of the event E_{6i} can be bounded using the union bound, as

$$
\begin{aligned}
\Pr\Big(E_{6i} \mid \bigcap_{k=1}^{5} E_{ki}^{c}\Big) &\leq MM_1 JJ_1 2^{-n(I(U, U_1; Y_3 | X_2) - \epsilon)} 2^{-n(I(X_2; Y_3) - \epsilon)} \\
&= 2^{-n(I(U, U_1, X_2; Y_3) - I(U_1; S | X_2) - (R_r + R_d) + 4\epsilon)}.
\end{aligned}
$$
(A.10)

Thus, $\Pr(E_{6i} \mid \cap_{k=1}^{5} E_{ki}^{c}) \to 0$ as $n \to \infty$ if $R_r + R_d \leq I(U, U_1, X_2; Y_3) - I(U, U_1; S \mid X_2)$.

For decoding the triple $(w_{r,i-1}, j^{\star}[i-1], w_{d,i-1})$ at the destination, let E_{7i} be the event

$$
\begin{aligned}
E_{7i} = \Big\{ & \exists\ w'_{d,i-1} \in \{1, \ldots, M_1\}, j_1 \in \{1, \ldots, J_1\} \\
& \text{s.t. } w'_{d,i-1} \neq w_{d,i-1}, \\
& \{ (\mathbf{x}_2(w_{r,i-2}), \mathbf{u}(w_{r,i-2}, w_{r,i-1}, j^{\star}[i-1]), \\
& \quad \mathbf{u}_{1 w'_{d,i-1}, j_1}\big(\mathbf{u}(w_{r,i-2}, w_{r,i-1}, j^{\star}[i-1])\big), \\
& \quad \mathbf{y}_3[i-1]) \in T_{\epsilon}^{n}\big(P_{X_2, U, U_1, Y_3}\big) \}, \\
& \{ (\mathbf{x}_2(w_{r,i-1}), \mathbf{y}_3[i]) \in T_{\epsilon}^{n}\big(P_{X_2, Y_3}\big) \} \Big\}.
\end{aligned}
$$
(A.11)

Conditioned on $\cap_{k=1}^{6} E_{ki}^{c}$, the probability of the event E_{7i} can be bounded using the union bound, as

$$
\begin{aligned}
\Pr\Big(E_{7i} \mid \bigcap_{k=1}^{6} E_{ki}^{c}\Big) &\leq M_1 J_1 2^{-n(I(U_1; Y_3 | U, X_2) - \epsilon)} \\
&\quad \times \Pr\{ (\mathbf{x}_2(w_{r,i-1}), \mathbf{y}_3[i]) \in T_{\epsilon}^{n}(P_{X_2, Y_3}) \} \\
&\leq 2^{-n(I(U_1; Y_3 | U, X_2) - I(U_1; S | U, X_2) - R_d + 3\epsilon)}.
\end{aligned}
$$
(A.12)

Thus, $\Pr(E_{7i} \mid \cap_{k=1}^{6} E_{ki}^{c}) \to 0$ as $n \to \infty$ if $R_d \leq I(U_1; Y_3 \mid U, X_2) - I(U_1; S \mid U, X_2)$.

For decoding the triple $(w_{r,i-1}, j^{\star}[i-1], w_{d,i-1})$ at the destination, let E_{8i} be the event

$$
\begin{aligned}
E_{8i} = \Big\{ & \exists\ w'_{d,i-1} \in \{1, \ldots, M_1\},\ j_1 \in \{1, \ldots, J_1\}, \\
& j' \in \{1, \ldots, J\} \text{ s.t. } w'_{d,i-1} \neq w_{d,i-1}, j' \neq j^{\star}[i-1], \\
& \{ (\mathbf{x}_2(w_{r,i-2}), \mathbf{u}(w_{r,i-2}, w_{r,i-1}, j'), \\
& \quad \mathbf{u}_{1 w'_{d,i-1}, j_1}\big(\mathbf{u}(w_{r,i-2}, w_{r,i-1}, j')\big), \\
& \quad \mathbf{y}_3[i-1]) \in T_{\epsilon}^{n}\big(P_{X_2, U, U_1, Y_3}\big) \}, \\
& \{ (\mathbf{x}_2(w_{r,i-1}), \mathbf{y}_3[i]) \in T_{\epsilon}^{n}\big(P_{X_2, Y_3}\big) \} \Big\}.
\end{aligned}
$$
(A.13)

Conditioned on $\cap_{k=1}^{7} E_{ki}^{c}$, the probability of the event E_{8i} can be bounded using the union bound, as

$$
\begin{aligned}
\Pr\Big(E_{8i} \mid \bigcap_{k=1}^{7} E_{ki}^{c}\Big) &\leq M_1 J_1 J 2^{-n(I(U, U_1; Y_3 | X_2) - \epsilon)} \\
&\quad \times \Pr\{ (\mathbf{x}_2(w_{r,i-1}), \mathbf{y}_3[i]) \in T_{\epsilon}^{n}(P_{X_2, Y_3}) \} \\
&\leq 2^{-n(I(U, U_1; Y_3 | X_2) - I(U, U_1; S | X_2) - R_d + \epsilon)}.
\end{aligned}
$$
(A.14)

Thus, $\Pr(E_{8i} \mid \cap_{k=1}^{7} E_{ki}^{c}) \to 0$ as $n \to \infty$ if $R_d \leq I(U, U_1; Y_3 \mid X_2) - I(U, U_1; S \mid X_2)$.

From the above, we have that the average probability of error goes to zero for sufficiently large n if the rate R is chosen to satisfy

$$
R = R_r + R_d\ : R_r \geq 0, R_d \geq 0,
$$
(A.15a)

$$
R_r \leq I(U; Y_2 \mid X_2) - I(U; S \mid X_2),
$$
(A.15b)

$$
R_d \leq I(U_1; Y_3 \mid U, X_2) - I(U_1; S \mid U, X_2),
$$
(A.15c)

$$
R_d \leq I(U, U_1; Y_3 \mid X_2) - I(U, U_1; S \mid X_2),
$$
(A.15d)

$$
R \leq I(U, U_1, X_2; Y_3) - I(U, U_1; S \mid X_2).
$$
(A.15e)

The set of rates defined by (A.15a), (A.15b), (A.15c), (A.15d), (A.15e) does not change if one adds the two additional rate bounds

$$
\begin{aligned}
R &\leq I(U; Y_2 \mid X_2) + I(U_1; Y_3 \mid U, X_2) \\
&\quad - I(U, U_1; S \mid X_2)
\end{aligned}
$$
(A.16a)

$$
\begin{aligned}
R &\leq I(U; Y_2 \mid X_2) - I(U; S \mid X_2) \\
&\quad + I(U, U_1; Y_3 \mid X_2) - I(U, U_1; S \mid X_2),
\end{aligned}
$$
(A.16b)

as (A.16a) can be obtained by adding (A.15b) and (A.15d); and (A.16b) can be obtained by adding (A.15b) and (A.15e). Then, we obtain the rate in Theorem 1 by applying *Fourier-Motzkin elimination* (FME) (see, e.g., [44]) to eliminate the variables R_r and R_d from the obtained system of rate

inequalities. Note that eliminating the variable R_r by FME algorithm adds the constraint (7a), and eliminating the variable R_d adds the constraints (7b) and (7c).

This proves the achievability of the rate (9) in Theorem 1 for every measure $P_{S,U,U_1,X_1,X_2,Y_2,Y_3}$ of the form (6) that satisfies (7a), (7b), (7c), that is, $P_{S,U,U_1,X_1,X_2,Y_2,Y_3} \in \mathcal{P}$.

It remains to show that the rate (9) is not altered if one restricts the random variables U and U_1 to have their alphabet sizes limited as in (8a), (8b). This is done by invoking the support lemma [45, page 310], as follows.

Fix a distribution $\mu \in \mathcal{P}$ of $(S, U, U_1, X_1, X_2, Y_2, Y_3)$ on $\mathcal{P}(\mathcal{S} \times \mathcal{U} \times \mathcal{U}_1 \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}_2 \times \mathcal{Y}_3)$. To prove the bound (8a) on $|\mathcal{U}|$, note that we have

$$I_\mu(U; Y_2 \mid X_2) + I_\mu(U_1; Y_3 \mid U, X_2) - I_\mu(U, U_1; S \mid X_2)$$

$$= I_\mu(U; Y_2, X_2) - I_\mu(U; S, X_2) + I_\mu(U_1; Y_3, X_2 \mid U)$$

$$\quad - I_\mu(U_1; S, X_2 \mid U)$$

$$= H_\mu(X_2, Y_2) - H_\mu(X_2, S)$$

$$\quad + H_\mu(X_2, Y_3 \mid U) - H_\mu(X_2, Y_2 \mid U)$$

$$\quad + H_\mu(U_1, S, X_2 \mid U) - H_\mu(U_1, X_2, Y_3 \mid U).$$

$$(A.17)$$

Similarly, we have

$$I_\mu(U; Y_2 \mid X_2) - I_\mu(U; S \mid X_2) + I_\mu(U, U_1; Y_3 \mid X_2)$$

$$\quad - I_\mu(U, U_1; S \mid X_2)$$

$$= I_\mu(U; Y_2 \mid X_2) + I_\mu(U_1; Y_3 \mid U, X_2)$$

$$\quad - I_\mu(U, U_1; S \mid X_2) + I_\mu(U; Y_3 \mid X_2) - I_\mu(U; S \mid X_2)$$

$$= H_\mu(X_2, Y_2) + H_\mu(X_2, Y_3) - 2H_\mu(X_2, S)$$

$$\quad + H_\mu(U_1, X_2, Y_3 \mid U) - H_\mu(X_2, Y_2 \mid U)$$

$$\quad + H_\mu(U_1, S, X_2 \mid U) + H_\mu(S, X_2 \mid U),$$

$$(A.18)$$

$$I_\mu(U, U_1, X_2; Y_3) - I_\mu(U, U_1; S \mid X_2)$$

$$= I_\mu(U; Y_3) + I_\mu(U_1, X_2; Y_3 \mid U) + I_\mu(S; X_2) - I_\mu(S; U)$$

$$\quad - I_\mu(U_1, X_2; S \mid U)$$

$$= H_\mu(Y_3) - H_\mu(S \mid X_2) + H_\mu(U_1, X_2, S \mid U)$$

$$\quad - H_\mu(U_1, X_2, Y_3 \mid U).$$

$$(A.19)$$

Hence, it suffices to show that the following functionals of $\mu(S, U_1, X_1, X_2, Y_2, Y_3)$

$$r_{s,x,x'}(\mu) = \mu(s, x, x') \quad \forall (s, x, x') \in \mathcal{S} \times \mathcal{X}_1 \times \mathcal{X}_2,$$

$$r_1(\mu) = \int_u d_\mu(u) \big[H_\mu(X_2, Y_3 \mid u) - H_\mu(X_2, Y_2 \mid u)$$

$$\quad + H_\mu(U_1, S, X_2 \mid u)$$

$$\quad - H_\mu(U_1, X_2, Y_3 \mid u) \big],$$

$$r_2(\mu) = \int_u d_\mu(u) \big[H_\mu(U_1, X_2, Y_3 \mid u) - H_\mu(X_2, Y_2 \mid u)$$

$$\quad + H_\mu(U_1, S, X_2 \mid u) + H_\mu(S, X_2 \mid u) \big],$$

$$r_3(\mu) = \int_u d_\mu(u) \big[H_\mu(U_1, X_2, S \mid u) - H_\mu(U_1, X_2, Y_3 \mid u) \big]$$

$$(A.20)$$

can be preserved with another measure $\mu' \in \mathcal{P}$—for given $(s, x, x') \in \mathcal{S} \times \mathcal{X}_1 \times \mathcal{X}_2$, $r_{s,x,x'}$ is obtained by marginalizing the measure μ. Observing that there is a total of $|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2| + 2$ functionals in (A.20), this is ensured by a standard application of the support lemma; and this shows that the cardinality of the alphabet of the auxiliary random variable U can be limited as indicated in (8a) without altering the rate (9).

Once the alphabet of U is fixed, we apply similar arguments to bound the alphabet of U_1, where this time $|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2|(|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2| + 2) - 1$ functionals must be satisfied in order to preserve the joint distribution of S, U, X_1, X_2, and two more functionals to preserve

$$I_\mu(U; Y_2 \mid X_2) + I_\mu(U_1; Y_3 \mid U, X_2) - I_\mu(U, U_1; S \mid X_2)$$

$$= I_\mu(U; X_2, Y_2) - I_\mu(U; X_2, S) + I_\mu(U_1; U, X_2, Y_3)$$

$$\quad - I_\mu(U_1; U, X_2, S)$$

$$= H_\mu(X_2, Y_2) + H_\mu(U, X_2, Y_3) - H_\mu(U, X_2, Y_2)$$

$$\quad - H_\mu(S, X_2) - H_\mu(U, X_2, Y_3 U_1) + H_\mu(U, X_2, SU_1),$$

$$I_\mu(U, U_1, X_2; Y_3) - I_\mu(U, U_1; S \mid X_2)$$

$$= H_\mu(Y_3) - H_\mu(S \mid X_2) + H_\mu(U, X_2, S \mid U_1)$$

$$\quad - H_\mu(U, X_2, Y_3 U_1),$$

$$(A.21)$$

yielding the bound on $|U_1|$ indicated in (8b).

B. Proof of Theorem 3

In this proof, we compute the lower bound in Theorem 1 using an appropriate jointly Gaussian distribution on S, U, U_1, X_1, X_2. We assume that X_1 and X_2 are zero-mean Gaussian with variance P_1 and P_2, respectively. The random variable X_2 is independent from S as shown by the distribution in Theorem 1. The random variable X_1 is decomposed as $X_1 = X_{1r} + X_{1d}$, where: X_{1r} is zero mean Gaussian with variance P_{1r}, is independent from X_{1d} and is jointly Gaussian with both S and X_2 with $\mathbb{E}[X_{1r}S] = \rho_{1s}\sqrt{P_{1r}Q}$ and $\mathbb{E}[X_{1r}X_2] = \rho_{12}\sqrt{P_{1r}P_2}$, for some $\rho_{1s} \in [-1, 1]$, $\rho_{12} \in [-1, 1]$; and X_{1d} is zero mean Gaussian with variance $P_{1d} = P_1 - P_{1r}$, and is independent from both the relay input X_2 and the state S. In what follows we will also use the covariances $\sigma_{12} = \mathbb{E}[X_1X_2]$ and $\sigma_{1s} = \mathbb{E}[X_1S]$, satisfying

$$\rho_{12} = \frac{\sigma_{12}}{\sqrt{P_{1r}P_2}}, \qquad \rho_{1s} = \frac{\sigma_{1s}}{\sqrt{P_{1r}Q}}. \qquad (B.1)$$

As we outlined after Theorem 3, we consider

$$U = X_{1r} + \alpha S,$$

$$U_1 = X_{1d} + \frac{P_{1d}}{P_{1d} + N_3}(1 - \alpha)S, \tag{B.2}$$

where α denotes a scale parameter the range of which will be specified below.

(i) Let us first compute the first term of the minimization in (9). Consider the term $[I(U; Y_2 \mid X_2) - I(U; S \mid X_2)]$. To evaluate the conditional mutual information $I(U; Y_2 \mid X_2)$, let us denote by $\mathbb{E}[Y_2 \mid U, X_2]$ the optimal linear estimator of Y_2 given (U, X_2) under minimum mean square error criterion. That is, $\mathbb{E}[Y_2 \mid U, X_2]$ is given by

$$\mathbb{E}[Y_2 \mid U, X_2] = \gamma_1 U + \gamma_2 X_2 \tag{B.3}$$

with

$$\gamma_1 = \frac{P_{1r} - (\sigma_{12}^2/P_2) + (\alpha + 1)\sigma_{1s} + \alpha Q}{P_{1r} - (\sigma_{12}^2/P_2) + 2\alpha\sigma_{1s} + \alpha^2 Q},$$

$$\gamma_2 = (1 - \gamma_1)\frac{\sigma_{12}}{P_2}. \tag{B.4}$$

For convenience let us define Δ as the denominator of the expression of γ_1 (for notational convenience, we omit the dependency of Δ on parameters α, ρ_{12} and ρ_{1s}), that is,

$$\Delta := P_{1r} - \frac{\sigma_{12}^2}{P_2} + 2\alpha\sigma_{1s} + \alpha^2 Q. \tag{B.5}$$

Then, we have

$I(U; Y_2 \mid X_2)$

$$= h(Y_2 \mid X_2) - h(Y_2 \mid U, X_2)$$

$$\overset{(a)}{=} \frac{1}{2}\log(\mathbb{E}[Y_2^2] - \mathbb{E}[Y_2\mathbb{E}[Y_2 \mid X_2]])$$

$$\quad - \frac{1}{2}\log(\mathbb{E}[Y_2^2] - \mathbb{E}[Y_2\mathbb{E}[Y_2 \mid U, X_2]])$$

$$\overset{(b)}{=} \frac{1}{2}\log\left(\frac{\mathbb{E}[Y_2^2] - (\sigma_{12}^2/P_2)}{\mathbb{E}[Y_2^2] - (\sigma_{12}^2/P_2) - \gamma_1^2\Delta}\right)$$

$$\overset{(c)}{=} \frac{1}{2}\log\left(\frac{P_{1r} - (\sigma_{12}^2/P_2) + 2\sigma_{1s} + Q + N_2 + P_{1d}}{N_2 + P_{1d} + P_{1r}Q(1 - \rho_{12}^2 - \rho_{1s}^2)(1 - \alpha)^2/\Delta}\right)$$

$$= \frac{1}{2}\log\left(\frac{(P_{1r} - (\sigma_{12}^2/P_2) + 2\sigma_{1s} + Q + N_2 + P_{1d})\Delta}{(N_2 + P_{1d})\Delta + P_{1r}Q(1 - \rho_{12}^2 - \rho_{1s}^2)(1 - \alpha)^2}\right), \tag{B.6}$$

where (a) follows since the vector (U, X_2, Y_2) is Gaussian; in (b) we used the fact that $\mathbb{E}[Y_2 \mid X_2] = (\sigma_{12}/P_2)X_2$ and

$$\mathbb{E}[Y_2\mathbb{E}[Y_2 \mid U, X_2]]$$

$$= \gamma_1(P_{1r} + (\alpha + 1)\sigma_{1s} + \alpha Q) + (1 - \gamma_1)\frac{\sigma_{12}^2}{P_2} \tag{B.7}$$

$$= \frac{\sigma_{12}^2}{P_2} + \gamma_1^2\Delta,$$

and (b) follows by substituting γ_1 using (B.4) and using straightforward algebra to obtain

$$\mathbb{E}[Y_2^2] - \frac{\sigma_{12}^2}{P_2} - \gamma_1^2\Delta$$

$$= N_2 + P_{1d} + \frac{1}{\Delta}\Big[(P_{1r}(1 - \rho_{12}^2) + 2\sigma_{1s} + Q)\Delta$$

$$\quad - (P_{1r}(1 - \rho_{12}^2) + (\alpha + 1)\sigma_{1s} + \alpha Q)^2\Big]$$

$$= N_2 + P_{1d} + \frac{P_{1r}Q}{\Delta}\Big[(1 + \alpha^2)(1 - \rho_{12}^2) + 4\alpha\rho_{1s}^2$$

$$\quad - (1 + \alpha)^2\rho_{1s}^2 - 2\alpha(1 - \rho_{12}^2)\Big]$$

$$= N_2 + P_{1d} + \frac{P_{1r}Q(1 - \rho_{12}^2 - \rho_{1s}^2)(1 - \alpha)^2}{\Delta}. \tag{B.8}$$

Similarly, to evaluate the conditional mutual information $I(U; S \mid X_2)$, let $\mathbb{E}[S \mid U, X_2]$ be the optimal linear estimator of S given (U, X_2) under minimum mean square error criterion, that is,

$$\mathbb{E}[S \mid U, X_2] = \lambda_1 U + \lambda_2 X_2 \tag{B.9}$$

with

$$\lambda_1 = \frac{\sigma_{1s} + \alpha Q}{\Delta},$$

$$\lambda_2 = -\frac{\sigma_{12}}{P_2}\frac{\sigma_{1s} + \alpha Q}{\Delta}. \tag{B.10}$$

Then, we have

$I(U; S \mid X_2)$

$$= h(S \mid X_2) - h(S \mid U, X_2)$$

$$= h(S) - h(S \mid U, X_2)$$

$$= \frac{1}{2}\log\mathbb{E}[S^2] - \frac{1}{2}\log(\mathbb{E}[S^2] - \mathbb{E}[S\mathbb{E}[S \mid U, X_2]])$$

$$= \frac{1}{2}\log Q - \frac{1}{2}\log(Q - \lambda_1(\sigma_{1s} + \alpha Q))$$

$$= \frac{1}{2}\log Q - \frac{1}{2}\log\left(Q - \frac{(\sigma_{1s} + \alpha Q)^2}{\Delta}\right)$$

$$= \frac{1}{2}\log\left(\frac{\Delta}{P_{1r}(1 - \rho_{12}^2 - \rho_{1s}^2)}\right). \tag{B.11}$$

Subtracting (B.11) from (B.6) and substituting $\xi := 1 - \rho_{12}^2 - \rho_{1s}^2$ and $\rho := \rho_{1s}$, we obtain

$$I(U; Y_2 \mid X_2) - I(U; S \mid X_2) = \frac{1}{2} \log \left(\frac{P_{1r} \xi \left[P_{1r} \xi + (\rho \sqrt{P_{1r}} + \sqrt{Q})^2 + N_2 + P_{1d} \right]}{P_{1r} Q \xi (1-\alpha)^2 + (N_2 + P_{1d}) \left[P_{1r} \xi + (\rho \sqrt{P_{1r}} + \alpha \sqrt{Q})^2 \right]} \right) \tag{B.12}$$

$$:= R_1(P_{1r}, Q, N_2 + P_{1d}, \xi, \rho, \alpha).$$

For the computation of the first term of the minimization in (9), it remains to compute $[I(U_1; Y_3 \mid U, X_2) - I(U_1; S \mid U, X_2)]$. Noting that $Y_3 = X_{1d} + U + X_2 + (1-\alpha)S + Z_3$, it can be easily shown that

$$I(U_1; Y_3 \mid U, X_2) - I(U_1; S \mid U, X_2) = \frac{1}{2} \log \left(1 + \frac{P_{1d}}{N_3} \right). \tag{B.13}$$

(ii) Consider now the second term of the minimization in (9). This term can be written as the sum of the first term of the minimization in (9) plus $[I(U; Y_3 \mid X_2) - I(U; S \mid X_2)]$. The term $[I(U; Y_3 \mid X_2) - I(U; S \mid X_2)]$ can be computed in a very similar way to $[I(U; Y_2 \mid X_2) - I(U; S \mid X_2)]$ above, by noting that

$$I(U; Y_3 \mid X_2) - I(U; S \mid X_2) = I(U; Y_3 - X_2 \mid X_2) - I(U; S \mid X_2) \tag{B.14}$$

and the two terms $Y_3 - X_2 = X_1 + S + Z_2$ and Y_2 differ only through the noise terms. That is,

$$I(U; Y_3 \mid X_2) - I(U; S \mid X_2) = R_1(P_{1r}, Q, N_3 + P_{1d}, \xi, \rho, \alpha), \tag{B.15}$$

which is readily obtained by substituting N_2 with N_3 on the RHS of (B.12).

(iii) Finally, we compute the third term of the minimization in (12). Using (B.13), we obtain

$$I(U, U_1, X_2; Y_3) - I(U, U_1; S \mid X_2)$$

$$= I(U, X_2; Y_3) - I(U; S \mid X_2) + \frac{1}{2} \log \left(1 + \frac{P_{1d}}{N_3} \right). \tag{B.16}$$

The term $I(U, X_2; Y_3) - I(U; S \mid X_2)$ on the RHS of (B.16) can be written as

$$I(U, X_2; Y_3) - I(U; S \mid X_2)$$

$$= I(X_2; Y_3) + I(U; Y_3 \mid X_2) - I(U; S \mid X_2) \tag{B.17}$$

$$\overset{(d)}{=} I(X_2; Y_3) + R_1(P_{1r}, Q, N_3 + P_{1d}, \xi, \rho, \alpha),$$

where (d) follows by using (B.15). The mutual information $I(X_2; Y_3)$ in (B.17) can be computed as

$$I(X_2; Y_3) = h(Y_3) - h(Y_3 \mid X_2)$$

$$\overset{(e)}{=} \frac{1}{2} \log \mathbb{E}[Y_3^2] - \frac{1}{2} \log (\mathbb{E}[Y_3^2] - \mathbb{E}[Y_3 \mathbb{E}[Y_3 \mid X_2]])$$

$$\overset{(f)}{=} \frac{1}{2} \log \mathbb{E}[Y_3^2] - \frac{1}{2} \log \left(P_1 + P_2 + 2\sigma_{12} + 2\sigma_{1s} + Q + N_3 - \frac{(\sigma_{12} + P_2)^2}{P_2} \right)$$

$$= \frac{1}{2} \log \left(\frac{P_{1r} \xi + \left(\sqrt{1 - \xi - \rho^2} \sqrt{P_{1r}} + \sqrt{P_2} \right)^2 + (\rho \sqrt{P_{1r}} + \sqrt{Q})^2 + N_3 + P_{1d}}{P_{1r} \xi + (\rho \sqrt{P_{1r}} + \sqrt{Q})^2 + N_3 + P_{1d}} \right) \tag{B.18}$$

$$= \frac{1}{2} \log \left(1 + \frac{\left(\sqrt{1 - \xi - \rho^2} \sqrt{P_{1r}} + \sqrt{P_2} \right)^2}{P_{1r} \xi + (\rho \sqrt{P_{1r}} + \sqrt{Q})^2 + N_3 + P_{1d}} \right),$$

and (e) follows since the vector (X_2, Y_3) is Gaussian, and in (f) we used the fact that $\mathbb{E}[Y_3 \mid X_2] = (1 + \sigma_{12}/P_2)X_2$.

Then, substituting $I(X_2; Y_3)$ in (B.17) using (B.18), we obtain

$$I(U, X_2; Y_3) - I(U; S \mid X_2) = \frac{1}{2} \log \left(\frac{P_{1r}\xi \left[P_{1r}\xi + \left(\sqrt{1 - \xi - \rho^2}\sqrt{P_{1r}} + \sqrt{P_2} \right)^2 + \left(\rho\sqrt{P_{1r}} + \sqrt{Q} \right)^2 + N_3 + P_{1d} \right]}{P_{1r}Q\xi(1-\alpha)^2 + (N_3 + P_{1d})\left[P_{1r}\xi + \left(\rho\sqrt{P_{1r}} + \alpha\sqrt{Q} \right)^2 \right]} \right) \quad \text{(B.19)}$$

$$= R_2(P_{1r}, P_2, Q, N_3 + P_{1d}, \xi, \rho, \alpha).$$

Finally, we obtain the rate achievable with the considered jointly Gaussian distribution on (S, U, U_1, X_1, X_2) by maximizing the minimum among two terms: (a) the sum of the RHS of (B.12), the RHS of (B.13) and the minimum among zero and the RHS of (B.15); and (b) the sum of the RHSs of (B.13) and (B.19). For the choice of coding parameters for which the RHS of (B.15) has nonnegative values, we obtain the rate in (22). (Note that the minimization is in fact among three terms, but two of which can be replaced by one single term as in (a).) The maximization is over all values of $\rho_{12} \in [-1, 1]$ and $\rho_{1s} \in [-1, 1]$ such that the covariance matrix $\Lambda_{S, X_1, X_2, Z_2, Z_3}$ of (S, X_1, X_2, Z_2, Z_3) has a nonnegative discriminant, that is, for $Q > 0$,

$$\xi = 1 - \rho_{12}^2 - \rho_{1s}^2 \geq 0, \quad \text{(B.20)}$$

and all values of α such that $R_1(P_{1r}, Q, N_2 + P_{1d}, \xi, \rho, \alpha)$ and $R_1(P_{1r}, Q, N_3 + P_{1d}, \xi, \rho, \alpha)$ have nonnegative values. (Note that if the RHS of (B.15) has nonnegative values, so does the RHS of (B.19); this can be easily seen from (B.17).) Furthermore, investigating $R_1(P_{1r}, Q, N_2 + P_{1d}, \xi, \rho, \alpha)$ and $R_2(P_{1r}, P_2, Q, N_3 + P_{1d}, \xi, \rho, \alpha)$ (with $\xi = 1 - \rho_{12}^2 - \rho_{1s}^2$), one can easily see that it suffices to consider $\rho_{12} \in [0, 1]$ and $\rho_{1s} \in [-1, 0]$. Thus, $\xi \in [0, 1]$ and $\rho \in [-\sqrt{1 - \xi}, 0]$. This concludes the proof of Theorem 3.

C. Proof of Theorem 5

As we outlined after Theorem 5, we divide the total transmission time into two periods of duration νT and $\bar{\nu}T$, respectively. Also we decompose the message W to be transmitted into two submessages W_a and W_d. During the period νT, the source transmits message W_a to both the relay and the destination and message W_d to only the relay. The relay decodes-and-forwards both messages during this period. During the period $\bar{\nu}T$, the source transmits message W_a to both the relay and the destination and message W_d to only the destination. The relay decodes-and-forwards only message W_a during this period. The relay operates in a full-duplex mode during both transmission periods.

Fix nonnegative $P_{1a}, P_{1d}, P_{2a}, P_{2d}$ such that $P_{1a} + P_{1d} \leq P_1$ and $P_{2a} + P_{2d} \leq P_2$; $\rho_{12} \in [-1, 1], \rho_{1s} \in [-1, 1]$ and $\nu \in [0, 1]$. Let $R_a^{(1)}$ and $R_a^{(2)}$ denote the rates at which message W_a is transmitted to the destination during the period νT and the period $\bar{\nu}T$, respectively. Similarly, let $R_d^{(1)}$ and $R_d^{(2)}$ denote the rates at which message W_d is transmitted to the destination during these periods. The total rate during the period νT is $T_1 = R_a^{(1)} + R_d^{(1)}$, and the total rate during the period $\bar{\nu}T$ is $T_2 = R_a^{(2)} + R_d^{(2)}$.

We assume that the channel states $S_2^{(1)}$ and $S_3^{(1)}$ during the period νT, and the channel states $S_2^{(2)}$ and $S_3^{(2)}$ during the period $\bar{\nu}T$, are zero-mean Gaussian with variance Q. Also, we let $S_a^{(k)} = (S_2^{(k)} + S_3^{(k)})/2$ and $S_d^{(k)} = (S_2^{(k)} - S_3^{(k)})/2$, $k = 1, 2$.

The encoding and transmission scheme during the two periods is as follows.

(1) Period νT. During this period the source sends a superposition of two Gaussian signals, $X_{1,i}^{(1)} = X_{1a,i}^{(1)} + X_{1d,i}^{(1)}$. The signal $X_{1a,i}^{(1)}$ carries message W_a and is obtained via a GDPC considering $(S_a^{(1)})^n$ as noncausal CSI. For this GDPC, we use the following auxiliary random variable to generate the auxiliary codewords $U_{1a,i}^{(1)}$,

$$U_{1a}^{(1)} = X_{1a}^{(1)} + \alpha_a S_a^{(1)}, \quad \text{(C.1)}$$

where $X_{1a}^{(1)} \sim \mathcal{N}(0, P_{1a})$ is jointly Gaussian with $S_a^{(1)}$, with $\mathbb{E}[X_{1a}^{(1)} S_a^{(1)}] = (\rho_{1s}/\sqrt{2})\sqrt{P_{1a}Q}$, and is independent from $S_d^{(1)}$; and α_a is the GDPC scale parameter. Thus, using this GDPC, $X_{1a,i}^{(1)}$ is obtained as

$$X_{1a,i}^{(1)} = U_{1a,i}^{(1)} - \alpha_a S_{a,i}^{(1)}, \quad \text{(C.2)}$$

where $U_{1a,i}^{(1)}$ is independently drawn with $U_{1a}^{(1)}$.

The signal $X_{1d,i}^{(1)}$ carries message W_d and is obtained via a standard DPC considering $[(1 - \alpha_a)(S_a^{(1)})^n + (S_d^{(1)})^n]$ as noncausal CSI. For this standard DPC, we use the following auxiliary random variable to generate the auxiliary codewords $U_{1d,i}^{(1)}$,

$$U_{1d}^{(1)} = X_{1d}^{(1)} + \frac{P_{1d}}{P_{1d} + N_2}\left[(1 - \alpha_a)S_a^{(1)} + S_d^{(1)} \right], \quad \text{(C.3)}$$

where $X_{1d}^{(1)} \sim \mathcal{N}(0, P_{1d})$ is independent from $X_{1a}^{(1)}$, $S_a^{(1)}$ and $S_d^{(1)}$. Thus, using this DPC, $X_{1d,i}^{(1)}$ is obtained as

$$X_{1d,i}^{(1)} = U_{1d,i}^{(1)} - \frac{P_{1d}}{P_{1d} + N_2}\left[(1 - \alpha_a)S_{a,i}^{(1)} + S_{d,i}^{(1)} \right], \quad \text{(C.4)}$$

where $U_{1d,i}^{(1)}$ is independently drawn with $U_{1d}^{(1)}$.

Suppose that the relay decodes correctly messages W_a and W_d (this will be justified below), then the relay sends a superposition of two Gaussian signals, $X_{2,i}^{(1)} = X_{2a,i}^{(1)} + X_{2d,i}^{(1)}$, where $X_{2a,i}^{(1)}$ carries message W_a and is independently drawn with $X_{2a}^{(1)} \sim \mathcal{N}(0, P_{2a})$, and $X_{2d,i}^{(1)}$ carries message W_d and is independently drawn with $X_{2d}^{(1)} \sim \mathcal{N}(0, P_{2d})$. Here $X_{2a}^{(1)}$ and $X_{2d}^{(1)}$ are mutually independent and are independent

from both $S_a^{(1)}$ and $S_d^{(1)}$. Furthermore, $X_{2a}^{(1)}$ is independent from $X_{1d}^{(1)}$ and is correlated with $X_{1a}^{(1)}$ with $\mathbb{E}[X_{1a}^{(1)}X_{2a}^{(1)}] = \rho_{12}\sqrt{P_{1a}P_{2a}}$; and $X_{2d}^{(1)}$ is independent from both $X_{1a}^{(1)}$ and $X_{1d}^{(1)}$.

(2) Period $\bar{\nu}T$. During this period, the source sends a superposition of two Gaussian signals, $X_{1,i}^{(2)} = X_{1a,i}^{(2)} + X_{1d,i}^{(2)}$, where signals $X_{1a,i}^{(2)}$ and $X_{1d,i}^{(2)}$ are generated in a very similar way to those during the period νT, except that now signal $X_{1d,i}^{(2)}$ is obtained as

$$X_{1d,i}^{(2)} = U_{1d,i}^{(2)} - \frac{P_{1d}}{P_{1d}+N_3}\Big[(1-\alpha_a)S_{a,i}^{(2)} - S_{d,i}^{(2)}\Big]. \qquad \text{(C.5)}$$

The relay sends only a Gaussian signal $X_{2a,i}^{(2)}$ that carries message W_a, where $X_{2a,i}^{(2)}$ is independently drawn with $X_{2a}^{(2)} \sim \mathcal{N}(0, P_2)$.

Decoding. Decoding at the relay during the period νT and decoding at the destination during both transmission periods exploit successive cancellation. The details of the computation of some of the mutual information terms in this section are very similar to those in the proof of Theorem 3 in Appendix B, and hence we omit them for brevity. Also, since all the random variables are i.i.d., we sometimes omit the time index. Furthermore, we use the functions defined in Definition 2 and the substitutions: $\xi := 1 - \rho_{12}^2 - \rho_{1s}^2$, $\rho := \rho_{1s}$, $\tilde{N}_2 := N_2 + P_{1d} + Q/2$ and $\tilde{N}_3 := N_3 + P_{1d} + Q/2$.

(1) Period νT. The relay receives

$$Y_2^{(1)} = X_{1a}^{(1)} + S_a^{(1)} + \Big[X_{1d}^{(1)} + S_d^{(1)} + Z_2\Big]. \qquad \text{(C.6)}$$

The relay first decodes the codeword $U_{1a}^{(1)}$ *fully*, that is, not only the bin index but also the correct sequence in the bin. In doing so, the decoder at the relay treats $[X_{1d}^{(1)} + S_d^{(1)}]$ as part of the noise. Note that the noise $[X_{1d}^{(1)} + S_d^{(1)} + Z_2]$ is Gaussian with variance $\tilde{N}_2$ and is independent from the inputs $X_{1a}^{(1)}$, $X_{2a}^{(1)}$ and the CSI $S_a^{(1)}$. The decoding in this step is reliable if

$$0 < I\Big(U_{1a}^{(1)}; Y_2^{(1)}X_{2a}^{(1)}\Big) - I\Big(U_{1a}^{(1)}; S_a^{(1)}X_{2a}^{(1)}\Big), \qquad \text{(C.7)}$$

$$R_a^{(1)} \le \nu\Big[I\Big(U_{1a}^{(1)}; Y_2^{(1)} \mid X_{2a}^{(1)}\Big) - I\Big(U_{1a}^{(1)}; S_a^{(1)} \mid X_{2a}^{(1)}\Big)\Big]$$
$$= \nu R_1\Big(P_{1a}, Q/2, \tilde{N}_2, \xi, \rho, \alpha_a\Big). \qquad \text{(C.8)}$$

The relay then subtracts out the decoded $U_{1a}^{(1)}$ from $Y_2^{(1)}$ to obtain

$$\tilde{Y}_2^{(1)} = X_{1d}^{(1)} + \big((1-\alpha_a)S_a^{(1)} + S_d^{(1)}\big) + Z_2, \qquad \text{(C.9)}$$

from which it can also decode message W_d reliably as long as

$$R_d^{(1)} \le \nu\Big[I\Big(U_{1d}^{(1)}; Y_2^{(1)} \mid U_{1a}^{(1)}, X_{2a}^{(1)}\Big)$$
$$- I\Big(U_{1d}^{(1)}; S_a^{(1)}, S_d^{(1)} \mid U_{1a}^{(1)}, X_{2a}^{(1)}\Big)\Big] \qquad \text{(C.10)}$$
$$= \frac{\nu}{2}\log\Big(1 + \frac{P_{1d}}{N_2}\Big).$$

The destination first decodes the codeword $U_{1a}^{(1)}$ fully from

$$Y_3^{(1)} = X_{1a}^{(1)} + X_{2a}^{(1)} + S_a^{(1)} + \Big[X_{1d}^{(1)} + X_{2d}^{(1)} - S_d^{(1)} + Z_3\Big]. \quad \text{(C.11)}$$

In doing so, the decoder at the destination treats $[X_{1d}^{(1)} + X_{2d}^{(1)} - S_d^{(1)}]$ as part of the noise. Note that the noise $[X_{1d}^{(1)} + X_{2d}^{(1)} - S_d^{(1)} + Z_3]$ is Gaussian with variance $\tilde{N}_3 + P_{2d}$, and is independent from the inputs $X_{1a}^{(1)}$, $X_{2a}^{(1)}$ and the CSI $S_a^{(1)}$. The decoding in this step is reliable if

$$R_a^{(1)} \le \nu\Big[I\Big(U_{1a}^{(1)}, X_{2a}^{(1)}; Y_3^{(1)}\Big) - I\Big(U_{1a}^{(1)}; S_a^{(1)} \mid X_{2a}^{(1)}\Big)\Big]$$
$$= \nu R_2\Big(P_{1a}, P_{2a}, \frac{Q}{2}, \tilde{N}_3 + P_{2d}, \xi, \rho, \alpha_a\Big), \qquad \text{(C.12)}$$

$$0 < I\Big(U_{1a}^{(1)}; Y_3^{(1)} \mid X_{2a}^{(1)}\Big) - I\Big(U_{1a}^{(1)}; S_a^{(1)} \mid X_{2a}^{(1)}\Big)$$
$$= R_1\Big(P_{1a}, Q/2, \tilde{N}_3 + P_{2d}, \xi, \rho, \alpha_a\Big). \qquad \text{(C.13)}$$

The destination then subtracts out the decoded $U_{1a}^{(1)}$ as well as $X_{2a}^{(1)}$ from $Y_3^{(1)}$, and it decodes message W_d from the residual

$$\tilde{Y}_3^{(1)} = X_{2d}^{(1)} + \Big[X_{1d}^{(1)} + (1-\alpha_a)S_a^{(1)} - S_d^{(1)} + Z_3\Big]. \qquad \text{(C.14)}$$

In doing so, the decoder at destination treats $[X_{1d}^{(1)} + (1-\alpha_a)S_a^{(1)} - S_d^{(1)}]$ as part of the noise. The decoding in this step is reliable if

$$R_d^{(1)} \le \nu I\Big(X_{2d}^{(1)}; Y_3^{(1)} \mid U_{1a}^{(1)}, X_{2a}^{(1)}\Big). \qquad \text{(C.15)}$$

Through straightforward algebra, which is omitted here for brevity, it can be shown that (C.23) gives

$$R_d^{(1)} \le \frac{\nu}{2}\log\Bigg(1 + \frac{P_{2d}}{N_3 + P_{1d} + (Q/2)\varsigma(P_{1a}, Q/2, \xi, \rho, \alpha_a)}\Bigg),$$
$$\text{(C.16)}$$

where $\varsigma(\cdot)$ is defined as in Definition 2.

Summary 1. From the above we get that, under the conditions (C.7), (C.13), the rate delivered to the destination during this period is given by

$$T_1 = R_a^{(1)} + R_d^{(1)}$$

$$\leq \nu \min\left\{ R_1\left(P_{1a}, \frac{Q}{2}, \tilde{N}_2, \xi, \rho, \alpha_a\right), \right.$$

$$\left. R_2\left(P_{1a}, P_{2a}, \frac{Q}{2}, \tilde{N}_3 + P_{2d}, \xi, \rho, \alpha_a\right) \right\}$$

$$+ \nu \min\left\{ \frac{1}{2}\log\left(1 + \frac{P_{1d}}{N_2}\right), \right.$$

$$\left. \frac{1}{2}\log\left(1 + \frac{P_{2d}}{N_3 + P_{1d} + (Q/2)\varsigma(P_{1a}, Q/2, \xi, \rho, \alpha_a)}\right) \right\}.$$
(C.17)

(2) Period $\bar{\nu}T$. During this period, the relay decodes only message W_a from

$$Y_2^{(2)} = X_{1a}^{(2)} + S_a^{(2)} + \left[X_{1d}^{(2)} + S_d^{(2)} + Z_2\right]. \quad \text{(C.18)}$$

In doing so, the decoder at the relay treats $[X_{1d}^{(2)} + S_d^{(2)}]$ as part of the noise. Again, this can be done reliably as long as

$$R_a^{(2)} \leq \bar{\nu}\left[I\left(U_{1a}^{(2)}; Y_2^{(2)} \mid X_{2a}^{(2)}\right) - I\left(U_{1a}^{(2)}; S_a^{(2)} \mid X_{2a}^{(2)}\right)\right]$$

$$= \bar{\nu} R_1\left(P_{1a}, \frac{Q}{2}, \tilde{N}_2, \xi, \rho, \alpha_a\right).$$
(C.19)

Decoding at the destination also is similar to that during the period νT. The destination first decodes $U_{1a}^{(1)}$ fully from

$$Y_3^{(2)} = X_{1a}^{(2)} + X_{2a}^{(2)} + S_a^{(2)} + \left[X_{1d}^{(2)} - S_d^{(2)} + Z_3\right]. \quad \text{(C.20)}$$

For the decoding of $U_{1a}^{(2)}$, the decoder at the destination decodes both the bin index and also the correct sequence in the bin. In doing so, it observes a total noise of variance $\tilde{N}_3$. The decoding in this step is reliable if

$$R_a^{(2)} \leq \bar{\nu}\left[I\left(U_{1a}^{(2)}, X_{2a}^{(2)}; Y_3^{(2)}\right) - I\left(U_{1a}^{(2)}; S_a^{(2)} \mid X_{2a}^{(2)}\right)\right]$$

$$= \bar{\nu} R_2\left(P_{1a}, P_2, \frac{Q}{2}, \tilde{N}_3, \xi, \rho, \alpha_a\right)$$

$$0 < I\left(U_{1a}^{(2)}; Y_3^{(2)} \mid X_{2a}^{(2)}\right) - I\left(U_{1a}^{(2)}; S_a^{(2)} \mid X_{2a}^{(2)}\right)$$

$$= R_1\left(P_{1a}, \frac{Q}{2}, \tilde{N}_3, \xi, \rho, \alpha_a\right).$$
(C.21)

The destination then subtracts out the decoded $U_{1a}^{(2)}$ and $X_{2a}^{(2)}$ from $Y_3^{(2)}$ and it decodes message W_d from the residual

$$\tilde{Y}_3^{(2)} = X_{1d}^{(2)} + \left[(1 - \alpha_a)S_a^{(2)} - S_d^{(2)}\right] + Z_3. \quad \text{(C.22)}$$

The decoding in this step is reliable if

$$R_d^{(2)} \leq \bar{\nu}\left[I\left(U_{1d}^{(2)}; Y_3^{(2)} \mid U_{1a}^{(2)}, X_{2a}^{(2)}\right) \right.$$

$$\left. - I\left(U_{1d}^{(2)}; S_a^{(2)}, S_d^{(2)} \mid U_{1a}^{(2)}, X_{2a}^{(2)}\right)\right]$$
(C.23)

$$= \frac{\bar{\nu}}{2}\log\left(1 + \frac{P_{1d}}{N_3}\right).$$

Summary 2. From the above we get that, under the condition (C.22), the rate delivered to the destination during this period is

$$T_2 = R_a^{(2)} + R_d^{(2)}$$

$$\leq \bar{\nu} \min\left\{ R_1\left(P_{1a}, \frac{Q}{2}, \tilde{N}_2, \xi, \rho, \alpha\right), \right.$$

$$\left. R_2\left(P_{1a}, P_2, \frac{Q}{2}, \tilde{N}_3, \xi, \rho, \alpha_a\right) \right\} + \frac{\bar{\nu}}{2}\log\left(1 + \frac{P_{1d}}{N_3}\right).$$
(C.24)

Let us choose the GDPC scale parameter α_a such that the inequalities (C.7), (C.13) and (C.22) are satisfied, that is, $\alpha_a \in \mathcal{B}_3(\xi, \rho)$. Then, for fixed $\nu \in [0, 1]$, and fixed $(P_{1a}, P_{1d}, P_{2a}, P_{2d})$ and (ξ, ρ, α_a) chosen as specified above, we obtain the rate delivered to the destination during the total transmission time by adding the rates T_1 and T_2.

Finally, we obtain the rate on the RHS of (29) by maximizing $T_1 + T_2$, over all values of $\nu \in [0, 1]$; $P_{1a}, P_{1d}, P_{2a}, P_{2d}$ such that $P_{1a} + P_{1d} \leq P_1$ and $P_{2a} + P_{2d} \leq P_2$; $\rho_{12} \in [-1, 1]$, $\rho_{1s} \in [-1, 1]$, $\alpha_a \in \mathcal{B}_3(\xi, \rho)$ such that the covariance matrices of $(S_a^{(1)}, S_d^{(1)}, X_1^{(1)}, X_2^{(1)}, Z_2, Z_3)$ and $(S_a^{(2)}, S_d^{(2)}, X_1^{(2)}, X_2^{(2)}, Z_2, Z_3)$ have nonnegative discriminants, that is, for $Q > 0$, $\xi = 1 - \rho_{12}^2 - \rho_{1s}^2 \geq 0$. Furthermore, investigating T_1 and T_2, one can easily see that it suffices to consider $\rho_{12} \in [0, 1]$ and $\rho_{1s} \in [-1, 0]$. Thus, $\xi \in [0, 1]$ and $\rho \in [-\sqrt{1 - \xi}, 0]$. This concludes the proof of Theorem 5.

Acknowledgment

References

[1] C. E. Shannon, "Channels with side information at the transmitter," *IBM Journal of Research and Development*, vol. 2, pp. 289–293, 1958.

[2] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[3] C. D. Heegard and A. El Gamal, "On the capacity of computer memory with defects," *IEEE Transactions on Information Theory*, vol. 29, no. 5, pp. 731–739, 1983.

[4] M. H. M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, 1983.

[5] G. Keshet, Y. Steinberg, and N. Merhav, "Channel coding in the presence of side information: subject review," *Foundations and Trends in Communications and Information Theory*, 2008.

[6] S. I. Gel'fand and M. S. Pinsker, "On Gaussian channel with random parameters," in *Proceedings of the IEEE International Symposium on Information Theory*, Tashkent, Uzbekistan, 1984.

[7] Y. Steinberg, "Coding for the degraded broadcast channel with random parameters, with causal and noncausal side

information," *IEEE Transactions on Information Theory*, vol. 51, no. 8, pp. 2867–2877, 2005.

[8] Y. Cemal and Y. Steinberg, "The multiple-access channel with partial state information at the encoders," *IEEE Transactions on Information Theory*, vol. 51, no. 11, pp. 3992–4003, 2005.

[9] Y.-H. Kim, A. Sutivong, and S. Sigurjonsson, "Multiple user writing on dirty paper," in *Proceedings of the IEEE International Symposium on Information Theory*, p. 534, Chicago, Ill, USA, June 2004.

[10] A. Zaidi and L. Vandendorpe, "Rate regions for the partially-cooperative relay broadcast channel with non-causal side information," in *Proceedings of the IEEE International Symposium on Information Theory*, pp. 1246–1250, Nice, France, June 2007.

[11] S. Kotagiri and J. N. Laneman, "Achievable rates for multiple access channels with state information known at one encoder," in *Proceedings of the Allerton Conference on Communications, Control and Computing*, Monticello, Ill, USA, 2004.

[12] S. Kotagiri and J. N. Laneman, "Multiaccess channels with state known to one encoder: a case of degraded message sets," in *Proceedings of the IEEE International Symposium on Information Theory*, pp. 1566–1570, Nice, France, June 2007.

[13] S. P. Kotagiri and J. N. Laneman, "Multiaccess channels with state known to some encoders and independent messages," *EURASIP Journal on Wireless Communications and Networking*, vol. 2008, Article ID 450680, 14 pages, 2008.

[14] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdù, "Cooperative encoding with asymmetric state information at the transmitters," in *Proceedings of the Allerton Conference on Communications, Control and Computing*, Monticello, Ill, USA, September 2006.

[15] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdù, "Cooperative multiple access encoding with states available at one transmitter," in *Proceedings of the IEEE International Symposium on Information Theory*, pp. 1556–1560, Nice, France, June 2007.

[16] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdù, "Cooperative multiple-access encoding with states available at one transmitter," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4448–4469, 2008.

[17] A. Zaidi, S. Kotagiri, J. N. Laneman, and L. Vandendorpe, "Multiaccess channels with state known to one encoder: another case of degraded message sets," in *Proceedings of the IEEE International Symposium on Information Theory*, pp. 2376–2380, Seoul, Korea, June-July 2009.

[18] A. Zaidi, S. Kotagiri, J. N. Laneman, and L. Vandendorpe, "Cooperative relaying with state available at the relay," in *Proceedings of the IEEE Information Theory Workshop*, pp. 139–143, Porto, Portugal, May 2008.

[19] A. Zaidi, S. Kotagiri, J. N. Laneman, and L. Vandendorpe, "Cooperative relaying with state available non-causally at the relay," submitted to *IEEE Transactions on Information Theory*.

[20] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Transactions on Information Theory*, vol. 49, no. 3, pp. 563–593, 2003.

[21] A. S. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1639–1667, 2002.

[22] A. Zaidi, P. Piantanida, and P. Duhamel, "Broadcast- and MAC-aware coding strategies for multiple user information embedding," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2974–2992, 2007.

[23] A. Zaidi and L. Vandendorpe, "Coding schemes for relay-assisted information embedding," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 1, pp. 70–85, 2009.

[24] N. Merhav, "On random coding error exponents of watermarking systems," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 420–430, 2000.

[25] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 1813–1827, 2006.

[26] N. Devroye, P. Mitran, and V. Tarokh, "Limits on communications in a cognitive radio channel," *IEEE Communications Magazine*, vol. 44, no. 6, pp. 44–49, 2006.

[27] A. Jovičić and P. Viswanath, "Cognitive radio: an information-theoretic perspective," in *Proceedings of the IEEE International Symposium on Information Theory*, pp. 2413–2417, July 2006.

[28] I. Marić, A. Goldsmith, G. Kramer, and S. Shamai (Shitz), "On the capacity of interference channels with one cooperating transmitter," *European Transactions on Telecommunications*, vol. 19, no. 4, pp. 405–420, 2008.

[29] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdù, "Cognitive interference channels with state information," in *Proceedings of the IEEE International Symposium on Information Theory*, pp. 1353–1357, Toronto, Canada, July 2008.

[30] O. Sahin and E. Erkip, "Achievable rates for the Gaussian interference relay channel," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '07)*, pp. 1627–1631, 2007.

[31] R. Dabora, I. Marić, and A. Goldsmith, "Relay strategies in multiuser networks," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '08)*, pp. 1–5, New Orleans, La, USA, December 2008.

[32] S. Sridharan, S. Vishwanath, S. A. Jafar, and S. Shamai (Shitz), "On the capacity of cognitive relay assisted Gaussian interference channel," in *Proceedings of the IEEE International Symposium on Information Theory*, pp. 549–553, Toronto, Canada, July 2008.

[33] A. B. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Transactions on Information Theory*, vol. 28, no. 6, pp. 841–850, 1982.

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

[35] A. El Gamal and M. Aref, "The capacity of the semideterministic relay channel," *IEEE Transactions on Information Theory*, vol. 28, p. 536, 1982.

[36] A. Khisti, U. Erez, A. Lapidoth, and G. W. Wornell, "Carbon copying onto dirty paper," *IEEE Transactions on Information Theory*, vol. 53, no. 5, pp. 1814–1827, 2007.

[37] A. Høst-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2020–2040, 2005.

[38] B. Schein and R. Gallager, "The Gaussian parallel relay network," in *Proceedings of the IEEE International Symposium on Information Theory*, p. 22, Sorrento, Italy, June 2000.

[39] M. Gastpar and M. Vetterli, "On the capacity of large Gaussian relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 765–779, 2005.

[40] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[41] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, New York, NY, USA, 1968.

[42] U. Erez, S. Shamai (Shitz), and R. Zamir, "Capacity and lattice strategies for cancelling known interference," *IEEE Transactions on Information Theory*, vol. 51, no. 11, pp. 3820–3833, 2005.

[43] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 1991.

[44] S. Lall, *Advanced Topics in Computation for Control*, Lecture Notes for Engr210b, Stanford University, Stanford, Calif, USA.

[45] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, London, UK, 1981.