Bounds on the Capacity of the Relay Channel with Noncausal State at the Source

Abdellatif Zaidi Shlomo Shamai (Shitz) Pablo Piantanida Luc Vandendorpe

Abstract

We consider a three-terminal state-dependent relay channel with the channel state available noncausally at only the source. Such a model may be of interest for node cooperation in the framework of cognition, i.e., collaborative signal transmission involving cognitive and non-cognitive radios. We study the capacity of this communication model. One principal problem is caused by the relay's not knowing the channel state. For the discrete memoryless (DM) model, we establish two lower bounds and an upper bound on channel capacity. The first lower bound is obtained by a coding scheme in which the source describes the state of the channel to the relay and destination, which then exploit the gained description for a better communication of the source's information message. The coding scheme for the second lower bound remedies the relay's not knowing the states of the channel by first computing, at the source, the appropriate input that the relay would send had the relay known the states of the channel, and then transmitting this appropriate input to the relay. The relay simply guesses the sent input and

The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Austin, TX, June 2010. This work has been supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless Communications (NEWCOM++). The work of S. Shamai has also been supported by the CORNET consortium.

Abdellatif Zaidi is with Université Paris-Est Marne La Vallée, 77454 Marne la Vallée Cedex 2, France. Email: abdellatif.zaidi@univ-mlv.fr

Shlomo Shamai is with the Department of Electrical Engineering, Technion Institute of Technology, Technion City, Haifa 32000, Israel. Email: sshlomo@ee.technion.ac.il

Pablo Piantanida is with the Department of Telecommunications, SUPELEC, 91190 Gif-sur-Yvette, France. Email: pablo.piantanida@supelec.fr

Luc Vandendorpe is with École Polytechnique de Louvain, Université Catholique de Louvain, Louvain-la-Neuve 1348, Belgium. Email: luc.vandendorpe@uclouvain.be

sends it in the next block. The upper bound accounts for not knowing the state at the relay and destination. For the general Gaussian model, we derive lower bounds on the channel capacity by exploiting ideas in the spirit of those we use for the DM model; and we show that these bounds are optimal for small and large noise at the relay irrespective to the strength of the interference. Furthermore, we also consider a relay model with orthogonal channels from the the source to the relay and from the source and relay to the destination in which the source input component that is heard by the relay does not depend on the channel states. We establish a better upper bound for both DM and Gaussian cases and we also characterize the capacity in a number of special cases.

Index Terms

User cooperation, relay channel, cognitive radio, channel state information, dirty paper coding.

I. Introduction

We consider a three-terminal state-dependent relay channel (RC) in which, as shown in Figure 1, the source wants to communicate a message W to the destination through the state-dependent RC in n uses of the channel, with the help of the relay. The channel outputs Y_2^n and Y_3^n for the relay and the destination, respectively, are controlled by the channel input X_1^n from the source, the relay input X_2^n and the channel state S^n, through a given memoryless probability law $W_{Y_2,Y_3|X_1,X_2,S}$. The channel state S^n is generated according to the n-product of a given memoryless probability law Q_S. It is assumed that the channel state is known, noncausally, to only the source. The destination estimates the message sent by the source from the received channel output. In this paper we study the capacity of this communication system. We will refer to the model in Figure 1 as *general state-dependent RC with informed source*.

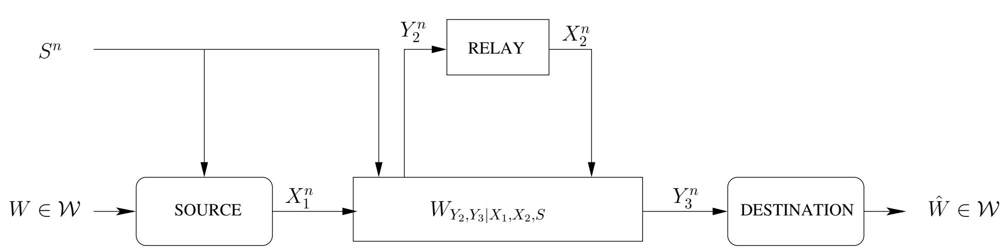


Fig. 1. General state-dependent relay channel with state information S^n available noncausally at only the source.

We shall also study an important class of state-dependent relay channels with orthogonal channels from the source to the relay and from the source and relay to the destination, shown in Figure 2. In this model, the source alphabet $\mathcal{X}_1 = \mathcal{X}_{1R} \times \mathcal{X}_{1D}$, $X_1^n = (X_{1R}^n, X_{1D}^n)$ and only the component X_{1D}^n knows the states S^n. Furthermore, the memoryless conditional law $W_{Y_2,Y_3|X_{1R},X_{1D},X_2,S}$ factorizes as

$$W_{Y_2,Y_3|X_{1R},X_{1D},X_2,S} = W_{Y_2|X_{1R},S} W_{Y_3|X_{1D},X_2,S}. \tag{1}$$

Note that this definition differs from the original definition of relay channels with orthogonal components in the classic setup of channels without states by El Gamal and Zahedi [1] through the presence of the state parameter and the fact that $X_2 \leftrightarrow (X_{1R}, S) \leftrightarrow Y_2$ forms a Markov chain. Perhaps somehow misleadingly, throughout this paper we will continue to refer to this class of state-dependent relay channels as *state-dependent RC with orthogonal components*, omitting explicitly mentioning the aforementioned Markov chain restriction and the fact that only one component of the source encoder components knows the channel states.

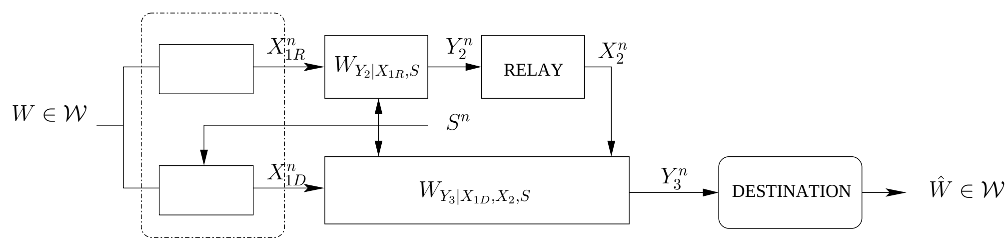


Fig. 2. State-dependent relay channel with the source input $X_1^n = (X_{1R}^n, X_{1D}^n)$, and only the component X_{1D}^n knowing the states of the channel noncausally.

One can think of the two source encoder components in Figure 2 as being two non-colocated base stations transmitting a common message to some destination with the help of a relay – the common message may be obtained by means of message cognition at the encoder whose input is heard at the relay.

A. 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. Examples of applications include information embedding [2], interference imposed by adjacent users, certain storage applications such as computer memories [3], coding for certain broadcast channels [4]–[6], dispersive (ISI) channels [7], block fading in wireless environments [8], cooperation in the realm of cognition [9] and others. The random state sequence may be known in a *causal* or *noncausal* manner. For single user models, the concept of channel state available at only the transmitter dates back to Shannon [10] for the causal channel state case, and to Gel'fand and Pinsker [11] for the noncausal channel state case. In [12], 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 to both the encoder and the decoder. In [13], 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, 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 [14].

A growing body of work studies multi-user state-dependent models. Recent advances in this regard can be found in [14]–[36], and many other works. Key to the investigation of a state-dependent model is whether the parameters controlling the channel are known to *all* or *only some* of the users in the communication model. If the parameters of the channel are known to only some of the users, the problem exhibits some *asymmetry* which makes its investigation more difficult in general. Also, in this case one has to expect some rate penalty due to the lack of knowledge of the state at the uninformed encoders, relative to the case in which all encoders would be informed.

The state-dependent multiaccess channel (MAC) with only one informed encoder and degraded message sets is considered in [15], [16], [37]–[40]; and the state-dependent relay channel (RC) with only informed relay is considered in [20], [21]. For all these models, the authors develop non-trivial outer or upper bounds that permit to characterize the rate loss due to not knowing the state at the uninformed encoders. Key feature to the development of these outer or upper bounding techniques is that, in all these models, the uninformed encoder not only does not know the channel state but can learn no information about it.

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 [20], [21], and it also connects with the MAC with asymmetric channel state and degraded message sets considered in [15]–[17]. However, establishing a non-trivial upper bound for the present model is more involved, comparatively. Partly, this is because, here, one uninformed encoder (the relay) is also a receiver; and, so, it can potentially get some information about the channel states from directly observing the past received sequence from the source. That is, at time i, the input $X_{2,i}$ of the relay can potentially depend on the channel states through its past output $Y_2^{i-1} = (Y_{2,1}, \ldots, Y_{2,i-1})$. For the general model in Figure 1, the relay can even know the states *noncausally*, potentially. This is because Y_2^{i-1} may depend on future values of the state through past source inputs $X_{1,j}(W, S^n)$, $j = 1, \ldots, i-1$. For the model of Figure 2, the relay can know the states only *strictly-causally*, but upper bounding the capacity seems still not easy. In our recent work [41], [42], we have shown that, in a multiaccess channel, strictly causal knowledge of the state at one encoder can be beneficial in general for the other encoder *even if the latter is informed noncausally* (in [42] we characterize the capacity region fully). Studying networks in which a subset of the nodes know the states noncausally and another subset know these states only strictly causally, i.e., networks with mixed – noncausal and strictly causal, states appears to be more challenging in general, and is likely to capture additional interest, especially after recent results on the utility of strictly causally known states in multiaccess channels [26], [27].

B. Main Contributions

For the general state-dependent RC with informed source shown in Figure 1, we derive two lower bounds and an upper bound on the channel capacity. In the discrete memoryless (DM) case, the first lower bound is obtained by a block Markov coding scheme in which the source describes the channel state to the relay and destination *ahead of time*. The source sends a two-layer description of the state consisting of two (possibly correlated) individual descriptions intended to be recovered at the relay and destination respectively. The relay recovers the individual description intended to it and then utilizes the estimated state as noncausal state information at the transmitter to implement collaborative source-relay binning in subsequent blocks, through a combined decode-and-forward [43, Theorem 5] and Gel'fand-Pinsker binning [11]. The destination guesses the source's message sent cooperatively by the source and relay and the individual description which is intended to it from its output and the previously recovered state. The rationale for the coding scheme which we use for the first lower bound is that, had the relay known the state with negligible distortion, then efficient cooperative source-relay binning in the spirit of [44] can be realized (recall that the model in [44] assumes availability of the state at both source and relay).

We obtain the second lower bound by a block Markov coding scheme in which, rather than the channel state itself, the source describes to the relay the appropriate input that the relay would send had the relay known the channel states, assuming a decode-and-forward relaying strategy. The source sends this description to the relay ahead of time. The relay recovers the sent input and retransmits it in the appropriate subsequent block. The rationale for the coding scheme which we use for the second lower bound is that, if the input is produced at the source using binning against the known state and if the relay recovers it with negligible error, then all would appear as if the relay were informed of the channel state. This is because, from an operational point-of-view, the relay actually need not know the channel state, but, rather, the appropriate input that it would send had it known this state.

For the state-dependent general model, we also establish an upper bound on the capacity. This upper bound accounts for not knowing the state at the relay and the destination. Then, considering the relay model of Figure 2, we derive a better upper bound that accounts also for the loss incurred by not knowing the state at one of the source encoder components. We show that this upper bound is strictly tighter than the max-flow min cut or cut-set upper bound obtained by assuming that the state is available at all nodes. We note that upper-bounding techniques for related models with asymmetric channel states, i.e., models with states known only at some of the encoders have been developed recently in our previous work [21] for a relay channel with states known only at the relay, and in [15]–[17] for a MAC with degraded message sets and states known only at one encoder. However, as we mentioned previously, the model that we study in this paper is more involved comparatively, essentially because, as a receiver the relay can get information about the unknown state. From this angle, our upper bounding techniques here are more linked to our recent works [41], [42].

Next, we also consider a memoryless Gaussian model in which the noise and the state are additive and Gaussian. The state represents an external interference and is known noncausally to only the source. We derive lower bounds on the capacity of the general Gaussian RC with informed source by applying the concepts that we develop for the DM case. Similar to the discrete case, one lower bound is based on the idea of describing the state to the relay beforehand; the relay recovers it and then utilizes it for collaborative binning in subsequent blocks. The other lower bound consists in transmitting to the relay a quantized version of the appropriate input that the relay would send had the relay known the channel state. We show that these lower bounds perform well in general and are optimal for large and small noise at the relay, respectively, irrespective of the strength of the interference.

Furthermore, considering a Gaussian version of the model shown in Figure 2, we also develop an upper bound on the capacity that is strictly better than the max-flow min cut or cut-set upper bound. We point out the rate loss in the upper bound incurred by the availability of the channel state at only the one source encoder component. Using this upper bound, we characterize the channel capacity in a number of cases, including when the interference corrupts transmission to the destination but not to the relay.

C. *Outline and Notation*

An outline of the remainder of this paper is as follows. Section II describes in more detail the communication models that we consider in this work. Sections III and IV are devoted to studying the discrete memoryless models, providing lower and upper bounds on channel capacity for the state-dependent General RC in Section III and for the state-dependent RC with orthogonal components in Section IV. Sections V and VI contain the corresponding Gaussian models, providing lower and upper bound on the capacity; and characterizing the channel capacity in some cases. Section VII contains some numerical results and discussions. Finally, Section VIII concludes the paper.

We use the following notations throughout the paper. Upper case letters are used to denote random variables, e.g., X; lower case letters are used to denote realizations of random variables, e.g., x; and calligraphic letters designate alphabets, i.e., $\mathcal{X}$. The probability distribution of a random variable X is denoted by $P_X(x)$. Sometimes, for convenience, we write it as P_X. We use the notation $\mathbb{E}_X[\cdot]$ to denote the expectation of random variable X. A probability distribution of a random variable Y given X is denoted by $P_{Y|X}$. The set of probability distributions defined on an alphabet $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$. The cardinality of a set $\mathcal{X}$ is denoted by $|\mathcal{X}|$. For convenience, the length n vector x^n will occasionally be denoted in boldface notation $\mathbf{x}$. The Gaussian distribution with mean μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$. Finally, throughout the paper, logarithms are taken to base 2, and the complement to unity of a scalar $u \in [0,1]$ is denoted by $\bar{u}$, i.e., $\bar{u} = 1 - u$.

II. System Model and Definitions

In this section, we formally present our communication model and the related definitions. 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 $Y_2^n \in \mathcal{Y}_2^n$ and $Y_3^n \in \mathcal{Y}_3^n$ for the relay and the destination, respectively, are controlled by the channel inputs $X_1^n \in \mathcal{X}_1^n$ from the source and $X_2^n \in \mathcal{X}_2^n$ from the relay, along with random states $S^n \in \mathcal{S}^n$. 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 only 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 $n^{-1} \log M$ bits per transmission.

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

$$\phi_1^n : \{1, \ldots, M\} \times \mathcal{S}^n \to \mathcal{X}_1^n, \tag{2}$$

a sequence of encoding functions at the relay

$$\phi_{2,i} : \mathcal{Y}_{2,1}^{i-1} \to \mathcal{X}_{2,i}, \tag{3}$$

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

$$\psi^n : \mathcal{Y}_3^n \to \{1, \ldots, M\}. \tag{4}$$

Let a (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 conditional probability distribution $W_{Y_2,Y_3|X_1,X_2,S}$. The joint probability mass function on $\mathcal{W}\times\mathcal{S}^n\times\mathcal{X}_1^n\times\mathcal{X}_2^n\times\mathcal{Y}_2^n\times\mathcal{Y}_3^n$ is given by

$$P(w, s^n, x_1^n, x_2^n, y_2^n, y_3^n) = P(w)\prod_{i=1}^{n} Q_S(s_i)P(x_{1,i}|w, s^n)P(x_{2,i}|y_2^{i-1})$$
$$\cdot W_{Y_2,Y_3|X_1,X_2,S}(y_{2,i}, y_{3,i}|x_{1,i}, x_{2,i}, s_i). \tag{5}$$

The destination estimates the message sent by the source from the channel output Y_3^n. The average probability of error is defined as $P_e^n = \mathbb{E}_S\Big[\Pr\Big(\psi^n(Y_3^n) \neq W|S^n = s^n\Big)\Big]$.

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 $(\epsilon_n, n, R)-$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.

We shall also study the relay model shown in Figure 2, in which the source alphabet $\mathcal{X}_1 = \mathcal{X}_{1R}\times\mathcal{X}_{1D}$, $X_1^n = (X_{1R}^n, X_{1D}^n)$ with the input component X_{1R}^n function of only the message W and the input component X_{1D}^n function of (W, S^n), i.e., $X_{1R}^n = \phi_{1R}^n(W)$ and $X_{1D}^n = \phi_{1D}^n(W, S^n)$ — ϕ_{1R}^n and ϕ_{1D}^n are the source encoding functions, and the conditional distribution $W_{Y_2,Y_3|X_{1R},X_{1D},X_2,S}$ factorizing as (1). The encoding at the relay and the decoding at the destination remain as in the model of Figure 1, i.e., given by (3) and (4), respectively.

III. The Discrete Memoryless RC with Informed Source

In this section, we consider the general state-dependent RC model of Figure 1. We assume that the alphabets $\mathcal{S}$, $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{Y}_2$ and $\mathcal{Y}_3$ in the model are all discrete and finite.

A. Lower Bounds on Channel Capacity: State Description

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

***Theorem** 1:* The capacity of the state-dependent discrete memoryless relay channel with informed source is lower bounded by

$$R^{\mathrm{lo}} = \max\ \min\,\{I(U; Y_2|V, \hat{S}_R) - I(U; S, \hat{S}_D|V, \hat{S}_R),$$
$$I(U, V; Y_3|\hat{S}_D) - I(U, V; S, \hat{S}_R|\hat{S}_D)\} \tag{6}$$

subject to the constraints

$$I(S;\hat{S}_R) \leq I(U_R;Y_2,\hat{S}_R|U,V) - I(U_R;S,\hat{S}_R,\hat{S}_D|U,V) \tag{7a}$$

$$I(S;\hat{S}_D) \leq I(U_D;Y_3,\hat{S}_D|U,V) - I(U_D;S,\hat{S}_R,\hat{S}_D|U,V) + [I(U;Y_3,\hat{S}_D|V) - I(U;S,\hat{S}_R,\hat{S}_D|V)]_- \tag{7b}$$

$$\begin{aligned} I(S;\hat{S}_R,\hat{S}_D) + I(\hat{S}_R;\hat{S}_D) \leq\ & I(U_R;Y_2,\hat{S}_R|U,V) - I(U_R;S,\hat{S}_R,\hat{S}_D|U,V) \\ & + I(U_D;Y_3,\hat{S}_D|U,V) - I(U_D;S,\hat{S}_R,\hat{S}_D|U,V) + [I(U;Y_3,\hat{S}_D|V) - I(U;S,\hat{S}_R,\hat{S}_D|V)]_- \\ & - I(U_R;U_D|U,V,S,\hat{S}_R,\hat{S}_D) \end{aligned} \tag{7c}$$

where $[x]_- \triangleq \min(x,0)$, and the maximization is over all joint measures on $\mathcal{S}\times\hat{\mathcal{S}}_R\times\hat{\mathcal{S}}_D\times\mathcal{U}_R\times\mathcal{U}_D\times\mathcal{U}\times\mathcal{V}\times\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{Y}_2\times\mathcal{Y}_3$ of the form

$$\begin{aligned} & P_{S,\hat{S}_R,\hat{S}_D,U_R,U_D,U,V,X_1,X_2,Y_2,Y_3} \\ & \quad = Q_S P_{\hat{S}_R,\hat{S}_D|S} P_{V|\hat{S}_R} P_{U|V,S,\hat{S}_R,\hat{S}_D} P_{U_R,U_D|V,U,S,\hat{S}_R,\hat{S}_D} P_{X_1|U_R,U_D,U,V,S,\hat{S}_R,\hat{S}_D} P_{X_2|V,\hat{S}_R} W_{Y_2,Y_3|X_1,X_2,S}. \end{aligned} \tag{8}$$

and satisfying

$$I(V;Y_3,\hat{S}_D) - I(V;\hat{S}_R) > 0. \tag{9}$$

Proof: An outline of the proof of Theorem 1 will follow, and complete error analysis appears in Appendix A.

In Theorem 1, the random variables $\hat{S}_R$ and $\hat{S}_D$ represent two descriptions $\hat{\mathbf{S}}_R$ and $\hat{\mathbf{S}}_D$ of the state $\mathbf{S}$ that are sent by the source *ahead of time* and meant to be recovered at the relay and destination, respectively. The random variables U_R and U_D are associated with the codewords $\mathbf{U}_R$ and $\mathbf{U}_D$ that are used by the source to carry these state descriptions to the relay and destination, respectively. The random variables U and V represent respectively the Gel'fand-Pinsker auxiliary vector $\mathbf{U}$ used to precode the information message at the source against the known state $(\mathbf{S},\hat{\mathbf{S}}_R,\hat{\mathbf{S}}_D)$ and the Gel'fand-Pinsker auxiliary vector $\mathbf{V}$ used to precode the information message at the relay against the state $\hat{\mathbf{S}}_R$. The allowed measure (8) implies the following Markov chains

$$V \leftrightarrow \hat{S}_R \leftrightarrow (S,\hat{S}_D), \qquad (U,V,U_R,U_D) \leftrightarrow (X_1,X_2,S) \leftrightarrow (Y_2,Y_3). \tag{10}$$

The first Markov chain reflects the fact that the input at the relay depends on the state only through the description that is recovered at the relay. The second Markov chain reflects the memoryless nature of the channel, and the fact the outputs at the relay and destination depend on all other codewords only through the inputs of the source and relay and the channel state.

The following remarks are useful for a better understanding of the coding scheme which we use to achieve the lower bound in Theorem 1.

Remark 1: The intuition for the coding scheme which we use to establish the lower bound in Theorem 1 is as follows. Had the relay known the state, the source and the relay could implement collaborative binning against that state for transmission to the destination [44]. Since the source knows the state of the channel noncausally, it can transmit a description of it to the relay *ahead of time*. The relay recovers the state (with a certain distortion), and then utilizes it in the relevant subsequent block through a collaborative binning scheme. The hope is that the benefit that the source can get from being assisted by a more capable relay will compensate the loss caused by the source's spending some of its resources to make the relay learn the state.

In general, it may also turn out to be useful to send a dedicated description of the state to the destination. The destination utilizes the recovered state as side information at the receiver. In the coding scheme that we employ to establish the lower bound in Theorem 1, in addition to its message, the source also sends a two-layer description of the state to the relay and destination; one layer description dedicated for each. The two layers are possibly correlated. The relay guesses the source's message and the individual state description which is dedicated to it from the source transmission and the previously recovered state description. It then utilizes the new state estimate as noncausal state at the encoder for collaborative source-relay binning over the next block, through a combined decode-and-forward and Gel'fand-Pinsker binning. The destination guesses the source's message sent cooperatively by the source and relay and the individual state description which is dedicated to it from its output and the previously recovered state description.

Remark 2: As it can be seen from the proof in Appendix A, the source sends the descriptions intended to the relay and destination *two blocks* ahead of time. That is, at the beginning of block i the source describes the state vector $\mathbf{s}[i+2]$ to the relay and destination. While one block delay is sufficient to describe the state to the relay, a minimum of two blocks is necessary for the state reconstruction at the destination because of the used window decoding technique. In the following remark, we will comment onto the relevance of sliding window for decoding at the destination for our model.

Outline of 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 scheme is based on an appropriate combination of block Markov encoding [43], Gel'fand-Pinsker binning [11], multiple descriptions [45] and Marton's coding for general broadcast channels [46]–[48]. Next, we outline the encoding and decoding procedures.

We transmit in $B+1$ blocks, each of length n. Let $\mathbf{s}[i]$ denote the state sequence controlling the channel in block i, with $i = 1, \dots, B+1$. During each of the first B blocks, the source encodes a message $w_i \in [1, 2^{nR}]$ and sends it over the channel. In addition, during each of the first $B-1$ blocks, the source also sends two individual descriptions of $\mathbf{s}[i+2]$ intended to be recovered at the relay and destination, respectively. We denote by $\hat{\mathbf{s}}_R[\iota_{Ri}]$, $\iota_{Ri} \in [1, 2^{n\hat{R}_R}]$, the description of $\mathbf{s}[i+2]$ intended to be recovered at the relay in block i, at rate $\hat{R}_R$, and by $\hat{\mathbf{s}}_D[\iota_{Di}]$, $\iota_{Di} \in [1, 2^{n\hat{R}_D}]$, the description of $\mathbf{s}[i+2]$ intended to be recovered at the destination in block i, at rate $\hat{R}_D$. For the last two blocks, for convenience, we set $w_{B+1} = 1$, $(\iota_{RB}, \iota_{DB}) = (1,1)$ and $(\iota_{RB+1}, \iota_{DB+1}) = (1,1)$. For fixed n, the average (channel coding) rate $R(B/(B+1))$ of the information message over $B+1$ blocks approaches R as $B \longrightarrow +\infty$, and the average (source coding) rates $\hat{R}_R((B-1)/(B+1))$ and $\hat{R}_D((B-1)/(B+1))$ approach $\hat{R}_R$ and $\hat{R}_D$, respectively, as $B \longrightarrow +\infty$.

Codebook generation: Fix a measure $P_{S,\hat{S}_R,\hat{S}_D,U_R,U_D,U,V,X_1,X_2,Y_2,Y_3}$ of the form (8). Calculate the marginals $P_{\hat{S}_R}$ and $P_{\hat{S}_D}$ induced by this measure. Fix $\epsilon > 0$, and let $M = 2^{n[R-\epsilon]}$,

$$J_V = 2^{n[I(V;\hat{S}_R)+\epsilon]} \qquad M_R = 2^{n[R_R-5\epsilon]} \qquad J_R = 2^{n[I(U_R;S,\hat{S}_R,\hat{S}_D|U,V)+\epsilon]}$$

$$J_U = 2^{n[I(U;S,\hat{S}_R,\hat{S}_D|V)+\epsilon]} \qquad M_D = 2^{n[R_D-5\epsilon]} \qquad J_D = 2^{n[I(U_D;S,\hat{S}_R,\hat{S}_D|U,V)+\epsilon]} \tag{11}$$

with

$$R_R = I(U_R; Y_2, \hat{S}_R|U, V) - I(U_R; S, \hat{S}_R, \hat{S}_D|U, V) - \epsilon$$

$$R_D = I(U_D; Y_3, \hat{S}_D|U, V) - I(U_D; S, \hat{S}_R, \hat{S}_D|U, V) + [I(U; Y_3, \hat{S}_D|V) - I(U; S, \hat{S}_R, \hat{S}_D|V)]_- - \epsilon \tag{12}$$

where $[x]_-$ denotes $\min(x, 0)$.

We may assume that first term of the minimization in (6) is non-negative, i.e., $I(U; Y_2, \hat{S}_R|V) - I(U; S, \hat{S}_R, \hat{S}_D|V) \geq 0$. We generate two statistically independent codebooks (codebooks 1 and 2) by following the steps outlined below twice. We shall use these codebooks for blocks with odd and even indices, respectively.

1) Generate $2^{n\hat{R}_R}$ n-vectors $\hat{\mathbf{s}}_R[1], \ldots, \hat{\mathbf{s}}_R[2^{n\hat{R}_R}]$ independently according to a uniform distribution over the set $T^n_\epsilon(P_{\hat{S}_R})$ of $\epsilon-$typical $\hat{\mathbf{S}}_R$ $n-$ vectors.
2) Generate $2^{n\hat{R}_D}$ n-vectors $\hat{\mathbf{s}}_D[1], \ldots, \hat{\mathbf{s}}_D[2^{n\hat{R}_D}]$ independently according to a uniform distribution over the set $T^n_\epsilon(P_{\hat{S}_D})$ of $\epsilon-$typical $\hat{\mathbf{S}}_D$ $n-$ vectors.
3) Generate $J_V M$ independent and identically distributed (i.i.d.) codewords $\{\mathbf{v}(w', j_V)\}$ indexed by $w' = 1, \ldots, M$, $j_V = 1, \ldots, J_V$. Each codeword $\mathbf{v}(w', j_V)$ is with i.i.d. components drawn according to P_V.
4) For each codeword $\mathbf{v}(w', j_V)$, generate a collection of $J_U M$ codewords $\{\mathbf{u}(w', j_V, w, j_U)\}$ indexed by $w = 1, \ldots, M$, $j_U = 1, \ldots, J_U$. Each codeword $\mathbf{u}(w', j_V, w, j_U)$ is with i.i.d. components drawn according to $P_{U|V}$.
5) For each codeword $\mathbf{v}(w', j_V)$, for each codeword $\mathbf{u}(w', j_V, w, j_U)$, generate a collection of $J_R M_R$ codewords $\{\mathbf{u}_R(w', j_V, w, j_U, k, j_R)\}$ indexed by $k = 1, \ldots, M_R$, $j_R = 1, \ldots, J_R$. Each codeword $\mathbf{u}_R(w', j_V, w, j_U, k, j_R)$ is with i.i.d. components drawn according to $P_{U_R|V,U}$.
6) For each codeword $\mathbf{v}(w', j_V)$, for each codeword $\mathbf{u}(w', j_V, w, j_U)$, generate a collection of $J_D M_D$ codewords $\{\mathbf{u}_D(w', j_V, w, j_U, l, j_D)\}$ indexed by $l = 1, \ldots, M_D$, $j_D = 1, \ldots, J_D$. Each codeword $\mathbf{u}_D(w', j_V, w, j_U, l, j_D)$ is with i.i.d. components drawn according to $P_{U_D|V,U}$.
7) (Binning à-la Marton [46], [47]): For $\iota_R \in [1, 2^{n\hat{R}_R}]$, define the cells

$$\mathcal{B}_{\iota_R} = [(\iota_R - 1)2^{n[R_R - \hat{R}_R - \epsilon]} + 1, \iota_R 2^{n[R_R - \hat{R}_R - \epsilon]}].$$

Similarly, for $\iota_D \in [1, 2^{n\hat{R}_D}]$, define the cells

$$\mathcal{C}_{\iota_D} = [(\iota_D - 1)2^{n[R_D - \hat{R}_D - \epsilon]} + 1, \iota_D 2^{n[R_D - \hat{R}_D - \epsilon]}],$$

where without loss of generality $2^{n[R_R - \hat{R}_R - \epsilon]}$ and $2^{n[R_D - \hat{R}_D - \epsilon]}$ are considered to be integer valued.

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 pick up the story in block i. Let w_i be the new message to be sent from the source node at the beginning of block i, and w_{i-1} the message sent in the previous block $i-1$. The encoding at the beginning of block i is as follows. The source finds, if possible, a pair $(\iota_{Ri}, \iota_{Di}) \in [1, 2^{n\hat{R}_R}] \times [1, 2^{n\hat{R}_D}]$ such that $(\mathbf{s}[i+2], \hat{\mathbf{s}}_R[\iota_{Ri}], \hat{\mathbf{s}}_D[\iota_{Di}])$ are jointly typical. If such (ι_{Ri}, ι_{Di}) does not exist, simply set $(\iota_{Ri}, \iota_{Di}) = (1, 1)$. We shall show that a successful encoding of $\mathbf{s}[i+2]$ at the source is accomplished with high probability provided that n is sufficiently large and

$$\begin{aligned}\hat{R}_R &> I(S;\hat{S}_R)\\ \hat{R}_D &> I(S;\hat{S}_D)\\ \hat{R}_R+\hat{R}_D &> I(S;\hat{S}_R,\hat{S}_D)+I(\hat{S}_R;\hat{S}_D). \end{aligned}\tag{13}$$

The source will send the quadruple $(w_{i-1}, w_i, \iota_{Ri}, \iota_{Di})$ over the channel. First, let us assume that the relay has decoded correctly message w_{i-1} and the indices $(\iota_{Ri-2}, \iota_{Ri-1})$, and the destination has decoded correctly message w_{i-2} and the index ι_{Di-2}. We shall show that our code construction allows the relay to decode correctly message w_i and the index ι_{Ri} and the destination to decode correctly message w_{i-1} and the index ι_{Di-1} at the end of block i (with a probability of error $\leq \epsilon$). Thus, the information state $(w_{i-2}, w_{i-1}, \iota_{Ri-1}, \iota_{Di-2})$ propagates forward and a recursive calculation of the probability of error can be made, yielding a probability of error $\leq (B+1)\epsilon$.

We continue with the strategy at the beginning of block i.

1) The relay knows w_{i-1} and ι_{Ri-2} and searches for the smallest $j_V \in J_V$ such that $\mathbf{v}(w_{i-1}, j_V)$ is jointly typical with $\hat{\mathbf{s}}_R[\iota_{Ri-2}]$ (the properties of jointly typical sequences guarantee that, with probability close to one, there exists one such j_V). Denote this j_V by $j^{\star}_{Vi} = j_V(\hat{\mathbf{s}}_R[\iota_{Ri-2}], w_{i-1})$. (Note that since $V \leftrightarrow \hat{S}_R \leftrightarrow (S, \hat{S}_D)$ forms a Markov chain, chosen as such, $\mathbf{v}(w_{i-1}, j^{\star}_{Vi})$ will be also jointly typical with $(\mathbf{s}[i], \hat{\mathbf{s}}_D[\iota_{Di-2}])$, by the Markov Lemma [49, p. 436]). Then the relay sends a vector $\mathbf{x}_2[i]$ with i.i.d. components given $\mathbf{v}(w_{i-1}, j^{\star}_{Vi})$ and $\hat{\mathbf{s}}_R[\iota_{Ri-2}]$, drawn according to the marginal $P_{X_2|V,\hat{S}_R}$ induced by the distribution (8). (For $i = 1,2$, the relay does not know an estimate of the channel state and so it sends some default codeword).
2) The source first searches for the smallest $j_U \in J_U$ such that $\mathbf{u}(w_{i-1}, j^{\star}_{Vi}, w_i, j_U)$ is jointly typical with the vector $\mathbf{s}[i], \hat{\mathbf{s}}_R[\iota_{Ri-2}], \hat{\mathbf{s}}_D[\iota_{Di-2}])$ given $\mathbf{v}(w_{i-1}, j^{\star}_{Vi})$. (Again, the properties of jointly typical sequences guarantee that there exists one such j_U). Denote this j_U by $j^{\star}_{Ui} = j_U(\mathbf{s}[i], \hat{\mathbf{s}}_R[\iota_{Ri-2}], \hat{\mathbf{s}}_D[\iota_{Di-2}], w_{i-1}, w_i)$.
3) Next, the source searches for one pair

$$\Big(\mathbf{u}_R(w_{i-1}, j^{\star}_{Vi}, w_i, j^{\star}_{Ui}, k_i, j_{Ri}), \mathbf{u}_D(w_{i-1}, j^{\star}_{Vi}, w_i, j^{\star}_{Ui}, l_i, j_{Di})\Big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}},$$

where

$$\begin{aligned}\mathcal{D}_{\iota_{Ri}\iota_{Di}} = \Big\{&\Big(\mathbf{u}_R(w_{i-1}, j^{\star}_{Vi}, w_i, j^{\star}_{Ui},k_i, j_{Ri}), \mathbf{u}_D(w_{i-1}, j^{\star}_{Vi}, w_i, j^{\star}_{Ui}, l_i, j_{Di})\Big) \ s.t.: \\ &k_i \in \mathcal{B}_{\iota_{Ri}}, l_i \in \mathcal{C}_{\iota_{Di}}, j_{Ri} \in J_R, j_{Di} \in J_D\\ &\Big(\mathbf{u}_R(w_{i-1}, j^{\star}_{Vi}, w_i, j^{\star}_{Ui}, k_i, j_{Ri}), \mathbf{s}[i], \hat{\mathbf{s}}_R[\iota_{Ri-2}], \hat{\mathbf{s}}_D[\iota_{Di-2}]\Big) \in T^n_\epsilon(P_{U_R S\hat{S}_R\hat{S}_D|UV})\\ &\Big(\mathbf{u}_D(w_{i-1}, j^{\star}_{Vi}, w_i, j^{\star}_{Ui}, l_i, j_{Di}), \mathbf{s}[i], \hat{\mathbf{s}}_R[\iota_{Ri-2}], \hat{\mathbf{s}}_D[\iota_{Di-2}]\Big) \in T^n_\epsilon(P_{U_D S\hat{S}_R\hat{S}_D|UV})\\ &\Big(\mathbf{u}_R(w_{i-1}, j^{\star}_{Vi}, w_i, j^{\star}_{Ui}, k_i, j_{Ri}), \mathbf{u}_D(w_{i-1}, j^{\star}_{Vi}, w_i, j^{\star}_{Ui}, l_i, j_{Di})\Big) \in T^n_\epsilon(P_{U_R,U_D|UVS\hat{S}_R\hat{S}_D})\Big\}. \end{aligned}\tag{14}$$

We shall show that, with high probability, the source will find one such pair provided that n is sufficiently large and

$$\hat{R}_R+\hat{R}_D < R_R+R_D - I(U_R;U_D|U,V,S,\hat{S}_R,\hat{S}_D).\tag{15}$$

Denote the found pair as $\left(\mathbf{u}_R(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui}, k_i, j^\star_{Ri}), \mathbf{u}_D(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui}, l_i, j^\star_{Di})\right)$.

4) The source then sends a vector $\mathbf{x}_1[i]$ with i.i.d. components given the vectors $\mathbf{v}(w_{i-1}, j^\star_{Vi})$, $\mathbf{u}(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui})$, $\mathbf{u}_R(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui}, k_i, j^\star_{Ri})$, $\mathbf{u}_D(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui}, l_i, j^\star_{Di})$ and $(\mathbf{s}[i], \hat{\mathbf{s}}_R[\iota_{Ri-2}], \hat{\mathbf{s}}_D[\iota_{Di-2}])$, drawn according to the marginal $P_{X_1|V,U,U_R,U_D,S,\hat{S}_R,\hat{S}_D}$ induced by the distribution (8).

Decoding: Decoding and state reconstruction at the relay are based on classical joint typicality. Decoding and state reconstruction at the destination are based on joint typicality and window-decoding. The decoding and reconstruction procedures at the end of block i are as follows.

1) The relay knows w_{i-1} and ι_{Ri-2} (in fact, the relay knows also ι_{Ri-1} but does not use it for decoding in this step). It declares that $(\hat{w}_i, \hat{\iota}_{Ri})$ are sent if there exists a unique triple $(\hat{w}_i, \hat{j}_{Ui}, \hat{k}_i)$, $\hat{w}_i \in [1, M]$, $\hat{j}_{Ui} \in J_U$, $\hat{k}_i \in [1, M_R]$, such that $\mathbf{u}(w_{i-1}, j^\star_{Vi}, \hat{w}_i, \hat{j}_{Ui})$, $\mathbf{u}_R(w_{i-1}, j^\star_{Vi}, \hat{w}_i, \hat{j}_{Ui}, \hat{k}_i, j_{Ri})$ are jointly typical with $(\mathbf{y}_2[i], \hat{\mathbf{s}}_R[\iota_{Ri-2}])$ given $\mathbf{v}(w_{i-1}, j^\star_{Vi})$, for some $j_{Ri} \in J_R$, where $j^\star_{Vi} = j_V(\hat{\mathbf{s}}_R[\iota_{Ri-2}], w_{i-1})$. One can show that, with the choice (12), the decoding error in this step is small for sufficiently large n if

$$R < I(U; Y_2, \hat{S}_R|V) - I(U; S, \hat{S}_R, \hat{S}_D|V). \tag{16}$$

If (16) is satisfied, the estimate $\hat{\iota}_{Ri}$ of ι_{Ri} at the relay is the index of the $\mathcal{B}_{\hat{\iota}_{Ri}}$ containing the found $\hat{k}_i$, i.e., $\hat{k}_i \in \mathcal{B}_{\hat{\iota}_{Ri}}$.

2) The destination knows the pair (w_{i-2}, l_{i-2}) and the index $j^\star_{Vi-1} = j_V(\hat{\mathbf{s}}_R[\iota_{Ri-3}], w_{i-2})$ and decodes the pair (w_{i-1}, l_{i-1}) based on the information received in block $i-1$ and block i. It declares that $(\hat{w}_{i-1}, \hat{l}_{i-1})$ is sent if there is a unique triple $(\hat{w}_{i-1}, \hat{j}_{Ui-1}, \hat{l}_{i-1})$, $\hat{w}_{i-1} \in [1, M]$, $\hat{j}_{Ui-1} \in J_U$, $\hat{l}_{i-1} \in [1, M_D]$, and a unique $\hat{j}_{Vi} \in J_V$, such that $\mathbf{u}(w_{i-2}, j^\star_{Vi-1}, \hat{w}_{i-1}, \hat{j}_{Ui-1})$, $\mathbf{u}_D(w_{i-2}, j^\star_{Vi-1}, \hat{w}_{i-1}, \hat{j}_{Ui-1}, \hat{l}_{i-1}, j_{Di-1})$ are jointly typical with $(\mathbf{y}_3[i-1], \hat{\mathbf{s}}_D[\iota_{Di-3}])$ given $\mathbf{v}(w_{i-2}, j^\star_{Vi-1})$ *and* $\mathbf{v}(\hat{w}_{i-1}, \hat{j}_{Vi})$ is jointly typical with $(\mathbf{y}_3[i], \hat{\mathbf{s}}_D[\iota_{Di-2}])$. One can show that, with the choice (12), the decoding error in this step is small for sufficiently large n if

$$\begin{aligned} R &< I(V, U; Y_3, \hat{S}_D) - I(V, U; S, \hat{S}_R, \hat{S}_D) \\ 0 &< I(V; Y_3, \hat{S}_D) - I(V; \hat{S}_R). \end{aligned} \tag{17}$$

If (17) is satisfied, the estimate $\hat{\iota}_{Di-1}$ of ι_{Di-1} at the destination is the index of the $\mathcal{C}_{\hat{\iota}_{Di-1}}$ containing the found $\hat{l}_{i-1}$, i.e., $\hat{l}_{i-1} \in \mathcal{C}_{\hat{\iota}_{Di-1}}$. Also, the destination obtains the correct index $j^\star_{Vi} = j_V(\hat{\mathbf{s}}_R[\iota_{Ri-2}], w_{i-1})$.

□

The achievable rate in Theorem 1 requires the relay to decode the message sent by the source *fully*, and this can be rather a severe constraint. We can generalize Theorem 1 by allowing the relay to decode the message sent by the source only *partially* [50]. This can be done by splitting the information sent by the source into two independent parts, one part is sent through the relay and the other part is sent directly to the destination. In the following theorem, the random variables V, U, U_R and U_D play the same roles as in Theorem 1 and U_1 is a new random variable that represents the information sent directly to the destination.

***Theorem** 2:* The capacity of the state-dependent discrete memoryless relay channel with informed source is lower bounded by

$$\begin{aligned} R^{\mathrm{lo}} = \max\ \min\ \{ & I(U; Y_2|V, \hat{S}_R) - I(U; S, \hat{S}_D|V, \hat{S}_R), \\ & I(U, V; Y_3|\hat{S}_D) - I(U, V; S, \hat{S}_R|\hat{S}_D)\} + I(U_1; Y_3|U, V, \hat{S}_D) - I(U_1; S, \hat{S}_R|U, V, \hat{S}_D) \end{aligned} \tag{18}$$

subject to the constraints

$$I(S;\hat{S}_R) \leq I(U_R;Y_2,\hat{S}_R|U,V) - I(U_R;S,\hat{S}_R,\hat{S}_D|U,V) \tag{19a}$$

$$I(S;\hat{S}_D) \leq I(U_D;Y_3,\hat{S}_D|U_1,U,V) - I(U_D;S,\hat{S}_R,\hat{S}_D|U_1,U,V) + [I(U_1,U;Y_3,\hat{S}_D|V) - I(U_1,U;S,\hat{S}_R,\hat{S}_D|V)]_- \tag{19b}$$

$$\begin{aligned} I(S;\hat{S}_R,\hat{S}_D) + I(\hat{S}_R;\hat{S}_D) \leq\ & I(U_R;Y_2,\hat{S}_R|U,V) - I(U_R;S,\hat{S}_R,\hat{S}_D|U,V) \\ & + I(U_D;Y_3,\hat{S}_D|U_1,U,V) - I(U_D;S,\hat{S}_R,\hat{S}_D|U_1,U,V) + [I(U_1,U;Y_3,\hat{S}_D|V) - I(U_1,U;S,\hat{S}_R,\hat{S}_D|V)]_- \\ & - I(U_R;U_D|U_1,U,V,S,\hat{S}_R,\hat{S}_D) \end{aligned} \tag{19c}$$

where $[x]_- \triangleq \min(x,0)$, and the maximization is over all joint measures on $\mathcal{S} \times \hat{\mathcal{S}}_R \times \hat{\mathcal{S}}_D \times \mathcal{U}_R \times \mathcal{U}_D \times \mathcal{U}_1 \times \mathcal{U} \times \mathcal{V} \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}_2 \times \mathcal{Y}_3$ of the form

$$\begin{aligned} & P_{S,\hat{S}_R,\hat{S}_D,U_R,U_D,U,V,X_1,X_2,Y_2,Y_3} \\ & \quad = Q_S P_{\hat{S}_R,\hat{S}_D|S} P_{V|\hat{S}_R} P_{U|V,S,\hat{S}_R,\hat{S}_D} P_{U_1|V,U,S,\hat{S}_R,\hat{S}_D} P_{U_R,U_D|V,U,U_1,S,\hat{S}_R,\hat{S}_D} P_{X_1|U_R,U_D,U,V,S,\hat{S}_R,\hat{S}_D} P_{X_2|V,\hat{S}_R} W_{Y_2,Y_3|X_1,X_2,S} \end{aligned} \tag{20}$$

and satisfying $U_1 \leftrightarrow (V,U,S,\hat{S}_R,\hat{S}_D) \leftrightarrow U_R$ is a Markov chain and

$$\begin{aligned} & 0 < I(V;Y_3,\hat{S}_D) - I(V;\hat{S}_R) \\ & 0 \leq I(U;Y_2|V,\hat{S}_R) - I(U;S,\hat{S}_D|V,\hat{S}_R) \\ & 0 \leq I(U_1;Y_3|U,V,\hat{S}_D) - I(U_1;S,\hat{S}_R|U,V,\hat{S}_D). \end{aligned} \tag{21}$$

The proof of Theorem 2 follows by a fair extension of that of Theorem 1, and so we omit it here for brevity.

Remark 3: In the coding scheme of Theorem 2, if the source sends no descriptions of the state to the relay and destination, i.e., $\hat{S}_R = \hat{S}_D = \emptyset$, the coding scheme reduces to a generalized Gel'fand-Pinsker binning scheme at the source that is combined with partial DF. In this case, the relay sends codewords that carry part of the information message and are independent of the channel states. The following achievable rate[1] is obtained from Theorem 2 by setting $\hat{S}_R = \hat{S}_D = \emptyset$, $U_R = U_D = \emptyset$ and $V = X_2$ independent of S, as

$$R = \max\min\Big\{ I(U;Y_2|X_2) + I(U_1;Y_3|U,X_2) - I(U,U_1;S|X_2),\ \ I(U,U_1,X_2;Y_3) - I(U,U_1;S|X_2) \Big\} \tag{22}$$

with the maximization over joint measures of the form

$$P_{S,U,U_1,X_1,X_2,Y_2,Y_3} = Q_S P_{X_2} P_{U|S,X_2} P_{U_1,X_1|U,S,X_2} W_{Y_2,Y_3|X_1,X_2,S} \tag{23}$$

and satisfying

$$\begin{aligned} & 0 \leq I(U;Y_2|X_2) - I(U;S|X_2) \\ & 0 \leq I(U_1;Y_3|U,X_2) - I(U_1;S|U,X_2) \\ & 0 \leq I(U,U_1;Y_3|X_2) - I(U,U_1;S|X_2). \end{aligned} \tag{24}$$

[1]We note that the achievable rate (22) is slightly larger than that of [23, Theorem 1] which contains one more term in the minimization.

B. Lower Bound on Channel Capacity: Analog Input Description

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

***Theorem** 3:* The capacity of the state-dependent discrete memoryless relay channel with informed source is lower bounded by

$$R^{\mathrm{lo}} = \max\ I(U;Y_3) - I(U;S) \tag{25}$$

subject to the constraint

$$I(X;\hat{X}) < I(U_R;Y_2) - I(U_R;S) - I(U_R;U|S) \tag{26}$$

where maximization is over all joint measures on $\mathcal{S}\times\mathcal{U}\times\mathcal{U}_R\times\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{X}\times\hat{\mathcal{X}}\times\mathcal{Y}_2\times\mathcal{Y}_3$ of the form

$$\begin{aligned} &P_{S,U,U_R,X_1,X_2,X,\hat{X},Y_2,Y_3} \\ &\qquad = Q_S P_{U,U_R|S} P_{X_1|U,U_R,S} P_{X|U,S} P_{\hat{X}|X} \mathbb{1}_{X_2=\hat{X}} W_{Y_2,Y_3|X_1,X_2,S}. \end{aligned} \tag{27}$$

Proof: The proof of Theorem 3 appears in Appendix B.

In Theorem 3, the random variable X represents an auxiliary vector $\mathbf{X}$ that is obtained by binning the information message at the source against the state $\mathbf{S}$. The random variable $\hat{X}$ represents a description $\hat{\mathbf{X}}$ of $\mathbf{X}$ that is sent by the source *ahead of time* and meant to be recovered only at the relay. The random variable U_R represents the information that carries the description $\hat{\mathbf{X}}$ of $\mathbf{X}$ to the relay, on top of the information message. The codeword $\mathbf{U}_R$ is binned against $(\mathbf{U},\mathbf{S})$. The allowed measure (27) implies the following Markov chains

$$(X_1,U_R) \leftrightarrow (U,S) \leftrightarrow X, \qquad (U,U_R,X,\hat{X}) \leftrightarrow (X_1,X_2,S) \leftrightarrow (Y_2,Y_3). \tag{28}$$

Remark 4: The rationale for the coding scheme which we use to obtain the lower bound in Theorem 3 is as follows. Had the relay known the message to be sent in each block and the state that corrupts the transmission in that block, then the relay generates its input using a collaborative Gel'fand-Pinsker scheme as in [44].
For our model, the source knows the message that the relay should optimally send in each block (if the relay gets the message correctly). It also knows the state sequence that corrupts the transmission in that block. It can then generate the appropriate relay input vector that the relay would send had the relay known the message and the state. The source can send this vector to the relay *ahead of time*, and if the relay can estimate it to high accuracy, then collaborative source-relay binning in the sense of [44] is readily realized for transmission from the source and relay to the destination.
More precisely, a block Markov encoding is used to establish Theorem 3. Let us consider transmission in two adjacent blocks i and $i+1$. In the beginning of block i, the source sends the information w_i of the current block, and, in addition, describes to the relay the input $\mathbf{x}[i+1]$ that the relay should send in the next block $i+1$ had the relay known the message w_{i+1} and the state $\mathbf{s}[i+1]$. Let $\hat{\mathbf{x}}[m_i]$ be a description of $\mathbf{x}[i+1]$. The message w_i and the index m_i which the source sends in block i are precoded using binning against the state that controls transmission in the current block i. The vector $\mathbf{x}[i+1]$, however, is the input that the relay would send in the next block $i+1$ had the

relay known the state $\mathbf{s}[i+1]$, and so is generated at the source using binning against the state $\mathbf{s}[i+1]$. The vector $\mathbf{x}[i+1]$, and its description which is sent to the relay during block i, are intended to combine coherently with the source transmission in block $i+1$.

Remark 5: In the scheme we described briefly in Remark 4, the relay needs only estimate the code vector $\mathbf{x}[i+1]$ sent by the source in block i, and transmit the obtained estimate in the next block $i+1$. For instance, the relay does not need to know the information message w_{i+1} that the estimated vector actually carries, let alone the state sequence $\mathbf{s}[i+1]$ that controls the channel in block $i+1$. Thus, from a practical viewpoint, this may be particularly convenient for communication with an oblivious relay. Transmission from the source terminal to the relay terminal can be regarded as that of an analog source which, in block i, produces a sequence $\mathbf{x}[i+1]$. This source has to be transmitted by the source terminal over a state-dependent channel and reconstructed at the relay terminal. The reconstruction error at the relay terminal influences the rate at which information can be decoded reliably at the destination by acting as an additional noise term.

C. Upper Bound on Channel Capacity

As we mentioned in Section I, the relay does not know the states of the channel directly in our model, but it can potentially get some information about S^n from the past received sequence from the informed source. More precisely, the input of the relay $X_{2,i}$ at time i depends on the channel states through $Y_2^{i-1} = (Y_{2,1}, \ldots, Y_{2,i-1})$ which in turn depends on these states through S^{i-1} and the past source inputs $X_{1,j}(W, S^n)$, $j = 1, \ldots, i-1$. Further, because the source knows the states noncausally this dependence may even be noncausal. This aspect makes establishing non-trivial upper bounds on the capacity, i.e., bounds that are strictly better than the cut-set upper bound

$$R_{\mathrm{triv}}^{\mathrm{up}} = \max_{p(x_1,x_2|s)} \min\Big\{I(X_1; Y_2, Y_3|S, X_2),\ \ I(X_1, X_2; Y_3|S)\Big\} \tag{29}$$

not easy.

The following theorem provides an upper bound on the capacity of the state-dependent general discrete memoryless RC with informed source.

***Theorem** 4:* The capacity of the state-dependent discrete memoryless relay channel with informed source is upper-bounded by

$$R^{\mathrm{up}} = \max\ \min\left\{I(V; Y_2, Y_3|U, X_2) - I(V; S|U, X_2),\ \ I(V; Y_3) - I(V; S)\right\} \tag{30}$$

where the maximization is over measures of the form

$$P_{S,U,V,X_1,X_2,Y_2,Y_3} = Q_S P_{U|S} P_{X_2|U,S} P_{V,X_1|U,S} W_{Y_2,Y_3|X_1,X_2,S}. \tag{31}$$

and $U \in \mathcal{U}$, $V \in \mathcal{V}$ are auxiliary random variables with

$$|\mathcal{U}| \leq |\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2| \tag{32a}$$

$$|\mathcal{V}| \leq \Big(|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2|\Big)^2, \tag{32b}$$

respectively.

Proof: The proof of Theorem 4 appears in Appendix C.

Note that the relay input X_2 depends on the state S in the measure (31), and this reflects our discussion above.

IV. The DM Model with Orthogonal Components

In this section, we consider the state-dependent RC with orthogonal components of Figure 2. This model has the source encoder component X_{1R}^n, which is the only encoder component heard by the relay, restricted to be independent of the channel states. For this reason, the coding schemes of Section III do not apply directly. Also, since in this model the relay input can depend on the states only strictly-causally, a better upper bound can be established.

A. Bounds on Channel Capacity

The following proposition provides a lower bound on the capacity of the state-dependent discrete memoryless RC with orthogonal components of Figure 2.

***Proposition** 1:* The capacity of the state-dependent discrete memoryless relay channel with orthogonal components of Figure 2 is lower bounded by

$$R_{\text{orth}}^{\text{lo}} = \max\min\Big\{ I(X_{1R};Y_2|X_2),\ \ I(X_{1R},X_2;Y_3) \Big\} + [I(U_1;Y_3|X_{1R},X_2) - I(U_1;S|X_{1R},X_2)]^+ \tag{33}$$

where $[x]^+ := \max(x,0)$ and the maximization is over all measures of the form

$$P_{S,U_1,X_{1R},X_{1D},X_2,Y_2,Y_3} = Q_S P_{X_2} P_{X_{1R}|X_2} P_{U_1,X_{1D}|S,X_2} W_{Y_2|S,X_{1R}} W_{Y_3|X_{1D},X_2,S}. \tag{34}$$

The proof of Proposition 1 follows by an easy extension of the generalized block-Markov scheme of [1] by allowing the source encoder component that is sent directly to the destination to be generated through a generalized Gel'fand-Pinsker binning scheme. For this reason, we only outline its proof.

In the rate (33), the variable U_1 represents the Gel'fand-Pinsker auxiliary random variable associated with the information sent directly to the destination. More specifically, the message W from the source is split into two independent parts, one of which is transmitted through the relay at rate R_r and the other is transmitted directly to the destination without the help of the relay at rate R_d. The total rate is $R = R_r + R_d$. The message that is transmitted through the relay can be decoded correctly if the rate R_r satisfies [43, Theorem 1]

$$R_r < \min\Big\{I(X_{1R},Y_2|X_2), I(X_1,X_2;Y_3)\Big\}. \tag{35}$$

The additional information which is transmitted through binning, on top of the information transmitted through the relay, can be decoded correctly at the destination if rate R_d satisfies

$$R_d < I(U_1;Y_3|X_{1R},X_2) - I(U_1;S|X_{1R},X_2). \tag{36}$$

This shows that message W can be sent reliably at the rate (33).

We now turn to establish an upper bound on the capacity of the model of Figure 2. We note although the output Y_2^{i-1} at the relay at time i can convey information only about the strictly causal part S^{i-1} of the state, upper bounding the channel capacity is non trivial *even* in this case. By better exploiting the fact that the input component X_{1R}^n that is heard at the relay does not know the state S^n at all in this model, we derive an upper bound which does not depend on auxiliary random variables. The result is stated in the following theorem.

***Theorem** 5:* The capacity of the state-dependent discrete memoryless relay channel with orthogonal components of Figure 2 is upper-bounded by

$$R_{\text{orth}}^{\text{up}} = \max \min \left\{ I(X_{1R}; Y_2|X_2, S),\ I(X_2; Y_3) \right\} + I(X_{1D}; Y_3|X_2, S) \tag{37}$$

where the maximization is over all joint measures of the form

$$P_{S,X_{1R},X_{1D},X_2,Y_2,Y_3} = Q_S P_{X_2} P_{X_{1R}|X_2} P_{X_{1D}|X_2,S} W_{Y_2|X_{1R},S} W_{Y_3|X_{1D},X_2,S} \tag{38}$$

Proof: The proof of Theorem 5 appears in Appendix D.

Observe that the second term of the minimization in (37) upper-bounds the information that the source and the relay can send to the destination by

$$I(X_2; Y_3) + I(X_{1D}; Y_3|X_2, S) = I(X_{1D}, X_2; Y_3|S) - I(X_2; S|Y_3), \tag{39}$$

which is strictly better than the corresponding term in the cut-set upper bound (29).

B. Comments and Digression

There is a connection between the state-dependent relay model with orthogonal components of Figure 2 and a state-dependent two-user multiaccess model with degraded message sets that we treated recently in [41], [42]. In the multiaccess model of [41], [42], the channel states are known noncausally to one of the encoders and only strictly causally to the other encoder. Also, both encoders transmit a common message and, in addition, the encoder that knows the states noncausally transmits an individual message. In [41] we derive bounds on the capacity region; and in [42] we characterize the full capacity region of this multiaccess model. In [41], [42], we show that the knowledge of the states only strictly causally at the encoder that sends only the common message *can increase* the capacity region in general. We also observe that the capacity region is increased even in the extreme case in which the encoder that knows the states only strictly causally has no message to transmit (i.e., common-message rate equal to zero). This suggests that in the relay model of Figure 2, although it can only know the states strictly causally, the relay can potentially help the source combat the effect of the state (in addition to its classic role of relaying the information message). Although it is not clear yet how the relay could exploit optimally the information about the strictly causal part of the state sequence that it can get by observing its output, the upper bound in Theorem 5 makes one step ahead towards this end; by bounding the information that the source and relay can transmit cooperatively; and so, in a sense, the capacity increase that the source can get through the relay's help.

V. The Memoryless Gaussian RC with Informed Source

A. System Model

In this section, we consider a full-duplex state-dependent RC informed source in which the channel state and the noise are additive and Gaussian. In this model, the channel state can model an additive Gaussian interference which is assumed to be known (noncausally) to only the source. 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 state S_i, by

$$Y_{2,i} = X_{1,i} + S_i + Z_{2,i} \tag{40a}$$

$$Y_{3,i} = X_{1,i} + X_{2,i} + S_i + Z_{3,i}. \tag{40b}$$

The channel state S_i is zero mean Gaussian random variable with variance Q; and only the source knows the state sequence S^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 state sequence S^n and the channel inputs (X_1^n, X_2^n). Also, we consider the following individual power constraints on the average transmitted power at the source and the relay,

$$\sum_{i=1}^{n} X_{1,i}^2 \leq nP_1, \qquad \sum_{i=1}^{n} X_{2,i}^2 \leq nP_2. \tag{41}$$

The definition of a code for this Gaussian model is the same as that given in the discrete case of Section III, with the additional constraint that the channel inputs should satisfy the power constraint (41).

B. Bounds on Channel Capacity

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

***Theorem** 6:* The capacity of the state-dependent Gaussian RC with informed source is lower-bounded by

$$R_{\mathrm{G}}^{\mathrm{lo}} = \max \frac{1}{2} \log\Big(1 + \frac{(\sqrt{\bar{\gamma} P_1} + \sqrt{P_2 - D})^2}{N_3 + D + \gamma P_1}\Big), \tag{42}$$

where

$$D := P_2 \frac{N_2}{N_2 + \gamma P_1} \tag{43}$$

and the maximization is over $\gamma \in [0, 1]$.

Remark 6: It is insightful to observe that the rate in Theorem 6 does not depend on the strength of the state S. This makes the coding scheme appreciable, particularly for the case of arbitrary strong interference in which classical coding schemes suffer greatly from the strong interference unknown at the relay.

Outline of Proof of Theorem 6: The result in Theorem 3 for the DM case can be extended to memoryless channels with discrete time and continuous alphabets using standard techniques [51, Chapter 7]. The proof of Theorem 6 follows through evaluation of the lower bound of Theorem 3 using the following jointly Gaussian input distribution. For $0 \leq \gamma \leq 1$, we let $X \sim \mathcal{N}(0, P_2)$ and $X_{1R} \sim \mathcal{N}(0, \gamma P_1)$, with X jointly Gaussian with S with $\mathbb{E}[XS] = 0$; and X_{1R} jointly Gaussian with (S, X), with $\mathbb{E}[X_{1R}S] = \mathbb{E}[X_{1R}X] = 0$. Also, for $0 \leq D \leq P_2$ given, we consider the test channel $\hat{X} = aX + \tilde{X}$, where $a := 1 - D/P_2$ and $\tilde{X}$ is a Gaussian random variable with zero mean and variance $\tilde{P}_2 = D(1 - D/P_2)$, independent from X and S. Using this test channel, we calculate $\mathbb{E}[(X - \hat{X})^2] = D$ and $\mathbb{E}[\hat{X}^2] = P_2 - D$.

We use the following choices of the auxiliary random variables in Theorem 3,

$$U = \Big(\sqrt{\frac{\bar{\gamma}P_1}{P_2}} + \sqrt{\frac{P_2 - D}{P_2}}\Big)X + \alpha S \tag{44}$$

$$U_R = X_{1R} + \alpha_R\Big(S + \frac{\sqrt{\bar{\gamma}P_1}}{\sqrt{\bar{\gamma}P_1} + \sqrt{P_2 - D}}X\Big), \tag{45}$$

where

$$\alpha = \frac{(\sqrt{\bar{\gamma}P_1} + \sqrt{P_2 - D})^2}{(\sqrt{\bar{\gamma}P_1} + \sqrt{P_2 - D})^2 + (N_3 + D + \gamma P_1)} \quad \text{and} \quad \alpha_R = \frac{\gamma P_1}{\gamma P_1 + N_2}. \tag{46}$$

Through straightforward algebra, which we omit here for brevity, it can be shown that the evaluation of the lower bound of Theorem 3 using the above choice gives the lower bound in Theorem 6.

Alternative Proof of Theorem 6: The encoding and transmission scheme is as follows. For $0 \le \gamma \le 1$, let $X \sim \mathcal{N}(0, P_2)$ and $X_{1R} \sim \mathcal{N}(0, \gamma P_1)$, with X jointly Gaussian with S with $\mathbb{E}[XS] = 0$; and X_{1R} jointly Gaussian with (S, X), with $\mathbb{E}[X_{1R}S] = \mathbb{E}[X_{1R}X] = 0$. Also, let $0 \le D \le P_2$ be given, and consider the test channel $\hat{X} = aX + \tilde{X}$, where $a := 1 - D/P_2$ and $\tilde{X}$ is a Gaussian random variable with zero mean and variance $\tilde{P}_2 = D(1 - D/P_2)$, independent from X and S. Using this test channel, we calculate $\mathbb{E}[(X - \hat{X})^2] = D$ and $\mathbb{E}[\hat{X}^2] = P_2 - D$.

We use the two random variables U and U_R given by (50) to generate the auxiliary codewords U_i and $U_{R,i}$ which we will use in the sequel.

As in the discrete case, a block Markov encoding is used. For each block i, let $\mathbf{x}[i]$ be a Gaussian signal which carries message $w_i \in [1, 2^{nR}]$ and is obtained via a DPC considering $\mathbf{s}[i]$ as noncausal channel state information, as

$$\Big(\sqrt{\frac{\bar{\gamma}P_1}{P_2}} + \sqrt{\frac{P_2 - D}{P_2}}\Big)\mathbf{x}[i] = \mathbf{u}[i] - \alpha\mathbf{s}[i], \tag{47}$$

where the components of $\mathbf{u}[i]$ are generated i.i.d. using the auxiliary random variable U.

For every block i, the source quantizes $\mathbf{x}[w_i]$ into $\hat{\mathbf{x}}[m_i]$, where $m_i \in [1, 2^{n\hat{R}}]$. Using the above test channel, the source can encode $\mathbf{x}[w_i]$ successfully at the quantization rate

$$\begin{aligned}\hat{R} &= I(X; \hat{X}) \\ &= \frac{1}{2}\log(\frac{P_2}{D}).\end{aligned} \tag{48}$$

Let m_i be the index associated with $\mathbf{x}[w_{i+1}]$. In the beginning of block i, the source sends a superposition of two Gaussian vectors,

$$\mathbf{x}_1[i] = \mathbf{x}_{1R}[m_i] + \sqrt{\frac{\bar{\gamma}P_1}{P_2}}\mathbf{x}[w_i]. \tag{49}$$

In equation (49), the signal $\mathbf{x}_{1R}[m_i]$ carries message m_i and is obtained via a DPC considering $(s[i], \mathbf{x}[w_i])$ as noncausal channel state information, as

$$\mathbf{x}_{1R}[m_i] = \mathbf{u}_R[i] - \alpha_R\Big(\mathbf{s}[i] + \sqrt{\frac{\bar{\gamma}P_1}{P_2}}\mathbf{x}[w_i]\Big), \tag{50}$$

where the components of $\mathbf{u}_R[i]$ are generated i.i.d. using the auxiliary random variable U_R.
In the beginning of block i, the relay has decoded message m_{i-1} correctly (this will be justified below) and sends

$$\mathbf{x}_2[i] = \frac{\sqrt{P_2}}{\sqrt{P_2 - D}}\hat{\mathbf{x}}[m_{i-1}]. \tag{51}$$

For the decoding arguments at the source and the relay, we give simple arguments based on intuition (the rigorous decoding uses joint typicality). Also, since all the random variables are i.i.d., we sometimes omit the time index. The relay decodes the index m_i from the received $\mathbf{y}_2[i]$ at the end of block i. Since signal $\mathbf{x}_{1R}[m_i]$ is precoded at the source against the interference caused by the information message w_i, decoding at the relay can be done reliably as long as n is large and

$$\hat{R} \leq \frac{1}{2}\log\Big(1 + \frac{\gamma P_1}{N_2}\Big). \tag{52}$$

The destination decodes message w_i from the received $\mathbf{y}_3[i]$ at the end of block i, considering signal $\mathbf{x}_{1R}[m_i]$ as unknown noise, with

$$\begin{aligned}\mathbf{y}_3[i] &= \mathbf{x}_1[i] + \mathbf{x}_2[i] + \mathbf{s}[i] + \mathbf{z}_3[i]\\ &= \Big(\sqrt{\frac{\bar{\gamma}P_1}{P_2}}\mathbf{x}[w_i] + \sqrt{\frac{P_2}{P_2 - D}}\hat{\mathbf{x}}[m_{i-1}]\Big) + \mathbf{s}[i] + (\mathbf{z}_3[i] + \mathbf{x}_{1R}[m_i]).\end{aligned} \tag{53}$$

Let now $\mathbf{x}'[i]$ be the optimal linear estimator of $\Big(\sqrt{\frac{\bar{\gamma}P_1}{P_2}}\mathbf{x}[w_i] + \sqrt{\frac{P_2}{P_2-D}}\hat{\mathbf{x}}[m_{i-1}]\Big)$ given $\mathbf{x}[w_i]$ under minimum mean square error criterion, and $\mathbf{e}_{\mathbf{x}}[i]$ the resulting estimation error. The estimator $\hat{\mathbf{x}}[i]$ and the estimation error $\mathbf{e}_{\mathbf{x}}[i]$ are given by

$$\begin{aligned}\mathbf{x}'[i] &= \mathbb{E}\Big[\sqrt{\frac{\bar{\gamma}P_1}{P_2}}\mathbf{x}[w_i] + \sqrt{\frac{P_2}{P_2 - D}}\hat{\mathbf{x}}[m_{i-1}]|\mathbf{x}[i]\Big]\\ &= \Big(\sqrt{\frac{\bar{\gamma}P_1}{P_2}} + \sqrt{\frac{P_2 - D}{P_2}}\Big)\mathbf{x}[w_i]\end{aligned} \tag{54}$$

$$\mathbf{e}_{\mathbf{x}}[i] = \sqrt{\frac{P_2}{P_2 - D}}\hat{\mathbf{x}}[m_{i-1}] - \sqrt{\frac{P_2 - D}{P_2}}\mathbf{x}[w_i]. \tag{55}$$

We can alternatively write the output $\mathbf{y}_3[i]$ in (53) as

$$\mathbf{y}_3[i] = \xi\mathbf{x}[w_i] + \mathbf{s}[i] + \Big(\mathbf{z}_3[i] + \mathbf{e}_{\mathbf{x}}[i] + \mathbf{x}_{1R}[m_i]\Big), \tag{56}$$

where

$$\xi := \sqrt{\frac{\bar{\gamma}P_1}{P_2}} + \sqrt{\frac{P_2 - D}{P_2}} \tag{57}$$

and $\mathbf{e}_{\mathbf{x}}[i]$ is Gaussian with variance D and is independent of $\mathbf{x}[w_i]$ and $\mathbf{s}[i]$.
Now, considering the equivalent form (56) of the output $\mathbf{y}_3[i]$, it is easy to see that the destination can decode message w_i correctly at the end of block i as long as n is large and

$$\begin{aligned}R &\leq I(U;Y_3) - I(U;S)\\ &= \frac{1}{2}\log\Big(1 + \frac{(\sqrt{\bar{\gamma}P_1} + \sqrt{P_2 - D})^2}{N_3 + D + \gamma P_1}\Big).\end{aligned} \tag{58}$$

Furthermore, combining (48) and (52) we get

$$D \geq P_2 \frac{N_2}{N_2 + \gamma P_1}. \tag{59}$$

Finally, observing that the RHS of (58) decreases with D, we obtain (42) by taking the equality in (59) and maximizing the RHS of (58) over $\gamma \in [0,1]$. This completes the proof.

□

We now turn to establish a lower bound on the capacity of the state-dependent Gaussian RC using the idea of state transmission. In this section, the source describes the channel state to only the relay. The relay guesses the information message and the transmitted state description and then transmits the message cooperatively with the source using binning against the state estimate, in a manner similar to that we described for the coding scheme for Theorem 1.

For convenience we define the following quantities $\tilde{Q}_S(\cdot)$ and $R(\cdot)$ which we will use throughout the remaining sections.

***Definition** 1:* Let

$$\begin{aligned}\tilde{Q}_S(t,Q,D) &:= (1-t)^2 Q - t(t-2)D\\ R(\alpha,P,Q,N) &:= \frac{1}{2}\log\Big(\frac{P(P+Q+N)}{PQ(1-\alpha)^2 + N(P+\alpha^2 Q)}\Big)\end{aligned}$$

for non-negative t, D, P, Q, N, and $\alpha \in \mathbb{R}$.

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

***Theorem** 7:* The capacity of the state-dependent Gaussian RC with informed source is lower-bounded by

$$\begin{aligned}R_{\mathrm{G}}^{\mathrm{lo}} = \max\ \min\ \Big\{&R\Big(\alpha, (1-\rho_{12}^2-\rho_{1s}^2)\bar{\theta}P_{1r}, \xi^2\tilde{Q}, N_2+\theta P_{1r}+P_{1d}\Big),\\ &R\Big(\alpha, (1-\rho_{12}^2-\rho_{1s}^2)\bar{\theta}P_{1r}, \xi^2\tilde{Q}, N_3+\theta P_{1r}+P_{1d}\Big) + \frac{1}{2}\log\Big(1+\frac{(\rho_{12}\sqrt{\bar{\theta}P_{1r}}+\sqrt{P_2})^2}{N_3+\xi^2 D+\theta P_{1r}+(1-\rho_{12}^2-\rho_{1s}^2)\bar{\theta}P_{1r}+P_{1d}}\Big)\Big\}\\ &+\frac{1}{2}\log\Big(1+\frac{P_{1d}}{N_3+\theta P_{1r}}\Big)\end{aligned} \tag{60}$$

where

$$D = Q\frac{N_2+P_{1d}}{N_2+\theta P_{1r}+P_{1d}} \tag{61}$$

$$\tilde{Q} = \tilde{Q}_S(\alpha_2, Q, D), \quad \xi = 1+\rho_{1s}\sqrt{\frac{\bar{\theta}P_{1r}}{Q}} \tag{62}$$

$$\alpha_2 = \frac{(\rho_{12}\sqrt{\bar{\theta}P_{1r}}+\sqrt{P_2})^2}{(\rho_{12}\sqrt{\bar{\theta}P_{1r}}+\sqrt{P_2})^2+(1-\rho_{12}^2-\rho_{1s}^2)\bar{\theta}P_{1r}+(N_3+\xi^2 D+\theta P_{1r}+P_{1d})} \tag{63}$$

and the maximization is over $P_{1r} \geq 0$, $P_{1d} \geq 0$ such that $0 \leq P_{1r}+P_{1d} \leq P_1$, $\theta \in [0,1]$, $\rho_{12} \in [0,1]$ and $\rho_{1s} \in [-1,0]$ such that $0 \leq \rho_{12}^2+\rho_{1s}^2 \leq 1$ and $\alpha \in \mathbb{R}$ such that $R((1-\rho_{12}^2-\rho_{1s}^2)\bar{\theta}P_{1r}, \xi^2\tilde{Q}, N_2+\theta P_{1r}+P_{1d}) \geq 0$ and $R((1-\rho_{12}^2-\rho_{1s}^2)\bar{\theta}P_{1r}, \xi^2\tilde{Q}, N_3+\theta P_{1r}+P_{1d}) + 1/2\log(1+P_{1d}/(N_3+\theta P_{1r})) \geq 0$.

Proof: A formal proof of Theorem 7 appears in Appendix E.

An outline of proof of Theorem 7 is as follows. The result in Theorem 1 for the DM case can be extended to memoryless channels with discrete time and continuous alphabets using standard techniques [51, Chapter 7]. For the state-dependent Gaussian relay channel (40), we evaluate the rate (6) with the following choice of input distribution. We choose $\hat{S}_D = \emptyset$, $U_D = \emptyset$. Furthermore, we consider the test channel $\hat{S}_R = aS + \tilde{S}_R$, where $a := 1 - D/Q$ and $\tilde{S}_R$ is a Gaussian random variable with zero mean and variance $\sigma^2_{\tilde{S}_R} = D(1 - D/Q)$, independent from S. The random variable X_2 is Gaussian with zero mean and variance P_2, independent of S and of $\hat{S}_R$. The random variable X_1 is composed of three parts, $X_1 = X_{SR} + X_{WR} + X_{WD}$, where X_{SR} is Gaussian with zero mean and variance θP_{1r}, for some $\theta \in [0, 1]$, is independent of S, $\hat{S}_R$, X_2; and $X_{WR} = \rho_{1s}\sqrt{\bar{\theta}P_{1r}/Q}S + \rho_{12}\sqrt{\bar{\theta}P_{1r}/P_2}X_2 + X'_{WR}$, where X'_{WR} is Gaussian with zero mean and variance $(1 - \rho^2_{12})\bar{\theta}P_{1r}$, for some $\rho_{12} \in [0, 1]$ and $\rho_{1s} \in [-1, 0]$ and is independent of X_{SR}, X_2 and $(S, \hat{S}_R)$; and X_{WD} is a Gaussian with zero mean and variance P_{1d}, chosen independently from all the other variables. The auxiliary random variables are chosen as

$$V = \Big(\rho_{12}\sqrt{\frac{\bar{\theta}P_{1r}}{P_2}} + 1\Big)X_2 + \alpha_2\Big(\rho_{1s}\sqrt{\frac{\bar{\theta}P_{1r}}{Q}} + 1\Big)\hat{S}_R \tag{64a}$$

$$U = X'_{WR} + \alpha\xi(S - \alpha_2\hat{S}_R) \tag{64b}$$

$$U_1 = X_{WD} + \frac{P_{1d}}{P_{1d} + N_3 + \theta P_{1r}}\xi(1 - \alpha)(S - \alpha_2\hat{S}_R) \tag{64c}$$

$$U_R = X_{SR} + \frac{\theta P_{1r}}{\theta P_{1r} + N_2 + P_{1d}}(1 - \alpha)S \tag{64d}$$

with

$$\alpha_2 = \frac{(\rho_{12}\sqrt{\bar{\theta}P_{1r}} + \sqrt{P_2})^2}{(\rho_{12}\sqrt{\bar{\theta}P_{1r}} + \sqrt{P_2})^2 + (1 - \rho^2_{12} - \rho^2_{1s})\bar{\theta}P_{1r} + (N_3 + \xi^2 D + \theta P_{1r} + P_{1d})} \tag{65a}$$

$$D := Q\frac{N_2 + P_{1d}}{N_2 + \theta P_{1r} + P_{1d}} \quad \text{and} \quad \xi = 1 + \rho_{1s}\sqrt{\frac{\bar{\theta}P_{1r}}{Q}}. \tag{65b}$$

Through straightforward algebra which is omitted for brevity, it can be shown that the evaluation of (6) with the aforementioned input distribution gives (60).

Remark 7: The parameter α in Theorem 7 stands for DPC's scale factor in precoding the information message against the interference on its way to the relay and to the destination. Because the model (40) has the links to the relay and to the destination corrupted by noise terms with distinct variances, one cannot remove the effect of the interference on the two links simultaneously via one single DPC as in [18]. This explains why the parameter α is left to be optimized over in (60). However, in the spirit of [18], one can improve the rate of Theorem 7 by time sharing coding schemes that are similar to the one we employed for Theorem 7 but with different inflation parameters tailored respectively for the link to the relay and the link to the destination, as in [23].

Similar to the general DM model of Section III, in the general Gaussian model (40) the relay does not know the states of the channel directly but can potentially get information about S^n from the observed output sequence Y_2^{i-1}. Also, Y_2^{i-1} may even contain information about future values of the state, and this makes establishing upper bounds on the capacity that are strictly better than the cut-set upper bound

$$R_{\mathrm{G}}^{\mathrm{up}} = \max_{p(x_1,x_2|s)} \min\Big\{I(X_1;Y_2,Y_3|S,X_2),\;\; I(X_1,X_2;Y_3|S)\Big\} \tag{66}$$

more difficult. Note that the cut-set upper bound is in general non-tight essentially because both X_1 and X_2 know the state S in (66).

C. *Analysis of Some Extreme Cases*

We now summarize the behavior of some of the developed lower and upper bounds in some extreme cases.

1) If $N_2 \longrightarrow 0$, e.g, the relay is located spatially very close to the source, the lower bound of Theorem 6 and the cut-set upper bound (66) tend asymptotically to the same value

$$\mathcal{C}_{\mathrm{G}} = \frac{1}{2}\log\Big(1 + \frac{(\sqrt{P_1} + \sqrt{P_2})^2}{N_3}\Big) - o(1) \tag{67}$$

where $o(1) \longrightarrow 0$ as $N_2 \longrightarrow 0$.

Equation (67) reflects the rationale for our coding scheme for the lower bound in Theorem 6 which is tailored to be asymptotically optimal whenever the relay can learn with negligible distortion the input that it should send. In this case, the rate (67) can be interpreted as the information between two transmit antennas which both know the channel state and one receive antenna. (For comparison, note that the coding scheme of Theorem 7 achieves rate smaller than that of Theorem 6 if $N_2 \longrightarrow 0$, because even though with the coding scheme of Theorem 7 as well the relay obtains the state estimate at almost no expense if N_2 is arbitrarily small, it also needs to know the information message to perform binning, however).

2) *Arbitrarily strong channel state:* In the asymptotic case $Q \to \infty$, the lower bound of Theorem 7 tends to

$$R_{\mathrm{G}}^{\mathrm{lo}} = \frac{1}{2}\log\Big(1 + \frac{P_1}{\max(N_2, N_3)}\Big). \tag{68}$$

The lower bound of Theorem 6 does not depend on the strength of the channel state, as we indicated previously.

3) If $N_2 \longrightarrow \infty$, i.e., the link to the relay is broken or too noisy, the cut-set upper bound (66) and the lower of Theorem 7 agree and give the channel capacity

$$\mathcal{C}_{\mathrm{G}} = \frac{1}{2}\log(1 + \frac{P_1}{N_3}). \tag{69}$$

Note that, for the Gaussian model (40), the lower of Theorem 6 is suboptimal if $N_2 \longrightarrow \infty$, and tends to

$$R_{\mathrm{G}}^{\mathrm{lo}} = \frac{1}{2}\log(1 + \frac{P_1}{N_3 + P_2}) \tag{70}$$

This is because the distortion in Theorem 6 is equal to its maximum value P_2 in this case. Equation (70) reflects a limitation of our coding scheme for the lower bound in Theorem 6 if the relay fails to reconstruct the input described by the source. In this case, the input from the relay acts as additional noise at the destination, thus causing the cooperative transmission to perform worse than simple direct transmission. The achievable rate (70) is, however, still better than had the state been merely treated as unknown noise if $P_2 \leq Q$. (For comparison, note that the lower bound of Theorem 7 vanishes if $N_2 \longrightarrow \infty$).

VI. The Memoryless Gaussian Model with Orthogonal Components

In this section we study an important class of state-dependent Gaussian relay channels with orthogonal components. In this model, the source input $X_{1,i} = (X_{1R,i}, X_{1D,i})$ with $X_{1R,i}$ independent of the channel state S^n, and 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 inputs from the source and relay and the channel state S_i by

$$Y_{2,i} = X_{1R,i} + S_i + Z_{2,i} \tag{71a}$$

$$Y_{3,i} = X_{1D,i} + X_{2,i} + S_i + Z_{3,i}. \tag{71b}$$

We consider separate power constraints on the average transmitted power at the encoder components,

$$\sum_{i=1}^{n} X_{1R,i}^2 \leq nP_{1R}, \quad \sum_{i=1}^{n} X_{1D,i}^2 \leq nP_{1D}, \quad \sum_{i=1}^{n} X_{2,i}^2 \leq nP_2. \tag{72}$$

The definition of a code for this Gaussian model follows that for the discrete case of Section IV, with the additional constraint that the channel inputs should satisfy the power constraint (72).

A. Bounds on Channel Capacity

The following proposition provides a lower bound on the capacity of the state-dependent Gaussian relay model (71).

***Proposition** 2:* The capacity of the state-dependent Gaussian relay model (71) is lower-bounded by

$$\begin{aligned} R^{\text{lo}}_{\text{G-orth}} = \max\min\Big\{&\frac{1}{2}\log\Big(1+\frac{P_{1R}}{N_2+Q}\Big)+\frac{1}{2}\log\Big(1+\frac{P_{1D}(1-\rho_{12}^2-\rho_{1s}^2)}{N_3}\Big),\\ &\frac{1}{2}\log\Big(1+\frac{(\sqrt{P_2}+\rho_{12}\sqrt{P_{1D}})^2}{P_{1D}(1-\rho_{12}^2-\rho_{1s}^2)+(\sqrt{Q}+\rho_{1s}\sqrt{P_{1D}})^2+N_3}\Big)+\frac{1}{2}\log\Big(1+\frac{P_{1D}(1-\rho_{12}^2-\rho_{1s}^2)}{N_3}\Big)\Big\}, \end{aligned} \tag{73}$$

where the maximization is over parameters $\rho_{12} \in [0,1]$ and $\rho_{1s} \in [-1,0]$ such that

$$\rho_{12}^2 + \rho_{1s}^2 \leq 1. \tag{74}$$

Proof: The proof of Proposition 2 appears in Appendix F.

We now turn to establish an upper bound on the capacity of the Gaussian model (71). It is easy to show that the cut-set-upper bound (66) can be written as

$$R^{\text{up}}_{\text{G-orth}} = \max_{p(x_2,s)p(x_{1R}|s,x_2)p(x_{1D}|s,x_2)} \min\Big\{I(X_{1R};Y_2|S,X_2)+I(X_{1D};Y_3|S,X_2),\ \ I(X_{1D},X_2;Y_3|S)\Big\} \tag{75}$$

in this case. In what follows we establish an upper bound that is *strictly* better than (75) by accounting for that the source input component $X_{1R,i}$ at time i does not know the state S^n at all and that the relay output Y_2^{i-1} is function of only the strictly causal part of the state in this case. The following theorem states the corresponding result.

Theorem 8: The capacity of the state-dependent Gaussian relay model (71) is upper-bounded by

$$R^{\text{up}}_{\text{G-orth}} = \max\min\Big\{\frac{1}{2}\log\Big(1+\frac{P_{1R}}{N_2}\Big)+\frac{1}{2}\log\Big(1+\frac{P_{1D}(1-\rho_{12}^2-\rho_{1s}^2)}{N_3}\Big),$$
$$\frac{1}{2}\log\Big(1+\frac{(\sqrt{P_2}+\rho_{12}\sqrt{P_{1D}})^2}{P_{1D}(1-\rho_{12}^2-\rho_{1s}^2)+(\sqrt{Q}+\rho_{1s}\sqrt{P_{1D}})^2+N_3}\Big)+\frac{1}{2}\log\Big(1+\frac{P_{1D}(1-\rho_{12}^2-\rho_{1s}^2)}{N_3}\Big)\Big\}, \quad (76)$$

where the maximization is over parameters $\rho_{12}\in[0,1]$, $\rho_{1s}\in[-1,0]$ such that

$$\rho_{12}^2+\rho_{1s}^2\leq 1. \quad (77)$$

Proof: The proof of Theorem 8 appears in Appendix G.

Remark 8: Similar to the DM case, the upper bound in Theorem 8 improves upon the cut-set upper bound through the second term of the minimization. The second term of the minimization is strictly tighter than that of the cut-set upper bound because it accounts for the rate loss incurred by not knowing the state S^n at all at the source encoder component $X_{1R,i}$ that is heard at the relay and that the relay output Y_2^{i-1} can depend on the state only strictly-causally in this case. Further, investigating closely the proof in Appendix G, it can be seen that, by opposition to the corresponding DM case, the relay ignores completely any information about the state in the multiaccess part of (76).

B. Capacity for Some Special Cases

In this section, we characterize the capacity for some special Gaussian models.

The achievable rate of Proposition 2 differs from the upper bound of Theorem 8 only through the first logarithm term in (73) in which the state is taken as unknown noise in the lower bound. Substituting $\rho := \rho_{1s}$ and $\zeta := 1-\rho_{12}^2-\rho_{1s}^2$ in (76) and (73), it is easy to see that if P_{1R}, P_{1D}, P_2, Q, N_2 and N_3 satisfy

$$N_2 \leq \max_{\zeta\in[0,1],\,\rho\in[-1,0]} \frac{P_{1R}[P_{1D}\zeta+(\sqrt{Q}+\rho\sqrt{P_{1D}})^2+N_3]}{(\sqrt{P_2}+\sqrt{1-\zeta-\rho^2}\sqrt{P_{1D}})^2}-Q \quad (78)$$

then the two bounds meet; and, so give the channel capacity

$$\mathcal{C}_{\text{G-orth}} = \max_{\zeta\in[0,1],\,\rho\in[-1,0]} \frac{1}{2}\log\Big(1+\frac{(\sqrt{P_2}+\sqrt{1-\zeta-\rho^2}\sqrt{P_{1D}})^2}{P_{1D}\zeta+(\sqrt{Q}+\rho\sqrt{P_{1D}})^2+N_3}\Big)+\frac{1}{2}\log\Big(1+\frac{P_{1D}\zeta}{N_3}\Big). \quad (79)$$

Let us now consider an important special case of (71) in which the interference affects only the channel to the destination, i.e.,

$$Y_{2,i} = X_{1R,i}+Z_{2,i} \quad (80a)$$

$$Y_{3,i} = X_{1D,i}+X_{2,i}+S_i+Z_{3,i}. \quad (80b)$$

In this case, the upper bound in Theorem 8 is tight. The following theorem characterizes the channel capacity in this case.

***Theorem** 9:* The capacity of the state-dependent Gaussian relay model (80) is given by

$$C_{\text{G-orth}} = \max\min\Big\{\frac{1}{2}\log\Big(1+\frac{P_{1R}}{N_2}\Big)+\frac{1}{2}\log\Big(1+\frac{P_{1D}(1-\rho_{12}^2-\rho_{1s}^2)}{N_3}\Big),$$
$$\frac{1}{2}\log\Big(1+\frac{(\sqrt{P_2}+\rho_{12}\sqrt{P_{1D}})^2}{P_{1D}(1-\rho_{12}^2-\rho_{1s}^2)+(\sqrt{Q}+\rho_{1s}\sqrt{P_{1D}})^2+N_3}\Big)+\frac{1}{2}\log\Big(1+\frac{P_{1D}(1-\rho_{12}^2-\rho_{1s}^2)}{N_3}\Big)\Big\}, \quad (81)$$

where the maximization is over parameters $\rho_{12} \in [0,1]$ and $\rho_{1s} \in [-1,0]$ such that

$$\rho_{12}^2 + \rho_{1s}^2 \leq 1. \quad (82)$$

Proof: The proof of Theorem 9 appears in Appendix H.

Another important special case of the state-dependent Gaussian relay model of Figure 2 is one such that $Y_3 = (Y_3^{(1)}, Y_3^{(2)})$ and the conditional distribution $W_{Y_3|X_{1D},S,X_2}$ factorizes as $W_{Y_3^{(1)}|X_2}W_{Y_3^{(2)}|X_{1D},S}$,

$$Y_{2,i} = X_{1R,i} + S_i + Z_{2,i} \quad (83a)$$

$$Y_{3,i}^{(1)} = X_{1D,i} + S_i + Z_{3,i}^{(1)} \quad (83b)$$

$$Y_{3,i}^{(2)} = X_{2,i} + Z_{3,i}^{(2)}, \quad (83c)$$

where the noises $Z_{3,i}^{(1)}$ and $Z_{3,i}^{(2)}$ are zero mean Gaussian random variables with variances N_3, and are mutually independent and independent from the state sequence S^n, the source input $X_1^n = (X_{1R}^n, X_{1D}^n)$ and the relay input X_2^n. Considering average power constraint $\sum_{i=1}^n X_{1,i}^2 \leq nP_1$ on X_1^n and $\sum_{i=1}^n X_{2,i}^2 \leq nP_2$ on X_2^n, the following corollary states the capacity of this model.

***Corollary** 1:* The capacity of the state-dependent Gaussian relay model (83) is given by

$$\mathcal{C}_{\text{G-orth}} = \max\ \min\Big\{\frac{1}{2}\log\Big(1+\frac{\gamma P_1}{N_2}\Big), \frac{1}{2}\log(1+\frac{P_2}{N_3})\Big\} + \frac{1}{2}\log\Big(1+\frac{(1-\gamma)P_1}{N_3}\Big), \quad (84)$$

where the maximization is over $\gamma \in [0,1]$.

The proof of Corollary 1 follows by specializing the cut-set upper bound to the model (83) and then observing that this upper bound can actually be attained using a combination of binning and generalized block Markov scheme where we let X_{1R} and X_{1D} to be zero-mean Gaussian with variances γP_1 and $(1-\gamma)P_1$, respectively, for some $0 \leq \gamma \leq 1$, independent of S and X_2; X_2 is zero-mean Gaussian with variance P_2 independent of S; and X_{1R} and X_{1D} obtained with standard DPCs for the links to the relay and to the receiver component $Y_2^{(3)}$, respectively. The source sends information to the receiver via the relay through the dirty paper coded X_{1R}, and independent information via the direct link through the dirty paper coded X_{1D}.

Extreme cases:

1) *Arbitrarily strong channel state:* In the asymptotic case $Q \to \infty$, the capacity of the model (71) is given by

$$\mathcal{C}_{\text{G-orth}} = \frac{1}{2}\log\Big(1+\frac{P_{1D}}{N_3}\Big). \quad (85)$$

This can be easily seen since both the upper bound of Theorem 9 and the lower bound (73) tend to the RHS of (85) in this case. The RHS of (85) is also clearly achievable by turning the relay off and applying standard DPC at the source.

2) If $N_2 \longrightarrow \infty$, i.e., the link to the relay is broken or too noisy, the lower and upper bounds on the capacity of the model (71) agree and give the channel capacity as the RHS of (85).

VII. Numerical Examples and Discussion

In this section we discuss some numerical examples, for the general Gaussian RC with informed source (40), the model (71) and the special case (80). We illustrate the results of Theorems 5, 6, 7 and 8 and, for the model (40), we also include comparisons with previously known achievable rates for this model such as that obtained using compress-and-forward (CF) and binning in [32, Theorem 4] and that with partial decode-and-forward and binning in [23, Theorem 3].

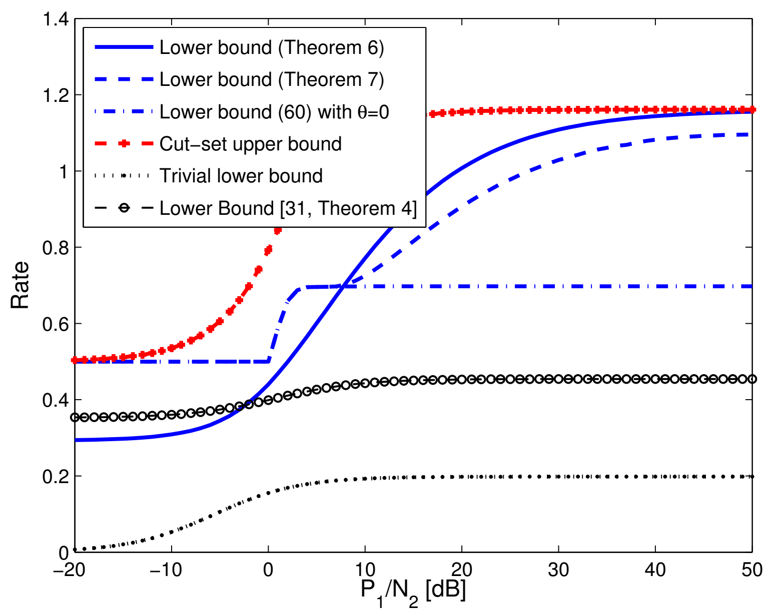


Fig. 3. Illustration of the lower bound of Theorem 6 and lower bound of Theorem 7 for the state-dependent General Gaussian RC with informed source (40) versus the SNR in the link source-to-relay. Numerical values are: $P_1 = P_2 = N_3 = 10$ dB and $Q = 15$ dB.

Figure 3 illustrates the lower bound of Theorem 6 and the lower bound of Theorem 7 for the model (40), as functions of the signal-to-noise-ratio (SNR) at the relay, i.e., SNR $= P_1/N_2$ (in decibels). Also shown for comparison are the lower bound obtained using CF and binning in [32, Theorem 4], the cut-set upper bound had the state been known also at the relay and the destination, i.e., (66), and the trivial lower bound obtained by considering the channel state as unknown noise and implementing full-DF at the relay. In order to show the effect of describing the state to the relay, the figure also shows a special case of the lower bound of Theorem 7 obtained by setting $\theta = 0$ in (60), i.e., a Gaussian version of the achievable rate (22) that we mentioned in Remark 3, and is a (slightly) improved version of [23, Theorem3].

The figure shows that the lower bound of Theorem 6 is asymptotically optimal at large SNR, and the lower bound of Theorem 7 is asymptotically optimal at small SNR. This shows the relevance of transmitting to the relay only a description of the appropriate input that it should send upon sending to it a description of the state itself at large SNR. At moderate SNR, however, sending a description of the state to the relay may improve upon sending to it a description of the appropriate Gel'fand-Pinsker binned codeword that it should send — (How the two bounds compare depends essentially on the strength of the state. For example, at large SNR, the stronger the state the larger the advantage of the lower bound of Theorem 6 upon that of Theorem 7). Furthermore, the figure also shows that the lower bound of Theorem 7 is better than that of [23, Theorem3], thereby reflecting the utility of describing the state to the relay (recall that the coding scheme that we employed for the lower bound of Theorem 7 involves also a partial cancellation of the state by the source to the relay, so that the relay benefits from it and the source benefits in turn). Figure 4 shows similar bounds computed for an example degraded Gaussian RC.

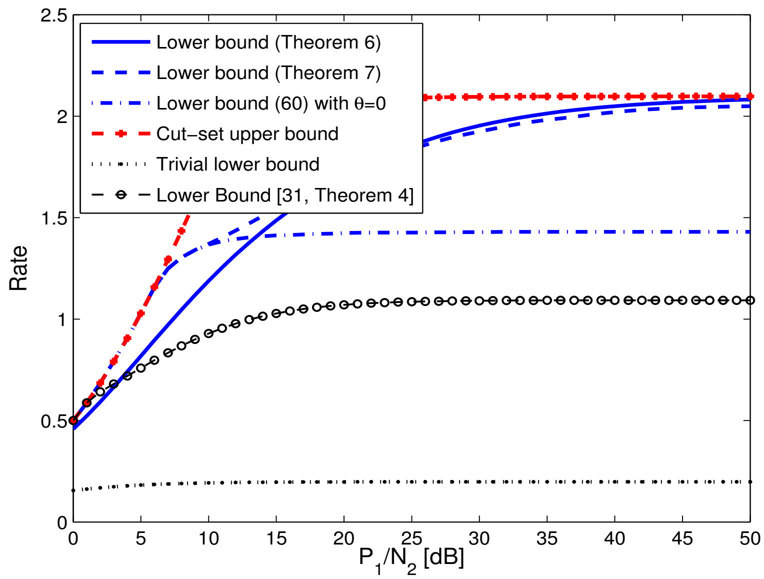
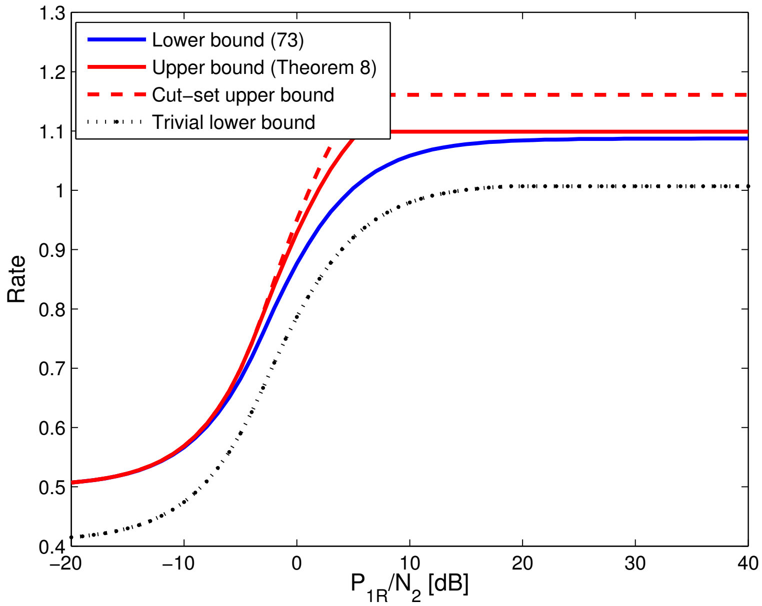
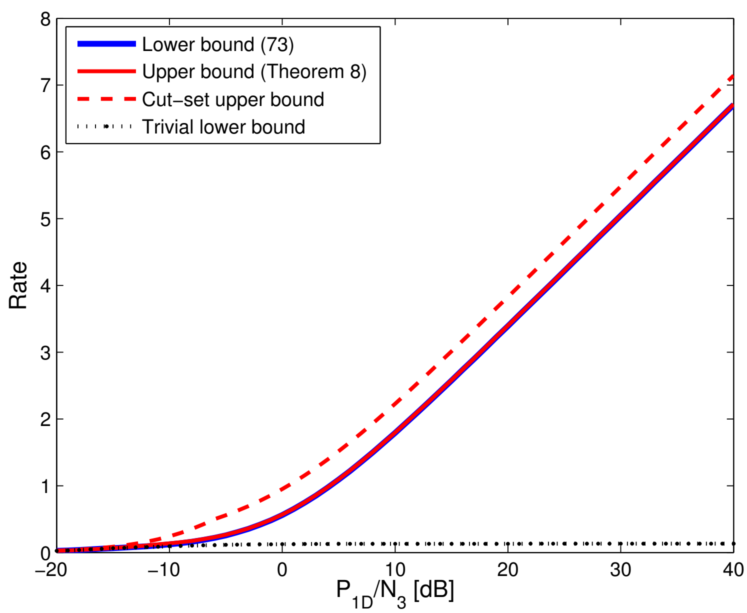


Fig. 4. Illustration of the lower bound of Theorem 6 and lower bound of Theorem 7 for an example state-dependent degraded Gaussian RC with informed source of (40), versus the SNR in the link source-to-relay. Numerical values are: $P_1 = 10$ dB, $P_2 = 20$ dB, $Q = 15$ dB, $N_3 = 10$ dB.

Remark 9: The lower bound of Theorem 6 is asymptotically close to optimal in SNR as we mentioned in the "Extremes Cases Analysis" section and is visible from Figure 3. This is because the appropriate relay input, which is precoded at the source against the state and is encoded in a manner that it should combine coherently with the source transmission in next block, can be sent by the source to the relay at almost no expense in power and can be learned by the relay with negligible distortion in this case. One can be tempted to expect a similar behavior for the lower bound of Theorem 7 since, for the latter as well, the relay can learn a "good" estimate of the state at almost no expense in source's power and with negligible distortion. This should not be, however, since our coding scheme for Theorem 7 requires the relay to also decode the source's information message. Related to this aspect, the effect of the limitation which we mentioned in Remark 7 is visible at large SNR for this lower bound. □



(a)



(b)

Fig. 5. Lower and upper bounds on the capacity of the state-dependent Gaussian RC with informed source (71). (a) bounds versus the SNR $P_{1R}/N2$ in the link source-to-relay, for numerical values $P_{1R} = P_{1D} = P_2 = N3 = 10$ dB, $Q = 5$ and (b) bounds versus the SNR $P_{1D}/N3$ in the link source-to-destination $P_{1R} = P_{1D} = P_2 = N2 = 10$ dB, $Q = 20$ dB.

Figure 5 illustrates the upper bound (76) of Theorem 8 and the lower bound (73) for the model (71). For comparison, the figure shows also the cut-set upper bound had the state been known also at the relay and the destination, i.e., (75), and the trivial lower bound obtained by considering the channel state as unknown noise and using a generalized block Markov coding scheme as in [1]. The curves are plotted against the signal-to-noise-ratio (SNR) at the relay, i.e., SNR = P_{1R}/N_2 (in decibels). Observe that the upper bound (76) is strictly better than the cut-set upper bound. The improvement is due to that the upper bound (76) accounts for some inevitable rate loss which is caused by not knowing the state at the relay, as we mentioned previously. Also, the improvement is visible mainly at small to relatively large values of SNR.

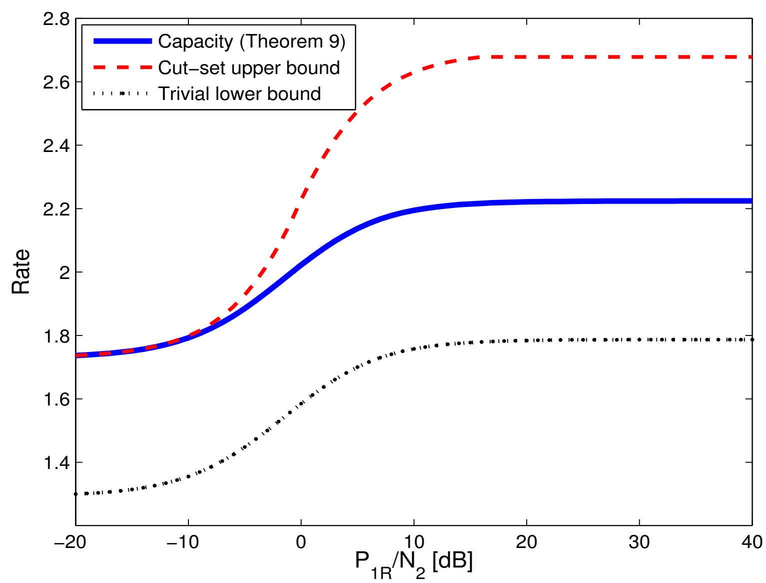


Fig. 6. Capacity of the state-dependent Gaussian RC model (80), versus the SNR in the link source-to-relay. Numerical values are: P_{1R} = 10 dB, $P_{1D} = P_2$ = 20 dB, Q = 10 dB, N_3 = 10 dB.

Figure 6 illustrates the capacity result of (80) as given by Theorem 9, as function the SNR in the link source-to-relay of P_{1R}/N_2 (in decibels). Also shown for comparison are the cut-set upper bound and the trivial lower bound obtained by considering the channel state as unknown noise and using a generalized block Markov coding scheme as in [1].

VIII. Conclusions and Discussion

In this paper, we consider a state-dependent relay channel with the channel states available noncausally at only the source, i.e., 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 some scenarios of node cooperation over wireless networks with some of the terminals equipped with cognition capabilities that enable estimating to high accuracy the states of the channel.

We investigate this problem in the discrete memoryless (DM) case and in the Gaussian case. For both cases, we derive lower and upper bounds on the channel capacity. A key feature of the model we study is that, assuming decode-and-forward relaying, the input of the relay should be generated using binning against the state that controls the channel in order to combat its effect and, at the same time, combine coherently with the source transmission. We develop two lower bounds on the capacity by using coding schemes which achieve this goal differently. In the first coding scheme, the source describes the channel state to the relay and to the destination, through a combined coding for multiple descriptions, binning and decode-and-forward scheme. The relay guesses an estimate of the transmitted information message and of the channel state and then utilizes the state estimate to perform cooperative binning with the source for sending the information message. The destination utilizes its output and the already recovered state to guess an estimate of the currently transmitted message and state description. In the second coding scheme, the source describes to the relay the appropriate input that the relay would send had the relay known the channel state. The relay then simply guesses this input and sends it in the appropriate subsequent block. The lower bound obtained with this scheme achieves close to optimal for some special cases.

Furthermore, the upper bounds that we establish in the discrete memoryless and the memoryless Gaussian cases account for not knowing the state at the relay and destination. Also, considering a class of relay channels with orthogonal channels from the source to the relay and from the source and relay to the destination in which the source input that is heard by the relay is independent of the channel state, we show that our upper bound is strictly tighter than that obtained by assuming that the channel state is also available at the relay and the destination, i.e., the max-flow min-cut or cut-set upper bound, and it helps characterizing the rate loss due to the asymmetry caused by having the channel state available at only one source encoder component. Also, we characterize the channel capacity fully in some cases, including when the state does not affect the channel to the relay.

We close this paper with a discussion on related aspects. Our coding scheme of Theorem 1 is, in essence, of decode-and-forward relaying type (though the relay also sends a compression version of the state on top of the decoded information message). Our coding scheme of Theorem 3 can be seen as being more of a non-standard compress-and-forward relaying type, since the relay sends a compressed version of the input produced at the source. Although not optimal in general, these schemes are tailored specifically to deal (at least partially) with the presence of the channel state in our model. The relay can of course employ other relaying schemes to assist the source, such as estimate-and-forward, amplify-and-forward or combinations of theses. However, while these schemes may outperform the schemes that we described in this paper for certain channel parameters, in general they do not really offer inherently better mechanisms of dealing with the presence of the channel state and exploiting its full knowledge at the source. In the case of states known causally or only strictly causally, the new noisy networking coding scheme by Lim *et al.* [52], which implements standard compression without Wyner-Ziv binning, has been proved to in general offer better rates for certain related relay [30] and multiaccess [30], [41], [42] models. For the model at hand, however, like for the standard state-independent three-terminal relay channel, noisy-network coding offers exactly the same rate as classic compress-and-forward at the relay, but no better, as observed recently in [53].

Acknowledgement

The authors would like to thank the Associate Editor and the anonymous reviewers for their useful comments and suggestions.

Appendix

Throughout this section we denote the set of strongly jointly ϵ-typical sequences [49, Chapter 14.2] with respect to the distribution $P_{X,Y}$ as $\mathcal{T}_\epsilon^n(P_{X,Y})$.

A. Proof of Theorem 1

Consider the random coding scheme that we outlined in Section III. We now analyse the average probability of error.

Analysis of Probability of Error: The average probability of error is given by

$$\begin{aligned}\Pr(\text{Error}) &= \sum_{\mathbf{s}\in\mathcal{S}^n} \Pr(\mathbf{s})\Pr(\text{error}|\mathbf{s})\\ &\leq \sum_{\mathbf{s}\notin\mathcal{T}_\epsilon^n(Q_S)} \Pr(\mathbf{s}) + \sum_{\mathbf{s}\in\mathcal{T}_\epsilon^n(Q_S)} \Pr(\mathbf{s})\Pr(\text{error}|\mathbf{s}). \qquad \text{(A-1)}\end{aligned}$$

The first term, $\Pr(\mathbf{s} \notin \mathcal{T}_\epsilon^n(Q_S))$, on the RHS of (A-1) goes to zero as $n \to +\infty$, by the strong asymptotic equipartition property (AEP) [49, p. 384]. 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_{ki}, $k = 3, \ldots, 6$, correspond to decoding errors at the relay at block i; and the events E_{ki}, $k = 7, \ldots, 13$, correspond to decoding errors at the destination at block i.

- Let $E_{1i} = E_{1i}^{(1)} \cup E_{1i}^{(2)} \cup E_{1i}^{(3)}$, with

$$\begin{aligned}E_{1i}^{(1)} &= \Big\{(\mathbf{s}[i+2], \hat{\mathbf{s}}_R[\iota_{Ri}]) \notin T_\epsilon^n(P_{S,\hat{S}_R}), \quad \textit{for all} \quad \iota_{Ri} \in [1, 2^{n\hat{R}_R}]\Big\}\\ E_{1i}^{(2)} &= \Big\{(\mathbf{s}[i+2], \hat{\mathbf{s}}_D[\iota_{Di}]) \notin T_\epsilon^n(P_{S,\hat{S}_D}), \quad \textit{for all} \quad \iota_{Di} \in [1, 2^{n\hat{R}_D}]\Big\}\\ E_{1i}^{(3)} &= \Big\{(\mathbf{s}[i+2], \hat{\mathbf{s}}_R[\iota_{Ri}], \hat{\mathbf{s}}_D[\iota_{Di}]) \notin T_\epsilon^n(P_{S,\hat{S}_R,\hat{S}_D}), \quad \textit{for all} \quad (\iota_{Ri}, \iota_{Di}) \in [1, 2^{n\hat{R}_R}] \times [1, 2^{n\hat{R}_D}]\Big\}. \qquad \text{(A-2)}\end{aligned}$$

  From known results in rate distortion theory [49, p. 336], it follows that $P(E_{1i}^{(1)}) \longrightarrow 0$ exponentially with n if $\hat{R}_R > I(S; \hat{S}_R)$. Similarly, $P(E_{1i}^{(2)}) \longrightarrow 0$ exponentially with n if $\hat{R}_D > I(S; \hat{S}_D)$. It remains to show that $P(E_{1i}^{(3)}) \longrightarrow 0$ exponentially with n if $\hat{R}_R + \hat{R}_D > I(S; \hat{S}_R, \hat{S}_D) + I(\hat{S}_R; \hat{S}_D)$, and this can be proved by following straightforwardly the arguments and algebra in [45].

- Let E_{2i} be the event that there is no pair $\big(\mathbf{u}_R(w_{i-1}, j_{Vi}^\star, w_i, j_{Ui}^\star, k_i, j_{Ri}), \mathbf{u}_D(w_{i-1}, j_{Vi}^\star, w_i, j_{Ui}^\star, l_i, j_{Di})\big)$ satisfying (14), i.e., the set $\mathcal{D}_{\iota_{Ri}\iota_{Di}}$ is empty.

  Using Chebychev's inequality, it is easy to see that

$$\begin{aligned}P\Big(\|\mathcal{D}_{\iota_{Ri}\iota_{Di}}\| = 0\Big) &\leq P\Big(\|\|\mathcal{D}_{\iota_{Ri}\iota_{Di}}\| - \mathbb{E}[\mathcal{D}_{\iota_{Ri}\iota_{Di}}]\| > \epsilon\mathbb{E}[\mathcal{D}_{\iota_{Ri}\iota_{Di}}]\Big)\\ &\leq \frac{\text{var}(\|\mathcal{D}_{\iota_{Ri}\iota_{Di}}\|)}{\epsilon^2(\mathbb{E}[\mathcal{D}_{\iota_{Ri}\iota_{Di}}])^2}. \qquad \text{(A-3)}\end{aligned}$$

We obtain bounds on $\mathbb{E}[\mathcal{D}_{\iota_{Ri}\iota_{Di}}]$ and $\mathrm{var}(\|\mathcal{D}_{\iota_{Ri}\iota_{Di}}\|)$ by proceeding in a way similar to [47]. We define the indicator functions,

$$\mathbb{1}\Big(\big(\mathbf{u}_R(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui}, k_i, j_{Ri}), \mathbf{u}_D(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui}, l_i, j_{Di})\big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}\Big) = \begin{cases} 1, & \text{if } \big(\mathbf{u}_R(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui}, k_i, j_{Ri}), \mathbf{u}_D(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui}, l_i, j_{Di})\big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}} \\ 0, & \text{otherwise.} \end{cases} \tag{A-4}$$

The cardinality of the set $\mathcal{D}_{\iota_{Ri}\iota_{Di}}$ is given by

$$\|\mathcal{D}_{\iota_{Ri}\iota_{Di}}\| = \sum_{k_i\in\mathcal{B}_{\iota_{Ri}}, l_i\in\mathcal{C}_{\iota_{Di}}} \sum_{j_{Ri}\in J_R, j_{Di}\in J_D} \mathbb{1}\Big(\big(\mathbf{u}_R(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui}, k_i, j_{Ri}), \mathbf{u}_D(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui}, l_i, j_{Di})\big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}\Big). \tag{A-5}$$

Thus,

$$\begin{aligned}\mathbb{E}[\mathcal{D}_{\iota_{Ri}\iota_{Di}}] &= \sum_{k_i\in\mathcal{B}_{\iota_{Ri}}, l_i\in\mathcal{C}_{\iota_{Di}}} \sum_{j_{Ri}\in J_R, j_{Di}\in J_D} \mathbb{E}\mathbb{1}\Big(\big(\mathbf{u}_R(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui}, k_i, j_{Ri}), \mathbf{u}_D(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui}, l_i, j_{Di})\big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}\Big) \\ &\geq \|\mathcal{B}_{\iota_{Ri}}\|\|\mathcal{C}_{\iota_{Di}}\| J_R J_D 2^{-n[I(U_R;S,\hat{S}_R,\hat{S}_D|U,V)+I(U_D;S,\hat{S}_R,\hat{S}_D|U,V)-I(U_R;U_D|U,V,S,\hat{S}_R,\hat{S}_D)+o(1)]} \\ &= 2^{n[R_R+R_D-\hat{R}_R-\hat{R}_D-I(U_R;U_D|U,V,S,\hat{S}_R,\hat{S}_D)-o(1)]}\end{aligned} \tag{A-6}$$

where $o(1) \longrightarrow 0$ as $n \longrightarrow \infty$.

Evaluating the variance, it can be shown (see Lemma 1 below) that

$$\mathrm{var}(\|\mathcal{D}_{\iota_{Ri}\iota_{Di}}\|) \leq 2^{n[R_R+R_D-\hat{R}_R-\hat{R}_D-I(U_R;U_D|U,V,S,\hat{S}_R,\hat{S}_D)+o(1)]}. \tag{A-7}$$

Therefore, for sufficiently large n

$$P\Big(\|\mathcal{D}_{\iota_{Ri}\iota_{Di}}\| = 0\Big) \leq \epsilon \tag{A-8}$$

provided that (15) is true.

Lemma 1:

$$\mathrm{var}(\|\mathcal{D}_{\iota_{Ri}\iota_{Di}}\|) \leq 2^{n[R_R+R_D-\hat{R}_R-\hat{R}_D-I(U_R;U_D|U,V,S,\hat{S}_R,\hat{S}_D)+o(1)]}. \tag{A-9}$$

Proof: For notational convenience, let us use temporarily in the proof of this lemma the shorthand notation $\mathbf{u}_R(k_i, j_{Ri}) := \mathbf{u}_R(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui}, k_i, j_{Ri})$ and $\mathbf{u}_D(l_i, j_{Di}) := \mathbf{u}_D(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui}, l_i, j_{Di})$. Then, we have

$$\begin{aligned}&\|\mathcal{D}_{\iota_{Ri}\iota_{Di}}\|^2 \\ &= \Bigg(\sum_{k_i\in\mathcal{B}_{\iota_{Ri}}, l_i\in\mathcal{C}_{\iota_{Di}}} \sum_{j_{Ri}\in J_R, j_{Di}\in J_D} \mathbb{1}\Big(\big(\mathbf{u}_R(k_i, j_{Ri}), \mathbf{u}_D(l_i, j_{Di})\big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}\Big)\Bigg)^2 \\ &= \sum_{(k_i,j_{Ri})=(k'_i,j'_{Ri})} \sum_{(l_i,j_{Di})=(l'_i,j'_{Di})} \mathbb{1}\Big(\big(\mathbf{u}_R(k_i, j_{Ri}), \mathbf{u}_D(l_i, j_{Di})\big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}\Big) \\ &\quad + \sum_{(k_i,j_{Ri})=(k'_i,j'_{Ri})} \sum_{(l_i,j_{Di})\neq(l'_i,j'_{Di})} \mathbb{1}\Big(\big(\mathbf{u}_R(k_i, j_{Ri}), \mathbf{u}_D(l_i, j_{Di})\big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}, \big(\mathbf{u}_R(k_i, j_{Ri}), \mathbf{u}_D(l'_i, j'_{Di})\big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}\Big) \\ &\quad + \sum_{(k_i,j_{Ri})\neq(k'_i,j'_{Ri})} \sum_{(l_i,j_{Di})=(l'_i,j'_{Di})} \mathbb{1}\Big(\big(\mathbf{u}_R(k_i, j_{Ri}), \mathbf{u}_D(l_i, j_{Di})\big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}, \big(\mathbf{u}_R(k'_i, j'_{Ri}), \mathbf{u}_D(l_i, j_{Di})\big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}\Big) \\ &\quad + \sum_{(k_i,j_{Ri})\neq(k'_i,j'_{Ri})} \sum_{(l_i,j_{Di})\neq(l'_i,j'_{Di})} \mathbb{1}\Big(\big(\mathbf{u}_R(k_i, j_{Ri}), \mathbf{u}_D(l_i, j_{Di})\big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}, \big(\mathbf{u}_R(k'_i, j'_{Ri}), \mathbf{u}_D(l'_i, j'_{Di})\big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}\Big).\end{aligned} \tag{A-10}$$

Taking the expectation and dividing by $\|\mathcal{B}_{\iota_{Ri}} \times \mathcal{C}_{\iota_{Di}} \times \mathcal{J}_R \times \mathcal{J}_D\|$ in both sides of (A-10), we get

$$\begin{aligned}\frac{\mathbb{E}[\|\mathcal{D}_{\iota_{Ri}\iota_{Di}}\|^2]}{\|\mathcal{B}_{\iota_{Ri}}\|\|\mathcal{C}_{\iota_{Di}}\|J_R J_D} &= \Pr\Big\{\Big(\mathbf{u}_R(k_i, j_{Ri}), \mathbf{u}_D(l_i, j_{Di})\Big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}\Big\} \\ &+ (J_D\|\mathcal{C}_{\iota_{Di}}\| - 1)\Pr\Big\{\Big(\mathbf{u}_R(k_i, j_{Ri}), \mathbf{u}_D(l_i, j_{Di})\Big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}, \Big(\mathbf{u}_R(k_i, j_{Ri}), \mathbf{u}_D(l'_i, j'_{Di})\Big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}\Big\} \\ &+ (J_R\|\mathcal{B}_{\iota_{Ri}}\| - 1)\Pr\Big\{\Big(\mathbf{u}_R(k_i, j_{Ri}), \mathbf{u}_D(l_i, j_{Di})\Big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}, \Big(\mathbf{u}_R(k'_i, j'_{Ri}), \mathbf{u}_D(l_i, j_{Di})\Big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}\Big\} \\ &+ (J_R\|\mathcal{B}_{\iota_{Ri}}\| - 1)(J_D\|\mathcal{C}_{\iota_{Di}}\| - 1)\Pr\Big\{\Big(\mathbf{u}_R(k_i, j_{Ri}), \mathbf{u}_D(l_i, j_{Di})\Big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}, \Big(\mathbf{u}_R(k'_i, j'_{Ri}), \mathbf{u}_D(l'_i, j'_{Di})\Big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}\Big\}.\end{aligned} \tag{A-11}$$

Let $\Delta := I(U_R; S, \hat{S}_R, \hat{S}_D|U, V) + I(U_D; S, \hat{S}_R, \hat{S}_D|U, V) - I(U_R; U_D|U, VS, \hat{S}_R, \hat{S}_D)$. It can be shown easily that

i) For $(k_i, j_{Ri}) = (k'_i, j'_{Ri})$ and $(l_i, j_{Di}) = (l'_i, j'_{Di})$,

$$\Pr\Big\{\Big(\mathbf{u}_R(k_i, j_{Ri}), \mathbf{u}_D(l_i, j_{Di})\Big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}\Big\} \leq 2^{-n(\Delta - \delta(\epsilon))}. \tag{A-12}$$

ii) For $(k_i, j_{Ri}) = (k'_i, j'_{Ri})$ and $(l_i, j_{Di}) \neq (l'_i, j'_{Di})$,

$$\Pr\Big\{\Big(\mathbf{u}_R(k_i, j_{Ri}), \mathbf{u}_D(l_i, j_{Di})\Big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}, \Big(\mathbf{u}_R(k_i, j_{Ri}), \mathbf{u}_D(l'_i, j'_{Di})\Big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}\Big\} \leq 2^{-2n(\Delta - \delta(\epsilon))}. \tag{A-13}$$

iii) For $(k_i, j_{Ri}) \neq (k'_i, j'_{Ri})$ and $(l_i, j_{Di}) = (l'_i, j'_{Di})$,

$$\Pr\Big\{\Big(\mathbf{u}_R(k_i, j_{Ri}), \mathbf{u}_D(l_i, j_{Di})\Big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}, \Big(\mathbf{u}_R(k'_i, j'_{Ri}), \mathbf{u}_D(l_i, j_{Di})\Big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}\Big\} \leq 2^{-2n(\Delta - \delta(\epsilon))}. \tag{A-14}$$

iv) For $(k_i, j_{Ri}) \neq (k'_i, j'_{Ri})$ and $(l_i, j_{Di}) \neq (l'_i, j'_{Di})$,

$$\Pr\Big\{\Big(\mathbf{u}_R(k_i, j_{Ri}), \mathbf{u}_D(l_i, j_{Di})\Big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}, \Big(\mathbf{u}_R(k'_i, j'_{Ri}), \mathbf{u}_D(l'_i, j'_{Di})\Big) \in \mathcal{D}_{\iota_{Ri}\iota_{Di}}\Big\} \leq 2^{-2n(\Delta - \delta(\epsilon))}. \tag{A-15}$$

Finally, substituting i-iv in the RHS of (A-11) and using (A-6), we obtain

$$\begin{aligned}\mathrm{var}(\|\mathcal{D}_{\iota_{Ri}\iota_{Di}}\|) &= \mathbb{E}[\|\mathcal{D}_{\iota_{Ri}\iota_{Di}}\|^2] - \mathbb{E}^2[\|\mathcal{D}_{\iota_{Ri}\iota_{Di}}\|] \\ &\leq \|\mathcal{B}_{\iota_{Ri}}\|\|\mathcal{C}_{\iota_{Di}}\|J_R J_D 2^{-n(\Delta - o(1))} \\ &= 2^{n[R_R + R_D - \hat{R}_R - \hat{R}_D - I(U_R; U_D|U, V, S, \hat{S}_R, \hat{S}_D) + o(1)]}\end{aligned} \tag{A-16}$$

This completes the proof of Lemma 1. ∎

- Let E_{3i} be the event that $\mathbf{u}(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui})$, $\mathbf{u}_R(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui}, k_i, j^\star_{Ri})$ are not jointly typical with $(\mathbf{y}_2[i], \hat{\mathbf{s}}_R[\iota_{Ri-2}])$ given $\mathbf{v}(w_{i-1}, j^\star_{Vi})$. That is

$$E_{3i} = \Big\{\Big(\mathbf{v}(w_{i-1}, j^\star_{Vi}), \mathbf{u}(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui}), \mathbf{u}_R(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui}, k_i, j^\star_{Ri}), \mathbf{y}_2[i], \hat{\mathbf{s}}_R[\iota_{Ri-2}]\Big) \notin \mathcal{T}^n_\epsilon(P_{V,U,U_R,Y_2,\hat{S}_R})\Big\}. \tag{A-17}$$

For $\mathbf{v}(w_{i-1}, j^\star_{Vi})$, $\mathbf{u}(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui})$, $\mathbf{u}_R(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui}, k_i, j^\star_{Ri})$, $\mathbf{u}_D(w_{i-1}, j^\star_{Vi}, w_i, j^\star_{Ui}, l_i, j^\star_{Di})$ jointly typical with $\mathbf{s}[i]$, $\hat{\mathbf{s}}_R[\iota_{Ri-2}]$, $\hat{\mathbf{s}}_D[\iota_{Di-2}]$ and with the source input $\mathbf{x}_1[i]$ and the relay input $\mathbf{x}_2[i]$, we have $\Pr(E_{3i}|E^c_{1i}, E^c_{2i}) \longrightarrow 0$ as $n \longrightarrow \infty$ by the Markov Lemma [49, p. 436].

- Let E_{4i} be the event that $\mathbf{u}(w_{i-1}, j^\star_{Vi}, w'_i, j_{Ui})$, $\mathbf{u}_R(w_{i-1}, j^\star_{Vi}, w'_i, j_{Ui}, k_i, j_{Ri})$ are jointly typical with $(\mathbf{y}_2[i], \hat{\mathbf{s}}_R[\iota_{Ri-2}])$ given $\mathbf{v}(w_{i-1}, j^\star_{Vi})$, for some $w'_i \in [1, M]$, $j_{Ui} \in J_U$, $k_i \in [1, M_R]$ and $j_{Ri} \in J_R$, with $w'_i \neq w_i$. That is,

$$\begin{aligned}E_{4i} = \Big\{&\exists\, w'_i \in [1, M],\ j_{Ui} \in J_U, k_i \in [1, M_R], j_{Ri} \in J_R \text{ s.t.: } w'_i \neq w_i, \\ &\Big(\mathbf{v}(w_{i-1}, j^\star_{Vi}), \mathbf{u}(w_{i-1}, j^\star_{Vi}, w'_i, j_{Ui}), \mathbf{u}_R(w_{i-1}, j^\star_{Vi}, w'_i, j_{Ui}, k_i, j_{Ri}), \mathbf{y}_2[i], \hat{\mathbf{s}}_R[\iota_{Ri-2}]\Big) \in \mathcal{T}^n_\epsilon(P_{V,U,U_R,Y_2,\hat{S}_R})\Big\}.\end{aligned} \tag{A-18}$$

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

$$\begin{aligned}\Pr(E_{4i}|E_{1i}^c, E_{2i}^c, E_{3i}^c) &\leq M J_U M_R J_R 2^{-n[I(U,U_R;Y_2,\hat{S}_R|V)-\epsilon]} \\ &= 2^{-n[I(U;Y_2|V,\hat{S}_R)-I(U;S,\hat{S}_D|V,\hat{S}_R)-R+4\epsilon]}. \end{aligned} \tag{A-19}$$

Thus, $\Pr(E_{4i}|E_{1i}^c, E_{2i}^c, E_{3i}^c) \longrightarrow 0$ as $n \longrightarrow \infty$ if $R < I(U;Y_2|V,\hat{S}_R) - I(U;S,\hat{S}_D|V,\hat{S}_R)$.

- Let E_{5i} be the event that $\mathbf{u}(w_{i-1}, j_{Vi}^\star, w_i, j_{Ui}')$, $\mathbf{u}_R(w_{i-1}, j_{Vi}^\star, w_i, j_{Ui}', k_i, j_{Ri})$ are jointly typical with $(\mathbf{y}_2[i], \hat{\mathbf{s}}_R[\iota_{Ri-2}])$ given $\mathbf{v}(w_{i-1}, j_{Vi}^\star)$, for some $j_{Ui}' \in J_U$, $k_i \in [1, M_R]$, $j_{Ri} \in J_R$ with $j_{Ui}' \neq j_{Ui}^\star$. That is,

$$\begin{aligned} E_{5i} = \Big\{ &\exists\, j_{Ui}' \in J_U, k_i \in [1, M_R], j_{Ri} \in J_R \text{ s.t. } j_{Ui}' \neq j_{Ui}^\star, \\ &\Big(\mathbf{v}(w_{i-1}, j_{Vi}^\star), \mathbf{u}(w_{i-1}, j_{Vi}^\star, w_i, j_{Ui}'), \mathbf{u}_R(w_{i-1}, j_{Vi}^\star, w_i, j_{Ui}', k_i, j_{Ri}), \mathbf{y}_2[i], \hat{\mathbf{s}}_R[\iota_{Ri-2}]\Big) \in \mathcal{T}_\epsilon^n(P_{V,U,U_R,Y_2,\hat{S}_R})\Big\}. \end{aligned} \tag{A-20}$$

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

$$\begin{aligned}\Pr(E_{5i}|E_{1i}^c, E_{2i}^c, E_{3i}^c, E_{4i}^c) &\leq J_U M_R J_R 2^{-n[I(U,U_R;Y_2,\hat{S}_R|V)-\epsilon]} \\ &= 2^{-n[I(U;Y_2|V,\hat{S}_R)-I(U;S,\hat{S}_D|V,\hat{S}_R)+3\epsilon]}. \end{aligned} \tag{A-21}$$

Thus, $\Pr(E_{5i}|E_{1i}^c, E_{2i}^c, E_{3i}^c, E_{4i}^c) \longrightarrow 0$ as $n \longrightarrow \infty$.

- Let E_{6i} be the event that $\mathbf{u}_R(w_{i-1}, j_{Vi}^\star, w_i, j_{Ui}^\star, k_i', j_{Ri})$ is jointly typical with $(\mathbf{y}_2[i], \hat{\mathbf{s}}_R[\iota_{Ri-2}])$ given $\mathbf{v}(w_{i-1}, j_{Vi}^\star)$, $\mathbf{u}(w_{i-1}, j_{Vi}^\star, w_i, j_{Ui}^\star)$, for some $k_i' \in [1, M_R]$, $j_{Ri} \in J_R$ with $k_i' \neq k_i$. That is,

$$\begin{aligned} E_{6i} = \Big\{ &\exists\, k_i' \in [1, M_R], j_{Ri} \in J_R \text{ s.t. } k_i' \neq k_i, \\ &\Big(\mathbf{v}(w_{i-1}, j_{Vi}^\star), \mathbf{u}(w_{i-1}, j_{Vi}^\star, w_i, j_{Ui}^\star), \mathbf{u}_R(w_{i-1}, j_{Vi}^\star, w_i, j_{Ui}^\star, k_i', j_{Ri}), \mathbf{y}_2[i], \hat{\mathbf{s}}_R[\iota_{Ri-2}]\Big) \in \mathcal{T}_\epsilon^n(P_{V,U,U_R,Y_2,\hat{S}_R})\Big\}. \end{aligned} \tag{A-22}$$

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

$$\begin{aligned}\Pr(E_{6i}|E_{1i}^c, E_{2i}^c, E_{3i}^c, E_{4i}^c, E_{5i}^c) &\leq M_R J_R 2^{-n[I(U_R;Y_2,\hat{S}_R|U,V)-\epsilon]} \\ &= 2^{-n(4\epsilon)}. \end{aligned} \tag{A-23}$$

Thus, $\Pr(E_{6i}|E_{1i}^c, E_{2i}^c, E_{3i}^c, E_{4i}^c, E_{5i}^c) \longrightarrow 0$ as $n \longrightarrow \infty$.

- For decoding the triple $(\hat{w}_{i-1}, \hat{j}_{Ui-1}, \hat{l}_{i-1})$ and the index $\hat{j}_{Vi}$ at the destination, let E_{7i} be the union of the following two events

$$\begin{aligned} E_{7i}^{(1)} &= \Big\{\Big(\mathbf{v}(w_{i-2}, j_V(\hat{\mathbf{s}}_R[\iota_{Ri-3}], w_{i-2})), \mathbf{u}(w_{i-2}, j_V(\hat{\mathbf{s}}_R[\iota_{Ri-3}], w_{i-2}), w_{i-1}, j_{Ui-1}^\star), \\ &\qquad \mathbf{u}_D(w_{i-2}, j_{Vi-1}^\star, w_{i-1}, j_{Ui-1}^\star, l_{i-1}, j_{Di-1}^\star), \mathbf{y}_3[i-1], \hat{\mathbf{s}}_D[\iota_{Di-3}]\Big) \notin \mathcal{T}_\epsilon^n(P_{V,U,U_D,Y_3,\hat{S}_D})\Big\} \\ E_{7i}^{(2)} &= \Big\{\Big(\mathbf{v}(w_{i-1}, j_V(\hat{\mathbf{s}}_R[\iota_{Ri-2}], w_{i-1})), \mathbf{y}_3[i], \hat{\mathbf{s}}_D[\iota_{Di-2}]\Big) \notin \mathcal{T}_\epsilon^n(P_{V,Y_3,\hat{S}_D})\Big\}. \end{aligned}$$

For $\mathbf{v}(w_{i-2}, j_{Vi-1}^\star)$, $\mathbf{u}(w_{i-2}, j_{Vi-1}^\star, w_{i-1}, j_{Ui-1}^\star)$, $\mathbf{u}_R(w_{i-2}, j_{Vi-1}^\star, w_{i-1}, j_{Ui-1}^\star, k_{i-1}, j_{Ri-1}^\star)$, $\mathbf{u}_D(w_{i-2}, j_{Vi-1}^\star, w_{i-1}, j_{Ui-1}^\star, l_{i-1}, j_{Di-1}^\star)$ jointly typical with $\mathbf{s}[i-1]$, $\hat{\mathbf{s}}_R[\iota_{Ri-3}]$, $\hat{\mathbf{s}}_D[\iota_{Di-3}]$ and with the source input $\mathbf{x}_1[i-1]$ and the relay input $\mathbf{x}_2[i-1]$, we have $\Pr(E_{7i}^{(1)}|\cap_{k=1}^{6} E_{ki}^c) \longrightarrow 0$ as $n \longrightarrow \infty$ by the Markov Lemma. Similarly, $\Pr(E_{7i}^{(2)}|\cap_{k=1}^{6} E_{ki}^c) \longrightarrow 0$ as $n \longrightarrow \infty$. Thus, $\Pr(E_{7i}|\cap_{k=1}^{6} E_{ki}^c) \longrightarrow 0$ as $n \longrightarrow \infty$.

- For decoding the triple $(\hat{w}_{i-1}, \hat{j}_{Ui-1}, \hat{l}_{i-1})$ and the index $\hat{j}_{Vi}$ at the destination, let E_{8i} be the event

$$E_{8i} = \Big\{\exists\ w'_{i-1} \in [1, M], j_{Ui-1} \in J_U, l_{i-1} \in [1, M_D], j_{Di-1} \in J_D, j_{Vi} \in J_V \text{ s.t.: } w'_{i-1} \neq w_{i-1},$$
$$\Big(\mathbf{v}(w_{i-2}, j^\star_{Vi-1}), \mathbf{u}(w_{i-2}, j^\star_{Vi-1}, w'_{i-1}, j_{Ui-1}), \mathbf{u}_D(w_{i-2}, j^\star_{Vi-1}, w'_{i-1}, j_{Ui-1}, l_{i-1}, j_{Di-1}), \mathbf{y}_3[i-1], \hat{\mathbf{s}}_D[\iota_{Ri-3}]\Big) \in \mathcal{T}^n_\epsilon(P_{V,U,U_D,Y_3,\hat{S}_D}),$$
$$\Big(\mathbf{v}(w'_{i-1}, j_{Vi}), \mathbf{y}_3[i], \hat{\mathbf{s}}_D[\iota_{Di-2}]\Big) \in \mathcal{T}^n_\epsilon(P_{V,Y_3,\hat{S}_D})\Big\}.$$

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

$$\begin{aligned}\Pr(E_{8i}|\cap_{k=1}^{7} E_{ki}^c) &\leq M J_U M_D J_D J_V 2^{-n[I(U,U_D;Y_3,\hat{S}_D|V)-\epsilon]}2^{-n[I(V;Y_3,\hat{S}_D)-\epsilon]}\\ &= 2^{-n[I(V,U;Y_3|\hat{S}_D)-I(V,U;S,\hat{S}_R|\hat{S}_D)-R-[I(U;Y_3,\hat{S}_D|V)-I(U;S,\hat{S}_R,\hat{S}_D|V)]_-+2\epsilon]}.\end{aligned} \tag{A-24}$$

Thus, $\Pr(E_{8i}|\cap_{k=1}^{7} E_{ki}^c) \longrightarrow 0$ as $n \longrightarrow \infty$ if $R < I(U, V; Y_3, \hat{S}_D) - I(U, V; S, \hat{S}_R, \hat{S}_D)$.

- For decoding the triple $(\hat{w}_{i-1}, \hat{j}_{Ui-1}, \hat{l}_{i-1})$ and the index $\hat{j}_{Vi}$ at the destination, let E_{9i} be the event

$$E_{9i} = \Big\{\exists\ j_{Vi} \in J_V \text{ s.t.: } j_{Vi} \neq j^\star_{Vi},$$
$$\Big(\mathbf{v}(w_{i-2}, j^\star_{Vi-1}), \mathbf{u}(w_{i-2}, j^\star_{Vi-1}, w_{i-1}, j^\star_{Ui-1}), \mathbf{u}_D(w_{i-2}, j^\star_{Vi-1}, w_{i-1}, j^\star_{Ui-1}, l_{i-1}, j_{Di-1}), \mathbf{y}_3[i-1], \hat{\mathbf{s}}_D[\iota_{Ri-3}]\Big) \in \mathcal{T}^n_\epsilon(P_{V,U,U_D,Y_3,\hat{S}_D}),$$
$$\Big(\mathbf{v}(w_{i-1}, j_{Vi}), \mathbf{y}_3[i], \hat{\mathbf{s}}_D[\iota_{Di-2}]\Big) \in \mathcal{T}^n_\epsilon(P_{V,Y_3,\hat{S}_D})\Big\}.$$

Conditioned on $\cap_{k=1}^{8} E_{ki}^c$, the probability of the event E_{9i} can be bounded using the union bound, as

$$\begin{aligned}\Pr(E_{9i}|\cap_{k=1}^{8} E_{ki}^c) &\leq J_V 2^{-n[I(V;Y_3,\hat{S}_D)-\epsilon]}\\ &= 2^{-n[I(V;Y_3,\hat{S}_D)-I(V;S,\hat{S}_R,\hat{S}_D)-2\epsilon]}.\end{aligned} \tag{A-25}$$

Thus, $\Pr(E_{9i}|\cap_{k=1}^{8} E_{ki}^c) \longrightarrow 0$ as $n \longrightarrow \infty$ if $I(V; Y_3, \hat{S}_D) - I(V; S, \hat{S}_R, \hat{S}_D) > 2\epsilon$.

- For decoding the triple $(\hat{w}_{i-1}, \hat{j}_{Ui-1}, \hat{l}_{i-1})$ and the index $\hat{j}_{Vi}$ at the destination, let E_{10i} be the event

$$E_{10i} = \Big\{\exists\ j'_{Ui-1} \in J_U, l_{i-1} \in [1, M_D], j_{Di-1} \in J_D, j_{Vi} \in J_V \text{ s.t.: } j'_{Ui-1} \neq j^\star_{Ui-1},\ j_{Vi} \neq j^\star_{Vi}$$
$$\Big(\mathbf{v}(w_{i-2}, j^\star_{Vi-1}), \mathbf{u}(w_{i-2}, j^\star_{Vi-1}, w_{i-1}, j'_{Ui-1}), \mathbf{u}_D(w_{i-2}, j^\star_{Vi-1}, w_{i-1}, j'_{Ui-1}, l_{i-1}, j_{Di-1}), \mathbf{y}_3[i-1], \hat{\mathbf{s}}_D[\iota_{Ri-3}]\Big) \in \mathcal{T}^n_\epsilon(P_{V,U,U_D,Y_3,\hat{S}_D}),$$
$$\Big(\mathbf{v}(w_{i-1}, j_{Vi}), \mathbf{y}_3[i], \hat{\mathbf{s}}_D[\iota_{Di-2}]\Big) \in \mathcal{T}^n_\epsilon(P_{V,Y_3,\hat{S}_D})\Big\}.$$

Conditioned on $\cap_{k=1}^{9} E_{ki}^c$, the probability of the event E_{10i} can be bounded using the union bound, as

$$\begin{aligned}\Pr(E_{10i}|\cap_{k=1}^{9} E_{ki}^c) &\leq J_U M_D J_D J_V 2^{-n[I(U,U_D;Y_3,\hat{S}_D|V)-\epsilon]}2^{-n[I(V;Y_3,\hat{S}_D)-\epsilon]}\\ &= 2^{-n[I(U,V;Y_3|\hat{S}_D)-I(U,V;S,\hat{S}_R|\hat{S}_D)-[I(U;Y_3,\hat{S}_D|V)-I(U;S,\hat{S}_R,\hat{S}_D|V)]_-+\epsilon]}.\end{aligned} \tag{A-26}$$

Thus, $\Pr(E_{10i}|\cap_{k=1}^{9} E_{ki}^c) \longrightarrow 0$ as $n \longrightarrow \infty$.

- For decoding the triple $(\hat{w}_{i-1}, \hat{j}_{Ui-1}, \hat{l}_{i-1})$ and the index $\hat{j}_{Vi}$ at the destination, let E_{11i} be the event

$$E_{11i} = \Big\{\exists\ j'_{Ui-1} \in J_U, l_{i-1} \in [1, M_D], j_{Di-1} \in J_D, j_{Vi} \in J_V \text{ s.t.: } j'_{Ui-1} \neq j^\star_{Ui-1},$$
$$\Big(\mathbf{v}(w_{i-2}, j^\star_{Vi-1}), \mathbf{u}(w_{i-2}, j^\star_{Vi-1}, w_{i-1}, j'_{Ui-1}), \mathbf{u}_D(w_{i-2}, j^\star_{Vi-1}, w_{i-1}, j'_{Ui-1}, l_{i-1}, j_{Di-1}), \mathbf{y}_3[i-1], \hat{\mathbf{s}}_D[\iota_{Ri-3}]\Big) \in \mathcal{T}^n_\epsilon(P_{V,U,U_D,Y_3,\hat{S}_D}),$$
$$\Big(\mathbf{v}(w_{i-1}, j^\star_{Vi}), \mathbf{y}_3[i], \hat{\mathbf{s}}_D[\iota_{Di-2}]\Big) \in \mathcal{T}^n_\epsilon(P_{V,Y_3,\hat{S}_D})\Big\}.$$

Conditioned on $\cap_{k=1}^{10} E_{ki}^c$, the probability of the event E_{11i} can be bounded using the union bound, as

$$\begin{aligned}\Pr(E_{11i}| \cap_{k=1}^{10} E_{ki}^c) &\leq J_U M_D J_D 2^{-n[I(U,U_D;Y_3,\hat{S}_D|V)-\epsilon]} \\ &= 2^{-n[I(U;Y_3|V,\hat{S}_D)-I(U;S,\hat{S}_R|V,\hat{S}_D)-[I(U;Y_3,\hat{S}_D|V)-I(U;S,\hat{S}_R,\hat{S}_D|V)]_-+3\epsilon]}. \end{aligned} \tag{A-27}$$

Thus, $\Pr(E_{11i}| \cap_{k=1}^{10} E_{ki}^c) \longrightarrow 0$ as $n \longrightarrow \infty$.

- For decoding the triple $(\hat{w}_{i-1}, \hat{j}_{Ui-1}, \hat{l}_{i-1})$ and the index $\hat{j}_{Vi}$ at the destination, let E_{12i} be the event

$$\begin{aligned}E_{12i} = \Big\{&\exists\, l'_{i-1} \in [1, M_D], j_{Di-1} \in J_D, j_{Vi} \in J_V \text{ s.t.: } l'_{i-1} \neq l_{i-1},\ \ j_{Vi} \neq j^\star_{Vi}, \\ &\Big(\mathbf{v}(w_{i-2}, j^\star_{Vi-1}), \mathbf{u}(w_{i-2}, j^\star_{Vi-1}, w_{i-1}, j^\star_{Ui-1}), \mathbf{u}_D(w_{i-2}, j^\star_{Vi-1}, w_{i-1}, j^\star_{Ui-1}, l'_{i-1}, j_{Di-1}), \mathbf{y}_3[i-1], \hat{\mathbf{s}}_D[\iota_{Ri-3}]\Big) \in \mathcal{T}_\epsilon^n(P_{V,U,U_D,Y_3,\hat{S}_D}), \\ &\Big(\mathbf{v}(w_{i-1}, j_{Vi}), \mathbf{y}_3[i], \hat{\mathbf{s}}_D[\iota_{Di-2}]\Big) \in \mathcal{T}_\epsilon^n(P_{V,Y_3,\hat{S}_D})\Big\}.\end{aligned}$$

Conditioned on $\cap_{k=1}^{11} E_{ki}^c$, the probability of the event E_{12i} can be bounded using the union bound, as

$$\begin{aligned}\Pr(E_{12i}| \cap_{k=1}^{11} E_{ki}^c) &\leq M_D J_D J_V 2^{-n[I(U_D;Y_3,\hat{S}_D|U,V)-\epsilon]} 2^{-n[I(V;Y_3,\hat{S}_D)-\epsilon]} \\ &= 2^{-n[I(V;Y_3,\hat{S}_D)-I(V;S,\hat{S}_R,\hat{S}_D)-[I(U;Y_3,\hat{S}_D|V)-I(U;S,\hat{S}_R,\hat{S}_D|V)]_-+2\epsilon]}. \end{aligned} \tag{A-28}$$

Thus, $\Pr(E_{12i}| \cap_{k=1}^{11} E_{ki}^c) \longrightarrow 0$ as $n \longrightarrow \infty$.

- For decoding the triple $(\hat{w}_{i-1}, \hat{j}_{Ui-1}, \hat{l}_{i-1})$ and the index $\hat{j}_{Vi}$ at the destination, let E_{13i} be the event

$$\begin{aligned}E_{13i} = \Big\{&\exists\, l'_{i-1} \in [1, M_D], j_{Di-1} \in J_D, j_{Vi} \in J_V \text{ s.t.: } l'_{i-1} \neq l_{i-1}, \\ &\Big(\mathbf{v}(w_{i-2}, j^\star_{Vi-1}), \mathbf{u}(w_{i-2}, j^\star_{Vi-1}, w_{i-1}, j^\star_{Ui-1}), \mathbf{u}_D(w_{i-2}, j^\star_{Vi-1}, w_{i-1}, j^\star_{Ui-1}, l'_{i-1}, j_{Di-1}), \mathbf{y}_3[i-1], \hat{\mathbf{s}}_D[\iota_{Ri-3}]\Big) \in \mathcal{T}_\epsilon^n(P_{V,U,U_D,Y_3,\hat{S}_D}), \\ &\Big(\mathbf{v}(w_{i-1}, j^\star_{Vi}), \mathbf{y}_3[i], \hat{\mathbf{s}}_D[\iota_{Di-2}]\Big) \in \mathcal{T}_\epsilon^n(P_{V,Y_3,\hat{S}_D})\Big\}.\end{aligned}$$

Conditioned on $\cap_{k=1}^{12} E_{ki}^c$, the probability of the event E_{13i} can be bounded using the union bound, as

$$\begin{aligned}\Pr(E_{13i}| \cap_{k=1}^{12} E_{ki}^c) &\leq M_D J_D 2^{-n[I(U_D;Y_3,\hat{S}_D|U,V)-\epsilon]} \\ &= 2^{-n[-|I(U;Y_3,\hat{S}_D|V)-I(U;S,\hat{S}_R,\hat{S}_D|V)+4\epsilon]}. \end{aligned} \tag{A-29}$$

Thus, $\Pr(E_{13i}| \cap_{k=1}^{12} E_{ki}^c) \longrightarrow 0$ as $n \longrightarrow \infty$.

This concludes the proof of Theorem 1.

B. Proof of Theorem 3

First we generate a random codebook that we use to obtain the lower bound in Theorem 3. This scheme is based on a combination of block Markov coding [43], Gel'fand-Pinsker binning [11], and classic rate distortion theory [49, Chapter 13]. Next, we outline the encoding and decoding procedures.

We transmit in B blocks, each of length n. During each of the first B blocks, the source encodes a message $w_i \in [1, 2^{nR}]$ and sends it over the channel, where $i = 1, \ldots, B$ denotes the index of the block. For convenience we let $w_{B+1} = 1$. For fixed n, the average rate $R\frac{B}{B+1}$ over $B+1$ blocks approaches R as $B \longrightarrow +\infty$.

Codebook generation: Fix a measure $P_{S,U,U_R,X_1,X_2,X,\hat{X},Y_2,Y_3}$ of the form (27). Calculate the marginal $P_{\hat{X}}$ induced by this measure. Fix $\epsilon > 0$ and let

$$J = 2^{n[I(U;S)+2\epsilon]} \qquad J_R = 2^{n[I(U_R;U,S)+2\epsilon]} \tag{B-1a}$$

$$M = 2^{n[R-4\epsilon]} \qquad M_R = 2^{n[\hat{R}-4\epsilon]}. \tag{B-1b}$$

1) We generate JM independent and identically distributed (i.i.d.) codewords $\{\mathbf{u}(w,j)\}$ indexed by $w = 1,\dots,M$, $j = 1,\dots,J$, each with i.i.d. components drawn according to P_U.
2) We generate $J_R M_R$ i.i.d. codewords $\{\mathbf{u}_R(m, j_R)\}$ indexed by $m = 1,\dots,M_R$, $j_R = 1,\dots,J_R$, each with i.i.d. components drawn according to P_{U_R}.
3) Independently, we randomly generate a rate distortion codebook consisting of M_R sequences $\hat{\mathbf{x}}$ drawn i.i.d. according to the $n-$product of the marginal $P_{\hat{X}}$. We index these sequences as $\hat{\mathbf{x}}[m], m = 1,\dots,M_R$.

Encoding: We pick up the story in block i. Let $w_i \in \{1,\dots,M\}$ be the new message to be sent from the source node at the beginning of block i, and $w_{i+1} \in \{1,\dots,M\}$ the message to be sent in the next block $i+1$ (note that we can assume that $w_i \neq w_{i+1}$, as the indices $\{w_k\}$ are assumed i.i.d. on $\{1,\dots,2^{nR}\}$, and so $\Pr(w_i = w_{i+1}) = 2^{-2nR} \to 0$ as $n \to +\infty$). The encoding at the beginning of block i is as follows.

i) The source searches for the smallest $j \in \{1,\cdots,J\}$ such that $\mathbf{u}(w_i, j)$ is jointly typical with $\mathbf{s}[i]$. (The properties of strongly typical sequences guarantee that there exists one such j). Denote this j by $j_i^\star = j(\mathbf{s}[i], w_i)$.
ii) Similarly, the source finds $j_{i+1}^\star = j(\mathbf{s}[i+1], w_{i+1})$ such that $\mathbf{u}(w_{i+1}, j_{i+1}^\star)$ is jointly typical with $\mathbf{s}[i+1]$ and then generates a vector $\mathbf{x}[w_{i+1}]$ with i.i.d. components given $\mathbf{u}(w_{i+1}, j_{i+1}^\star)$ and $\mathbf{s}[i+1]$, drawn according to the marginal $P_{X|U,S}$.
iii) Then, the source indices $\mathbf{x}[w_{i+1}]$ by m_i if there exists an $m_i \in \{1,\dots,M_R\}$ such that $\mathbf{x}[w_{i+1}]$ and $\hat{\mathbf{x}}[m_i]$ are jointly strongly typical. If there is more than one such m_i, the source selects the first in lexicographic order. If there is no such m_i, let $m_i = 1$. Shannon's rate-distortion theory [49, Chapter 13] ensures that the encoding of $\mathbf{x}[w_{i+1}]$ is accomplished successfully with high probability provided that n is sufficiently large and

$$\hat{R} > I(X;\hat{X}). \tag{B-2}$$

iv) Next, the source looks for the smallest $j_R \in \{1,\cdots,J_R\}$ such that $\mathbf{u}_R(m_i, j_R)$ is jointly typical with $(\mathbf{s}[i], \mathbf{u}(w_i, j_i^\star))$. (Again, the properties of strongly typical sequences guarantee that there exists one such j_R). Denote this j_R by $j_{Ri}^\star = j_R(\mathbf{s}[i], \mathbf{u}(w_i, j_i^\star))$.

Continuing with the strategy. Let $m_0 = 1$. The encoding at the beginning of block i is as follows.

1) The relay knows m_{i-1} (this will be justified below), and sends $\mathbf{x}_2[i] = \hat{\mathbf{x}}[m_{i-1}]$.
2) The source transmits the pair (w_i, m_i). It sends a vector $\mathbf{x}_1[i]$ with i.i.d. components given the vectors $\mathbf{u}(w_i, j_i^\star)$, $\mathbf{u}_R(m_i, j_{Ri}^\star)$ and $\mathbf{s}[i]$, drawn according to the marginal $P_{X_1|U,U_R,S}$ induced by the distribution (27).

   Decoding: The reconstruction of the vector $\mathbf{x}[w_{i+1}]$ at the relay and the decoding procedure at destination at the end of block i, are as follows.

1) The relay knows m_{i-1} and estimates m_i from the received $\mathbf{y}_2[i]$. It declares that $\hat{m}_i$ is sent if there is a unique $\hat{m}_i \in \{1, \dots, M_R\}$ such that $\mathbf{u}_R(\hat{m}_i, j_{Ri})$ and $\mathbf{y}_2[i]$ are jointly typical for some $j_{Ri} \in \{1, \dots, J_R\}$. One can show that the decoding error in this step is small for sufficiently large n if

$$\begin{aligned}\hat{R} &< I(U_R; Y_2) - I(U_R; U, S) \\ &= I(U_R; Y_2) - I(U_R; S) - I(U_R; U|S). \end{aligned} \tag{B-3}$$

2) The destination estimates w_i from the received $\mathbf{y}_3[i]$. It declares that $\hat{w}_i$ is sent if there is a unique $\hat{w}_i \in \{1, \dots, M\}$ such that $\mathbf{u}(\hat{w}_i, j_i)$ and $\mathbf{y}_3[i]$ are jointly typical for some $j_i \in \{1, \dots, J\}$. One can show that the decoding error in this step is small for sufficiently large n if

$$R < I(U; Y_3) - I(U; S). \tag{B-4}$$

Analysis of Probability of Error: Fix a probability distribution $P_{S,U,U_R,X_1,X_2,X,\hat{X},Y_2,Y_3}$ satisfying (27). Let $\mathbf{s}[i]$ and (w_i, m_i) be the state sequence in block i and the message pair sent from the source node in block i, respectively. As we already mentioned above, at the beginning of block i the source transmits $\mathbf{x}_1(w_i, m_i)$ and the relay transmits $\mathbf{x}_2[i] = \hat{\mathbf{x}}[m_{i-1}]$.

The average probability of error is such that

$$\Pr(\text{Error}) \leq \sum_{\mathbf{s} \notin \mathcal{T}_\epsilon^n(Q_S)} \Pr(\mathbf{s}) + \sum_{\mathbf{s} \in \mathcal{T}_\epsilon^n(Q_S)} \Pr(\mathbf{s}) \Pr(\text{error}|\mathbf{s}). \tag{B-5}$$

The first term, $\Pr(\mathbf{s} \notin \mathcal{T}_\epsilon^n(Q_S))$, on the RHS of (B-5) goes to zero as $n \to \infty$, by the asymptotic equipartition property (AEP) [49, p. 384]. Thus, it is sufficient to upper bound the second term on the RHS of (B-5).

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}, E_{2i} and E_{3i} correspond to encoding errors at block i; the events E_{4i} and E_{5i} correspond to decoding errors at the relay at block i; and the events E_{6i} and E_{7i} correspond to decoding errors at the destination at block i.

- Let E_{1i} be the event that there is no sequence $\mathbf{u}(w_i, j)$ jointly typical with $\mathbf{s}[i]$, i.e.,

$$E_{1i} = \Big\{ \nexists\, j \in \{1, \dots, J\} \text{ s.t. } \Big(\mathbf{u}(w_i, j), \mathbf{s}[i]\Big) \in \mathcal{T}_\epsilon^n(P_{U,S}) \Big\}.$$

  To bound the probability of the event E_{1i}, we use a standard argument [11]. More specifically, for $\mathbf{u}(w_i, j)$ and $\mathbf{s}[i]$ generated independently with i.i.d. components drawn according to P_U and Q_S, respectively, the probability that $\mathbf{u}(w_i, j)$ is jointly typical with $\mathbf{s}[i]$ is greater than $(1-\epsilon)2^{-n(I(U;S)+\epsilon)}$ 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

$$\Pr(E_{1i}) \leq [1 - (1-\epsilon)2^{-n(I(U;S)+\epsilon)}]^J. \tag{B-6}$$

  Taking the logarithm on both sides of (B-6) and substituting J using (B-1) we obtain $\ln(\Pr(E_{1i})) \leq -(1-\epsilon)2^{n\epsilon}$. Thus, $\Pr(E_{1i}) \to 0 \quad \text{as} \quad n \to \infty$.

- Let E_{2i} be the event that there is no sequence $\mathbf{u}(w_{i+1}, j)$ jointly typical with $\mathbf{s}[i+1]$, and E_{3i} the event that there is no sequence $\mathbf{u}_R(m_i, j_R)$ jointly typical with $(\mathbf{s}[i], \mathbf{u}(w_i, j_i^\star))$. Proceeding similarly to for the event E_{1i}, it can be easily shown that, conditioned on E_{1i}^c and $E_{1i}^c \cap E_{2i}^c$, respectively, these tow events have vanishing probabilities as $n \to +\infty$.

- For the decoding at the relay, let E_{4i} be the event that $\mathbf{u}_R(m_i, j_{Ri}^\star)$ is not jointly typical with $\mathbf{y}_2[i]$. That is

$$E_{4i} = \Big\{\Big(\mathbf{u}_R(m_i, j_{Ri}^\star), \mathbf{y}_2[i]\Big) \notin \mathcal{T}_\epsilon^n(P_{U_R,Y_2,\hat{X}})\Big\}. \tag{B-7}$$

For $\mathbf{u}(w_i, j_i^\star)$, $\mathbf{u}_R(m_i, j_{Ri}^\star)$ jointly typical with $\mathbf{s}[i]$, and with the source input $\mathbf{x}_1[i]$ and the relay input $\mathbf{x}_2[i]$, we have $\Pr(E_{4i}|E_{1i}^c, E_{2i}^c, E_{3i}^c) \longrightarrow 0$ as $n \longrightarrow \infty$ by the Markov Lemma [49, p. 436].

- For the decoding at the relay, let E_{5i} be the event that $\mathbf{u}_R(m_i', j_{Ri})$ is jointly typical with $\mathbf{y}_2[i]$ for some $m_i' \in [1, M_R]$ and $j_{Ri} \in J_R$, with $m_i' \neq m_i$. That is,

$$\begin{aligned} E_{5i} = \Big\{\exists\ m_i' \in [1, M_R],\, j_{Ri} \in J_R \text{ s.t. } m_i' \neq m_i, \\ \Big(\mathbf{u}_R(m_i', j_{Ri}), \mathbf{y}_2[i]\Big) \in \mathcal{T}_\epsilon^n(P_{U_R,Y_2,\hat{X}})\Big\}. \end{aligned} \tag{B-8}$$

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

$$\begin{aligned} \Pr(E_{5i}|E_{1i}^c, E_{2i}^c, E_{3i}^c, E_{4i}^c) &\leq M_R J_R 2^{-n[I(U_R;Y_2)-\epsilon]} \\ &= 2^{-n[I(U_R;Y_2)-I(U_R;U,S)-\hat{R}+\epsilon]}. \end{aligned} \tag{B-9}$$

Thus, $\Pr(E_{3i}|E_{1i}^c, E_{2i}^c, E_{3i}^c, E_{4i}^c) \longrightarrow 0$ as $n \longrightarrow \infty$ if $R < I(U_R; Y_2) - I(U_R; S) - I(U_R; U|S)$.

- For the decoding at the destination, let E_{6i} be the event that $\mathbf{u}(w_i, j_i^\star)$ is not jointly typical with $\mathbf{y}_3[i]$. That is

$$E_{6i} = \Big\{\Big(\mathbf{u}(w_i, j_i^\star), \mathbf{y}_3[i]\Big) \notin \mathcal{T}_\epsilon^n(P_{U,Y_3})\Big\}. \tag{B-10}$$

For $\mathbf{u}(w_i, j_i^\star)$, $\mathbf{u}_R(m_i, j_{Ri}^\star)$ jointly typical with $\mathbf{s}[i]$, and with the source input $\mathbf{x}_1[i]$ and the relay input $\mathbf{x}_2[i]$, we have $\Pr(E_{6i}|E_{1i}^c, E_{2i}^c, E_{3i}^c, E_{4i}^c, E_{5i}^c) \longrightarrow 0$ as $n \longrightarrow \infty$ by the Markov Lemma [49, p. 436].

- For the decoding at the destination, let E_{7i} be the event that $\mathbf{u}(w_i', j_i)$ is jointly typical with $\mathbf{y}_3[i]$ for some $w_i' \in [1, M]$ and $j_i \in J$, with $w_i' \neq k_i$. That is,

$$\begin{aligned} E_{7i} = \Big\{\exists\ w_i' \in [1, M],\, j_i \in J \text{ s.t. } w_i' \neq k_i, \\ \Big(\mathbf{u}(w_i', j_i), \mathbf{y}_3[i]\Big) \in \mathcal{T}_\epsilon^n(P_{U,Y_3})\Big\}. \end{aligned} \tag{B-11}$$

Conditioned on the events E_{1i}^c, E_{2i}^c, E_{3i}^c, E_{4i}^c, E_{5i}^c and E_{6i}^c, the probability of the event E_{7i} can be bounded using the union bound, as

$$\begin{aligned} \Pr(E_{7i}|E_{1i}^c, E_{2i}^c, E_{3i}^c, E_{4i}^c, E_{5i}^c, E_{6i}^c) &\leq MJ 2^{-n[I(U;Y_3)-\epsilon]} \\ &= 2^{-n[I(U;Y_3)-I(U;S)-R+\epsilon]} \end{aligned} \tag{B-12}$$

Thus, $\Pr(E_{7i}|E_{1i}^c, E_{2i}^c, E_{3i}^c, E_{4i}^c, E_{5i}^c, E_{6i}^c) \longrightarrow 0$ as $n \longrightarrow +\infty$ if $R < I(U; Y_3) - I(U; S)$.

This concludes the proof of Theorem 3.

C. Proofs of Theorem 4

Let an (ϵ_n, n, R) code be given. By Fano's inequality, we have

$$\begin{aligned} nR &= H(W) \\ &\leq I(W; Y_3^n) + 1 + nR\epsilon_n. \end{aligned} \tag{C-1}$$

Let us define $\bar{U}_i = (S_{i+1}^n, Y_2^{i-1}, Y_3^{i-1})$ and $\bar{V}_i = (W, S_{i+1}^n, Y_3^{i-1})$, $i = 1, \ldots, n$.
We have

$$\begin{aligned} I(W; Y_3^n) &\leq I(W; Y_2^n, Y_3^n) \\ &\overset{(a)}{=} I(W; Y_2^n, Y_3^n) - I(W; S^n) \qquad \text{(C-2)} \\ &= \sum_{i=1}^n I(W; Y_{2,i}, Y_{3,i} | Y_2^{i-1}, Y_3^{i-1}) - I(W; S_i | S_{i+1}^n) \\ &= \sum_{i=1}^n I(W, S_{i+1}^n; Y_{2,i}, Y_{3,i} | Y_2^{i-1}, Y_3^{i-1}) - I(S_{i+1}^n; Y_{2,i}, Y_{3,i} | W, Y_2^{i-1}, Y_3^{i-1}) - I(W; S_i | S_{i+1}^n) \\ &\overset{(b)}{=} \sum_{i=1}^n I(W, S_{i+1}^n; Y_{2,i}, Y_{3,i} | Y_2^{i-1}, Y_3^{i-1}) - I(S_i; Y_2^{i-1}, Y_3^{i-1} | W, S_{i+1}^n) - I(W; S_i | S_{i+1}^n) \\ &= \sum_{i=1}^n I(W, S_{i+1}^n; Y_{2,i}, Y_{3,i} | Y_2^{i-1}, Y_3^{i-1}) - I(S_i; W, Y_2^{i-1}, Y_3^{i-1} | S_{i+1}^n) \\ &= \sum_{i=1}^n I(W; Y_{2,i}, Y_{3,i} | S_{i+1}^n, Y_2^{i-1}, Y_3^{i-1}) + I(S_{i+1}^n; Y_{2,i}, Y_{3,i} | Y_2^{i-1}, Y_3^{i-1}) - I(S_i; Y_2^{i-1}, Y_3^{i-1} | S_{i+1}^n) - I(S_i; W | S_{i+1}^n, Y_2^{i-1}, Y_3^{i-1}) \\ &\overset{(c)}{=} \sum_{i=1}^n I(W; Y_{2,i}, Y_{3,i} | S_{i+1}^n, Y_2^{i-1}, Y_3^{i-1}) - I(S_i; W | S_{i+1}^n, Y_2^{i-1}, Y_3^{i-1}) \\ &\overset{(d)}{=} \sum_{i=1}^n I(W; Y_{2,i}, Y_{3,i} | S_{i+1}^n, Y_2^{i-1}, Y_3^{i-1}, X_{2,i}) - I(S_i; W | S_{i+1}^n, Y_2^{i-1}, Y_3^{i-1}, X_{2,i}) \\ &= \sum_{i=1}^n I(\bar{V}_i; Y_{2,i}, Y_{3,i} | \bar{U}_i, X_{2,i}) - I(\bar{V}_i; S_i | \bar{U}_i, X_{2,i}) \qquad \text{(C-3)} \end{aligned}$$

where: (a) follows since message W is independent of the state S^n; (b) follows from Csiszar and Korner's "summation by parts"-lemma [54]

$$\sum_{i=1}^n I(S_{i+1}^n; Y_{2,i}, Y_{3,i} | W, Y_2^{i-1}, Y_3^{i-1}) = \sum_{i=1}^n I(S_i; Y_2^{i-1}, Y_3^{i-1} | W, S_{i+1}^n) \tag{C-4}$$

(c) follows similarly, from Csiszar and Korner's "summation by parts"

$$\sum_{i=1}^n I(S_{i+1}^n; Y_{2,i}, Y_{3,i} | Y_2^{i-1}, Y_3^{i-1}) = \sum_{i=1}^n I(S_i; Y_2^{i-1}, Y_3^{i-1} | S_{i+1}^n) \tag{C-5}$$

(d) follows from the fact that X_{2i} is a deterministic function of Y_2^{i-1}.
Similarly,

$$I(W; Y_3^n) \overset{(e)}{\leq} \sum_{i=1}^n I(W, S_{i+1}^n, Y_3^{i-1}; Y_{3,i}) - I(W, S_{i+1}^n, Y_3^{i-1}; S_i)$$

$$= \sum_{i=1}^{n} I(\bar{V}_i; Y_{3,i}) - I(\bar{V}_i; S_i) \tag{C-6}$$

where (e) follows exactly as in the converse part of the proof of the capacity of Gel'fand-Pinsker channel [11] by replacing Y^n with Y_3^n.

From the above, we have

$$R \leq \frac{1}{n} \sum_{i=1}^{n} I(\bar{V}_i; Y_{2,i}, Y_{3,i}|\bar{U}_i, X_{2,i}) - I(\bar{V}_i; S_i|\bar{U}_i, X_{2,i}) + 1 + nR\epsilon_n$$

$$R \leq \frac{1}{n} \sum_{i=1}^{n} I(\bar{V}_i; Y_{3,i}) - I(\bar{V}_i; S_i) + 1 + nR\epsilon_n \tag{C-7}$$

We introduce a random variable T which is uniformly distributed over $\{1, \cdots, n\}$. Set $S = S_T$, $\bar{U} = \bar{U}_T$, $\bar{V} = \bar{V}_T$, $X_1 = X_{1,T}$, $X_2 = X_{2,T}$, $Y_2 = Y_{2,T}$, and $Y_3 = Y_{3,T}$. We substitute T into the above bounds. Considering the first bound in (C-7), we have

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n} I(\bar{V}_i; Y_{2,i}, Y_{3,i}|\bar{U}_i, X_{2,i}) - I(\bar{V}_i; S_i|\bar{U}_i, X_{2,i}) \\
&= I(\bar{V}; Y_2, Y_3|\bar{U}, X_2, T) - I(\bar{V}; S|\bar{U}, X_2, T) \\
&= I(T, \bar{V}; Y_2, Y_3|\bar{U}, X_2) - I(T; Y_2, Y_3|\bar{U}, X_2) - I(T, \bar{V}; S|\bar{U}, X_2) + I(T; S|\bar{U}, X_2) \\
&\leq I(T, \bar{V}; Y_2, Y_3|\bar{U}, X_2) - I(T, \bar{V}; S|\bar{U}, X_2) + I(T; S|\bar{U}, X_2) \\
&= I(T, \bar{V}; Y_2, Y_3|\bar{U}, X_2) - I(T, \bar{V}; S|\bar{U}, X_2)
\end{aligned} \tag{C-8}$$

where in the last equality we used the fact that T is independent of all the other variables.
Similarly, considering the second bound in (C-7), we obtain

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n} I(\bar{V}_i; Y_{3,i}) - I(\bar{V}_i; S_i) \\
&= I(\bar{V}; Y_3|T) - I(\bar{V}; S|T) \\
&= I(T, \bar{V}; Y_3) - I(T; Y_3) - I(T, \bar{V}; S) + I(T; S) \\
&\leq I(T, \bar{V}; Y_3) - I(T, \bar{V}; S).
\end{aligned} \tag{C-9}$$

Let us now define $U = \bar{U}$ and $V = (T, \bar{V})$. Using (C-7), (C-8) and (C-9), we then get

$$R \leq I(V; Y_2, Y_3|U, X_2) - I(V; S|U, X_2) + 1 + nR\epsilon_n$$

$$R \leq I(V; Y_3) - I(V; S) + 1 + nR\epsilon_n. \tag{C-10}$$

So far we have shown that, for a given sequence of $(\epsilon_n, n, R)-$codes with ϵ_n going to zero as n goes to infinity, there exists a probability distribution of the form (31) such that the rate R essentially satisfies (30). This completes the proof of Theorem 4.

It remains to show that the rate (30) is not altered if one restricts the random variables U and U to have their alphabet sizes limited as indicated in (32). This is done by invoking the support lemma [55, p. 310]. Fix a distribution μ of $(S, U, V, X_1, X_2, Y_2, Y_3)$ on $\mathcal{P}(\mathcal{S}\times\mathcal{U}\times\mathcal{V}\times\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{Y}_2\times\mathcal{Y}_3)$ that has the form (31).

To prove the bound (32a) on $|\mathcal{U}|$, note that we have

$$\begin{aligned}
&I_\mu(V;Y_2,Y_3|U,X_2) - I_\mu(V;S|U,X_2)\\
&\quad= I_\mu(V,X_2;Y_2,Y_3|U) - I_\mu(X_2;Y_2,Y_3|U) - I_\mu(V,X_2;S|U) + I_\mu(X_2;S|U)\\
&\quad= H_\mu(Y_2,Y_3|U) - H_\mu(V,X_2,Y_2,Y_3|U) + H_\mu(V,X_2,S|U) + H_\mu(X_2|U) - H_\mu(X_2,S|U). \qquad \text{(C-11)}
\end{aligned}$$

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

$$r_{s,x,x'}(\mu) = \mu(s,x,x') \quad \forall\,(s,x,x') \in \mathcal{S}\times\mathcal{X}_1\times\mathcal{X}_2 \qquad \text{(C-12a)}$$

$$r_1(\mu) = \int_u d_\mu(u)[H_\mu(Y_2,Y_3|u) - H_\mu(V,X_2,Y_2,Y_3|u) + H_\mu(V,X_2,S|u) + H_\mu(X_2|u) - H_\mu(X_2,S|u)] \qquad \text{(C-12b)}$$

can be preserved with another measure μ' that has the form (31). Observing that there is a total of $|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2|$ functionals in (C-12), 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_1 can be limited as indicated in (32a) without altering the rate (30).

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

$$I_\mu(V;Y_3) - I_\mu(V;S) = H_\mu(Y_3) - H_\mu(S) - H_\mu(Y_3|V) + H_\mu(S|V), \qquad \text{(C-13)}$$

yielding the bound indicated in (32b). This completes the proof of Theorem 4.

D. Proof of Theorem 5

We prove that for any (ϵ,n,R) code consisting of a mapping $\phi_1^n = (\phi_{1R}^n, \phi_{1D}^n)$ at the hyper source with $\phi_{1R}^n : \mathcal{W} \longrightarrow \mathcal{X}_{1R}^n$ and $\phi_{1D}^n : \mathcal{W}\times\mathcal{S}^n \longrightarrow \mathcal{X}_{1D}^n$, a sequence of mappings $\phi_{2,i} : \mathcal{Y}_2^{i-1} \longrightarrow \mathcal{X}_2$, $i = 1,\dots,n$, at the relay, and a mapping $\psi^n : \mathcal{Y}^n \longrightarrow W$ at the decoder with average error probability $P_e^n \to 0$ as $n \to 0$, the rate R must satisfy (37).

By Fano's inequality, we have

$$H(W|Y_3^n) \le nR\epsilon_n + 1 \triangleq n\delta_n. \qquad \text{(D-1)}$$

Thus,

$$nR = H(W) \le I(W;Y_3^n) + n\delta_n \qquad \text{(D-2)}$$

We now upper bound $I(W;Y_3^n)$ as in the following lemma, the proof of which follows.

Lemma 2:

$$\text{i)}\;\; I(W;Y_3^n) \le \sum_{i=1}^n I(X_{1R,i};Y_{2,i}|S_i,X_{2,i}) + I(X_{1D,i};Y_{3,i}|S_i,X_{2,i}) \qquad \text{(D-3a)}$$

$$\text{ii)}\;\; I(W;Y_3^n) \le \sum_{i=1}^n I(X_{1D,i};Y_{3,i}|S_i,X_{2,i}) + I(X_{2,i};Y_{3,i}). \qquad \text{(D-3b)}$$

Proof: To simplify the notation, we use $S^i = (S_1, S_2, \cdots, S_i)$, $Y_k^i = (Y_{k,1}, Y_{k,2}, \cdots, Y_{k,i})$, $k = 2, 3$, and $X_j^i = (X_{j,1}, X_{j,2}, \cdots, X_{j,i})$, $j = 1R, 1D, 2$.

1) The proof of the bound on $I(W; Y_3^n)$ given in i) follows straightforwardly by revealing the state to the destination and using the channel structure (1).

$$
\begin{aligned}
I(W; Y_3^n) &\overset{(a)}{\leq} \sum_{i=1}^n I(X_{1R,i}, X_{1D,i}; Y_{2,i}, Y_{3,i} | X_{2,i}, S_i) && \text{(D-4)}\\
&= \sum_{i=1}^n I(X_{1R,i}, X_{1D,i}; Y_{2,i} | X_{2,i}, S_i) + I(X_{1R,i}, X_{1D,i}; Y_{3,i} | X_{2,i}, S_i, Y_{2,i}) && \text{(D-5)}\\
&= \sum_{i=1}^n I(X_{1R,i}; Y_{2,i} | X_{2,i}, S_i) + I(X_{1D,i}; Y_{2,i} | X_{1R,i}, X_{2,i}, S_i)\\
&\quad + I(X_{1R,i}, X_{1D,i}; Y_{3,i} | X_{2,i}, S_i, Y_{2,i}) && \text{(D-6)}\\
&\overset{(b)}{=} \sum_{i=1}^n I(X_{1R,i}; Y_{2,i} | X_{2,i}, S_i) + I(X_{1R,i}, X_{1D,i}; Y_{3,i} | X_{2,i}, S_i, Y_{2,i}) && \text{(D-7)}\\
&= \sum_{i=1}^n I(X_{1R,i}; Y_{2,i} | X_{2,i}, S_i) + H(Y_{3,i} | X_{2,i}, S_i, Y_{2,i}) - H(Y_{3,i} | X_{1R,i}, X_{1D,i}, X_{2,i}, S_i, Y_{2,i}) && \text{(D-8)}\\
&\overset{(c)}{=} \sum_{i=1}^n I(X_{1R,i}; Y_{2,i} | X_{2,i}, S_i) + H(Y_{3,i} | X_{2,i}, S_i, Y_{2,i}) - H(Y_{3,i} | X_{1D,i}, X_{2,i}, S_i) && \text{(D-9)}\\
&\overset{(d)}{\leq} \sum_{i=1}^n I(X_{1R,i}; Y_{2,i} | X_{2,i}, S_i) + H(Y_{3,i} | X_{2,i}, S_i) - H(Y_{3,i} | X_{1D,i}, X_{2,i}, S_i) && \text{(D-10)}\\
&= \sum_{i=1}^n I(X_{1R,i}; Y_{2,i} | X_{2,i}, S_i) + I(X_{1D,i}; Y_{3,i} | X_{2,i}, S_i) && \text{(D-11)}
\end{aligned}
$$

where:

(a) follows trivially by revealing the state to the destination; (b) follows since $X_{1D,i} \leftrightarrow (X_{1R,i}, X_{2,i}, S_i) \leftrightarrow Y_{2,i}$; (c) follows since $(X_{1R,i}, Y_{2,i}) \leftrightarrow (X_{1D,i}, X_{2,i}, S_i) \leftrightarrow Y_{3,i}$; and (d) follows since conditioning reduces entropy.

2) The proof of the bound on $I(W; Y_3^n)$ given in ii) follows as follows.

$$
\begin{aligned}
I(W; Y_3^n) &= I(W, S^n; Y_3^n) - I(S^n; Y_3^n | W)\\
&= \Big(\sum_{i=1}^n I(W, S^n; Y_{3,i} | Y_3^{i-1})\Big) - H(S^n | W) + H(S^n | W, Y_3^n)\\
&\overset{(e)}{=} \sum_{i=1}^n H(Y_{3,i} | Y_3^{i-1}) - H(Y_{3,i} | W, S^n, Y_3^{i-1}) - H(S_i) + H(S_i | W, Y_3^n, S^{i-1})\\
&\overset{(f)}{\leq} \sum_{i=1}^n H(Y_{3,i}) - H(Y_{3,i} | X_{1D,i}, X_{2,i}, S_i) - H(S_i) + H(S_i | W, Y_3^n, S^{i-1})\\
&\overset{(g)}{=} \sum_{i=1}^n H(Y_{3,i}) - H(Y_{3,i} | X_{1D,i}, X_{2,i}, S_i) - H(S_i) + H(S_i | W, Y_3^n, S^{i-1}, Y_2^{i-1})
\end{aligned}
$$

$$\stackrel{(h)}{=} \sum_{i=1}^{n} H(Y_{3,i}) - H(Y_{3,i}|X_{1D,i}, X_{2,i}, S_i) - H(S_i) + H(S_i|W, Y_3^n, S^{i-1}, Y_2^{i-1}, X_{2,i})$$

$$\stackrel{(i)}{\leq} \sum_{i=1}^{n} I(X_{1D,i}, X_{2,i}, S_i; Y_{3,i}) - H(S_i) + H(S_i|X_{2,i}, Y_{3,i})$$

$$= \sum_{i=1}^{n} I(X_{1D,i}, X_{2,i}, S_i; Y_{3,i}) - I(S_i; X_{2,i}, Y_{3,i})$$

$$= \sum_{i=1}^{n} I(X_{1D,i}; Y_{3,i}|S_i, X_{2,i}) + I(X_{2,i}; Y_{3,i}) - I(X_{2,i}; S_i)$$

$$\stackrel{(j)}{=} \sum_{i=1}^{n} I(X_{1D,i}; Y_{3,i}|S_i, X_{2,i}) + I(X_{2,i}; Y_{3,i}), \tag{D-12}$$

where: (e) follows from the fact that the state S^n is i.i.d. and is independent of the message W; (f) follows from $(W, S^n, Y_3^{i-1}) \leftrightarrow (X_{1D,i}, X_{2,i}, S_i) \leftrightarrow Y_{3,i}$ is a Markov chain; (g) follows from $Y_2^{i-1} \leftrightarrow (W, S^{i-1}, Y_3^n) \leftrightarrow S_i$ is a Markov chain; (h) follows from the fact that $X_{2,i}$ is a deterministic function of Y_2^{i-1}; (i) follows from the fact that conditioning reduces entropy; and (j) holds since $X_{2,i}$ is independent of S_i. ■

We introduce a random variable T which is uniformly distributed over $\{1, \cdots, n\}$. Set $S = S_T$, $X_{1R} = X_{1R,T}$, $X_{1D} = X_{1D,T}$, $X_2 = X_{2,T}$, $Y_2 = Y_{2,T}$, and $Y_3 = Y_{3,T}$. We substitute T into the above bounds. Considering the bound (D-12), we obtain

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} I(X_{1D,i}; Y_{3,i}|S_i, X_{2,i}) + I(X_{2,i}; Y_{3,i}) \\
&= I(X_{1D}; Y_3|S, X_2, T) + I(X_2; Y_3|T) \\
&= I(X_{1D}, X_2, S; Y_3|T) - I(S; X_2, Y_3|T)
\end{aligned} \tag{D-13}$$

and, similarly,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} I(X_{1R,i}; Y_{2,i}|X_{2,i}, S_i) + I(X_{1D,i}; Y_{3,i}|X_{2,i}, S_i) \\
&= I(X_{1R}; Y_2|S, X_2, T) + I(X_{1D}; Y_3|S, X_2, T)
\end{aligned} \tag{D-14}$$

where the distribution on $(T, S, X_{1R}, X_{1D}, X_2, Y_2, Y_3)$ from a given code is of the form

$$\begin{aligned}
P_{T,S,X_{1R},X_{1D},X_2,Y_2,Y_3} = {} & Q_S P_T P_{X_2|T} P_{X_{1R}|X_2,T} P_{X_{1D}|S,X_2,T} \\
& \times W_{Y_2|S,X_{1R}} W_{Y_3|S,X_{1D},X_2}.
\end{aligned} \tag{D-15}$$

We now eliminate the variable T from (D-13) and (D-14) as follows. The right-hand side of (D-13) can be bounded as

$$\begin{aligned}
&I(X_{1D}, X_2, S; Y_3|T) - I(S; X_2, Y_3|T) \\
&\stackrel{(k)}{\leq} H(Y_3) - H(Y_3|X_{1D}, X_2, S) - H(S|T) + H(S|X_2, Y_3, T) \\
&= I(X_{1D}, X_2, S; Y_3) - H(S|T) + H(S|X_2, Y_3, T) \\
&\stackrel{(l)}{\leq} I(X_{1D}, X_2, S; Y_3) - H(S) + H(S|X_2, Y_3)
\end{aligned}$$

$$= I(X_{1D}, X_2, S; Y_3) - I(S; X_2, Y_3)$$

$$= I(X_{1D}; Y_3|S, X_2) + I(X_2; Y_3), \tag{D-16}$$

where:

(k) holds since $H(Y_3|T) \leq H(Y_3)$ and $H(Y_3|X_{1D}, X_2, S, T) = H(Y_3|X_{1D}, X_2, S)$ (by the Markovian relation $T \leftrightarrow (X_{1D}, X_2, S) \leftrightarrow Y_3$); and

(l) holds since S is independent of T and $H(S|X_{1D}, Y_3, T) \leq H(S|X_{1D}, Y_3)$.

Similarly, right-hand side of (D-13) can be bounded as

$$I(X_{1R}; Y_2|S, X_2, T) + I(X_{1D}; Y_3|S, X_2, T) \leq I(X_{1R}; Y_2|S, X_2) + I(X_{1D}; Y_3|S, X_2). \tag{D-17}$$

Finally, combining (D-2), (D-12), (D-16) at one hand, and (D-2), (D-11), (D-17) at the other hand, we get

$$R \leq I(X_{1D}; Y_3|S, X_2) + I(X_2; Y_3) \tag{D-18a}$$

$$R \leq I(X_{1R}; Y_2|S, X_2) + I(X_{1D}; Y_3|S, X_2), \tag{D-18b}$$

where the distribution on $(S, X_{1R}, X_{1D}, X_2, Y_2, Y_3)$, obtained by marginalizing (D-15) over the variable T, has the form given in (38).

We conclude that, for a given sequence of $(\epsilon_n, n, R)-$codes with ϵ_n going to zero as n goes to infinity, there exists a probability distribution of the form (38) such that the rate R satisfies (D-18). This completes the proof of Theorem 5.

E. Proof of Theorem 7

The encoding and transmission scheme is as follows. Let $P_{1r} \geq 0$, $P_{1d} \geq 0$ and $D \geq 0$ be given such that $P_{1r} + P_{1d} \leq P_1$ and $0 \leq D \leq Q$. Also, consider the test channel $\hat{S}_R = aS + \tilde{S}_R$, where $a := 1 - D/Q$ and $\tilde{S}_R$ is a Gaussian random variable with zero mean and variance $\sigma^2_{\tilde{S}_R} = D(1 - D/Q)$, independent from S. Using this test channel, we calculate $\mathbb{E}[(S - \hat{S}_R)^2] = D$ and $\mathbb{E}[\hat{S}_R^2] = Q - D$. Let $X_2 \sim \mathcal{N}(0, P_2)$ be jointly Gaussian with $\hat{S}_R$ with $\mathbb{E}[X_2\hat{S}_R] = 0$ and independent from S, and $X_{SR} \sim \mathcal{N}(0, \theta P_{1r})$ jointly Gaussian with $(S, \hat{S}_R)$ with $\mathbb{E}[X_{SR}S] = 0$ and $\mathbb{E}[X_{SR}\hat{S}_R] = 0$, where $0 \leq \theta \leq 1$. Also, let $X_{WR} \sim \mathcal{N}(0, \bar{\theta}P_{1r})$ be jointly Gaussian with (X_2, S) and independent of X_{SR}, with $\mathbb{E}[X_{WR}S] = \sigma_{1s}$ and $\mathbb{E}[X_{WR}X_2] = \sigma_{12}$; and $X_{WD} \sim \mathcal{N}(0, P_{1d})$ jointly Gaussian with and independent of $(X_{WR}, X_{SR}, X_2, S, \hat{S}_R)$. In what follows, we use the random variables V, U, U_1 and U_R given by (64) to generate the auxiliary codewords V_i, U_i, U_{1i} and U_{Ri} which we will use in the sequel. Also, recall the definition of $\tilde{Q}$, ξ and α_2 in (62) and (63), respectively, which we will use in the rest of this proof.

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 into three independent parts, i.e., $X_1^n = X_{SR}^n + X_{wr}^n + X_{wd}^n$, where X_{SR}^n carries a description $\hat{S}_R^n$ of the state S^n that is intended to be recovered only at the relay and has power constraint $n\theta P_{1r}$, X_{wr}^n carries message W_r and has power constraint $n\bar{\theta}P_{1r}$ and X_{wd}^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 $R = R_r + R_d$.

As in the discrete case, a block Markov encoding is used. Let $w_i = (w_{ri}, w_{di}) \in [1, 2^{nR_r}]\times[1, 2^{nR_d}]$ denote the message to be transmitted in block i and $\mathbf{s}[i]$ denote the state controlling the channel in block i. The source quantizes $\mathbf{s}[i]$

into $\hat{\mathbf{s}}_R[\iota_{Ri-1}]$, where $\iota_{Ri-1} \in [1, 2^{n\hat{R}_R}]$. Using the aforementioned test channel, the source can encode $\mathbf{s}[i]$ successfully at the quantization rate

$$\begin{aligned}\hat{R}_R &= I(S;\hat{S}_R)\\ &= \frac{1}{2}\log(\frac{Q}{D}).\end{aligned} \tag{E-1}$$

In the beginning of block i, the relay has decoded correctly message w_{ri-1} and the index ι_{Ri-1} of the description $\hat{\mathbf{s}}_R[\iota_{Ri-1}]$ sent by the source in the previous block $i-1$ (this will be justified below) and sends a Gaussian signal $\mathbf{x}_2[w_{ri-1}]$ which carries message w_{ri-1} and is obtained via a DPC considering $\hat{\mathbf{s}}_R[\iota_{Ri-1}]$ as noncausal channel state information at the transmitter, as

$$\mathbf{x}_2[w_{ri-1}] = \frac{\sqrt{P_2}}{\rho_{12}\sqrt{\bar{\theta}P_{1r}} + \sqrt{P_2}}\Big(\mathbf{v}[i] - \alpha_2\xi\hat{\mathbf{s}}_R[\iota_{Ri-1}]\Big), \tag{E-2}$$

where the components of $\mathbf{v}[i]$ are generated i.i.d. using the auxiliary random variable V.

Let ι_{Ri} be the index associated with the state $\mathbf{s}[i+1]$ of the next block $i+1$. In the beginning of block i, the source sends a superposition of three Gaussian vectors,

$$\begin{aligned}\mathbf{x}_1[i] &= \mathbf{x}_{SR}[\iota_{Ri}] + \mathbf{x}_{wr}[w_{ri-1}, w_{ri}] + \mathbf{x}_{wd}[w_{di}]\\ \mathbf{x}_{wr}[w_{ri-1}, w_{ri}] &= \rho_{1s}\sqrt{\frac{\bar{\theta}P_{1r}}{Q}}\mathbf{s}[i] + \rho_{12}\sqrt{\frac{\bar{\theta}P_{1r}}{P_2}}\mathbf{x}_2[w_{ri-1}] + \mathbf{x}'_{wr}[w_{ri}].\end{aligned} \tag{E-3}$$

In (E-3), the vectors $\mathbf{x}_{SR}[\iota_{Ri}]$ and $\mathbf{x}_{wd}[w_{di}]$ are generated i.i.d. using the auxiliary random variables X_{SR} and X_{WD}, respectively; and the vector $\mathbf{x}'_{wr}[w_{ri}]$ has power $n(1-\rho_{12}^2-\rho_{1s}^2)\bar{\theta}P_{1r}$ and is independent of $\mathbf{s}[i]$, $\mathbf{x}_2[w_{ri-1}]$, $\mathbf{x}_{SR}[\iota_{Ri}]$ and $\mathbf{x}_{wd}[w_{di}]$. Furthermore, the vector $\mathbf{x}_{SR}[\iota_{Ri}]$ carries a description $\hat{\mathbf{s}}_R[\iota_{Ri}]$ of the state $\mathbf{s}[i+1]$ that affects transmission in the next block $i+1$, intended to be recovered only at the relay; the vector $\mathbf{x}_2[w_{ri-1}]$ carries cooperative information w_{ri-1}, and the vector $\mathbf{x}'_{wr}[w_{ri}]$ carries new information w_{ri}. The vectors $\mathbf{x}_{SR}[\iota_{Ri}]$, $\mathbf{x}_{wd}[w_{di}]$ and $\mathbf{x}'_{wr}[w_{ri}]$ are obtained via DPCs considering $(s[i], \hat{\mathbf{s}}_R[\iota_{Ri-1}])$ as noncausal channel state information at the transmitter, as

$$\mathbf{x}_{SR}[\iota_{Ri}] = \mathbf{u}_R[i] - \frac{\theta P_{1r}}{\theta P_{1r} + N_2 + P_{1d}}(1-\alpha)\mathbf{s}[i] \tag{E-4a}$$

$$\mathbf{x}_{wd}[w_{di}] = \mathbf{u}_1[i] - \frac{P_{1d}}{P_{1d} + N_3 + \theta P_{1r}}\xi(1-\alpha)\Big(\mathbf{s}[i] - \alpha_2\hat{\mathbf{s}}_R[\iota_{Ri-1}]\Big) \tag{E-4b}$$

$$\mathbf{x}'_{wr}[w_{ri}] = \mathbf{u}[i] - \alpha\xi\Big(\mathbf{s}[i] - \alpha_2\hat{\mathbf{s}}_R[\iota_{Ri-1}]\Big) \tag{E-4c}$$

where the components of $\mathbf{u}_R[i]$, $\mathbf{u}_1[i]$ and $\mathbf{u}[i]$ are generated i.i.d. using the auxiliary random variables U_R, U_1 and U respectively.

We now describe the decoding operations (we give simple arguments; the rigorous decoding uses joint typicality testing). Consider first the decoding at the relay. In block i, the relay receives

$$\mathbf{y}_2[i] = \mathbf{x}_{SR}[\iota_{Ri}] + \rho_{12}\sqrt{\frac{\bar{\theta}P_{1r}}{P_2}}\mathbf{x}_2[w_{ri-1}] + \mathbf{x}'_{wr}[w_{ri}] + \Big(1 + \rho_{1s}\sqrt{\frac{\bar{\theta}P_{1r}}{Q}}\Big)\mathbf{s}[i] + (\mathbf{z}_2[i] + \mathbf{x}_{wd}[w_{di}]). \tag{E-5}$$

The relay knows w_{ri-1} and ι_{Ri-1} and decodes the pair (w_{ri}, ι_{Ri}) from $\mathbf{y}_2[i]$. The relay decodes w_{ri} and ι_{Ri} successively, starting by w_{ri}. To decode w_{ri}, the relay subtracts out the quantity $\Big(\rho_{12}\sqrt{\bar{\theta}P_{1r}/P_2}\mathbf{x}_2[w_{ri-1}] + \alpha_2\xi\hat{\mathbf{s}}_R[\iota_{Ri-1}]\Big)$ from $\mathbf{y}_2[i]$

to make the channel equivalent to

$$\tilde{\mathbf{y}}_2[i] = \mathbf{x}'_{wr}[w_{ri}] + \xi\Big(\mathbf{s}[i] - \alpha_2\hat{\mathbf{s}}_R[\iota_{Ri-1}]\Big) + (\mathbf{z}_2[i] + \mathbf{x}_{SR}[\iota_{Ri}] + \mathbf{x}_{wd}[w_{di}]). \tag{E-6}$$

The relay decodes message w_{ri} from $\tilde{\mathbf{y}}_2[i]$ treating signals $\mathbf{x}_{SR}[\iota_{Ri}]$ and $\mathbf{x}_{wd}[w_{di}]$ as unknown independent noises. This can be done reliably as long as n is large and

$$\begin{aligned} R_r &\le I(U;\tilde{Y}_2) - I(U;S-\alpha_2\hat{S}_R) \\ &= R\Big(\alpha,(1-\rho_{12}^2-\rho_{1s}^2)\bar{\theta}P_{1r},\xi^2\tilde{Q},N_2+\theta P_{1r}+P_{1d}\Big) \end{aligned} \tag{E-7}$$

where the equality follows through straightforward algebra which we omit here for brevity (note that the variance of the additive state $\xi(S-\alpha_2\hat{S}_R)$ in (E-6) is $\xi^2\mathbb{E}[(S-\alpha_2\hat{S}_R)^2] = \xi^2[(1-\alpha_2)^2Q - \alpha_2(\alpha_2-2)D] := \xi^2\tilde{Q}$). Next, for the decoding of ι_{Ri}, the relay subtracts out the quantity $\Big(\mathbf{u}[i]-(1-\alpha)\alpha_2\xi\hat{\mathbf{s}}_R[\iota_{Ri-1}]\Big)$ from $\tilde{\mathbf{y}}_2[i]$ to make the channel equivalent to

$$\check{\mathbf{y}}_2[i] = \mathbf{x}_{SR}[\iota_{Ri}] + (1-\alpha)\mathbf{s}[i] + (\mathbf{z}_2[i] + \mathbf{x}_{wd}[w_{di}]). \tag{E-8}$$

The relay decodes the index ι_{Ri} from $\check{\mathbf{y}}_2[i]$ correctly as long as n is large and

$$\begin{aligned} \hat{R}_R &\le I(U_R;\check{Y}_2) - I(U_R;S) \\ &= \frac{1}{2}\log\Big(1+\frac{\theta P_{1r}}{N_2+P_{1d}}\Big). \end{aligned} \tag{E-9}$$

We now turn to the decoding at the destination at the end of block i. In block i, the destination receives

$$\begin{aligned} \mathbf{y}_3[i] &= \mathbf{x}_1[i] + \mathbf{x}_2[w_{ri-1}] + \mathbf{s}[i] + \mathbf{z}_3[i] \\ &= \Big(\rho_{12}\sqrt{\frac{\bar{\theta}P_{1r}}{P_2}}+1\Big)\mathbf{x}_2[w_{ri-1}] + \mathbf{x}'_{wr}[w_{ri}] + \mathbf{x}_{wd}[w_{di}] + \Big(\rho_{1s}\sqrt{\frac{\bar{\theta}P_{1r}}{Q}}+1\Big)\mathbf{s}[i] + (\mathbf{z}_3[i] + \mathbf{x}_{SR}[\iota_{Ri}]). \end{aligned} \tag{E-10}$$

At the end of block i, the destination knows message w_{ri-2} and decodes the pair (w_{ri-1}, w_{di-1}) successively, treating the signal that carries the state description as unknown independent noise. It starts by decoding message w_{ri-1}, using $(\mathbf{y}_3[i-1], \mathbf{y}_3[i])$. Note that w_{ri-1} is carried by both auxiliary vectors $\mathbf{v}[i]$ and $\mathbf{u}[i-1]$. If n is large, it can do so reliably at rate

$$\begin{aligned} R_r &\le I(V,U;Y_3) - I(V,U;S,\hat{S}_R) \\ &= [I(V;Y_3) - I(V;\hat{S}_R)] + [I(U;Y_3|V) - I(U;S,\hat{S}_R|V)] \end{aligned} \tag{E-11}$$

where the equality follows since the choice of $(V,\hat{S}_R)$ in (64) satisfies $V \leftrightarrow \hat{S}_R \leftrightarrow S$ is a Markov chain.
We first compute the term $[I(V;Y_3) - I(V;\hat{S}_R)]$. Let $\tilde{\mathbf{s}}[i]$ be the estimation error of $\xi\mathbf{s}[i]$ given $\hat{\mathbf{s}}_R[\iota_{Ri-1}]$ under minimum mean square error criterion. Since $\mathbf{s}[i]$ and $\hat{\mathbf{s}}_R[\iota_{Ri-1}]$ are jointly Gaussian, $\tilde{\mathbf{s}}[i]$ is i.i.d. Gaussian with variance $\mathbb{E}[(\xi S-\xi\hat{S}_R)^2] = \xi^2 D$ per element and is independent from $\hat{\mathbf{s}}_R[\iota_{Ri-1}]$. Thus, we can alternatively write the output $\mathbf{y}_3[i]$ as

$$\mathbf{y}_3[i] = \Big(\rho_{12}\sqrt{\frac{\bar{\theta}P_{1r}}{P_2}}+1\Big)\mathbf{x}_2[w_{ri-1}] + \mathbf{x}'_{wr}[w_{ri}] + \mathbf{x}_{wd}[w_{di}] + \xi\hat{\mathbf{s}}_R[\iota_{Ri-1}] + (\mathbf{z}_3[i] + \mathbf{x}_{SR}[\iota_{Ri}] + \tilde{\mathbf{s}}[i]). \tag{E-12}$$

With the choice of the auxiliary random variable V as in (64) and that of the associated Costa's scale factor α_2 set to its optimal value as in (63), the destination decodes the vector $\mathbf{v}[i]$ correctly from $\mathbf{y}_3[i]$ at rate

$$I(V;Y_3)-I(V;\hat{S}_R)=\frac{1}{2}\log\Big(1+\frac{(\rho_{12}\sqrt{\bar{\theta}P_{1r}}+\sqrt{P_2})^2}{N_3+\xi^2D+\theta P_{1r}+(1-\rho_{12}^2-\rho_{1s}^2)\bar{\theta}P_{1r}+P_{1d}}\Big) \tag{E-13}$$

where the equality follows through straightforward algebra. Let us now compute the term $[I(U;Y_3|V)-I(U;S,\hat{S}_R|V)]$. Observing that the destination can peel off $\mathbf{v}[i-1]$ from $\mathbf{y}_3[i-1]$ to make the channel equivalent to

$$\begin{aligned}\tilde{\mathbf{y}}_3[i-1]&=\mathbf{y}_3[i-1]-\Big(\Big(\rho_{12}\sqrt{\frac{\bar{\theta}P_{1r}}{P_2}}+1\Big)\mathbf{x}_2[w_{ri-2}]+\alpha_2\xi\hat{\mathbf{s}}_R[\iota_{Ri-2}]\Big)\\&=\mathbf{x}'_{wr}[w_{ri-1}]+\xi\mathbf{s}[i-1]-\alpha_2\xi\hat{\mathbf{s}}_R[\iota_{Ri-2}]+(\mathbf{z}_3[i-1]+\mathbf{x}_{SR}[\iota_{Ri-1}]+\mathbf{x}_{wd}[w_{di-1}]),\end{aligned} \tag{E-14}$$

it is easy to see that, if n is large and with the choice of the auxiliary random variable U as in (64), the destination obtains the vector $\mathbf{u}[i-1]$ correctly from $\mathbf{y}_3[i-1]$ at rate

$$\begin{aligned}I(U;Y_3|V)-I(U;S,\hat{S}_R|V)&=I(U;\tilde{Y}_3)-I(U;\xi(S-\alpha_2\hat{S}_R))\\&=R\Big(\alpha,(1-\rho_{12}^2-\rho_{1s}^2)\bar{\theta}P_{1r},\xi^2\tilde{Q},N_3+\theta P_{1r}+P_{1d}\Big)\end{aligned} \tag{E-15}$$

where the last equality follows through straightforward algebra.

Finally, the destination can peel off $\mathbf{u}[i-1]$ from $\tilde{\mathbf{y}}_3[i-1]$ to make the channel equivalent to

$$\begin{aligned}\check{\mathbf{y}}_3[i-1]&=\tilde{\mathbf{y}}_3[i-1]-\Big(\mathbf{x}'_{wr}[w_{ri-1}]+\alpha\xi(\mathbf{s}[i-1]-\alpha_2\hat{\mathbf{s}}_R[\iota_{Ri-2}])\Big)\\&=\mathbf{x}_{wd}[w_{di-1}]+\xi(1-\alpha)(\mathbf{s}[i-1]-\alpha_2\xi\hat{\mathbf{s}}_R[\iota_{Ri-2}])+(\mathbf{z}_3[i-1]+\mathbf{x}_{SR}[\iota_{Ri-1}]).\end{aligned} \tag{E-16}$$

From (E-16), it is easy to see that if n is large, and with the choice of the auxiliary random variable U_1 as in (64), the destination obtains the vector $\mathbf{u}_1[i-1]$ (which carries message w_{di-1}) correctly at rate

$$\begin{aligned}R_d&\leq I(U_1;\check{Y}_3)-I(U_1;\xi(1-\alpha)(S-\alpha_2\hat{S}_R))\\&=\frac{1}{2}\log(1+\frac{P_{1d}}{N_3+\theta P_{1r}}).\end{aligned} \tag{E-17}$$

Finally, for given D, adding (E-7) and (E-17), we obtain the first term of the minimization in (60); and adding (E-13), (E-15) and (E-17), we obtain the second term of the minimization in (60). Also, similar to in the proof of Theorem 6, observing that the rate terms in (60) decrease with D, we obtain the lower bound in Theorem 7 by taking the equality in (E-9) and maximizing the minimization in (60) over $P_{1r}\geq 0$, $P_{1d}\geq 0$ such that $0\leq P_{1r}+P_{1d}\leq P_1$, $\theta\in[0,1]$, $\rho_{12}\in[0,1]$ and $\rho_{1s}\in[-1,0]$ such that $0\leq\rho_{12}^2+\rho_{1s}^2\leq 1$ and $\alpha\in\mathbb{R}$ such that the RHS of (E-7) is non-negative and the sum of the RHS of (E-15) and the RHS of (E-17) is non-negative. This completes the proof.

F. *Proof of Proposition 2*

In the proof we compute the rate (33) of Proposition 1 using an appropriate jointly Gaussian distribution on $(S,U_1,X_{1R},X_{1D},X_2)$. The algebra in this section is similar to that in the proof of [21, Theorem 3] and [15, Theorem 6].

We first compute the term $[I(U_1; Y_3|X_{1R}, X_2) - I(U_1; S|X_{1R}, X_2)]$ in the RHS of (33) because this gives insights about the distribution that we should use to compute the lower bound. We assume that X_{1R}, X_{1D} and X_2 are jointly Gaussian random variables with zero-mean and variance P_{1R}, P_{1D} and P_2, respectively. The random variables X_{1R} and X_2 are independent and independent of the state S. The random variable X_{1D} is independent of X_{1R} and jointly Gaussian with (S, X_2), with $\mathbb{E}[X_{1D}X_2] = \rho_{12}\sqrt{P_{1D}P_2}$ and $\mathbb{E}[X_{1D}S] = \rho_{1s}\sqrt{P_{1D}Q}$, for some correlation coefficients $\rho_{12} \in [-1, 1]$ and $\rho_{1s} \in [-1, 1]$.

Let $\hat{X}_{1D} = \mathbb{E}[X_{1D}|S, X_{1R}, X_2]$ be the optimal linear estimator of X_{1D} given (S, X_{1R}, X_2) under minimum mean square error criterion, and X'_{1D} be the resulting estimation error (note that $\mathbb{E}[X_{1D}|S, X_{1R}, X_2] = \mathbb{E}[X_{1D}|S, X_2]$). The estimator $\hat{X}_{1D}$ and the estimation error X'_{1D} are given by

$$\hat{X}_{1D} = \rho_{12}\sqrt{\frac{P_{1D}}{P_2}}X_2 + \rho_{1s}\sqrt{\frac{P_{1D}}{Q}}S \tag{F-1}$$

$$X'_{1D} = X_{1D} - \hat{X}_{1D}. \tag{F-2}$$

We can then write Y_3 in (80) alternatively as

$$Y_3 = X'_{1D} + (1 + \rho_{12}\sqrt{\frac{P_{1D}}{P_2}})X_2 + (1 + \rho_{1s}\sqrt{\frac{P_{1D}}{Q}})S + Z_3. \tag{F-3}$$

Let now

$$Y'_3 := Y_3 - \mathbb{E}[Y_3|X_{1R}, X_2] = X'_{1D} + (1 + \rho_{1s}\sqrt{\frac{P_{1D}}{Q}})S + Z_3. \tag{F-4}$$

Noticing now that X'_{1D} is independent of the state S in (F-4), it is clear that an optimal choice of the associated auxiliary random variable U_1 is

$$U_1 = X'_{1D} + \alpha(1 + \rho_{1s}\sqrt{\frac{P_{1D}}{Q}})S, \tag{F-5}$$

where α is Costa's parameter given by

$$\alpha = \frac{\mathbb{E}[X'^2_{1D}]}{\mathbb{E}[X'^2_{1D}] + \mathbb{E}[Z_3^2]} = \frac{P_{1D}(1 - \rho_{12}^2 - \rho_{1s}^2)}{P_{1D}(1 - \rho_{12}^2 - \rho_{1s}^2) + N_3}. \tag{F-6}$$

Then we can easily show that

$$I(U_1; Y_3|X_{1R}, X_2) - I(U_1; S|X_{1R}, X_2) = I(U_1; Y'_3) - I(U_1; S) \tag{F-7}$$

By substituting X'_{1D} in (F-5), we get

$$U_1 = X_{1D} - \rho_{12}\sqrt{\frac{P_{1D}}{P_2}}X_2 + \alpha_{\text{opt}}S \tag{F-8}$$

with

$$\begin{aligned}\alpha_{\text{opt}} &= \alpha(1 + \rho_{1s}\sqrt{\frac{P_{1D}}{Q}}) - \rho_{1s}\sqrt{\frac{P_{1D}}{Q}}\\ &= \Big[\frac{P_{1D}(1 - \rho_{12}^2 - \rho_{1s}^2)}{P_{1D}(1 - \rho_{12}^2 - \rho_{1s}^2) + N_3}\Big(1 + \rho_{1s}\sqrt{\frac{P_{1D}}{Q}}\Big) - \rho_{1s}\sqrt{\frac{P_{1D}}{Q}}\Big].\end{aligned} \tag{F-9}$$

Now, it is easy to see that, with the choice (F-8), we have

$$I(U_1; Y_3|X_{1R}, X_2) - I(U_1; S|X_{1R}, X_2) = I(U_1; Y'_3) - I(U_1; S)$$

$$= \frac{1}{2}\log\left(1 + \frac{\mathbb{E}[X'^2_{1D}]}{N_3}\right)$$
$$= \frac{1}{2}\log\left(1 + \frac{P_{1D}(1-\rho_{12}^2-\rho_{1s}^2)}{N_3}\right). \tag{F-10}$$

We now compute the terms $I(X_{1R};Y_2|X_2)$ and $I(X_2;Y_3)$. It is easy to see that, with the aforementioned jointly Gaussian input distribution,

$$\begin{aligned} I(X_{1R};Y_2|X_2) &= I(X_{1R};Y_2) \\ &= \frac{1}{2}\log(1+\frac{P_{1R}}{N_2+Q}). \end{aligned} \tag{F-11}$$

Also, we have

$$\begin{aligned} I(X_{1R},X_2;Y_3) &\stackrel{(a)}{=} I(X_2;Y_3) \\ &= h(Y_3) - h(Y_3|X_2) \\ &= h(Y_3) - h(X'_{1D} + \mathbb{E}[X_{1D}|X_2] + \mathbb{E}[X_{1D}|S] + S + Z_3|X_2) \\ &\stackrel{(b)}{=} h(Y_3) - h(X'_{1D} + \mathbb{E}[X_{1D}|S] + S + Z_3) \\ &\stackrel{(c)}{=} \frac{1}{2}\log\Big(\frac{\mathbb{E}[(X_{1D}+X_2+S)^2]+\mathbb{E}[Z_3^2]}{\mathbb{E}[X'^2_{1D}]+\mathbb{E}[(S+\mathbb{E}[X_{1D}|S])^2]+\mathbb{E}[Z_3^2]}\Big) \\ &= \frac{1}{2}\log\Big(1+\frac{(\sqrt{P_2}+\rho_{12}\sqrt{P_{1D}})^2}{P_{1D}(1-\rho_{12}^2-\rho_{1s}^2)+(\sqrt{Q}+\rho_{1s}\sqrt{P_{1D}})^2+N_3}\Big). \end{aligned} \tag{F-12}$$

where: (a) holds since X_{1R} is independent of (X_2,Y_3), (b) holds since X'_{1D} and S are independent of X_2, and (c) follows through straightforward algebra.

Adding (F-10) and (F-11) we obtain the first term of the minimization in (81); and adding (F-10) and (F-12) we obtain the second term of the minimization in (81).

Finally, we obtain the capacity in Theorem 9 by maximizing the RHS of (81) over all possible values of $\rho_{12}\in[-1,1]$ and $\rho_{1s}\in[-1,1]$. Investigating the two terms of the minimization, we can easily see that it suffices to consider $\rho_{12}\in[0,1]$ and $\rho_{1s}\in[-1,0]$. This concludes the proof of Theorem 9.

G. *Proof of Theorem 8*

In this section, we first use the upper bound for the DM case in Theorem 5 to obtain a new upper bound on the capacity of the state-dependent additive Gaussian model (71). Then, we show that this new upper bound is maximized by jointly Gaussian $(S,X_{1R},X_{1D},X_2,Z_2,Z_3)$.

From Theorem 5, we have that, given any (ϵ_n,n,R) sequence of codes with average error probability $P_e^n\longrightarrow 0$ as $n\longrightarrow+\infty$, the transmission rate R satisfies

$$R \leq \min\Big\{I(X_{1R};Y_2|X_2,S),\ I(X_2;Y_3)\Big\} + I(X_{1D};Y_3|X_2,S) \tag{G-1}$$

for some joint measure of the form

$$P_{S,X_{1R},X_{1D},X_2,Y_2,Y_3} = Q_S P_{X_2} P_{X_{1R}|X_2} P_{X_{1D}|X_2,S} W_{Y_2|X_{1R},S} W_{Y_3|X_{1D},X_2,S}. \tag{G-2}$$

Since the channel structure (71) satisfies $W_{Y_2|X_{1R},X_2,S} = W_{Y_2|X_{1R},S}$, it follows that

$$\begin{aligned} I(X_{1R};Y_2|S,X_2) &= H(Y_2|S,X_2) - H(Y_2|S,X_2,X_{1R}) \\ &= H(Y_2|S,X_2) - H(Y_2|S,X_{1R}) \\ &\leq H(Y_2|S) - H(Y_2|S,X_{1R}) \\ &= I(X_{1R};Y_2|S). \end{aligned} \tag{G-3}$$

An upper bound on the capacity of the channel (71) is then given by

$$R \leq \min\Big\{I(X_{1R};Y_2|S),\ I(X_2;Y_3)\Big\} + I(X_{1D};Y_3|X_2,S) \tag{G-4}$$

for some joint measure of the form

$$P_{S,X_{1R},X_{1D},X_2,Y_2,Y_3} = Q_S P_{X_2} P_{X_{1R}} P_{X_{1D}|X_2,S} W_{Y_2|X_{1R},S} W_{Y_3|X_{1D},X_2,S}. \tag{G-5}$$

(Note that, in contrast to in Theorem 5 and (G-2), the inputs X_{1R} and X_2 are independent in (G-5)).

Fix a joint distribution on $(S,X_{1R},X_{1D},X_2,Y_2,Y_3)$ of the form (G-5) satisfying

$$\begin{aligned} &\mathbb{E}[X_{1R}^2] = \tilde{P}_{1R} \leq P_{1R}, \quad \mathbb{E}[X_{1D}^2] = \tilde{P}_{1D} \leq P_{1D}, \quad \mathbb{E}[X_2^2] = \tilde{P}_2 \leq P_2, \\ &\mathbb{E}[X_{1D}X_2] = \sigma_{12}, \quad \mathbb{E}[X_{1D}S] = \sigma_{1s}. \end{aligned} \tag{G-6}$$

We shall also use the correlation coefficients $\rho_{12} \in [-1,1]$, $\rho_{1s} \in [-1,1]$ defined as

$$\rho_{12} = \frac{\sigma_{12}}{\sqrt{\tilde{P}_{1D}\tilde{P}_2}}, \quad \rho_{1s} = \frac{\sigma_{1s}}{\sqrt{\tilde{P}_{1D}Q}}. \tag{G-7}$$

We first compute the first term in the minimization on the RHS of (G-4). We have

$$R \leq I(X_{1R};Y_2|S) + I(X_{1D};Y_3|X_2,S) \tag{G-8}$$

$$= h(X_{1R}+Z_2|S) - h(Z_2) + h(X_{1D}+Z_3|X_2,S) - h(Z_3) \tag{G-9}$$

$$\stackrel{(a)}{\leq} h(X_{1R}+Z_2) - h(Z_2) + h(X_{1D}+Z_3|X_2,S) - h(Z_3) \tag{G-10}$$

$$\stackrel{(b)}{\leq} \frac{1}{2}\log\Big(1+\frac{\tilde{P}_{1R}}{N_2}\Big) + \frac{1}{2}\log\Big(1+\frac{\tilde{P}_{1D}(1-\rho_{12}^2-\rho_{1s}^2)}{N_3}\Big), \tag{G-11}$$

where: (a) holds since conditioning reduces entropy; and (b) holds since the conditional differential entropy $h(X_{1R}+Z_2)$ is maximized if (X_{1R},Z_2) are jointly Gaussian and, by the *Maximum Conditional Differential Entropy Lemma* [56, Part I], the conditional differential entropy $h(X_{1D}+Z_3|X_2,S)$ is maximized if (S,X_{1D},X_2,Z_3) are jointly Gaussian.

We now compute the term $[I(X_2;Y_3)+I(X_{1D};Y_3|X_2,S)]$. We have

$$\begin{aligned} I(X_2;Y_3)+I(X_{1D};Y_3|X_2,S) &\stackrel{(c)}{=} I(X_{1D};Y_3|X_2,S) + I(X_2;Y_3) - I(X_2;S) \\ &= I(X_{1D};Y_3|X_2,S) + I(X_2;Y_3|S) - I(X_2;S|Y_3) \\ &= h(Y_3|S) - h(Y_3|S,X_{1D},X_2) - h(S|Y_3) + h(S|X_1,Y_3) \end{aligned}$$

$$= h(Y_3) - h(S) + h(S|X_2, Y_3) - h(Z_3) \tag{G-12}$$

where: (c) follows since X_2 and S are independent.

For fixed second moments (G-6), we have

$$h(Y_3) \leq \frac{1}{2}\log(2\pi e)(\tilde{P}_{1D} + \tilde{P}_2 + 2\sigma_{12} + 2\sigma_{1s} + Q + N_3), \tag{G-13}$$

where equality is attained if Y_3 is Gaussian. Similarly, the term $h(S|X_2, Y_3)$ is maximized if (S, X_2, Y_3) are jointly Gaussian. Let $\hat{S}(X_2, Y_3) = \mathbb{E}[S|X_2, Y_3]$ be the MMSE estimator of S given (X_2, Y_3), i.e.,

$$\begin{aligned}\hat{S}(X_2, Y_3) &= \mathbb{E}[S|X_2, X_{1D} + S + Z_3]\\ &= \gamma_1 X_2 + \gamma_2(X_{1D} + S + Z_3)\end{aligned} \tag{G-14}$$

with

$$\begin{aligned}\gamma_1 &= -\frac{\sigma_{12}(Q + \sigma_{1s})}{\tilde{P}_2(\tilde{P}_{1D} + 2\sigma_{1s} + Q + N_3) - \sigma_{12}^2}\\ \gamma_2 &= \frac{\tilde{P}_2(Q + \sigma_{1s})}{\tilde{P}_2(\tilde{P}_{1D} + 2\sigma_{1s} + Q + N_3) - \sigma_{12}^2}.\end{aligned} \tag{G-15}$$

$$\begin{aligned}h(S|X_2, Y_3) &= h(S - \hat{S}(X_2, Y_3)|X_2, Y_3)\\ &\leq h(S - \gamma_1 X_2 - \gamma_2(X_{1D} + S + Z_3))\\ &= \frac{1}{2}\log(2\pi e)\mathbb{E}\Big[\Big(S - \gamma_1 X_2 - \gamma_2(X_{1D} + S + Z_3)\Big)^2\Big]\\ &= \frac{1}{2}\log\Big((2\pi e)\frac{Q\tilde{P}_{1D}\tilde{P}_2 + \tilde{P}_2 N_3 Q - \sigma_{1s}^2\tilde{P}_2 - \sigma_{12}^2 Q}{\tilde{P}_2(\tilde{P}_{1D} + 2\sigma_{1s} + Q + N_3) - \sigma_{12}^2}\Big),\end{aligned} \tag{G-16}$$

where the inequality is attained with equality if S, X_{1D}, X_2, Y_3 are jointly Gaussian. Then, from (G-12), (G-13) and (G-16) and straightforward algebra, we obtain

$$\begin{aligned}I(X_2; Y_3) + I(X_{1D}; Y_3|S, X_2) = \frac{1}{2}\log\Big(1 + \frac{(\sqrt{\tilde{P}_2} + \rho_{12}\sqrt{\tilde{P}_{1D}})^2}{\tilde{P}_{1D}(1 - \rho_{12}^2 - \rho_{1s}^2) + (\sqrt{Q} + \rho_{1s}\sqrt{\tilde{P}_{1D}})^2 + N_3}\Big)\\ + \frac{1}{2}\log\Big(1 + \frac{\tilde{P}_{1D}(1 - \rho_{12}^2 - \rho_{2s}^2)}{N_3}\Big).\end{aligned} \tag{G-17}$$

For convenience, let us now define the function $\Theta_1(\tilde{P}_{1R}, \tilde{P}_{1D}, \rho_{12}, \rho_{1s})$ as the RHS of (G-11) and the function $\Theta_2(\tilde{P}_{1D}, \tilde{P}_2, \rho_{12}, \rho_{2s})$ as the RHS of (G-17). From the above analysis, the capacity of the channel is upper-bounded as

$$C \leq \max\ \min\{\Theta_1(\tilde{P}_{1R}, \tilde{P}_{1D}, \rho_{12}, \rho_{1s}), \Theta_2(\tilde{P}_{1D}, \tilde{P}_2, \rho_{12}, \rho_{1s})\} \tag{G-18}$$

where the maximization is over all covariance matrices of (X_{1R}, X_{1D}, X_2, S) of the form

$$\Lambda_{X_{1R}, X_{1D}, X_2, S} = \begin{pmatrix} \tilde{P}_{1R} & 0 & 0 & 0\\ 0 & \tilde{P}_{1R} & \rho_{12}\sqrt{\tilde{P}_{1D}\tilde{P}_2} & \rho_{1s}\sqrt{\tilde{P}_{1D}Q}\\ 0 & \rho_{12}\sqrt{\tilde{P}_{1D}\tilde{P}_2} & \tilde{P}_2 & 0\\ 0 & \rho_{1s}\sqrt{\tilde{P}_{1D}Q} & 0 & Q \end{pmatrix}, \tag{G-19}$$

that satisfy

$$\tilde{P}_{1R} \leq P_{1R}, \quad \tilde{P}_{1D} \leq P_{1D}, \quad \tilde{P}_2 \leq P_2 \tag{G-20}$$

and have non-negative discriminant,

$$Q\tilde{P}_{1R}\tilde{P}_{1D}\tilde{P}_2(1-\rho_{12}^2-\rho_{2s}^2) \geq 0, \tag{G-21}$$

i.e., for $Q > 0$,

$$\rho_{12}^2 + \rho_{2s}^2 \leq 1. \tag{G-22}$$

Investigating $\Theta_1(\tilde{P}_{1R}, \tilde{P}_{1D}, \rho_{12}, \rho_{1s})$ and $\Theta_2(\tilde{P}_{1D}, \tilde{P}_2, \rho_{12}, \rho_{1s})$, it can be seen that it suffices to consider $\rho_{12} \in [0, 1]$ and $\rho_{1s} \in [-1, 0]$ for the maximization in (G-18).

Also, it is easy to see that, for fixed $\tilde{P}_{1D}$, the functions $\Theta_1(\tilde{P}_{1R}, \tilde{P}_{1D}, \rho_{12}, \rho_{1s})$ and $\Theta_2(\tilde{P}_{1D}, \tilde{P}_2, \rho_{12}, \rho_{1s})$ increase monotonically with $\tilde{P}_{1R}$ and $\tilde{P}_2$. So, for fixed $\tilde{P}_{1D}$, they are maximized at $\tilde{P}_{1R} = P_{1R}$ and $\tilde{P}_2 = P_2$. To complete the proof, we should show that $\Theta_1(P_{1R}, \tilde{P}_{1D}, \rho_{12}, \rho_{1s})$ and $\Theta_2(\tilde{P}_{1D}, P_2, \rho_{12}, \rho_{1s})$ are also maximized at $\tilde{P}_{1D} = P_{1D}$.
It is clear that the function $\Theta_1(P_{1R}, \tilde{P}_{1D}, \rho_{12}, \rho_{1s})$ increases with $\tilde{P}_{1D}$. The term $\Theta_2(\tilde{P}_{1D}, P_2, \rho_{12}, \rho_{1s})$ can be seen as the sum rate of a two-user state-dependent MAC with state information S^n known to one encoder, both encoders sending a common message and the informed encoder sending, in addition, an individual message [15]. As argued in [15], this sum rate increases with the power of the informed encoder [15, Appendix E], i.e., $\tilde{P}_{1D}$ here. This concludes the proof of Theorem 8.

H. Proof of Theorem 9

1) *Converse Part:* the proof of the converse part of Theorem 9 follows by noticing that the computation of the upper bound (G-4) in the proof of Theorem 8 for the special case (80), and using the same jointly Gaussian distribution as in Appendix G, gives the RHS of (81).

2) *Achievability Part:* the proof of the direct part of Theorem 9 follows by computing the rate (33) using an appropriate jointly Gaussian distribution on $(S, U_1, X_{1R}, X_{1D}, X_2)$. The algebra is similar to that in the proof of Proposition 2 and is therefore omitted for brevity.

References

[1] A. El Gamal and S. Zahedi, "Capacity of a class of relay channels with orthogonal components," *IEEE Trans. Inf. Theory*, vol. IT-51, pp. 1815–1817, May 2005.

[2] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Trans. Inf. Theory*, vol. 49, pp. 563–593, Mar. 2003.

[3] A. V. Kusnetsov and B. S. Tsybakov, "Coding in a memory with defective cells," *Problems of Predach. Inform.*, vol. 10, pp. 52–60, 1974.

[4] G. Caire and S. Shamai (Shitz), "On the throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-49, pp. 1691–1706, Jul. 2003.

[5] S. Viswanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum rate capacity of Gaussian MIMO broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-49, pp. 2658–2668, Oct. 2003.

[6] P. Viswanath and D. N. Tse, "Sum capacity of the vector Gaussian MIMO broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-49, pp. 1912–1921, Aug. 2003.

[7] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multi-terminal binning," *IEEE Trans. Inf. Theory*, vol. IT-48, pp. 1250–1276, Jun. 2002.

[8] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communication aspects," *IEEE Trans. Inf. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.

[9] J. Mitola, *Cognitive Radio: an Integrated Agent Architecture for Software Defined Radio*. PhD thesis, KTH Royal Institute of Technology, Stockholm, Sweeden, 2000.

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

[11] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, pp. 19–31, 1980.

[12] C. D. Heegard and A. El Gamal, "On the capacity of computer memory with defects," *IEEE Trans. Inf. Theory*, vol. IT-29, pp. 731–739, Sep. 1983.

[13] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, pp. 439–441, May 1983.

[14] 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.

[15] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdù, "Cooperative multiple access encoding with states available at one transmitter," *IEEE Trans. Inf. Theory*, vol. 54, pp. 4448–4469, Oct. 2008.

[16] S. Kotagiri and J. N. Laneman, "Multiaccess channels with state known to some encoders and independent messages," *EURASIP Journal on Wireless Commnunications and Networking*, vol. Article ID 450680. doi:10.1155/2008/450680, 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 *Proc. IEEE Int. Symp. Information Theory*, Seoul, Korea, Jun.-Jul. 2009, pp. 2376–2380.

[18] A. Khisti, U. Erez, A. Lapidoth, and G. Wornell, "Carbon copying onto dirty paper," *IEEE Trans. Inf. Theory*, vol. 53, pp. 1814–1827, May 2007.

[19] T. Philosoph, A. Khisti, U. Erez, and R. Zamir, "Lattice strategies for the dirty multiple access channel," in *Proc. IEEE Int. Symp. Information Theory*, Nice, France, Jun. 2007, pp. 386–390.

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

[21] ——, "Cooperative relaying with state available non-causally at the relay," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2272–2298, May 2010.

[22] A. Zaidi and L. Vandendorpe, "Rate regions for the partially-cooperative relay-broadcast channel with non-causal side information," in *Proc. IEEE Int. Symp. Information Theory*, Nice, France, Jun. 2007, pp. 1246–1250.

[23] ——, "Lower bounds on the capacity of the relay channel with states at the source," *EURASIP Journal on Wireless Commnunications and Networking*, vol. Article ID 634296. doi:10.1155/2009/634296, 2009.

[24] Y. Cemal and Y. Steinberg, "The multiple-access channel with partial state information at the encoders," *IEEE Trans. Inf. Theory*, vol. IT-51, pp. 3992–4003, Nov. 2005.

[25] Y. Steinberg, "Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information," *IEEE Trans. Inf. Theory*, vol. IT-51, pp. 2867–2877, Aug. 2005.

[26] A. Lapidoth and Y. Steinberg, "The multiple access channel with causal and strictly causal side information at the encoders," in *Proc. Int. Zurich Seminar on Communications (IZS)*, Zurich, Switzerland, Mar. 2010, pp. 13–16.

[27] ——, "The multiple access channel with two independent states each known causally at one encoder," in *Proc. IEEE Int. Symp. Information Theory*, Austin, TX, USA, Jun. 2010, pp. 480–484.

[28] H. Permuter, S. Shamai (Shitz), and A. Somekh-Baruch, "Message and state cooperation in multiple access channels," *IEEE Trans. Inf. Theory*, vol. 57, pp. 6379–6396, Oct. 2011.

[29] M. Li, O. Simeone, and A. Yener, "Multiple access channels with states causally known at transmitters," *Submitted for publication in IEEE Trans. Inf. Theory. Available at http://arxiv.org/abs/1011.6639*, 2010.

[30] ——, "Message and state cooperation in a relay channel when only the relay knows the state," *Submitted for publication in IEEE Trans. Inf. Theory. Available at http://arxiv.org/abs/1102.0768*, 2011.

[31] S. I. Bross and A. Lapidoth, "The state-dependent multiple-access channel with states available at a cribbing encoder," in *Proc. of IEEE 26-th Convention of Electrical and Electronics Engineers in Israel*, Israel, 2010.

[32] B. Akhbari, M. Mirmohseni, and M. R. Aref, "Compress-and-forward strategy for the relay channel with non-causal state information," in *Proc. IEEE Int. Symp. Information Theory*, Seoul, Korea, Jun.-Jul. 2009, pp. 1169–1173.

[33] M. N. Khormuji and M. Skoglund, "On cooperative downlink transmission with frequency reuse," in *Proc. IEEE Int. Symp. Information Theory*, Seoul, Korea, Jun.-Jul. 2009, pp. 849–853.

[34] G. Como and S. Yüksel, "On the capacity of memoryless finite state multiple-access channels with asymmetric state information at the encoders," *Accepted for publication in IEEE Trans. Inf. Theory. Available at http://arxiv.org/abs/1011.1012.1912*, 2011.

[35] N. Şen, G. Como, S. Yüksel, and F. Alajaji, "On the capacity of memoryless finite-state multiple access channels with asymmetric noisy state information at the encoders," in *Proc. IEEE Int. Symp. Information Theory 2011, submitted for publication. Available at http://arxiv.org/abs/1103.3054*, 2011.

[36] R. Khosravi-Farsani and F. Marvasti, "Capacity bounds for multiuser channels with non-causal channel state information at the transmitters," in *Proc. IEEE Int. Symp. Information Theory 2011, submitted for publication. Available at http://arxiv.org/abs/1102.3410*, 2011.

[37] S. Kotagiri and J. N. Laneman, "Achievable rates for multiple access channels with state information known at one encoder," in *Proc. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, 2004.

[38] S. Kotagiri and J. Laneman, "Multiaccess channels with state known to one encoder: A case of degraded message sets," in *Proc. IEEE Int. Symp. Information Theory*, Nice, France, Jun. 2007, pp. 1566–1570.

[39] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdù, "Cooperative encoding with asymmetric state information at the transmitters," in *Proc. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Sep. 2006.

[40] ——, "Cooperative multiple access encoding with states available at one transmitter," in *Proc. IEEE Int. Symp. Information Theory*, Nice, France, Jun. 2007, pp. 1556–1560.

[41] A. Zaidi, P. Piantanida, and S. Shamai (Shitz), "Multiple access channel with states known noncausally at one encoder and only strictly causally at the other encoder," in *Proc. IEEE Int. Symp. Information Theory, submitted for publication*, 2011.

[42] ——, "The capacity region of multiple access channel with states known noncausally at one encoder and only strictly causally at the other encoder," *IEEE Trans. Inf. Theory, to be submitted for publication*, 2011.

[43] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, pp. 572–584, Sep. 1979.

[44] Y.-H. Kim, A. Sutivong, and S. Sigurjonsson, "Multiple user writing on dirty paper," in *Proc. IEEE Int. Symp. Information Theory*, Chicago-USA, Jun. 2004, p. 534.

[45] A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. IT-28, pp. 851–857, Nov. 1982.

[46] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-25, pp. 306–311, May. 1979.

[47] A. E. Gamal and E. C. van der Meulen, "A proof of marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-27, pp. 120–122, Jan. 1981.

[48] Y. Steinberg and S. Shamai (Shitz), "Achievable rates for the broadcast channel with states known at the transmitter," in *Proc. IEEE Int. Symp. Information Theory*, Adelaide, Australia, Sep. 2005, pp. 2184–2188.

[49] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Willey & Sons INC., 1991.

[50] A. El Gamal and M. Aref, "The capacity of the semideterministic relay channel," *IEEE Trans. Inf. Theory*, vol. IT-28, p. 536, May 1982.

[51] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Willey, 1968.

[52] S. H. Lim, Y.-H. Kim, A. E. Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, pp. 3132–3152, May 2011.

[53] M. N. Khormuji and M. Skoglund, "Noisy network coding approach to the relay channel with a random state," in *Proc. of 45th Annual Conf. on Information Sciences and Systems (CISS)*, Mar. 2011, pp. 1–5.

[54] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, pp. 339–348, 1978.

[55] ——, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. London, U. K.: Academic Press, 1981.

[56] A. E. Gamal and Y.-H. Kim, *Lecture Notes on Network Information Theory*. Available at http://arxiv.org/abs/1001.3404, 2010 [on line].