Cooperative Relaying with State Available Non-Causally at the Relay

Abdellatif Zaidi Shiva Prasad Kotagiri J. Nicholas Laneman Luc Vandendorpe

Abstract

We consider a three-terminal state-dependent relay channel with the channel state non-causally available at only the relay. In the framework of cooperative wireless networks, some specific terminals may be equipped with cognition capabilities, i.e., the relay in our setup. In the discrete memoryless (DM) case, we establish lower and upper bounds on channel capacity. The lower bound is obtained by a coding scheme at the relay that uses a combination of codeword splitting, Gel'fand-Pinsker binning, and decode-and-forward relaying. The upper bound improves upon that obtained by assuming that the channel state is available at the source, the relay, and the destination. For the Gaussian case, we also derive lower and upper bounds on the capacity. The lower bound is obtained by a coding scheme at the relay that uses a combination of codeword splitting, generalized dirty paper coding, and decode-and-forward relaying; the upper bound is also better than that obtained by assuming that the channel state is available at the source, the relay, and the destination. In the case of degraded Gaussian channels, the lower bound meets with the upper bound for some special cases and the capacity is obtained for these cases. Furthermore, in the Gaussian case, we also extend the results to the case in which the relay operates in a half-duplex mode.

Index Terms

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

Abdellatif Zaidi and Luc Vandendorpe are with École Polytechnique de Louvain, Université Catholique de Louvain, Louvain-la-Neuve 1348, Belgium. Email: {abdellatif.zaidi,luc.vandendorpe}@uclouvain.be

J. Nicholas Laneman is with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556; Email: jnl@nd.edu. Shiva Prasad Kotagiri was with the Department of Electrical Engineering, University of Notre Dame, and is now with SERDES Technology Group, Xilinx Inc., San Jose, CA, USA-95124; Email: skotagir@gmail.com

The work of A. Zaidi and L. Vandendorpe has been supported in part by the EU framework program COOPCOM and the network of excellence NEWCOM++. The work of S. P. Kotagiri and J. N. Laneman has been supported in part by NSF Grants CNS06-26595 and CCF05-46618.

I. INTRODUCTION

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

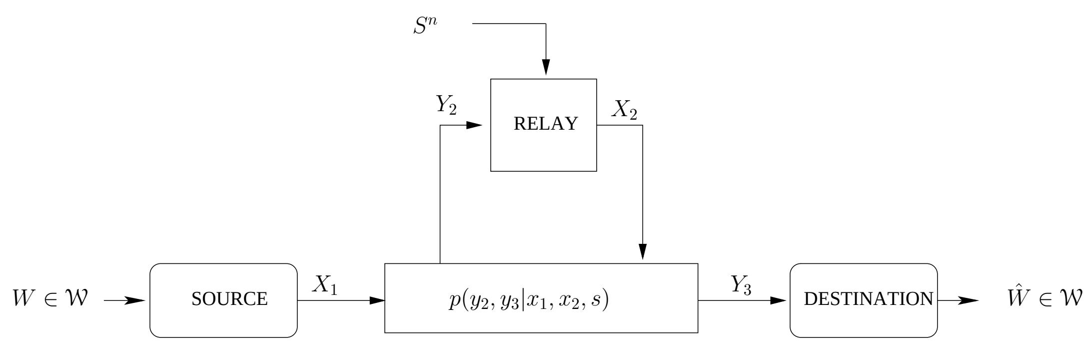


Fig. 1. Relay channel with state information S^n available non-causally at only the relay.

A. Background

Channels with random parameters or states have received considerable attention due to a wide range of possible applications. Shannon initiated the study of single-user models with state *causally* available at the encoder [1]. For the single-user discrete memoryless (DM) state-dependent models, Gel'fand and Pinsker derive the capacity for the setup in which the channel state is *non-causally* available at the encoder [2]. In this case, a random coding scheme based on binning, known as *Gel'fand-Pinsker coding*, achieves the capacity [2]. Costa considers an additive Gaussian channel with additive Gaussian state known at the encoder and shows that Gel'fand-Pinsker coding with a specific auxiliary random variable, widely known as *dirty paper* *coding* (DPC), achieves the trivial upper bound obtained by assuming the channel state available also at the decoder [3]. Interestingly, DPC eliminates 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. Also, since DPC achieves the trivial upper bound for this model, there is no need to derive tighter upper bounds for this model. In [4], models with channel state available non-causally at the encoder are studied from the view of memories with defects. Practical coding realizations using concepts of lattices for the models with non-causal encoder state information are studied, e.g., in [5], [6]. For a review on the subject of state-dependent channels and related work, the reader may refer to [7].

A growing body of work studies multi-user state-dependent models with non-causal encoder state information [8]–[20]. In the multi-user models, the channel state can be known to all, only some, or none of the users in the communication system. In the case of state-dependent DM models, the multiple access channels (MAC) are considered in [11] if partial channel state is available at all the encoders and the full channel state is available at the decoder, and the state dependent broadcast channels (BC) are considered in [12], [21] if the channel state is non-causally known at the encoder.

In the Gaussian case, the MAC with all informed encoders, the BC with informed encoder, the physically degraded relay channel (RC) with informed source and informed relay, and the physically degraded relay broadcast channel (RBC) with informed source and informed relay are studied in [8], [9], [18]. In all these cases, it is shown that some variants of DPC achieve the respective capacity or the respective capacity region. Also, since for all these models DPC achieves the trivial upper or outer bound obtained by assuming that the channel state is also available at the decoders, it is not necessary to derive tighter upper or outer bounds. For all these models, the key assumption that makes the problem relatively easy is the availability of the channel state at *all* the encoders in the communication model so that these encoders can remove the effect of the channel state on their respective communication using variants of DPC. It is interesting to study state-dependent multi-user models in which *some, but not all,* encoders are informed of the channel state, because the uninformed encoders cannot apply DPC.

The state-dependent MAC with some, but not all, encoders informed of the channel state is considered in [10], [13]–[16] and the state-dependent relay channel with informed source is considered in [18], [19]. For the Gaussian case of these models, the informed encoder applies a slightly generalized DPC (GDPC) in which the channel input random variable and the channel state random variable are negatively correlated. In these models, the uninformed encoders benefit from GDPC applied by the informed encoders because the negative correlation between the codewords at the informed encoders and the channel state can be interpreted as (partial) state cancellation. For the state-dependent MAC with one informed encoder and the case in which the message sets are degraded, the capacity region for the Gaussian case is obtained by deriving a non-trivial outer bound [16]. For the study of communication models in which only some encoders are informed, it is important to obtain non-trivial upper or outer bounds that show how tight are the achievable rates or rate regions obtained by application of variants of DPC. In this paper, we consider a state-dependent relay channel with the channel state known to only the relay. This model is conceptually different from the model considered in [18], [19] in which the channel state is non-causally known to only the source.

B. Motivation

Channels whose probabilistic input-output relationship depends on random parameters, called channel state, can model a large variety of problems, each related to some physical situation of interest. This includes information embedding [22]–[27], dispersive (ISI) channels [5], fading in wireless environments [28], and multiple-input multiple-output (MIMO) broadcast channels [29]–[31] where DPC is a central ingredient in achieving the capacity region [32].

More recently, driven by the growing demand for frequency spectrum, smart radio devices that are capable of obtaining the knowledge about the channel state, called cognitive radios, are introduced into communication systems in order to help those non-cognitive radios in terms of spectral efficiency [33]. In a wireless interference network in which some terminals compete and some others cooperate, equipping some specific terminals with cognition capabilities that allow them to learn the interference to high accuracy would help other non-cognitive terminals. These cognitive radios can exploit the knowledge of the interference or channel state to remove its effect on the transmission of their own messages and also that of the messages of the non-cognitive terminals as well. In order to better understand communication systems that involve cognitive radios, it is important to study fundamental performance limits, such as capacity or capacity regions, of the models with only some of the encoders being informed. For example, to increase system spectral efficiency, collaboration has been investigated in the realm of cognition in [34]–

[36]. Also, the problem of collaborative signal transmission in the presence of some cognizant terminals has been investigated for a MAC scenario in [13], [16] and for an interference channel scenario in [37]–[40]. The setup we consider in this paper also models the building block for collaborative wireless networks in which only the relays, but neither sources nor destinations, are cognizant of the channel state. An example of such scenario is shown in Figure 2.

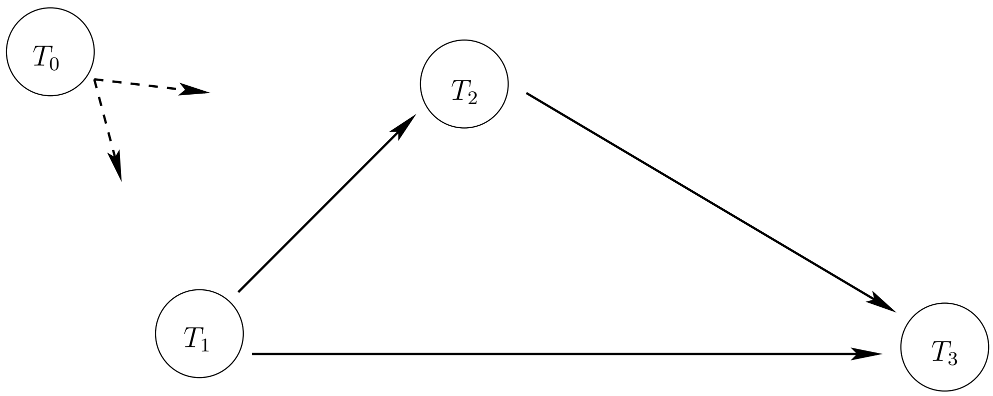


Fig. 2. Example wireless network with cognition capabilities. If the relay T_2 is cognizant of the competing source T_0, it can help the source T_1 cancel the effect of the interference from T_0.

C. Main Contributions

For the DM case, we derive lower and upper bounds on the capacity of the general state-dependent relay channel with informed relay. The lower bound is obtained by a coding scheme at the relay that uses a combination of codeword splitting, Gel'fand-Pinsker coding, and decode-and-forward (DF) relaying. For this model, designing a codebook at the relay is challenging since such a codebook should allow the source to generate codewords that are correlated with the channel input of the relay which exploits the available channel state. In this work, this is accomplished by codeword splitting at the relay. In codeword splitting, the channel input of the relay is generated from two codewords: the first of which is a function of the cooperative information and the channel state, and the second of which is a function of only the cooperative information. Since the source knows the cooperative information, it can generate its channel input in a way such that it is correlated with the latter codeword at the relay, which is a function of only the cooperative information.

The upper bound on the capacity is tighter than that obtained by assuming that the channel state is also available at the source and the destination. This upper bound is non-trivial and relates to the bounding technique developed in the context of multiple access channels with asymmetric channel state in [16, Theorem 2]; however, we note that the present upper bound is proved using techniques that are different from those in [16]. On a related note, we mention that at a high level there is a connection between the multiple access transmission part in the RC with informed relay in this work and the models in [13], [16]. However, there are also numerous conceptual differences that will be discussed whenever relevant. In particular, by opposition to [13], [16], in our setup, the uninformed encoder (the source) knows the message of the informed encoder (the relay).

Furthermore, we specialize the results to the case in which the channel is degraded. Also, we extend the lower bound for the DF relaying scheme to the case in which the relay employs a partial decode-and-forward relaying scheme.

We apply the concepts developed in the DM case to the Gaussian case in which both the noise and the state are additive Gaussian random variables. In our analysis for the Gaussian RC, we first allow the relay to operate in a *full-duplex* mode in which it can transmit and receive simultaneously, and then we constrain it to operate in a *half-duplex* mode in which it can either only transmit or only receive.

In the case of full-duplex transmission, we derive lower and upper bounds on the capacity of the Gaussian relay channel with informed relay. We obtain the lower bound by using the concepts of codeword splitting, generalized DPC (GDPC) [10], [41], and decode-and-forward relaying. Through codeword splitting, the channel input of the source is *partially* coherent with the channel input of the relay. We also point out the loss incurred by the availability of the channel state at only the relay in the upper bound. We show that the lower bound is in general close to the upper bound. In the case of degraded Gaussian channel, the lower bound meets with the upper bound for some special cases.

In the case of half-duplex transmission, we derive lower and upper bounds for the capacity of the Gaussian relay channel with informed relay. In this case, we focus on relaying protocols in which the relay either fully or partially decodes the source message, re-encodes and sends it to the destination, i.e., *full/partial decode-and-forward*.

D. Outline and Notation

An outline of the remainder of this paper is as follows. Section II describes the communication model that we consider in this work. Section III provides lower and upper bounds on the capacity of the general discrete memoryless RC with informed relay. Section IV provides lower and upper bounds on the capacity of the Gaussian RC with informed relay. This section also contains some numerical results and discussions. Finally, Section V 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. The short-hand notation X_i^j indicates a sequence of random variables $(X_i, X_{i+1}, \cdots, X_j)$ and x_i^j denotes a particular realization of a random sequence X_i^j. For convenience, the length n vector x^n will occasionally be denoted in boldface notation $\mathbf{x}$. We use the notation $\mathbb{E}_X[\cdot]$ to denote the expectation of random variable X. The set of probability distributions defined on an alphabet $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$. A probability distribution of a random variable Y given X is denoted by $P_{Y|X}$. The Gaussian distribution with mean μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$. Finally, throughout the paper, 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 definitions related to it. As shown in Figure 1, we consider a state-dependent relay channel denoted by $W_{Y_2,Y_3|X_1,X_2,S}$ whose outputs $Y_2 \in \mathcal{Y}_2$ and $Y_3 \in \mathcal{Y}_3$ for the relay and the destination, respectively, are controlled by the channel inputs $X_1 \in \mathcal{X}_1$ from the source and $X_2 \in \mathcal{X}_2$ from the relay, along with a channel state $S \in \mathcal{S}$. It is assumed that the channel state S_i at time instant i is independently drawn from a given distribution Q_S and the channel state S^n is non-causally known at the relay.

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 $\mathcal{W} = \{1, \cdots, M\}$. The information rate R is defined as $\log M/n$ bits per transmission.

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

$$\phi_1^n : \quad \{1, \cdots, M\} \rightarrow \mathcal{X}_1^n,$$

a sequence of encoding functions at the relay

$$\phi_{2,i} : \quad \mathcal{Y}_{2,1}^{i-1} \times \mathcal{S}^n \rightarrow \mathcal{X}_{2,i},$$

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

$$\psi^n : \quad \mathcal{Y}_3^n \rightarrow \{1, \cdots, M\}.$$

From a (M, n) code, the sequences X_1^n and X_2^n from the source and the relay, respectively, are transmitted across a state-dependent relay channel $W(y_2, y_3|x_1, x_2, s)$ modeled as a memoryless conditional probability distribution, so that

$$P_{Y_2^n, Y_3^n|X_1^n, X_2^n, S^n}(y_2^n, y_3^n|x_1^n, x_2^n, s^n) = \prod_{i=1}^{n} W_{Y_2,Y_3|X_1,X_2,S}(y_{2,i}, y_{3,i}|x_{1,i}, x_{2,i}, s_i). \quad (1)$$

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

An (ϵ, n, R) code for the state-dependent RC with informed relay 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 $\lim_{n\rightarrow\infty} \epsilon_n = 0$. The capacity C of the non-causal state-dependent RC with informed relay is the supremum of the set of achievable rates.

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

$$W_{Y_2,Y_3|X_1,X_2,S} = W_{Y_2|X_1,X_2,S} W_{Y_3|Y_2,X_2,S}. \quad (2)$$

III. The Discrete Memoryless RC with Informed Relay

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

A. Lower Bound on Capacity

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

Theorem 1: The capacity C of the state-dependent DM RC with informed relay satisfies $C \geq R^{\text{lo}}$, where

$$R^{\text{lo}} = \max \min \Big\{ I(X_1; Y_2|S, U_1), \\ I(X_1, U_1, U_2; Y_3) - I(U_2; S|U_1) \Big\}, \quad (3)$$

with the maximization over all probability distributions of the form

$$P_{S,U_1,U_2,X_1,X_2,Y_2,Y_3} = \\ Q_S P_{U_1} P_{X_1|U_1} P_{U_2|U_1,S} P_{X_2|U_1,U_2,S} W_{Y_2,Y_3|X_1,X_2,S} \quad (4)$$

and $U_1 \in \mathcal{U}_1$ and $U_2 \in \mathcal{U}_2$ are auxiliary random variables with

$$|\mathcal{U}_1| \leq |\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2| + 1 \quad (5a)$$

$$|\mathcal{U}_2| \leq \Big(|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2| + 1\Big)|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2|, \quad (5b)$$

respectively.

Remark 1: The lower bound (3) is based upon a technique at the relay we call *codeword splitting*, combining decode-and-forward (DF) relaying [42, Theorem 1] with Gel'fand-Pinsker coding [2]. In conventional DF strategies, the source knows the relay input, allowing the source and relay to utilize a joint codebook to transmit cooperative information. However, in our model there is a tension between the utility of a joint codebook for relaying and the utility of the relay's making use of the channel state, which is unknown to the source. To resolve this tension, we generate two codebooks at the relay. In one codebook, the random codewords U_1^n are generated using a random variable U_1 that is independent of the channel state S. The relay chooses the appropriate random codeword from this codebook using only the cooperative information. In the other codebook, the codewords U_2^n are generated using a random variable U_2 that is correlated with the channel state S and the variable U_1 through $P_{U_2|U_1,S}$. The relay chooses the appropriate codeword from this codebook using both the cooperative information and the channel state, in order to combat the effect of the channel state on the communication. Finally, the relay generates the channel input X_2^n from (U_1^n, U_2^n) using the conditional probability law $P_{X_2|U_1,U_2,S}$. The source knows U_1^n as this is a function of only the cooperative information, and, given U_1^n, it generates the random codeword X_1^n according to the conditional probability law $P_{X_1|U_1}$. Thus, the channel inputs of the source and the relay are correlated through U_1^n. A dependence diagram of the random variables that are involved in the coding scheme is shown in Figure 3. ∎

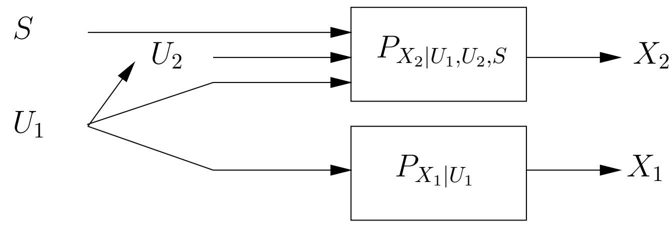


Fig. 3. Dependence diagram of the random variables for the lower bound in Theorem 1.

Remark 2: The term $[I(X_1, U_1, U_2; Y_3) - I(U_2; S|U_1)]$ in (3) can be interpreted as an achievable sum rate over a state-dependent MAC with one informed encoder and degraded messages, i.e., one common and one individual message. In our model, the informed encoder sends only the common message, i.e., the cooperative information of DF relaying, and the uninformed encoder sends both the common and individual messages. By contrast, [13], [16] derive the capacity region for the reverse situation in which the informed encoder sends both the common and individual messages, and the uninformed encoder sends only the common message. This swapping of roles makes coding at the relay more involved than in [13], [16] for the state-dependent MAC and the [18], [19] for the related state-dependent RC with informed source. ∎

Proof of Theorem 1:

First we generate a random codebook that we use to obtain the lower bound in Theorem 1. Next, we outline the encoding and decoding procedures at the source and the relay. The coding scheme is based on a combination of codeword splitting, regular-encoding backward decoding for DF [43], and a variation of Gelf'and-Pinsker binning. A formal proof with complete error analysis is given in Appendix A.

Codebook generation:

Fix a measure $P_{S,U_1,U_2,X_1,X_2,Y_2,Y_3}$ satisfying (4). Fix $\epsilon > 0$ and denote

$$J = 2^{n(I(U_2;S|U_1)+2\epsilon)} \tag{6a}$$

$$M = 2^{n(R-4\epsilon)}. \tag{6b}$$

1. We generate M independent and identically distributed (i.i.d.) codewords $\{\mathbf{u}_1(w')\}$ indexed by $w' = 1, \ldots, M$, each with i.i.d. components drawn according to P_{U_1}.
2. For each codeword $\mathbf{u}_1(w')$, we generate M i.i.d. codewords $\{\mathbf{x}_1(w', w)\}$ at the source indexed by $w = 1, \ldots, M$, and J auxiliary codewords $\{\mathbf{u}_2(w', j)\}$ at the relay indexed by $j = 1, \ldots, J$. The codewords $\mathbf{x}_1(w', w)$ and $\mathbf{u}_2(w', j)$ are with i.i.d. components given $\mathbf{u}_1(w')$ drawn according to $P_{X_1|U_1}$ and $P_{U_2|U_1}$, respectively.

Outline of the coding scheme:

We outline the coding scheme in the following. The message W to be sent from the source node is divided into B blocks $w_1, w_2, \ldots, w_B$ of nR bits each. For convenience we let $w_{B+1} = 1$. The transmission is performed in $B+1$ blocks. We denote by $\mathbf{s}[i]$ the channel state in block i, $i = 1, \ldots, B+1$.

Continuing with the strategy, in the first block, the source transmits $\mathbf{x}_1(1, w_1)$. The relay searches for the smallest $j \in \{1, \cdots, J\}$ such that $\mathbf{u}_1(1)$, $\mathbf{u}_2(1, j)$ and $\mathbf{s}[1]$ are jointly typical (the properties of strongly typical sequences guarantee that there exists one such j). Denote this j by $j^\star = j(\mathbf{s}[1], 1)$. Then, the relay transmits a vector $\mathbf{x}_2(1)$ with i.i.d. components given $(\mathbf{u}_1(1), \mathbf{u}_2(1, j^\star), \mathbf{s}[1])$ drawn according to the marginal $P_{X_2|U_1,U_2,S}$ induced by the distribution (4). The decoder at the relay uses joint typicality. It declares that message $\hat{w}_1$ is sent if there is a unique $\hat{w}_1$ such that $\mathbf{x}_1(1, \hat{w}_1)$ is jointly typical with $(\mathbf{y}_2[1], \mathbf{s}[1])$ given $\mathbf{u}_1(1)$, where $\mathbf{y}_2[1]$ denotes the information received at the relay in block 1. One can show that the relay can decode reliably as long as n is large and

$$R < I(X_1; Y_2|S, U_1). \tag{7}$$

So, suppose the relay correctly obtains w_1. In the second block, the source transmits $\mathbf{x}_1(w_1, w_2)$ and the relay transmits a vector $\mathbf{x}_2(w_1)$ with i.i.d. components given $\mathbf{u}_1(w_1)$, $\mathbf{u}_2(w_1, j(\mathbf{s}[2], w_1))$, $\mathbf{s}[2]$ drawn according to the marginal $P_{X_2|U_1,U_2,S}$; the sequence $\mathbf{u}_2(w_1, j(\mathbf{s}[2], w_1))$ is chosen such that $j(\mathbf{s}[2], w_1)$ is the smallest $j \in \{1, \cdots, J\}$ satisfying $\mathbf{u}_1(w_1)$, $\mathbf{u}_2(w_1, j)$ and $\mathbf{s}[2]$ are jointly typical. Upon observation of $\mathbf{y}_2[2]$, the decoder at the relay declares that $\hat{w}_2$ is sent if there is a unique $\hat{w}_2$ such that $\mathbf{x}_1(w_1, \hat{w}_2)$ is jointly typical with $(\mathbf{y}_2[2], \mathbf{s}[2])$ given $\mathbf{u}_1(w_1)$. Again, it can decode reliably as long as n is large and (7) is true. At the relay, one continues in this way until block $B+1$.

Consider now the destination, and let $\mathbf{y}_3[i]$ be the received information at the destination in block i. Suppose these information are collected until the last block of transmission is completed. The destination can then perform Willem's *backward decoding* [43], by first decoding w_B from $\mathbf{y}_3[B+1]$. Note that $\mathbf{y}_3[B+1]$ depends on $\mathbf{x}_1(w_B, 1)$, $\mathbf{u}_1(w_B)$ and $\mathbf{u}_2(w_B, j(\mathbf{s}[B+1], w_B))$, which in turn depends only on w_B. The decoder at the destination uses joint typicality. It declares that $\hat{w}_B$ is sent if there is a unique $\hat{w}_B$ such that $\mathbf{x}_1(\hat{w}_B, 1)$, $\mathbf{u}_1(\hat{w}_B)$, $\mathbf{u}_2(\hat{w}_B, j_B)$, $\mathbf{y}_3[B+1]$ are jointly typical, for some index $j_B \in \{1, \ldots, J\}$. One can show that the destination can decode reliably as long as n is large and

$$R < I(X_1, U_1, U_2; Y_3) - I(U_2; S|U_1). \tag{8}$$

So, suppose the destination correctly obtains w_B. Next, the destination decodes w_{B-1} from $\mathbf{y}_3[B]$, which depends on $\mathbf{x}_1(w_{B-1}, w_B)$, $\mathbf{u}_1(w_{B-1})$ and $\mathbf{u}_2(w_{B-1}, j(\mathbf{s}[B], w_{B-1}))$. Since the destination knows w_B, it can again decode reliably as long as n is large and (8) is true. At the destination, one continues in this fashion until all message blocks have been decoded. The average rate over the $B+1$ blocks is $RB/(B+1)$ bits per use, and by making B large one can get the rate as close to R as desired.

Remark 3: In the case of classic RC without state, one can consider three different decode-and-forward strategies: irregular encoding successive decoding [42], regular encoding sliding-window decoding [44] and regular encoding backward decoding. It is well known that these three strategies achieve the same rate in this case. In the state-dependent case with informed relay, one can show that backward decoding achieves rates higher than those of sliding-window decoding. More precisely, sliding window decoding at the destination at the end of block i is as follows (we use the notation in the proof of Theorem 1). The destination knows w_{i-2} and also the correct index $j(\mathbf{s}[i-1], w_{i-2})$, and decodes w_{i-1} based on the information received in the two adjacent blocks $i-1$ and i. It declares that the message $\hat{w}_{i-1}$ is sent if there is a unique pair $(\hat{w}_{i-1}, \hat{j}_{i-1})$ such that the vectors $\mathbf{u}_1(w_{i-2})$, $\mathbf{u}_2(w_{i-2}, j(\mathbf{s}[i-1], w_{i-2}))$, $\mathbf{x}_1(w_{i-2}, \hat{w}_{i-1})$, $\mathbf{y}_3[i-1]$ are jointly typical, *and* the vectors $\mathbf{u}_1(\hat{w}_{i-1})$, $\mathbf{u}_2(\hat{w}_{i-1}, \hat{j}_{i-1})$, $\mathbf{y}_3[i]$ are jointly typical. Thus, the destination obtains the message w_{i-1} if

$$R < I(X_1, U_1, U_2; Y_3) - I(U_2; S|U_1) \tag{9}$$

$$I(U_2; Y_3|U_1) - I(U_2; S|U_1) > 0. \tag{10}$$

Hence, with window decoding also, the achievable rate is obtained by maximizing the RHS of (3). However, unlike the above backward decoding scheme, the maximization is over a set of distributions of the form (4) that satisfy the constraint (10). Because of the additional constraint, this set is smaller than the one used in Theorem 1. Informally speaking, the additional constraint (10) guarantees that, in the decoding of the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$, the destination can actually decode the vector $\mathbf{u}_2$ *fully*, i.e., it can determine not only the bin index (i.e., the message w_{i-1}) but also the correct sequence in the bin (i.e., the index $j(\mathbf{s}[i], w_{i-1})$). ∎

The achievable rate in (3) requires the relay to *fully* decode the message sent by the source, and this can be rather a severe constraint. We can generalize Theorem 1 by allowing the relay to decode the source message *only partially* [45]. This can be done by introducing a new random variable U that represents the information decoded by the relay. The following corollary provides an achievable rate obtained by using a partial decode-and-forward (PDF) scheme at the informed relay.

Corollary 1: The capacity C of the state-dependent DM RC with informed relay satisfies $C \geq R^{\text{lo}}_{\text{PDF}}$, where

$$\begin{aligned} R^{\text{lo}}_{\text{PDF}} = \; & \max \min \Big\{ I(U; Y_2|S, U_1) + I(X_1; Y_3|U, U_1, U_2) \\ & + \min\{0, I(U_2; Y_3|U, U_1) - I(U_2; S|U_1)\}, I(X_1, U_1, U_2; Y_3) - I(U_2; S|U_1) \Big\}, \end{aligned} \tag{11}$$

with the maximization over all probability distributions of the form

$$\begin{aligned} & P_{S,U_1,U_2,U,X_1,X_2,Y_2,Y_3} = \\ & \quad Q_S P_{U_1} P_{U|U_1} P_{X_1|U_1,U} P_{U_2|U_1,S} P_{X_2|U_1,U_2,S} W_{Y_2,Y_3|X_1,X_2,S} \end{aligned} \tag{12}$$

and $U_1 \in \mathcal{U}_1$, $U_2 \in \mathcal{U}_2$, and $U \in \mathcal{U}$ are auxiliary random variables with

$$|\mathcal{U}_1| \leq |\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2| + 2 \tag{13a}$$

$$|\mathcal{U}_2| \leq \Big(|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2| + 2\Big)|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2| + 2 \tag{13b}$$

$$|\mathcal{U}| \leq \Big(|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2| + 2\Big)|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2| + 2, \tag{13c}$$

respectively.

The proof of Corollary 1 is similar to that of Theorem 1 and, hence, only an outline of it is given in Appendix B. For instance, the particular choice $U = X_1$ in Corollary 1 gives the lower bound in Theorem 1.

An informal interpretation of the rate (11) for the case in which $[I(U_2;Y_3|U,U_1) - I(U_2;S|U_1)] > 0$ is as follows. Since $I(U;Y_3|U_1,U_2,X_1) = 0$ for the distribution considered in (12), the second term of the minimization in (11) can be written as

$$I(U,U_1,U_2;Y_3) - I(U_2;S|U_1) + I(X_1;Y_3|U,U_1,U_2).$$

The rate (11) can then be interpreted as the rate achievable if the message W transmitted by the source is split into two independent parts, one of which is transmitted through the relay, say at rate R_r, and the other is transmitted directly to the destination without the help of the relay, say at rate R_d. The total rate is $R = R_r + R_d$. In (11) the auxiliary variable U stands for the information decoded by the relay and plays the role of X_1 in Theorem 1. Thus, it follows from (3) that the message transmitted through the relay can be decoded correctly at the destination if rate R_r satisfies

$$R_r \ < \ \min\Big\{I(U;Y_2|S,U_1), I(U,U_1,U_2;Y_3) - I(U_2;S|U_1)\Big\}. \tag{14}$$

It can also be easily argued (see Appendix B) that the additional information, which is sent on top of the information transmitted through the relay, can be decoded correctly at the destination if rate R_d satisfies

$$R_d \ < \ I(X_1;Y_3|U,U_1,U_2). \tag{15}$$

This shows that message W can be sent at the rate (11).

We close this section by noting the relay can employ other relaying schemes to assist the source, such as estimate-and-forward [42], amplify-and-forward [46]–[48] or combinations of these schemes. However, these schemes are beyond the scope of this paper.

B. Upper Bound on Capacity

The following theorem provides an upper bound on the capacity of the state-dependent DM RC with informed relay.

Theorem 2: The capacity C of the state-dependent DM RC with informed relay satisfies $C \leq R^{\text{up}}$, where

$$\begin{aligned} R^{\text{up}} \; = \; & \max\min\Big\{I(X_1;Y_2,Y_3|S,X_2), \\ & I(X_1,X_2;Y_3|S) - I(X_1;S|Y_3)\Big\} \end{aligned} \tag{16}$$

with the maximization over all probability distributions of the form

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

The proof of Theorem 2 appears in Appendix C.

In the second term of the minimum in (16), $I(X_1;S|Y_3)$ can be interpreted as the rate penalty caused by the source's not knowing the channel state. This rate loss makes the above upper bound tighter than the trivial upper bound obtained by assuming that the channel state is also available at the source and the destination, i.e.,

$$R^{\text{up}}_{\text{triv}} \; = \; \max\min\Big\{I(X_1;Y_2,Y_3|S,X_2), I(X_1,X_2;Y_3|S)\Big\} \tag{18}$$

with the maximization over all distributions of the form

$$P_{S,X_1,X_2,Y_2,Y_3} = Q_S P_{X_1|S} P_{X_2|X_1,S} W_{Y_2,Y_3|X_1,X_2,S}. \tag{19}$$

If the channel is physically degraded, the upper bound in Theorem 2 reduces to the one in the following corollary.

Corollary 2: The capacity of the state-dependent physically degraded RC with informed relay satisfies $C_{\text{D}} \leq R^{\text{up}}_{\text{D}}$, where

$$\begin{aligned} R^{\text{up}}_{\text{D}} \; = \; & \max\min\Big\{I(X_1;Y_2|S,X_2), \\ & I(X_1,X_2;Y_3|S) - I(X_1;S|Y_3)\Big\} \end{aligned} \tag{20}$$

with the maximization over all probability distributions of the form

$$P_{S,X_1,X_2,Y_2,Y_3} = Q_S P_{X_1} P_{X_2|X_1,S} W_{Y_2|X_1,X_2,S} W_{Y_3|Y_2,X_2,S}. \tag{21}$$

Similar to the general case in Theorem 2, the upper bound in Corollary 2 is tighter than the trivial upper bound in (18) for the degraded case.

IV. The Gaussian RC with Informed Relay

In this section, we consider a state-dependent Gaussian RC in which both the channel state and the noise are additive and Gaussian. We also assume that the additive channel state is non-causally known to only the relay. First, we consider full-duplex transmission at the relay, i.e., the relay transmits and receives at the same time, and we derive lower and upper bounds on channel capacity for this case. Then, we extend these results to the half-duplex mode in which the relay is constrained to operate in a time-division (TD) manner.

A. *Full-Duplex Channel Model*

For the full-duplex state-dependent Gaussian RC, the channel outputs $Y_{2,i}$ and $Y_{3,i}$ at time instant i for the relay and the destination, respectively, are related to the channel input $X_{1,i}$ from the source and $X_{2,i}$ from the relay, and the channel state S_i by

$$Y_{2,i} = X_{1,i} + S_i + Z_{2,i} \tag{22a}$$

$$Y_{3,i} = X_{1,i} + X_{2,i} + S_i + Z_{3,i}, \tag{22b}$$

where S_i is a zero mean Gaussian random variable with variance Q, $Z_{2,i}$ is zero mean Gaussian with variance N_2, and $Z_{3,i}$ is zero mean Gaussian with variance N_3. The random variables S_i, $Z_{2,i}$ and $Z_{3,i}$ at time instant $i \in \{1, 2, \ldots, n\}$ are mutually independent, and are independent of $(S_j, Z_{2,j}, Z_{3,j})$ for $j \neq i$. The random variables $Z_{2,i}$ and $Z_{3,i}$ are also independent of the channel inputs (X_1^n, X_2^n).

For the full-duplex degraded additive Gaussian RC, the channel outputs $Y_{2,i}$ and $Y_{3,i}$ for the relay and the destination, respectively, are related to the channel inputs $X_{1,i}$ and $X_{2,i}$ and the state S_i by

$$Y_{2,i} = X_{1,i} + S_i + Z_{2,i} \tag{23a}$$

$$Y_{3,i} = X_{2,i} + Y_{2,i} + Z'_{3,i}, \tag{23b}$$

where $(Z'_{3,1}, \cdots, Z'_{3,n})$ is a sequence of i.i.d. zero mean Gaussian random variables with variance $N'_3 = N_3 - N_2$ which is independent of Z_2^n.

The channel inputs from the source and the relay should satisfy the following average power constraints,

$$\sum_{i=1}^{n} X_{1,i}^2 \leq nP_1, \qquad \sum_{i=1}^{n} X_{2,i}^2 \leq nP_2. \tag{24}$$

As we indicated previously, we assume that the channel state S^n is non-causally known at only the relay. The definition of a code for this channel is the same as that given in Section II, with the additional constraint that the channel inputs should satisfy the power constraints (24).

1) Lower and Upper Bounds on Capacity: In this section, we derive lower and upper bounds on the capacity of the state-dependent full-duplex Gaussian RC with informed relay. The results obtained in Section IV for the DM case can be extended to memoryless channels with discrete time and continuous alphabets using standard techniques [49]. We use the bounds in Theorem 1 and Theorem 2 to compute bounds on the capacity for the Gaussian case.

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

Theorem 3: The capacity C_{G} of the state-dependent Gaussian RC with informed relay satisfies $C_{\mathrm{G}} \geq R_{\mathrm{G}}^{\mathrm{lo}}$, where

$$R_{\mathrm{G}}^{\mathrm{lo}} = \max_{\rho'_{12}} \min \Bigg\{ \frac{1}{2}\log(1+\frac{P_1(1-\rho'^2_{12})}{N_2}), \max_{\theta,\rho'_{2s}} \; \frac{1}{2}\log\left(1+\frac{P_1+\bar{\theta}P_2+2\rho'_{12}\sqrt{\bar{\theta}P_1P_2}}{\theta P_2+Q+N_3+2\rho'_{2s}\sqrt{\theta P_2 Q}}\right)+\frac{1}{2}\log(1+\frac{\theta P_2(1-\rho'^2_{2s})}{N_3}) \Bigg\}, \tag{25}$$

with the maximization over parameters $\rho'_{12} \in [0,1]$, $\theta \in [0,1]$, and $\rho'_{2s} \in [-1,0]$.

Proof: A formal proof of Theorem 3 is given in Appendix D.

Remarks:

- We compute the lower bound in (3) for an appropriate choice of the input distribution that will be specified in the sequel. By extension, Remark 1 also applies for the Gaussian case. More specifically, we should consider two important features in the design of an efficient coding scheme at the relay: obtaining correlation or coherence between the channel inputs from the source and the relay, and exploiting the channel state to remove the effect of the state on the communication. As we already mentioned, it is not obvious to accomplish these features because the channel state is not available at the source. Proceeding like for the code construction in the DM case, we split the relay input X_2^n into two parts, namely U_1^n and $\tilde{X}_2^n$. Furthermore, here we set U_1^n and $\tilde{X}_2^n$ to be *independent*. The first part, U_1^n, is a function of only the cooperative information, and is generated using standard coding. Since the source knows the cooperative information at the relay, it can generate its codeword X_1^n

in such a way that it is coherent with U_1^n, by allowing correlation between X_1^n and U_1^n. The second part, $\tilde{X}_2^n$, which is independent of the source input X_1^n, is a function of both the cooperative information and the channel state S^n at the relay, and is generated using a GDPC similar to that in [14], [16], [18].

- More formally, we decompose the relay input random variable X_2 as

$$X_2 = U_1 + \tilde{X}_2, \tag{26}$$

where: U_1 is zero mean Gaussian with variance $\bar{\theta}P_2$, is independent of both $\tilde{X}_2$ and S, and is correlated with X_1 with $\mathbb{E}[U_1X_1] = \rho'_{12}\sqrt{\bar{\theta}P_1P_2}$, for some $\theta \in [0,1]$, $\rho'_{12} \in [-1,1]$; and $\tilde{X}_2$ is zero mean Gaussian with variance θP_2, is independent of X_1, and is correlated with the channel state S with $\mathbb{E}[\tilde{X}_2S] = \rho'_{2s}\sqrt{\theta P_2 Q}$, for some $\rho'_{2s} \in [-1,1]$. Expressed in terms of the covariances $\sigma_{12} = \mathbb{E}[X_1X_2] = \mathbb{E}[X_1U_1]$ and $\sigma_{2s} = \mathbb{E}[X_2S] = \mathbb{E}[\tilde{X}_2S]$, the parameters ρ'_{12}, ρ'_{2s} are given by

$$\rho'_{12} = \frac{\sigma_{12}}{\sqrt{\bar{\theta}P_1P_2}}, \quad \rho'_{2s} = \frac{\sigma_{2s}}{\sqrt{\theta P_2 Q}}. \tag{27}$$

For the GDPC, we choose the auxiliary random variable U_2 as

$$U_2 = \tilde{X}_2 + \alpha_{\text{opt}} S \tag{28}$$

with

$$\alpha_{\text{opt}} = \frac{\theta P_2(1-\rho'^2_{2s}) - \rho'_{2s}\sqrt{\frac{\theta P_2}{Q}}N_3}{\theta P_2(1-\rho'^2_{2s}) + N_3}. \tag{29}$$

We now provide an upper bound on the capacity of the state-dependent general Gaussian RC with informed relay.

Theorem 4: The capacity C_{G} of the state-dependent general Gaussian RC with informed relay satisfies $C_{\text{G}} \leq R_{\text{G}}^{\text{up}}$, where

$$\begin{aligned} R_{\text{G}}^{\text{up}} &= \max\min\Bigg\{\frac{1}{2}\log\Big(1 + P_1(1 - \frac{\rho_{12}^2}{1-\rho_{2s}^2})(\frac{1}{N_2}+\frac{1}{N_3})\Big), \\ &\frac{1}{2}\log\Big(1 + \frac{(\sqrt{P_1}+\rho_{12}\sqrt{P_2})^2}{P_2(1-\rho_{12}^2-\rho_{2s}^2) + (\sqrt{Q}+\rho_{2s}\sqrt{P_2})^2 + N_3}\Big) + \frac{1}{2}\log(1 + \frac{P_2(1-\rho_{12}^2-\rho_{2s}^2)}{N_3})\Bigg\}, \end{aligned} \tag{30}$$

with the maximization over parameters $\rho_{12} \in [0,1]$ and $\rho_{2s} \in [-1,0]$ such that

$$\rho_{12}^2 + \rho_{2s}^2 \leq 1. \tag{31}$$

Proof: The proof of Theorem 4 is given in Appendix E. In the proof, we evaluate[1] the upper bound (16) using an appropriate joint distribution of S, X_1, X_2, Y_2, Y_3.

Following straightforwardly the proof of Theorem 4 in Appendix E, it can be easily shown that the capacity of the state-dependent degraded Gaussian RC is upper-bounded as in the following corollary.

Corollary 3: The capacity C_{DG} of the state-dependent degraded Gaussian RC with informed relay satisfies $C_{\text{DG}} \leq R^{\text{up}}_{\text{DG}}$, where

$$\begin{aligned} R^{\text{up}}_{\text{DG}} = &\max\min\Big\{\frac{1}{2}\log\Big(1+\frac{P_1(1-\rho_{12}^2-\rho_{2s}^2)}{N_2(1-\rho_{2s}^2)}\Big), \\ &\frac{1}{2}\log\Big(1+\frac{(\sqrt{P_1}+\rho_{12}\sqrt{P_2})^2}{P_2(1-\rho_{12}^2-\rho_{2s}^2)+(\sqrt{Q}+\rho_{2s}\sqrt{P_2})^2+N_3}\Big)+\frac{1}{2}\log(1+\frac{P_2(1-\rho_{12}^2-\rho_{2s}^2)}{N_3})\Big\}, \end{aligned} \tag{32}$$

with the maximization over parameters $\rho_{12} \in [0,1]$ and $\rho_{2s} \in [-1,0]$ such that

$$\rho_{12}^2 + \rho_{2s}^2 \leq 1. \tag{33}$$

2) Analysis of some Special Cases: We note that comparing the bounds in Theorem 3 and Theorem 4 analytically can be tedious in the general case. In what follows, we identify a few cases in which the lower bound and the upper bound meet, and so we obtain the capacity expression for these case.

In the following corollary we recast the lower bound (25) into an equivalent form by substituting $\varrho_{12} = \rho'_{12}\sqrt{\bar{\theta}}$ and $\varrho_{2s} = \rho'_{2s}\sqrt{\theta}$. Also, we recast the upper bound given in Corollary (3) into an equivalent form by substituting $\kappa = \rho_{12}/\sqrt{1-\rho_{2s}^2}$ and $\rho = \rho_{2s}$.

Corollary 4: For the Gaussian RC, the lower bound (25) in Theorem 3 can be written as

$$\begin{aligned} R^{\text{lo}}_{\text{G}} = &\max\min\Bigg\{\frac{1}{2}\log(1+\frac{P_1(1-\varrho_{12}^2-\theta)}{N_2(1-\theta)}), \\ &\frac{1}{2}\log\Big(1+\frac{(\sqrt{P_1}+\varrho_{12}\sqrt{P_2})^2+(\bar{\theta}-\varrho_{12}^2)P_2}{P_2(1-\bar{\theta}-\varrho_{2s}^2)+(\sqrt{Q}+\varrho_{2s}\sqrt{P_2})^2+N_3}\Big)+\frac{1}{2}\log(1+\frac{P_2(1-\bar{\theta}-\varrho_{2s}^2)}{N_3})\Bigg\}, \end{aligned} \tag{34}$$

[1]In Theorem 4, if the maximizing ρ_{2s} in (30) has absolute value equal to unity then (31) implies that ρ_{12} is zero. In this case, and also in the rest of this paper, we use the convention that $\frac{0}{0} = 0$.

where the maximization is over parameters $\theta \in [0,1]$, $\varrho_{12} \in [0,1]$, $\varrho_{2s} \in [-1,0]$ such that

$$\varrho_{12}^2 + \varrho_{2s}^2 \leq 1. \tag{35}$$

For the physically degraded case, the upper bound in Corollary 3 can be written as

$$\begin{aligned} R_{\text{DG}}^{\text{up}} = \max_{\kappa} \min \Bigg\{ & \frac{1}{2}\log\Big(1 + \frac{P_1(1-\kappa^2)}{N_2}\Big), \\ & \max_{\rho} \frac{1}{2}\log(1 + \frac{P_2(1-\kappa^2(1-\rho^2)-\rho^2)}{N_3}) \\ & + \frac{1}{2}\log\Big(1 + \frac{P_1 + \kappa^2(1-\rho^2)P_2 + 2\kappa\sqrt{1-\rho^2}\sqrt{P_1P_2}}{P_2(1-\kappa^2(1-\rho^2)) + Q + 2\rho\sqrt{P_2Q} + N_3}\Big)\Bigg\}, \end{aligned} \tag{36}$$

where the maximization is over parameters $\kappa \in [0,1]$ and $\rho \in [-1,0]$.

By investigating the bounds in Theorem 3 and Theorem 4, and the equivalent forms of these bounds in Corollary 4, it can be shown that the lower bound for the degraded case is tight for certain values of P_1, P_2, Q, N_2, N_3. The following observation states some cases for which the lower bound is tight.

Observation 1: For the physically degraded Gaussian RC, we have:

1) If P_1, P_2, Q, N_2, N_3 satisfy

$$N_2 \geq \max_{\zeta\in[-1,0]} \frac{P_1N_3(P_2 + Q + N_3 + 2\zeta\sqrt{P_2Q})}{P_1N_3 + P_2(1-\zeta^2)(P_1 + P_2 + Q + N_3 + 2\zeta\sqrt{P_2Q})}, \tag{37}$$

then channel capacity is given by

$$C_{\text{DG}} = \frac{1}{2}\log(1 + \frac{P_1}{N_2}), \tag{38}$$

which is the same as the interference-free capacity, i.e., the capacity if the channel state were not present in the model, or were also known to the source.

2) If the maximizing ρ_{12} and ρ_{2s} in the upper bound in Corollary 3 are such that condition (31) is met with equality, i.e., $\rho_{12}^2 + \rho_{2s}^2 = 1$, then the lower bound (34) is tight and gives the capacity.

Proof: The proof of observation 1 appears in Appendix F.

The condition in (37) specifies a range of values (P_1, P_2, Q, N_2, N_3) for which the lower bound for the degraded Gaussian case is tight. In this case, the capacity is the same as that of the degraded Gaussian RC with informed relay and informed source or interference-free capacity. Thus, the first statement in Observation 1 also provides an *sufficient* condition for the rate loss incurred by not knowing the interference at the source as well to be zero. At a high level, condition (37) means that there is no rate loss due to the asymmetry when capacity is constrained by the broadcast part in the model, i.e, transmission from the source to the relay and the destination. By investigating the upper bound (36) and comparing it with the interference-free capacity, it can be shown that this condition is also *necessary*. That is, the interference-free capacity is attained *only* if (37) is fulfilled. If the capacity of our model is constrained by the sum rate of the cooperative MAC part, i.e., the cooperative transmission from the source and the relay to the destination, the asymmetry resulting from not knowing the interference at the source as well causes an inevitable rate loss, i.e., the term $I(X_1;S|Y_3)$ in Corollary 2.

Extreme Cases

We now summarize the behavior of the above bounds in some extreme cases.

1. In the case of arbitrary strong channel state, i.e., $Q \to \infty$, capacity for degraded Gaussian case is given by

$$C_{\text{DG}} \;=\; \min\Big\{\frac{1}{2}\log(1+\frac{P_1}{N_2}),\; \frac{1}{2}\log(1+\frac{P_2}{N_3})\Big\}. \tag{39}$$

   In this case, the lower bound in (25) is maximized for $\theta = 1, \rho'_{2s} = 0, \rho'_{12} = 0$, and the upper bound in (36) is maximized for $\rho = 0, \kappa = 0$, and the two bounds meet. In this strong channel state case, we note that the direct transmission from the source to the destination is not possible because of the strong channel state and the transmission from the source to the destination only through the relay is possible. Also, (39) suggests that traditional multi-hop transmission achieves the capacity. A two-hop scheme allows to completely cancel the effect of the channel state by subtracting it out upon reception at the relay, and by applying standard DPC for transmission from the relay to the destination.

2. *Deaf helper problem:* In the case in which the relay is unable to *hear* the source (e.g., due to a very noisy or broken link source-to-relay) and $Q \to \infty$, the rate

$$C_{\text{G}} \;=\; \frac{1}{2}\log(1+\frac{P_1}{N_3}) \tag{40}$$

   can be achieved as follows. At time i, the source sends a Gaussian codeword $X_{1,i}$ which is independent of the state S_i. Independently, the relay generates its input $X_{2,i}$ using a *dummy* DPC as $X_2 = U_2 - S$, where $X_2 \sim \mathcal{N}(0, P_2)$ is independent of S and U_2 is Costa's auxiliary random variable. Upon reception of $Y_{3,i} = X_{1,i}+X_{2,i}+S_i+Z_{3,i}$ at the destination, the decoder first decodes the codeword $U_{2,i}$ fully, i.e., not only the bin index but also the correct sequence in the bin. This can be done reliably as long as $I(U_2;Y_3) - (U_2;S) > 0$. Then, the decoder at the destination subtracts out $U_{2,i}$ from $Y_{3,i}$ to obtain $\tilde{Y}_{3,i} = X_{1,i} + Z_{3,i}$ from which it decodes the source's message using standard decoding, at rate (40). A related scenario for a helper over a state-dependent Gaussian MAC is studied in [17].

3. For $Q = 0$, i.e., no channel state at all in the model, capacity for the degraded Gaussian case is given by

$$C_{\text{DG}} \;=\; \max_{0\leq\beta\leq 1} \min\left\{\frac{1}{2}\log\Big(1+\frac{P_1(1-\beta^2)}{N_2}\Big), \frac{1}{2}\log\Big(1+\frac{P_1+P_2+2\beta\sqrt{P_1P_2}}{N_3}\Big)\right\}, \quad (41)$$

which is the same as the capacity of the standard degraded Gaussian channel [42, Theorem 5]. This can be directly obtained by putting $Q = 0$ in (25) and (36). In this case, the maximizing parameters are $\theta = 0$, $\rho'_{2s} = 0$ for (25) and $\rho = 0$ for (36).

4. If $P_2 = 0$, capacity for the degraded Gaussian case, and capacity for the general Gaussian case if $Q + N_3 \geq N_2$, are given by

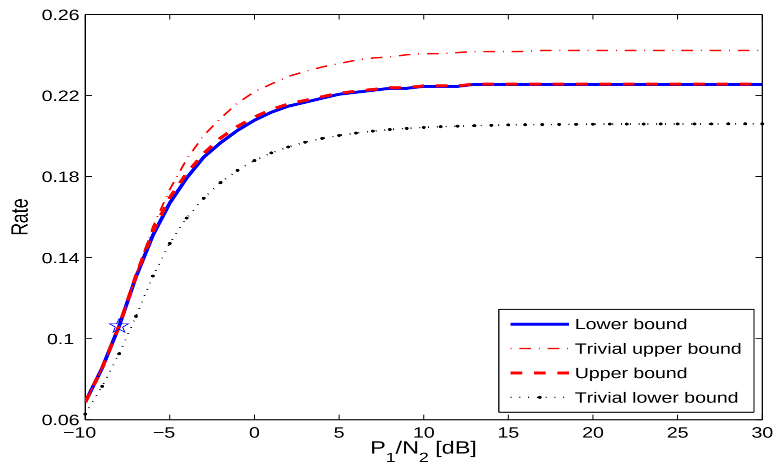
$$C_{\text{G}} = C_{\text{DG}} \;=\; \frac{1}{2}\log(1+\frac{P_1}{Q+N_3}). \quad (42)$$

In this case, the informed relay cannot help the source, and the interference is simply treated as an unknown noise.

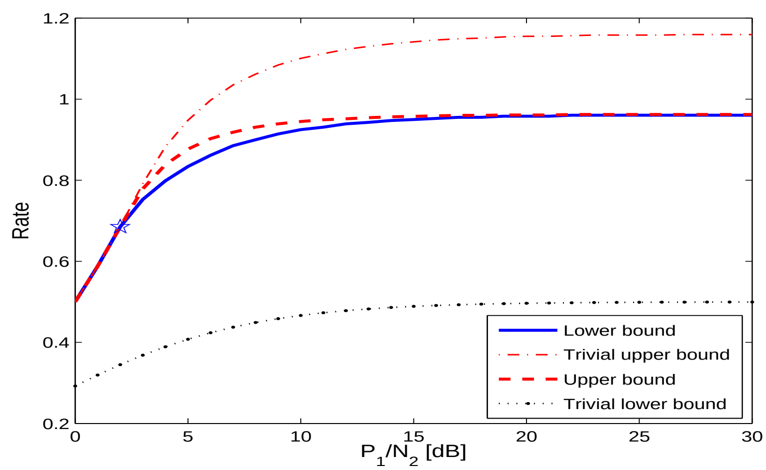
3) Numerical Examples and Discussion: In this section we discuss some numerical examples, for both the degraded Gaussian case and the general Gaussian case. We consider two numerical examples, a) $P_1 = P_2 = Q = 10$ dB, $N_3 = 20$ dB; and b)$P_1 = P_2 = Q = N_3 = 10$ dB.

Figure 4 illustrates the lower bound (34) and the upper bound (36) 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 trivial upper bound (18) computed for the degraded Gaussian case and the trivial lower bound obtained by considering the channel state as an unknown noise.

The curves show that the lower bound and the upper bound do not meet for all SNR regimes. However, as it is visible from the depicted numerical examples, the gap between the two bounds is small for the degraded case. Furthermore, the curves in Figure 4 also illustrate the results in observation 1, by showing that the lower bound and the upper bound meet for the cases stated in Observation 1. We note that the pentagram marker visible in Figure 4 indicates capacity when



(a)



(b)

Fig. 4. Lower and upper bounds on the capacity of the state-dependent degraded Gaussian RC with informed relay versus the SNR in the link source-to-relay, for two examples of numerical values (a) $P_1 = P_2 = Q = 10$ dB, $N_3 = 20$ dB, and (b) $P_1 = P_2 = Q = N_3 = 10$ dB.

the noise at the relay is equal to the RHS of (37); and this illustrates the first case for which the lower bound and the upper bound meet in Proposition 1. Also, Figure 5 depicts the variation of $\rho_{12}^2 + \rho_{2s}^2$, where ρ_{12} and ρ_{2s} are the maximizing for the upper bound, as a function of the SNR for the two numerical examples considered in Figure 4; and this illustrates the second case for which the lower and upper bounds meet in Proposition 1.

Figure 6 illustrates the lower bound (25) and the upper bound (30) as functions of the SNR

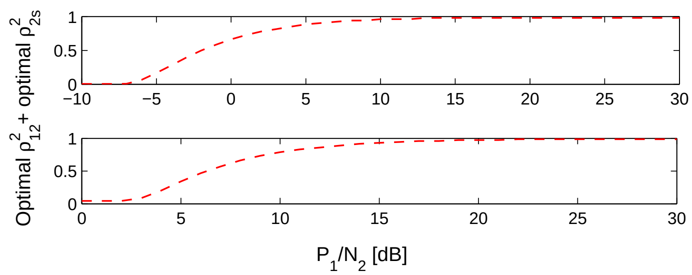


Fig. 5. The sum $\rho_{12}^2+\rho_{2s}^2$ in the constraint (31). Optimal ρ_{12} and ρ_{2s} are the maximizing for the upper bound for the numerical examples considered in Figure 4. The upper subfigure is for the upper bound curve in Figure 4(a), and the lower subfigure is for the upper bound curve in Figure 4(b).

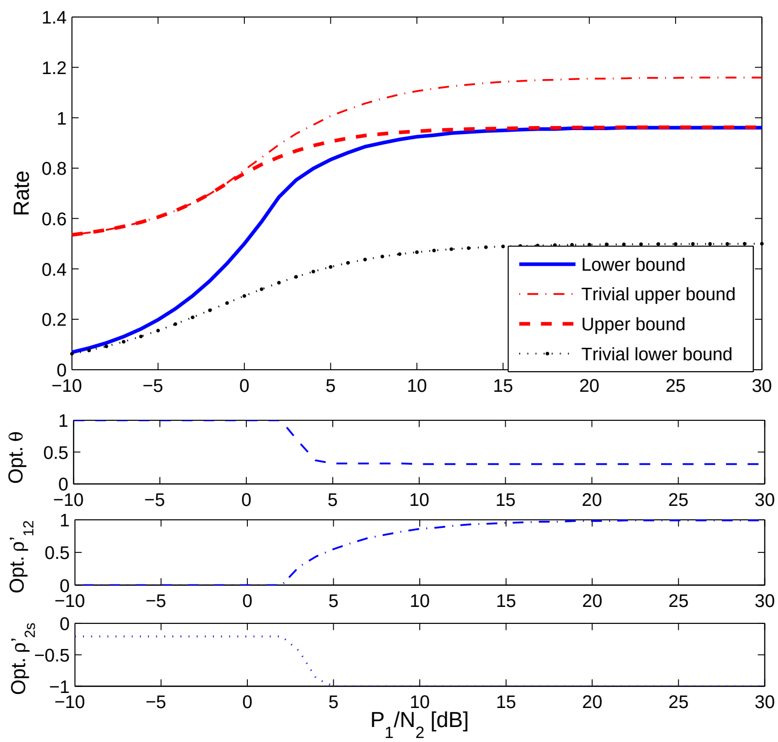


Fig. 6. Lower and upper bounds on the capacity of the state-dependent general Gaussian RC with informed relay and the maximizing $\theta, \rho_{12}', \rho_{2s}'$ in (25) as functions of the SNR at the relay. Numerical values are $P_1 = P_2 = Q = N_3 = 10$ dB.

at the relay for the general Gaussian channel. Also shown for comparison are the trivial upper bound (18) computed for the general Gaussian case and the trivial lower bound obtained by considering the channel state as an unknown noise. The curves show that the lower bound is close to the upper bound at large SNR, i.e., when capacity of the channel is determined by the sum rate of the MAC formed by transmission from the uninformed source and the informed relay to the destination. The gap between the lower bound and the upper bound which is visible at low SNR is due to the fact that DF relaying is not effective at this range of SNR and also to that our upper bounding technique is efficient on the MAC side, but not on the BC side of the relay channel.

Furthermore, Figure 6 also shows the variation of the maximizing θ, ρ'_{12}, ρ'_{2s} in (25) as function of the SNR at the relay. This shows how the informed relay allocates its power among combating the interference for the source (related to the value of ρ'_{2s}) and sending signals that are coherent with the transmission from the source (related to the values of θ and ρ'_{12}).

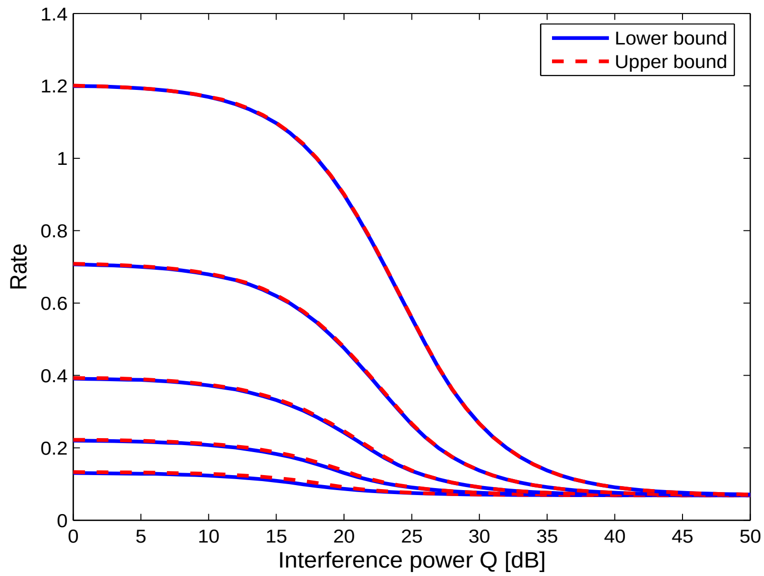
In Figure 7, the lower and upper bounds are plotted as function of the interference power Q, for fixed value of the power at the relay and several choices of the power at the source. The curves are depicted for two examples of noise configuration: $N_2 < N_3$ ($N_2 = 10$ dB and $N_3 = 20$ dB), and $N_2 = N_3 = 10$ dB. The curves illustrate the discussion in the above extreme cases analysis. For instance, for both noise configurations, that the rate achievable for very large values of Q is strictly positive illustrates that the transmission from the uninformed source to the uninformed destination is possible even in presence of an infinitely strong interference.

Furthermore, comparing the lower bound and the upper bound in the case $N_2 = N_3$, we observe that the lower bound is tight (only) for large values of Q. This corresponds to when the second term of the minimization in the upper bound (30) becomes operative, i.e., is smaller than the first term.

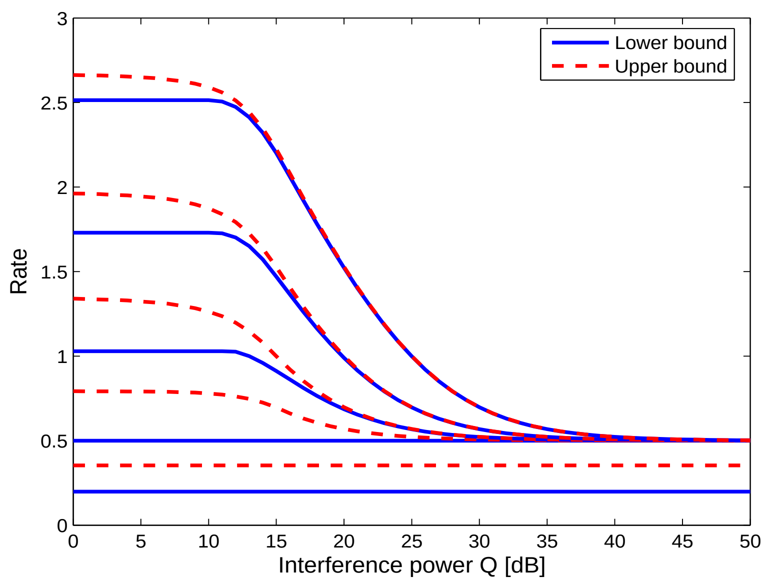
B. Half-Duplex Channel Model

In this section, we extend the results of Section IV-A to the case of half-duplex relaying, i.e., the relay can either transmit only or receive only. We consider a state-dependent Gaussian RC with informed relay, and we assume that the relay operates in a time-division (TD) relaying mode. In the TD mode, for a given time window, the relay is in the receive mode for a fraction of the given time and in the transmit mode for the remaining fraction of this time. Since the message from the source is transmitted to the destination in n channel uses, in the remaining of this section, we refer to the time indices from [2] 1 to $\lfloor \nu n \rfloor$ as the *relay-receive period* and the time indices from $\lfloor \nu n \rfloor + 1$ to n as the *relay-transmit period*, for some $\nu \in [0, 1]$. Furthermore, to generalize the model, we assume that the channel state $S^{(1)}$ is zero mean Gaussian with

[2] For a scalar x, $\lfloor x \rfloor$ stands for the largest integer small than or equal to x.



(a) $P_2 = N_2 = 10,\ N_3 = 20$ dB



(b) $P_2 = N_2 = N_3 = 10$ dB

Fig. 7. Bounds on channel capacity as function of the interference power Q. The curves correspond to different choices of power at the source: from bottom to top $P_1 = 5, 10, 15, 20, 25$ dB.

variance $Q^{(1)}$ during the relay-receive period, and the channel state $S^{(2)}$ is zero mean Gaussian with variance $Q^{(2)}$ during the relay-transmit period. The channel output $Y_{2,i}$ at time instant i at the relay is given by

$$Y_{2,i} = X_{1,i}^{(1)} + S_i^{(1)} + Z_{2,i},$$

during the relay-receive period, and is zero with probability one during the relay-transmit period. The channel output at time-instant i at the destination is given by

$$Y_{3,i}^{(1)} = X_{1,i}^{(1)} + S_i^{(1)} + Z_{3,i} \qquad \text{during the relay-receive period} \tag{43a}$$

$$Y_{3,i}^{(2)} = X_{1,i}^{(2)} + X_{2,i} + S_i^{(2)} + Z_{3,i} \qquad \text{during the relay-transmit period.} \tag{43b}$$

Furthermore, the source has average power constraint $P_1^{(1)}$ during the relay-receive period and average power constraint $P_1^{(2)}$ during the relay-transmit period ; the relay has average power constraint P_2.

For fixed values of ν, $P_1^{(1)}$, $P_1^{(2)}$ and P_2, we have the following upper and lower bounds on the capacity of the state-dependent half-duplex Gaussian RC with informed relay.

Proposition 1: The capacity of the state-dependent TD Gaussian RC with informed relay is upper-bounded by

$$R_{\text{G}}^{\text{up}}(\text{TD}) \;=\; \max\;\min\{R_1^{\text{up}},\; R_2^{\text{up}}\} \tag{44}$$

with

$$R_1^{\text{up}} = \frac{\nu}{2}\log\Big(1 + P_1^{(1)}(\frac{1}{N_2} + \frac{1}{N_3})\Big) + \frac{\bar{\nu}}{2}\log\Big(1 + \frac{P_1^{(2)}(1-\rho_{12}^2-\rho_{2s}^2)}{N_3(1-\rho_{2s}^2)}\Big), \tag{45a}$$

$$R_2^{\text{up}} = \bar{\nu}\Psi(P_1^{(2)}, P_2, Q^{(2)}, \rho_{12}, \rho_{2s}) + \frac{\nu}{2}\log\Big(1 + \frac{P_1^{(1)}}{N_3 + Q^{(1)}}\Big), \tag{45b}$$

where $\Psi(P_1, P_2, Q, \rho_{12}, \rho_{2s})$ is defined as the second term of the minimization in (30), and the maximization is over parameters $\rho_{12} \in [0, 1]$ and $\rho_{2s} \in [-1, 0]$ such that $\rho_{12}^2 + \rho_{2s}^2 \leq 1$.

Proposition 2: The capacity of the state-dependent TD Gaussian RC with informed relay is lower-bounded by

$$R_{\text{G}}^{\text{lo}}(\text{TD}) \;=\; \max\;\min\{R_1^{\text{lo}},\; R_2^{\text{lo}}\} \tag{46}$$

with

$$R_1^{\text{lo}} = \frac{\nu}{2}\log\Big(1+\frac{P_1^{(1)}}{N_2}\Big)+\frac{\bar{\nu}}{2}\log\Big(1+\frac{(1-\rho_{12}'^2)P_1^{(2)}}{N_3+\Phi(\alpha',\theta,\rho_{2s}')}\Big) \tag{47a}$$

$$\begin{aligned} R_2^{\text{lo}} &= \frac{\nu}{2}\log\Big(1+\frac{P_1^{(1)}}{N_3+Q^{(1)}}\Big) \\ &+\frac{\bar{\nu}}{2}\log\Big(1+\frac{P_1^{(2)}+\bar{\theta}P_2+2\rho_{12}'\sqrt{\bar{\theta}P_1^{(2)}P_2}}{\theta P_2+Q^{(2)}+2\rho_{2s}'\sqrt{\theta P_2 Q^{(2)}}+N_3}\Big) \\ &+\frac{\bar{\nu}}{2}\log\Big(\frac{P_2'(P_2'+Q^{(2)}+N_3)}{P_2'Q^{(2)}(1-\alpha')^2+N_3(P_2'+\alpha'^2Q^{(2)})}\Big), \end{aligned} \tag{47b}$$

where, maximization is over parameters $\theta \in [0,1]$, $\rho_{12}' \in [0,1]$, $\rho_{2s}' \in [-1,0]$ and $\alpha' \in \mathcal{A}(\rho_{12}',\theta,\rho_{2s}') := \{t \in \mathbb{R} \,:\, \Theta(t,\rho_{12}',\theta,\rho_{2s}') > 0\}$;

$$\Phi(t,\theta,\rho_{2s}') := \frac{(1-t)^2\theta P_2 Q^{(2)}(1-\rho_{2s}'^2)}{\theta P_2+2t\rho_{2s}'\sqrt{\theta P_2 Q^{(2)}}+t^2Q^{(2)}} \tag{48}$$

$$\begin{aligned} \Theta(t,\rho_{12}',\theta,\rho_{2s}') &:= \frac{1}{2}\log\Big(\frac{P_1^{(2)}(1-\rho_{12}'^2)+\theta P_2+Q^{(2)}+N_3}{\theta P_2+Q^{(2)}+2\rho_{2s}'\sqrt{\theta P_2Q^{(2)}}+N_3}\Big) \\ &+\frac{1}{2}\log\Big(\frac{P_2'(P_2'+Q^{(2)}+N_3)}{P_2'Q^{(2)}(1-t)^2+N_3(P_2'+t^2Q^{(2)})}\Big) \\ &-\frac{1}{2}\log\Big(1+\frac{(1-\rho_{12}'^2)P_1^{(2)}}{N_3+\Phi(t,\theta,\rho_{2s}')}\Big). \end{aligned} \tag{49}$$

and $P_2' := \theta P_2(1-\rho_{2s}'^2)$.

The proofs of Proposition 1 and Proposition 2 appear in Appendix G.

V. Conclusion

In this paper, we consider a state-dependent relay channel with the channel state available non-causally at only the relay, i.e., neither at the source nor at the destination. We refer to this communication model as *state-dependent RC with informed relay*. This setup may model the basic building block for node cooperation over wireless networks in which some of the terminals may be equipped with cognition capabilities that permit to know, i.e., estimate to high accuracy, the states of the channel.

We investigate this problem in the discrete memoryless (DM) case and in the Gaussian case, and we derive bounds on the channel capacity. For both cases, the upper bounds are tighter than those obtained by assuming that the channel state is also available at the source and the destination, and they characterize the rate loss due to the asymmetry, i.e., having the channel state available at the relay but not the source. The lower bounds are obtained by a coding scheme that splits the codeword at the informed relay into two parts: one part depends only on the cooperative information, not on the known channel state, and is used to enable coherent transmission from the source and the relay to the destination; another part is a function of both the cooperative information and the known channel state, and is used to combat the effects of the channel state on the communication through a generalized Gel'fand-Pinsker binning scheme. In the Gaussian case, we consider average power constraints at the source and the relay and power allocation at the relay among the two parts of the code, allowing for a tradeoff between the coherence gain obtained through the coherent transmission and the mitigation of the channel state.

Furthermore, we also derive bounds on the capacity for the case in which the channel is physically degraded. For the degraded Gaussian case, we show that the lower bound meets with the upper bound for some special cases, for which we obtain the expression of channel capacity. Also, for the general Gaussian case, we extend the results to the case in which the relay operates in a half-duplex mode.

Finally, we note that some of the concepts developed in this paper can be applied to a state-dependent multiple access channel (MAC) with degraded message sets in which the uninformed encoder knows the message to be sent by the informed encoder [50].

APPENDIX

Throughout this section we denote the set of strongly jointly ϵ-typical sequences [51, Chapter 14.2] with respect to the distribution $P_{X,Y}$ as $T_\epsilon^n(P_{X,Y})$.

A. *Proof of Theorem 1*

Consider the random coding scheme that we outlined in Section III. We now give a formal description of the coding scheme and analyse the average probability of error.

As we outlined after Theorem 1 we transmit in $B+1$ 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 fixed n, the average rate $R\frac{B}{B+1}$ over $B+1$ blocks approaches R as $B \longrightarrow +\infty$.

Encoding: Let w_i be the new message to be sent from the source node at the beginning of block i, and w_{i-1} be the message sent in the previous block $i-1$. At the beginning of block i, the relay has decoded the message w_{i-1} correctly, and the source sends $\mathbf{x}_1(w_{i-1}, w_i)$. The relay searches for the smallest $j \in \{1, \cdots, J\}$ such that $\mathbf{u}_1(w_{i-1})$, $\mathbf{u}_2(w_{i-1}, j)$ and $\mathbf{s}[i]$ are jointly typical. Denote this j by $j^\star = j(\mathbf{s}[i], w_{i-1})$. If such $j^\star$ is not found, or if the observed state is not typical, an error is declared and $j^\star$ is set to J. Then, the relay transmits a vector $\mathbf{x}_2(w_{i-1})$ with i.i.d. components given $(\mathbf{u}_1(w_{i-1}), \mathbf{u}_2(w_{i-1}, j^\star), \mathbf{s}[i])$ drawn according to the marginal $P_{X_2|U_1,U_2,S}$ induced by the distribution (4). For convenience, we list the codewords at the source and the relay that are used for transmission in the first four blocks in Figure 8.

	block 1	block 2	block 3	block 4
Source codewords	$\mathbf{x}_1(1, w_1)$	$\mathbf{x}_1(w_1, w_2)$	$\mathbf{x}_1(w_2, w_3)$	$\mathbf{x}_1(w_3, 1)$
Relay codewords	$\mathbf{u}_1(1)$	$\mathbf{u}_1(w_1)$	$\mathbf{u}_1(w_2)$	$\mathbf{u}_1(w_3)$
	$\mathbf{u}_2(1, j(\mathbf{s}[1], 1))$	$\mathbf{u}_2(w_1, j(\mathbf{s}[2], w_1))$	$\mathbf{u}_2(w_2, j(\mathbf{s}[3], w_2))$	$\mathbf{u}_2(w_3, j(\mathbf{s}[4], w_3))$
	$\mathbf{x}_2(1)$	$\mathbf{x}_2(w_1)$	$\mathbf{x}_2(w_2)$	$\mathbf{x}_2(w_3)$

Fig. 8. Regular encoding for DF for the state-dependent RC with informed relay. At the beginning of block i, the source transmits $\mathbf{x}_1(w_{i-1}, w_i)$ and the relay transmits a codeword $\mathbf{x}_2(w_{i-1})$ with i.i.d. components given $(\mathbf{u}_1(w_{i-1}), \mathbf{u}_2(w_{i-1}, j(\mathbf{s}[i], w_{i-1})), \mathbf{s}[i])$ drawn according to the marginal $P_{X_2|U_1,U_2,S}$.

Decoding: The decoding procedure at the relay is based on joint typicality. The decoding procedure at the destination is based on a combination of joint typicality and backward-decoding.

1. At the end of block i, the relay knows w_{i-1} and declares that $\hat{w}_i$ is sent if there is a unique $\hat{w}_i$ such that $\mathbf{x}_1(w_{i-1}, \hat{w}_i)$, $\mathbf{u}_1(w_{i-1})$, $\mathbf{y}_2[i]$ and $\mathbf{s}[i]$ are jointly typical, where $\mathbf{y}_2[i]$ denotes the output of the channel at the relay in block i. One can show that the decoding error in this step is small for sufficiently large n if

$$R < I(X_1; Y_2|S, U_1). \tag{A-1}$$

2. At the end of the transmission, the destination has collected all the blocks of channel outputs $\mathbf{y}_3[1], \mathbf{y}_3[2], \ldots, \mathbf{y}_3[B+1]$, and can then perform Willem's backward-decoding by first decoding w_B from $\mathbf{y}_3[B+1]$.

   First, the destination declares that $\hat{w}_B$ is sent if there is a unique $\hat{w}_B$ such that $\mathbf{u}_1(\hat{w}_B)$, $\mathbf{u}_2(\hat{w}_B, j_B)$, $\mathbf{x}_1(\hat{w}_B, 1)$, $\mathbf{y}_3[B+1]$ are jointly typical, for some $j_B \in \{1, \ldots, J\}$. One can show that the decoding error in this step is small for sufficiently large n if

$$R < I(X_1, U_1, U_2; Y_3) - I(U_2; S|U_1). \tag{A-2}$$

Next, for b ranging from B to 2, the destination knows w_b and decodes w_{b-1} based on the information received in block b. It declares that $\hat{w}_{b-1}$ is sent if there is a unique $\hat{w}_{b-1}$ such that $\mathbf{u}_1(\hat{w}_{b-1})$, $\mathbf{u}_2(\hat{w}_{b-1}, j_{b-1})$, $\mathbf{x}_1(\hat{w}_{b-1}, w_b)$, $\mathbf{y}_3[b]$ are jointly typical, for some $j_{b-1} \in \{1, \dots, J\}$. One can show that the decoding error in this step is small for sufficiently large n if (A-2) is true.

Analysis of Probability of Error:

Fix a probability distribution $P_{S,U_1,U_2,X_1,X_2,Y_2,Y_3}$ satisfying (4). Let $\mathbf{s}[i]$, w_{i-1} and w_i be the state sequence in block i, the message sent from the source node in block $i-1$ and the message sent in block i, respectively. As we already mentioned above, at the beginning of block i the source transmits $\mathbf{x}_1(w_{i-1}, w_i)$ and the relay transmits a vector $\mathbf{x}_2(w_{i-1})$ with i.i.d. components conditionally given $(\mathbf{u}_1(w_{i-1}), \mathbf{u}_2(w_{i-1}, j^\star), \mathbf{s}[i])$, with $j^\star = j(\mathbf{s}[i], w_{i-1})$, drawn according to the marginal $P_{X_2|U_1,U_2,S}$.

The average probability of error is such that

$$\begin{aligned} \text{Pr(Error)} \leq & \sum_{(\mathbf{s},\mathbf{u}_1)\notin T_\epsilon^n(Q_S P_{U_1})} \text{Pr}(\mathbf{s})\text{Pr}(\mathbf{u}_1) \\ & + \sum_{(\mathbf{s},\mathbf{u}_1)\in T_\epsilon^n(Q_S P_{U_1})} \text{Pr}(\mathbf{s})\text{Pr}(\mathbf{u}_1)\text{Pr}(\text{error}|\mathbf{s}, \mathbf{u}_1). \end{aligned} \tag{A-3}$$

The first term, $\text{Pr}((\mathbf{s}, \mathbf{u}_1) \notin T_\epsilon^n(Q_S P_{U_1}))$, on the RHS of (A-3) goes to zero as $n \to \infty$, by the strong asymptotic equipartition property (AEP) [51]. Thus, it is sufficient to upper bound the second term on the RHS of (A-3).

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 event E_{1i} corresponds to the encoding step at block i; the events E_{2i} and E_{3i} correspond to decoding at the relay at block i; the events E_{4B} and E_{5B} correspond to decoding at the destination at block $B+1$, and for b ranging from B to 2, the events $E_{6(b-1)}$ and $E_{7(b-1)}$ correspond to decoding at the destination at block b.

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

$\mathbf{u}_1(w_{i-1})$, i.e.,

$$E_{1i} = \Big\{ \nexists\, j \in \{1, \ldots, J\} \text{ s.t. } \Big(\mathbf{u}_1(w_{i-1}), \mathbf{u}_2(w_{i-1}, j), \mathbf{s}[i]\Big) \in T_\epsilon^n(P_{U_1,U_2,S}) \Big\}.$$

To bound the probability of the event E_{1i}, we use a standard argument [2]. More specifically, for $\mathbf{u}_2(w_{i-1}, j)$ and $\mathbf{s}[i]$ generated independently given $\mathbf{u}_1(w_{i-1})$, with i.i.d. components drawn according to $P_{U_2|U_1}$ and Q_S, respectively, the probability that $\mathbf{u}_2(w_{i-1}, j)$ is jointly typical with $\mathbf{s}[i]$ given $\mathbf{u}_1(w_{i-1})$ is greater than $(1-\epsilon)2^{-n(I(U_2;S|U_1)+\epsilon)}$ for sufficiently large n. There is a total of J such $\mathbf{u}_2$'s in each bin. The probability of the event E_{1i}, the probability that there is no such $\mathbf{u}_2$, is therefore bounded as

$$\text{Pr}(E_{1i}) \leq [1 - (1-\epsilon)2^{-n(I(U_2;S|U_1)+\epsilon)}]^J. \tag{A-4}$$

Taking the logarithm on both sides of (A-4) and substituting J using (6) we obtain that $\ln(\text{Pr}(E_{1i})) \leq -(1-\epsilon)2^{n\epsilon}$. Thus, $\text{Pr}(E_{1i}) \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty$.

- Let E_{2i} be the event that $\mathbf{x}_1(w_{i-1}, w_i)$, $\mathbf{y}_2[i]$, $\mathbf{s}[i]$ are not jointly typical given $\mathbf{u}_1(w_{i-1})$, i.e.,

$$E_{2i} = \Big\{ \Big(\mathbf{u}_1(w_{i-1}), \mathbf{x}_1(w_{i-1}, w_i), \mathbf{y}_2[i], \mathbf{s}[i]\Big) \notin T_\epsilon^n(P_{U_1,X_1,Y_2,S}) \Big\}.$$

Conditioned on E_{1i}^c, the event complement of E_{1i}, we have that $(\mathbf{s}[i], \mathbf{u}_1(w_{i-1}))$ is jointly typical with $\mathbf{u}_2(w_{i-1}, j(\mathbf{s}[i], w_{i-1}))$ and with the source input $\mathbf{x}_1(w_{i-1}, w_i)$ and the relay input $\mathbf{x}_2(w_{i-1})$, i.e.,

$$\Big(\mathbf{s}[i], \mathbf{u}_1(w_{i-1}), \mathbf{u}_2(w_{i-1}, j(\mathbf{s}[i], w_{i-1})), \mathbf{x}_1(w_{i-1}, w_i), \mathbf{x}_2(w_{i-1})\Big) \in T_\epsilon^n(Q_S P_{U_1} P_{X_1|U_1} P_{U_2,X_2|S,U_1,X_1}). \tag{A-5}$$

For $\mathbf{s}[i]$, $\mathbf{u}_1(w_{i-1})$, $\mathbf{x}_1(w_{i-1}, w_i)$ jointly typical, we have $\text{Pr}(E_{2i}|E_{1i}^c) \longrightarrow 0$ as $n \longrightarrow \infty$, by the Markov Lemma [51].

- Let E_{3i} be the event that $\mathbf{x}_1(w_{i-1}, w_i')$, $\mathbf{y}_2[i]$, $\mathbf{s}[i]$ are jointly typical given $\mathbf{u}_1(w_{i-1})$ for some $w_i' \neq w_i$, i.e.,

$$E_{3i} = \Big\{ \exists\, w_i' \in \{1, \ldots, M\} \text{ s.t. } w_i' \neq w_i, \Big(\mathbf{u}_1(w_{i-1}), \mathbf{x}_1(w_{i-1}, w_i'), \mathbf{y}_2[i], \mathbf{s}[i]\Big) \in T_\epsilon^n(P_{U_1,X_1,Y_2,S}) \Big\}.$$

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

$$\text{Pr}(E_{3i}|E_{1i}^c, E_{2i}^c) \leq M 2^{-n(I(X_1;Y_2,S|U_1)-\epsilon)} \tag{A-6a}$$

$$= 2^{-n(I(X_1;Y_2|S,U_1)-R+3\epsilon)}, \tag{A-6b}$$

where in (A-6b) we used the fact that $I(X_1;S|U_1)=0$ under the joint distribution (4). Thus, Pr$(E_{3i}|E_{1i}^c,E_{2i}^c)\longrightarrow 0$ as $n\longrightarrow\infty$ if $R<I(X_1;Y_2|S,U_1)$.

- For the decoding of message w_B at the destination, let E_{4B} be the event that $\mathbf{u}_1(w_B)$, $\mathbf{u}_2(w_B,j(\mathbf{s}[B+1],w_B))$, $\mathbf{x}_1(w_B,1)$, $\mathbf{y}_3[B+1]$ are not jointly typical, i.e.,

$$E_{4B}=\Big\{\Big(\mathbf{u}_1(w_B),\mathbf{u}_2(w_B,j(\mathbf{s}[B+1],w_B)),\mathbf{x}_1(w_B,1),\mathbf{y}_3[B+1]\Big)\notin T_\epsilon^n(P_{U_1,U_2,X_1,Y_3})\Big\}.$$

For $\mathbf{s}[B+1]$, $\mathbf{u}_1(w_B)$, $\mathbf{u}_2(w_B,j(\mathbf{s}[B+1],w_B))$, $\mathbf{x}_1(w_B,1)$ and $\mathbf{x}_2(w_B)$ jointly typical as shown by (A-5), Pr$(E_{4B}|E_{1i}^c,E_{2i}^c,E_{3i}^c)\longrightarrow 0$ as $n\longrightarrow\infty$, by the Markov Lemma.

- For the decoding of message w_B at the destination, let E_{5B} be the event that $\mathbf{u}_1(w'_B)$, $\mathbf{u}_2(w'_B,j'_B)$, $\mathbf{x}_1(w'_B,1)$, $\mathbf{y}_3[B+1]$ are jointly typical for some $w'_B\neq w_B$ and some $j'_B\in\{1,\dots,J\}$, i.e.,

$$E_{5B}=\Big\{\exists\, w'_B\in\{1,\dots,M\},j'_B\in\{1,\dots,J\}\text{ s.t. }w'_B\neq w_B,\\ \Big(\mathbf{u}_1(w'_B),\mathbf{u}_2(w'_B,j'_B),\mathbf{x}_1(w'_B,1),\mathbf{y}_3[B+1]\Big)\in T_\epsilon^n(P_{U_1,U_2,X_1,Y_3})\Big\}.$$

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

$$\text{Pr}(E_{5B}|E_{1i}^c,E_{2i}^c,E_{3i}^c,E_{4B}^c)\leq MJ2^{-n(I(X_1,U_1,U_2;Y_3)-\epsilon)} \quad \text{(A-7a)}$$

$$=2^{-n(I(X_1,U_1,U_2;Y_3)-I(U_2;S|U_1)-R+\epsilon)}. \quad \text{(A-7b)}$$

Thus Pr$(E_{5B}|\cap_{k=1}^3 E_{ki}^c,E_{4B}^c)\longrightarrow 0$ as $n\longrightarrow\infty$ if $R<I(X_1,U_1,U_2;Y_3)-I(U_2;S|U_1)$.

- For the decoding of message w_{b-1} at the destination, $b=B,\dots,2$, let $E_{6(b-1)}$ be the event that $\mathbf{u}_1(w_{b-1})$, $\mathbf{u}_2(w_{b-1},j(\mathbf{s}[b],w_{b-1}))$, $\mathbf{x}_1(w_{b-1},w_b)$, $\mathbf{y}_3[b]$ are not jointly typical, i.e.,

$$E_{6(b-1)}=\Big\{\Big(\mathbf{u}_1(w_{b-1}),\mathbf{u}_2(w_{b-1},j(\mathbf{s}[b],w_{b-1})),\mathbf{x}_1(w_{b-1},w_b),\mathbf{y}_3[b]\Big)\notin T_\epsilon^n(P_{U_1,U_2,X_1,Y_3})\Big\}.$$

For $\mathbf{s}[b]$, $\mathbf{u}_1(w_{b-1})$, $\mathbf{u}_2(w_{b-1},j(\mathbf{s}[b],w_{b-1}))$, $\mathbf{x}_1(w_{b-1},w_b)$ and $\mathbf{x}_2(w_{b-1})$ jointly typical as shown by (A-5), Pr$(E_{6(b-1)}|\cap_{k=1}^3E_{ki}^c,E_{4B}^c,E_{5B}^c)\longrightarrow 0$ as $n\longrightarrow\infty$, by the Markov Lemma.

- For the decoding of message w_{b-1} at the destination, let $E_{7(b-1)}$ be the event that $\mathbf{u}_1(w'_{b-1})$, $\mathbf{u}_2(w'_{b-1},j'_{b-1})$, $\mathbf{x}_1(w'_{b-1},w_b)$, $\mathbf{y}_3[b]$ are jointly typical for some $w'_{b-1}\neq w_{b-1}$ and some $j'_{b-1}\in\{1,\dots,J\}$, i.e.,

$$E_{7(b-1)}=\Big\{\exists\, w'_{b-1}\in\{1,\dots,M\},j'_{b-1}\in\{1,\dots,J\},\text{ s.t. }w'_{b-1}\neq w_{b-1},\\ \Big(\mathbf{u}_1(w'_{b-1}),\mathbf{u}_2(w'_{b-1},j'_{b-1}),\mathbf{x}_1(w'_{b-1},w_b),\mathbf{y}_3[b]\Big)\in T_\epsilon^n(P_{U_1,U_2,X_1,Y_3})\Big\}.$$

Proceeding like for the event E_{5B}, one can easily show that $\Pr(E_{7(b-1)}|\cap_{k=1}^{3}E_{ki}^c, E_{4B}^c, E_{5B}^c, E_{6(b-1)}^c)$ can be bounded similarly to in (A-7), and this shows that $\Pr(E_{7(b-1)}|\cap_{k=1}^{3}E_{ki}^c, E_{4B}^c, E_{5B}^c, E_{6(b-1)}^c)$ $\longrightarrow 0$ as $n \longrightarrow \infty$ if $R < I(X_1, U_1, U_2; Y_3) - I(U_2; S|U_1)$.

It remains to show that the rate (3) is not altered if one restricts the random variables U_1 and U_2 to have their alphabet sizes limited as indicated in (5). This is done by invoking the support lemma [52, p. 310]. Fix a distribution μ of $(S, U_1, U_2, X_1, X_2, Y_2, Y_3)$ on $\mathcal{P}(\mathcal{S}\times\mathcal{U}_1\times\mathcal{U}_2\times\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{Y}_2\times\mathcal{Y}_3)$ that has the form (4).

To prove the bound (5a) on $|\mathcal{U}_1|$, note that we have

$$I_\mu(X_1; Y_2|S, U_1) = I_\mu(X_1; Y_2, S|U_1) \tag{A-8a}$$

$$= H_\mu(X_1|U_1) + H_\mu(Y_2, S|U_1) - H_\mu(X_1, Y_2, S|U_1), \tag{A-8b}$$

where (A-8a) follows since $S \leftrightarrow U_1 \leftrightarrow X_1$ under the distribution μ. Also, we have

$$I_\mu(X_1, U_1, U_2; Y_3) - I_\mu(U_2; S|U_1) = I_\mu(U_1; Y_3) + I_\mu(X_1, U_2; Y_3|U_1) - I_\mu(U_2; S|U_1) \tag{A-9a}$$

$$= H_\mu(Y_3) - H_\mu(S) - H_\mu(U_2|U_1) + H_\mu(U_2, S|U_1)$$
$$+ H_\mu(X_1, U_2|U_1) - H_\mu(X_1, U_2, Y_3|U_1). \tag{A-9b}$$

Hence, it suffices to show that the following functionals of $\mu(S, U_1, U_2, X_1, X_2, Y_2, Y_3)$

$$r_{s,x,x'}(\mu) = \mu(s, x, x') \quad \forall\, (s, x, x') \in \mathcal{S}\times\mathcal{X}_1\times\mathcal{X}_2 \tag{A-10a}$$

$$r_1(\mu) = \int_u d_\mu(u)[H_\mu(X_1|u) + H_\mu(Y_2, S|u) - H_\mu(X_1, Y_2, S|u)] \tag{A-10b}$$

$$r_2(\mu) = \int_u d_\mu(u)[H_\mu(X_1, U_2|u) - H_\mu(X_1, U_2, Y_3|u) - H_\mu(U_2|u) + H_\mu(U_2, S|u)], \tag{A-10c}$$

can be preserved with another measure μ' that has the form (4). Observing that there is a total of $|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2| + 1$ functionals in (A-10), 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 (5a) without altering the rate (3).

Once the alphabet of U_1 is fixed, we apply similar arguments to bound the alphabet of U_2, where this time $|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2|(|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2|+1)-1$ functionals must be satisfied in order to preserve the joint distribution of S, U_1, X_1, X_2, and one more functional to preserve

$$I_\mu(X_1, U_1, U_2; Y_3) - I_\mu(U_2; S|U_1) = H_\mu(Y_3) - H_\mu(S) - H_\mu(U_1|U_2) + H_\mu(U_1, S|U_2)$$
$$+ H_\mu(X_1, U_1|U_2) - H_\mu(X_1, U_1, Y_3|U_2), \tag{A-11}$$

yielding the bound indicated in (5b).

B. *Proof of Corollary 1*

The proof combines rate-splitting [53] and the techniques used in the proof of Theorem 1. As we already mentioned in the discussion following Corollary 1, we split the message W to be transmitted from the source node into two independent parts $W = W_r + W_d$; the relay forwards only the part W_r, at rate R_r, and the part W_d is sent directly to the destination, at rate R_d. The total rate is $R = R_r + R_d$. We transmit in $B+1$ blocks, each of length n. During each of the first B blocks, the source sends a message $w_i = w_{r,i} + w_{d,i}$, with $w_{r,i} \in [1, 2^{nR_r}]$ and $w_{d,i} \in [1, 2^{nR_d}]$ and $i = 1, \ldots, B$ denotes the index of the block. For convenience, we let $w_{r,B+1} = w_{d,1} = 1$. 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 $P_{S,U_1,U_2,U,X_1,X_2,Y_2,Y_3}$ satisfying (12). Fix $\epsilon > 0$ and let

$$J = 2^{n(I(U_2;S|U_1)+2\epsilon)} \tag{B-12a}$$

$$M_r = 2^{n(R_r-2\epsilon)} \tag{B-12b}$$

$$M_d = 2^{n(R_d-4\epsilon)}. \tag{B-12c}$$

1. We generate M_r i.i.d. codewords $\{\mathbf{u}_1(w'_r)\}$ indexed by $w'_r = 1, \ldots, M$, each with i.i.d. components drawn according to P_{U_1}. For each $\mathbf{u}_1(w'_r)$, we generate M_r i.i.d. codewords $\{\mathbf{u}(w'_r, w_r)\}$ at the source indexed by $w_r = 1, \ldots, M_r$, and J auxiliary codewords $\{\mathbf{u}_2(w'_r, j)\}$ at the relay indexed by $j = 1, \ldots, J$. The codewords $\mathbf{u}(w'_r, w_r)$ and $\mathbf{u}_2(w'_r, j)$ are with i.i.d. components given $\mathbf{u}_1(w'_r)$ drawn according to $P_{U|U_1}$ and $P_{U_2|U_1}$, respectively.
2. For each $\mathbf{u}_1(w'_r)$, for each $\mathbf{u}(w'_r, w_r)$, we generate M_d i.i.d. codewords $\{\mathbf{x}_1(w'_r, w_r, w_d)\}$ indexed by $w_d = 1, \ldots, M_d$, each with i.i.d. components given $(\mathbf{u}_1(w'_r), \mathbf{u}(w'_r, w_r))$ drawn according to $P_{X_1|U_1,U}$.

Encoding: At the beginning of block i, let $w_i = w_{r,i} + w_{d,i}$ be the new message to be sent from the source and $w_{i-1} = w_{r,i-1} + w_{d,i-1}$ be the message sent in the previous block $i-1$.

At the beginning of block i, the relay has decoded $w_{r,i-1}$ correctly, and the source transmits $\mathbf{x}_1(w_{r,i-1}, w_{r,i}, w_{d,i})$. The relay searches for the smallest $j \in \{1, \cdots, J\}$ such that $\mathbf{u}_2(w_{r,i-1}, j)$ and $\mathbf{s}[i]$ are jointly typical given $\mathbf{u}_1(w_{r,i-1})$. Since the vectors $\mathbf{u}_2(w_{r,i-1}, j)$ and $\mathbf{s}[i]$ are generated independently given $\mathbf{u}_1(w_{r,i-1})$ according to the memoryless distributions defined by the n-product of $P_{U_2|U_1}$ and the n-product of Q_S, respectively; and there are J sequences in the bin indexed by $w_{r,i-1}$, the probability that there is no such sequence $\mathbf{u}_2$ goes to zero as $n \longrightarrow +\infty$. Denote the found j by $j^\star = j(\mathbf{s}[i], w_{r,i-1})$. The relay then transmits a vector $\mathbf{x}_2(w_{r,i-1})$ with i.i.d. components conditionally given $(\mathbf{u}_1(w_{r,i-1}), \mathbf{u}_2(w_{r,i-1}, j^\star), \mathbf{s}[i])$ drawn according to the marginal $P_{X_2|U_1,U_2,S}$ induced by (12).

Decoding: The decoding procedures at the source and the relay are as follows.

1. At the end of block i, the relay knows $w_{r,i-1}$ and declares that $\hat{w}_{r,i}$ is sent if there is a unique $\hat{w}_{r,i}$ such that $\mathbf{u}(w_{r,i-1}, \hat{w}_{r,i})$, $\mathbf{y}_2[i]$ and $\mathbf{s}[i]$ are jointly typical given $\mathbf{u}_1(w_{r,i-1})$. One can show that the decoding error in this step is small for sufficiently large n if
$$R_r < I(U; Y_2|S, U_1). \tag{B-13}$$
2. At the end of the transmission, the destination has collected all the blocks of channel outputs $\mathbf{y}_3[1], \mathbf{y}_3[2], \ldots, \mathbf{y}_3[B+1]$, and can then perform backward-decoding by first decoding $(w_{r,B}, w_{d,B+1})$ from $\mathbf{y}_3[B+1]$.
First, it declares that the pair $(\hat{w}_{r,B}, \hat{w}_{d,B+1})$ is sent if there is a unique pair $(\hat{w}_{r,B}, \hat{w}_{d,B+1})$, with $\hat{w}_{r,B} \in \{1, \ldots, M_r\}$ and $\hat{w}_{d,B+1} \in \{1, \ldots, M_d\}$, there is $j_B \in \{1, \ldots, J\}$, such that $\mathbf{u}_1(\hat{w}_{r,B})$, $\mathbf{u}_2(\hat{w}_{r,B}, j_B)$, $\mathbf{u}(\hat{w}_{r,B}, 1)$, $\mathbf{x}_1(\hat{w}_{r,B}, 1, \hat{w}_{d,B+1})$, $\mathbf{y}_3[B+1]$ are jointly typical. One can show that the decoding error in this step is small for sufficiently large n if
$$\begin{aligned} R_d &< I(X_1; Y_3|U, U_1, U_2) \\ R_d &< I(X_1, U_2; Y_3|U, U_1) - I(U_2; S|U_1) \\ R_r + R_d &< I(X_1, U, U_1, U_2; Y_3) - I(U_2; S|U_1) \\ &\overset{(a)}{=} I(X_1, U_1, U_2; Y_3) - I(U_2; S|U_1), \end{aligned} \tag{B-14}$$
where in (a) we used the fact that $I(U; Y_3|U_1, U_2, X_1) = 0$ under the distribution (12).
Next, for b ranging from B to 2, the destination knows $w_{r,b}$ and decodes $(w_{r,b-1}, w_{d,b})$ based on the information received in block b. It declares that the pair $(\hat{w}_{r,b-1}, \hat{w}_{d,b})$ is sent if there is a unique pair $(\hat{w}_{r,b-1}, \hat{w}_{d,b})$, with $\hat{w}_{r,b-1} \in \{1, \ldots, M_r\}$ and $\hat{w}_{d,b} \in \{1, \ldots, M_d\}$, there is $j_{b-1} \in \{1, \ldots, J\}$, such that $\mathbf{u}_1(\hat{w}_{r,b-1})$, $\mathbf{u}_2(\hat{w}_{r,b-1}, j_{b-1})$, $\mathbf{u}(\hat{w}_{r,b-1}, w_{r,b})$, $\mathbf{x}_1(\hat{w}_{r,b-1}, w_{r,b}, \hat{w}_{d,b})$, $\mathbf{y}_3[b]$ are jointly typical. One can show that the decoding error in this step is small for sufficiently large n if (B-14) is true.

It remains to show that the rate (11) is not altered if the sizes of the alphabets of the auxiliary random variables U, U_1 and U_2 are restricted as in (13). This can be easily done by following the steps in the proof of Theorem 1.

C. *Proof of Theorem 2*

Consider a sequence of $(\epsilon_n, n, R)-$codes with $\epsilon_n \to 0$ as $n \to +\infty$. By Fano's inequality, we have

$$H(W|Y_3^n) \leq nR\epsilon_n + 1 \triangleq n\delta_n. \tag{C-15}$$

Thus,

$$nR = H(W) \leq I(W; Y_3^n) + n\delta_n. \tag{C-16}$$

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

Lemma 1:

$$\text{i)}\quad I(W; Y_3^n) \leq \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; Y_{3,i}|S_i) - I(S_i; X_{1,i}|Y_{3,i}) \tag{C-17a}$$

$$\text{ii)}\quad I(W; Y_3^n) \leq \sum_{i=1}^{n} I(X_{1,i}; Y_{2,i}, Y_{3,i}|S_i, X_{2,i}). \tag{C-17b}$$

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 $X_j^i = (X_{j,1}, X_{j,2}, \cdots, X_{j,i})$, $j = 1, 2$.

We obtain the bound on $I(W; Y_3^n)$ given in (i) as follows.

$$\begin{aligned}
I(W; Y_3^n) &= I(W, S^n; Y_3^n) - I(S^n; Y_3^n|W) \\
&= \sum_{i=1}^{n} I(W, S^n; Y_{3,i}|Y_3^{i-1}) - H(S^n|W) + H(S^n|W, Y_3^n) \\
&= \sum_{i=1}^{n} [H(Y_{3,i}|Y_3^{i-1}) - H(Y_{3,i}|W, S^n, Y_3^{i-1}) \\
&\qquad - H(S_i) + H(S_i|W, Y_3^n, S^{i-1})] \\
&\overset{(a)}{\leq} \sum_{i=1}^{n} [H(Y_{3,i}) - H(Y_{3,i}|X_{1,i}, X_{2,i}, S_i) \\
&\qquad - H(S_i) + H(S_i|W, Y_3^n, S^{i-1}, X_{1,i})]
\end{aligned}$$

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

where

(a) follows from $(W, S^n, Y_3^{i-1}) \leftrightarrow (X_{1,i}, X_{2,i}, S_i) \leftrightarrow Y_{3,i}$ (a Markov chain); and the fact that $X_{1,i}$ is a deterministic function of W

(b) follows from the fact that conditioning reduces entropy.

We obtain the bound on $I(W; Y_3^n)$ given in (ii) as follows.

$$I(W; Y_3^n) \leq I(W; Y_2^n, Y_3^n)$$
$$= H(W) - H(W|Y_2^n, Y_3^n)$$
$$\overset{(c)}{\leq} H(W|S^n) - H(W|Y_2^n, Y_3^n, S^n)$$
$$= \sum_{i=1}^{n} I(W; Y_{2,i}, Y_{3,i}|Y_2^{i-1}, Y_3^{i-1}, S^n)$$
$$\overset{(d)}{=} \sum_{i=1}^{n} I(W; Y_{2,i}, Y_{3,i}|Y_2^{i-1}, Y_3^{i-1}, S^n, X_{2,i})$$
$$= \sum_{i=1}^{n} [H(Y_{2,i}, Y_{3,i}|Y_2^{i-1}, Y_3^{i-1}, S^n, X_{2,i})$$
$$- H(Y_{2,i}, Y_{3,i}|Y_2^{i-1}, Y_3^{i-1}, S^n, X_{2,i}, W)]$$
$$\overset{(e)}{=} \sum_{i=1}^{n} [H(Y_{2,i}, Y_{3,i}|Y_2^{i-1}, Y_3^{i-1}, S^n, X_{2,i})$$
$$- H(Y_{2,i}, Y_{3,i}|Y_2^{i-1}, Y_3^{i-1}, S^n, X_{2,i}, W, X_{1,i})]$$
$$\leq \sum_{i=1}^{n} [H(Y_{2,i}, Y_{3,i}|S_i, X_{2,i})$$
$$- H(Y_{2,i}, Y_{3,i}|Y_2^{i-1}, Y_3^{i-1}, S^n, X_{2,i}, W, X_{1,i})]$$
$$\overset{(f)}{=} \sum_{i=1}^{n} [H(Y_{2,i}, Y_{3,i}|S_i, X_{2,i}) - H(Y_{2,i}, Y_{3,i}|S_i, X_{2,i}, X_{1,i})]$$

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

where

(c) follows from the fact that W and S^n are independent; and $H(W|Y_2^n, Y_3^n) \geq H(W|Y_2^n, Y_3^n, S^n)$

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

(e) follows from the fact that $X_{1,i}$ is a deterministic function of W

(f) follows from the fact that the channel is discrete memoryless.

■

We introduce a random variable T which is uniformly distributed over $\{1, \cdots, n\}$. Set $S = S_T$, $X_1 = X_{1,T}$, $X_2 = X_{2,T}$, $Y_2 = Y_{2,T}$, and $Y_3 = Y_{3,T}$. Then we obtain

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; Y_{3,i}|S_i) - I(S_i; X_{1,i}|Y_{3,i}) \\
&= I(X_1, X_2; Y_3|S, T) - I(S; X_1|Y_3, T) \\
&= I(X_1, X_2, S; Y_3|T) - I(S; X_1, Y_3|T), \qquad \text{(C-18)}
\end{aligned}$$

and

$$\frac{1}{n} \sum_{i=1}^{n} I(X_{1,i}; Y_{2,i}, Y_{3,i}|S_i, X_{2,i}) = I(X_1; Y_2, Y_3|S, X_2, T), \qquad \text{(C-19)}$$

where the distribution on $(T, S, X_1, X_2, Y_2, Y_3)$ from a given code is of the form

$$P_{T,S,X_1,X_2,Y_2,Y_3} = P_S P_T P_{X_1|T} P_{X_2|X_1,S,T} W_{Y_2,Y_3|S,X_1,X_2}. \qquad \text{(C-20)}$$

We now eliminate the variable T from (C-18) and (C-19) as follows. The RHS of (C-18) can be bounded as

$$\begin{aligned}
&I(X_1, X_2, S; Y_3|T) - I(S; X_1, Y_3|T) \\
&\overset{(g)}{\leq} H(Y_3) - H(Y_3|X_1, X_2, S) - H(S|T) + H(S|X_1, Y_3, T) \\
&= I(X_1, X_2, S; Y_3) - H(S|T) + H(S|X_1, Y_3, T) \\
&\overset{(h)}{\leq} I(X_1, X_2, S; Y_3) - H(S) + H(S|X_1, Y_3) \\
&= I(X_1, X_2, S; Y_3) - I(S; X_1, Y_3) \\
&= I(X_1, X_2; Y_3|S) - I(S; X_1|Y_3), \qquad \text{(C-21)}
\end{aligned}$$

where

(g) holds since $H(Y_3|T) \leq H(Y_3)$ and $H(Y_3|X_1, X_2, S, T) = H(Y_3|X_1, X_2, S)$ (by the Markovian relation $T \leftrightarrow (X_1, X_2, S) \leftrightarrow Y_3$)

(h) holds since S is independent of T and $H(S|X_1, Y_3, T) \leq H(S|X_1, Y_3)$.

Similarly, the RHS of (C-19) can be bounded as

$$I(X_1; Y_2, Y_3|S, X_2, T) \leq I(X_1; Y_2, Y_3|S, X_2). \tag{C-22}$$

Finally, combining (C-16), (C-17a), (C-18), (C-21) at one hand, and (C-16), (C-17b), (C-19), (C-22) at the other hand, we get

$$R \leq I(X_1, X_2; Y_3|S) - I(S; X_1|Y_3) \tag{C-23a}$$

$$R \leq I(X_1; Y_2, Y_3|S, X_2), \tag{C-23b}$$

where the distribution on (S, X_1, X_2, Y_2, Y_3), obtained by marginalizing (C-20) over the variable T, has the form given in (17).

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 (17) such that the rate R satisfies (C-23). This completes the proof of Theorem 2.

D. *Proof of Theorem 3*

In this proof, we compute the lower bound in Theorem 1 using an appropriate jointly Gaussian distribution on S, X_1, U_1, U_2, X_2. The techniques used in this section rely strongly on those used in the proof of Theorem 6 in [16].

We first evaluate the second term of the minimization in (3) because this gives insights about the distribution that we should use to compute the lower bound. The second term of the minimization in (3) can be written as

$$I(X_1, U_1, U_2; Y_3) - I(U_2; S|U_1) =$$
$$I(X_1, U_1; Y_3) + I(U_2; Y_3|X_1, U_1) - I(U_2; S|X_1, U_1), \tag{D-24}$$

which follows from the fact that $I(U_2; S|U_1) = I(U_2; S|U_1, X_1)$ for the considered distribution.

We first focus on the evaluation of the term $[I(U_2; Y_3|X_1, U_1) - I(U_2; S|X_1, U_1)]$. To evaluate it, we assume that X_1 is zero mean Gaussian with variance P_1, U_1 is zero mean Gaussian with variance $\bar{\theta}P_2$, and X_1 and U_1 are jointly Gaussian with $\mathbb{E}[U_1X_1] = \rho'_{12}\sqrt{\bar{\theta}P_1P_2}$, for some $\theta \in [0,1]$, $\rho'_{12} \in [-1,1]$. The random variables X_1 and U_1 are independent of S as shown by the distribution given in Theorem 1. We also consider

$$X_2 = U_1 + \tilde{X}_2 \tag{D-25}$$

where, $\tilde{X}_2$ is zero mean Gaussian with variance θP_2, is independent of both X_1 and U_1, and is jointly Gaussian with S with $\mathbb{E}[\tilde{X}_2S] = \rho'_{2s}\sqrt{\theta P_2 Q}$, for some $\rho'_{2s} \in [-1,1]$. Then, from (22) and (D-25), we can write Y_3 as

$$Y_3 = X_1 + U_1 + \tilde{X}_2 + S + Z_3. \tag{D-26}$$

Let $\hat{\tilde{X}}_2 = \mathbb{E}[\tilde{X}_2|S]$ be the optimal linear estimator of $\tilde{X}_2$ given S under minimum mean square error criterion, and X'_2 be the resulting estimation error. The estimator $\hat{\tilde{X}}_2$ and the estimation error X'_2 are given by

$$\hat{\tilde{X}}_2 = \rho'_{2s}\sqrt{\frac{\theta P_2}{Q}}S \tag{D-27}$$

$$X'_2 = \tilde{X}_2 - \hat{\tilde{X}}_2. \tag{D-28}$$

We can alternatively write Y_3 in (D-26) as

$$\begin{aligned} Y_3 &= (\tilde{X}_2 - \hat{\tilde{X}}_2) + \hat{\tilde{X}}_2 + X_1 + U_1 + S + Z_3 \\ &= X'_2 + X_1 + U_1 + S' + Z_3, \end{aligned} \tag{D-29}$$

where

$$S' = \left(1 + \rho'_{2s}\sqrt{\frac{\theta P_2}{Q}}\right)S.$$

We now consider the following new channel output Y'_3 given by

$$Y'_3 := Y_3 - \mathbb{E}[Y_3|X_1, U_1] = X'_2 + S' + Z_3. \tag{D-30}$$

This new channel output Y'_3 is similar to the channel output considered in [3] because X'_2 is independent of the state S'. Hence, the capacity of this new channel is achieved if we use an auxiliary random variable

$$U_2 = X'_2 + \alpha S', \tag{D-31}$$

where α is Costa's parameter given by

$$\alpha = \frac{\mathbb{E}[X_2'^2]}{\mathbb{E}[X_2'^2] + \mathbb{E}[Z_3^2]} = \frac{\theta P_2(1-\rho_{2s}'^2)}{\theta P_2(1-\rho_{2s}'^2) + N_3}. \tag{D-32}$$

Then we can easily show that

$$[I(U_2;Y_3|X_1,U_1) - I(U_2;S|X_1,U_1)] = [I(U_2;Y_3') - I(U_2;S')].$$

The term $[I(U_2;Y_3') - I(U_2;S')]$ is maximized if U_2 is chosen as in (D-31). Thus, we obtain that

$$\begin{aligned} I(U_2;Y_3|X_1,U_1) - I(U_2;S|X_1,U_1) &= \frac{1}{2}\log\left(1 + \frac{\mathbb{E}[X_2'^2]}{N_3}\right) \\ &= \frac{1}{2}\log\left(1 + \frac{\theta P_2(1-\rho_{2s}'^2)}{N_3}\right). \end{aligned} \tag{D-33}$$

By substituting X_2' and S' in (D-31), we get

$$\begin{aligned} U_2 &= \tilde{X}_2 - \rho_{2s}'\sqrt{\frac{\theta P_2}{Q}}S + \alpha\Big(1 + \rho_{2s}'\sqrt{\frac{\theta P_2}{Q}}\Big)S \\ &= \tilde{X}_2 + \alpha_{\text{opt}}S, \end{aligned} \tag{D-34}$$

where

$$\begin{aligned} \alpha_{\text{opt}} &= \left(1 + \rho_{2s}'\sqrt{\frac{\theta P_2}{Q}}\right)\alpha - \rho_{2s}'\sqrt{\frac{\theta P_2}{Q}} \\ &= \frac{\theta P_2(1-\rho_{2s}'^2) - \rho_{2s}'\sqrt{\frac{\theta P_2}{Q}}N_3}{\theta P_2(1-\rho_{2s}'^2) + N_3}. \end{aligned} \tag{D-35}$$

The term $I(X_1,U_1;Y_3)$ on the RHS of (D-24) can be computed as

$$\begin{aligned} I(X_1,U_1;Y_3) &= h(Y_3) - h(Y_3|X_1,U_1) \\ &= h(Y_3) - h(\tilde{X}_2 + S + Z_3|X_1,U_1) \\ &\overset{(b)}{=} h(Y_3) - h(\tilde{X}_2 + S + Z_3) \\ &= \frac{1}{2}\log\Big(\frac{\mathbb{E}[(X_1+X_2+S)^2] + \mathbb{E}[Z_3^2]}{\mathbb{E}[(\tilde{X}_2+S)^2] + \mathbb{E}[Z_3^2]}\Big) \\ &= \frac{1}{2}\log\Big(1 + \frac{P_1 + \bar{\theta}P_2 + 2\rho_{12}'\sqrt{\bar{\theta}P_1P_2}}{\theta P_2 + Q + N_3 + 2\rho_{2s}'\sqrt{\theta P_2 Q}}\Big), \end{aligned} \tag{D-36}$$

where (b) follows from the fact that $\tilde{X}_2$ and S are independent of (X_1,U_1). Then, by adding (D-33) and (D-36) we get the second term of the minimization in (25).

The first term of the minimization in (3) can be written as

$$\begin{aligned} I(X_1;Y_2|S,U_1) &= h(Y_2|S,U_1) - h(Y_2|S,U_1,X_1) \\ &\overset{(a)}{=} h(X_1+Z_2|U_1) - h(Z_2) \\ &= \frac{1}{2}\log(1+\frac{P_1(1-\rho'^2_{12})}{N_2}), \end{aligned} \tag{D-37}$$

where (a) follows from the fact that the random variable S is independent of U_1 and X_1.

Finally, we obtain the rate on the RHS of (25) by maximization over all possible values of $\theta \in [0,1]$, $\rho'_{12} \in [-1,1]$ and $\rho'_{2s} \in [-1,1]$. Investigating the two terms of the minimization, we can easily see that it suffices to consider $\rho'_{12} \in [0,1]$ and $\rho'_{2s} \in [-1,0]$.

E. Proof of Theorem 4

In this section we use the upper bound for the DM case in Theorem 2 to compute the upper bound on the capacity of the state-dependent full-duplex Gaussian RC with informed relay.

Fix a joint distribution of X_1, X_2, S, Y_2, Y_3 of the form (17) satisfying

$$\begin{aligned} &\mathbb{E}[X_1^2] = \tilde{P}_1 \leq P_1, \quad \mathbb{E}[X_2^2] = \tilde{P}_2 \leq P_2, \\ &\mathbb{E}[X_1X_2] = \sigma_{12}, \quad \mathbb{E}[X_2S] = \sigma_{2s}, \quad \mathbb{E}[X_1S] = 0. \end{aligned} \tag{E-38}$$

We shall also use the correlation coefficients ρ_{12} and ρ_{2s} defined as

$$\rho_{12} = \frac{\sigma_{12}}{\sqrt{\tilde{P}_1\tilde{P}_2}}, \quad \rho_{2s} = \frac{\sigma_{2s}}{\sqrt{\tilde{P}_2 Q}}. \tag{E-39}$$

We first compute the first term in the minimization on the RHS of (16). Let $\mathbf{Y} = (X_1 + Z_2, X_1 + Z_3)^T$. We have

$$\begin{aligned} I(X_1;Y_2,Y_3|S,X_2) &= h(Y_2,Y_3|S,X_2) - h(Y_2,Y_3|S,X_1,X_2) \\ &= h(X_1+Z_2, X_1+Z_3|S,X_2) - h(Z_2,Z_3) \\ &\overset{(a)}{\leq} \frac{1}{2}\log\left|\mathbb{E}\Big[\Big(\mathbf{Y}-\mathbb{E}[\mathbf{Y}|S,X_2]\Big)\Big(\mathbf{Y}-\mathbb{E}[\mathbf{Y}|S,X_2]\Big)^T\Big]\right| \\ &\qquad - \frac{1}{2}\log(N_2N_3) \\ &= \frac{1}{2}\log\frac{\Big|\mathbb{E}[\mathbf{Y}\mathbf{Y}^T] - \mathbb{E}[\mathbb{E}[\mathbf{Y}|S,X_2]\mathbb{E}[\mathbf{Y}|S,X_2]^T]\Big|}{N_2N_3} \end{aligned}$$

$$\stackrel{(b)}{=} \frac{1}{2}\log\Big(1+\tilde{P}_1(1-\frac{\rho_{12}^2}{1-\rho_{2s}^2})(\frac{1}{N_2}+\frac{1}{N_3})\Big), \tag{E-40}$$

where, $|\cdot|$ denotes the determinant operator,

(a) follows from the fact that the conditional differential entropy $h(X_1+Z_2, X_1+Z_3|S, X_2)$ is maximized if (S, X_1, X_2, Z_2, Z_3) are jointly Gaussian, and

(b) follows from the fact the vector (S, X_1, X_2, Z_2, Z_3) is a jointly Gaussian vector and the MMSE estimator of $\mathbf{Y}$ given (S, X_2) is

$$\mathbb{E}[\mathbf{Y}|S, X_2] = (-\frac{\sigma_{12}\sigma_{2s}}{\tilde{P}_2 Q-\sigma_{2s}^2}S+\frac{\sigma_{12}Q}{\tilde{P}_2 Q-\sigma_{2s}^2}X_2)\times(1,1)^T. \tag{E-41}$$

We now compute the term $[I(X_1, X_2; Y_3|S) - I(X_1; S|Y_3)]$. We have

$$\begin{aligned} I(X_1, X_2; Y_3|S) - I(X_1; S|Y_3) &= h(Y_3|S) - h(Y_3|X_1, X_2, S) - h(S|Y_3) + h(S|X_1, Y_3) \\ &= h(Y_3) - h(S) + h(S|X_1, Y_3) - h(Z_3). \end{aligned} \tag{E-42}$$

For fixed second moments (E-38), we have

$$h(Y_3) \leq \frac{1}{2}\log(2\pi e)(\tilde{P}_1+\tilde{P}_2+2\sigma_{12}+2\sigma_{2s}+Q+N_3), \tag{E-43}$$

where equality is attained if Y_3 is Gaussian. Similarly, the term $h(S|X_1, Y_3)$ is maximized if (S, X_1, Y_3) are jointly Gaussian. Let $\hat{S}(X_1, Y_3) = \mathbb{E}[S|X_1, Y_3]$ be the MMSE estimator of S given (X_1, Y_3), i.e.,

$$\begin{aligned} \hat{S}(X_1, Y_3) &= \mathbb{E}[S|X_1, X_2+S+Z_3] \\ &= \gamma_1 X_1 + \gamma_2(X_2+S+Z_3) \end{aligned} \tag{E-44}$$

with

$$\begin{aligned} \gamma_1 &= -\frac{\sigma_{12}(Q+\sigma_{2s})}{\tilde{P}_1(\tilde{P}_2+2\sigma_{2s}+Q+N_3)-\sigma_{12}^2} \\ \gamma_2 &= \frac{\tilde{P}_1(Q+\sigma_{2s})}{\tilde{P}_1(\tilde{P}_2+2\sigma_{2s}+Q+N_3)-\sigma_{12}^2}. \end{aligned} \tag{E-45}$$

Then we have

$$\begin{aligned} h(S|X_1, Y_3) &= h(S-\hat{S}(X_1, Y_3)|X_1, Y_3) \\ &\leq h(S-\gamma_1 X_1-\gamma_2(X_2+S+Z_3)) \\ &= \frac{1}{2}\log(2\pi e)\mathbb{E}\Big[\Big(S-\gamma_1 X_1-\gamma_2(X_2+S+Z_3)\Big)^2\Big] \\ &= \frac{1}{2}\log\Big((2\pi e)\frac{Q\tilde{P}_1\tilde{P}_2+\tilde{P}_1 N_3 Q-\sigma_{2s}^2\tilde{P}_1-\sigma_{12}^2 Q}{\tilde{P}_1(\tilde{P}_2+2\sigma_{2s}+Q+N_3)-\sigma_{12}^2}\Big), \end{aligned} \tag{E-46}$$

where the inequality is attained with equality if S, X_1, X_2, Y_3 are jointly Gaussian. From (E-42), (E-43) and (E-46), we obtain

$$\begin{aligned} I(X_1, X_2; Y_3|S) - I(X_1; S|Y_3) &= \frac{1}{2}\log\Bigg(\frac{(\tilde{P}_1 + \tilde{P}_2 + 2\sigma_{12} + 2\sigma_{2s} + Q + N_3)}{(\tilde{P}_1\tilde{P}_2 + 2\tilde{P}_1\sigma_{2s} + \tilde{P}_1 Q + \tilde{P}_1 N_3 - \sigma_{12}^2)} \\ &\qquad \times \frac{(Q\tilde{P}_1\tilde{P}_2 + \tilde{P}_1 N_3 Q - \sigma_{2s}^2\tilde{P}_1 - \sigma_{12}^2 Q)}{QN_3}\Bigg) \\ &= \frac{1}{2}\log\Big(1 + \frac{(\sqrt{\tilde{P}_1} + \rho_{12}\sqrt{\tilde{P}_2})^2}{\tilde{P}_2(1-\rho_{12}^2-\rho_{2s}^2) + (\sqrt{Q} + \rho_{2s}\sqrt{\tilde{P}_2})^2 + N_3}\Big) \\ &\qquad + \frac{1}{2}\log\Big(1 + \frac{\tilde{P}_2(1-\rho_{12}^2-\rho_{2s}^2)}{N_3}\Big). \end{aligned} \tag{E-47}$$

For convenience, let us define the function $\Theta_1(\tilde{P}_1, \rho_{12}, \rho_{2s})$ as the RHS of (E-40) and the function $\Theta_2(\tilde{P}_1, \tilde{P}_2, \rho_{12}, \rho_{2s})$ as the RHS of (E-47). From the above analysis, the capacity of the channel is upper-bounded as

$$C \leq \max\ \min\{\Theta_1(\tilde{P}_1, \rho_{12}, \rho_{2s}), \Theta_2(\tilde{P}_1, \tilde{P}_2, \rho_{12}, \rho_{2s})\} \tag{E-48}$$

where the maximization is over all covariance matrices $\Lambda_{X_1,X_2,S,Z_2,Z_3}$ of (X_1, X_2, S, Z_2, Z_3),

$$\Lambda_{X_1,X_2,S,Z_2,Z_3} = \begin{pmatrix} \tilde{P}_1 & \rho_{12}\sqrt{\tilde{P}_1\tilde{P}_2} & 0 & 0 & 0 \\ \rho_{12}\sqrt{\tilde{P}_1\tilde{P}_2} & \tilde{P}_2 & \rho_{2s}\sqrt{\tilde{P}_2 Q} & 0 & 0 \\ 0 & \rho_{2s}\sqrt{\tilde{P}_2 Q} & Q & 0 & 0 \\ 0 & 0 & 0 & N_2 & 0 \\ 0 & 0 & 0 & 0 & N_3 \end{pmatrix}, \tag{E-49}$$

that satisfy

$$\tilde{P}_1 \leq P_1, \quad \tilde{P}_2 \leq P_2 \tag{E-50}$$

and have non-negative discriminant,

$$Q\tilde{P}_1\tilde{P}_2 N_2 N_3(1-\rho_{12}^2-\rho_{2s}^2) \geq 0, \tag{E-51}$$

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

$$\rho_{12}^2 + \rho_{2s}^2 \leq 1. \tag{E-52}$$

Furthermore, investigating $\Theta_1(\tilde{P}_1, \rho_{12}, \rho_{2s})$ and $\Theta_2(\tilde{P}_1, \tilde{P}_2, \rho_{12}, \rho_{2s})$, it can be seen that it suffices to consider $\rho_{12} \in [0, 1]$ and $\rho_{2s} \in [-1, 0]$ for the maximization in (E-48).

To complete the proof, we should show that $\Theta_1(\tilde{P}_1, \rho_{12}, \rho_{2s})$ and $\Theta_2(\tilde{P}_1, \tilde{P}_2, \rho_{12}, \rho_{2s})$ are maximized at $\tilde{P}_1 = P_1$ and $\tilde{P}_2 = P_2$. It is easy to show that $\Theta_1(\tilde{P}_1, \rho_{12}, \rho_{2s})$ and $\Theta_2(\tilde{P}_1, \tilde{P}_2, \rho_{12}, \rho_{2s})$ increase monotonically with $\tilde{P}_1$ for fixed ρ_{12}, ρ_{2s}, $\tilde{P}_2$. Then we can replace $\tilde{P}_1$ with P_1 in both $\Theta_1(\tilde{P}_1, \rho_{12}, \rho_{2s})$ and $\Theta_2(\tilde{P}_1, \tilde{P}_2, \rho_{12}, \rho_{2s})$. To show that $\tilde{P}_2$ can be replaced by P_2, we use the following intuitive argument. Since the term $\Theta_1(P_1, \rho_{12}, \rho_{2s})$ does not depend on $\tilde{P}_2$ for given ρ_{12} and ρ_{2s}, it remains to show that $\tilde{P}_2$ can be replaced with P_2 in only the term $\Theta_2(\tilde{P}_1, \tilde{P}_2, \rho_{12}, \rho_{2s})$. The term $\Theta_2(\tilde{P}_1, \tilde{P}_2, \rho_{12}, \rho_{2s})$ is the sum rate of a two-user MAC with asymmetric CSI in which the informed encoder knows the message of the informed encoder [16, Theorem 6]. Then, considering this MAC, it can be argued [16] that for the sum-rate to be maximized the informed encoder should use the entire power available, i.e., P_2. This concludes the proof of Theorem 4.

F. Proof of Observation 1

We first prove the first statement in Observation 1. Let us denote $N_2^\star$ as the RHS of (37). We have

$$\begin{aligned} R_{\mathrm{G}}^{\mathrm{lo}} &\overset{(a)}{\geq} \min\Big\{\frac{1}{2}\log(1+\frac{P_1}{N_2}), \\ &\qquad \max_{-1\leq\rho'_{2s}\leq 0} \frac{1}{2}\log\Big(1+\frac{P_1}{P_2+Q+N_3+2\rho'_{2s}\sqrt{P_2Q}}\Big)+\frac{1}{2}\log(1+\frac{P_2(1-\rho'^2_{2s})}{N_3})\Big\} \\ &\overset{(b)}{=} \frac{1}{2}\log(1+\frac{P_1}{N_2}) \\ &:= R_{\mathrm{DG}}, \end{aligned} \tag{F-53}$$

where (a) follows by putting $\rho'_{12} = 0$ and $\theta = 1$ in (25), and (b) follows if $N_2 \geq N_2^\star$. Then, it is easy to observe that

$$R_{\mathrm{DG}}^{\mathrm{up}} \leq R_{\mathrm{DG}}. \tag{F-54}$$

From (F-53) and (F-54), we get that

$$R_{\mathrm{DG}} \leq R_{\mathrm{G}}^{\mathrm{lo}} \leq C_{\mathrm{DG}} \leq R_{\mathrm{DG}}^{\mathrm{up}} \leq R_{\mathrm{DG}}. \tag{F-55}$$

Then we can conclude that the lower bound and upper bound meet if $N_2 \geq N_2^\star$.

Let us now prove the second statement in Observation 1. If the pair (ρ_{12}, ρ_{2s}) that maximizes the upper bound in Corollary 3 satisfies the condition in (31) with equality, i.e., $\rho_{12}^2 + \rho_{2s}^2 = 1$, then we choose $\varrho_{2s} = \rho_{2s}$, $\varrho_{12} = \rho_{12}$, and $\theta = \varrho_{2s}^2$ (i.e., $\bar{\theta} = \varrho_{12}^2$) in the lower bound (34) to achieve the upper bound, and thus obtain channel capacity in this case.

G. Proofs for Time Division Relaying

1) Proof of Proposition 1: Let $(X^{(1)}_{1,1}, X^{(1)}_{1,2}, \ldots, X^{(1)}_{1,\lfloor \nu n \rfloor})$ and $(X^{(2)}_{1,\lfloor \nu n \rfloor+1}, X^{(2)}_{1,\lfloor \nu n \rfloor+2}, \ldots, X^{(2)}_{1,n})$ be the transmitted sequences from the source during the relay-receive period and the relay-transmit period, respectively. The relay receives $Y_{2,1}, Y_{2,2}, \ldots, Y_{2,\lfloor \nu n \rfloor}$ during the relay-receive period and transmits a sequence $X_{2,\lfloor \nu n \rfloor+1}, X_{2,\lfloor \nu n \rfloor+2}, \cdots, X_{2,n}$ during the relay-transmit period. From Fano's inequality (C-16) and Lemma 1, we have the following

$$nR \leq \min \left\{ \sum_{i=1}^{n} I(X_{1,i}; Y_{2,i}, Y_{3,i} | S_i, X_{2,i}), \right.$$
$$\left. \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; Y_{3,i} | S_i) - I(X_{1,i}; S_i | Y_{3,i}) \right\} + n\delta_n. \tag{G-56}$$

We now specialize this bound to the TD mode for which we have $X_{2,i} = 0$ for $i \leq \lfloor \nu n \rfloor$ (as the relay does not transmit during the relay-receive period) and $Y_{2,i} = 0$ for $i \geq \lfloor \nu n \rfloor + 1$ (as the relay does not receive during the relay-transmit period). This gives

$$nR \leq \min \left\{ \sum_{i=1}^{\lfloor \nu n \rfloor} I(X^{(1)}_{1,i}; Y_{2,i}, Y^{(1)}_{3,i} | S^{(1)}_i, X_{2,i} = 0) + \sum_{i=\lfloor \nu n \rfloor+1}^{n} I(X^{(2)}_{1,i}; Y^{(2)}_{3,i} | S^{(2)}_i, X_{2,i}), \right.$$
$$\sum_{i=1}^{\lfloor \nu n \rfloor} I(X^{(1)}_{1,i}; Y^{(1)}_{3,i} | S^{(1)}_i, X_{2,i} = 0) - I(X^{(1)}_{1,i}; S^{(1)}_i | Y^{(1)}_{3,i})$$
$$\left. + \sum_{i=\lfloor \nu n \rfloor+1}^{n} I(X^{(2)}_{1,i}, X_{2,i}; Y^{(2)}_{3,i} | S^{(2)}_i) - I(X^{(2)}_{1,i}; S^{(2)}_i | Y^{(2)}_{3,i}) \right\} + n\delta_n. \tag{G-57}$$

By letting $n \to \infty$ and using standard arguments [51], we get the single letter upper bound on capacity

$$C \leq \max \min \left\{ \nu I(X^{(1)}_1; Y_2, Y^{(1)}_3 | S^{(1)}, X_2 = 0) + \bar{\nu} I(X^{(2)}_1; Y^{(2)}_3 | S^{(2)}, X_2), \right.$$
$$\nu I(X^{(1)}_1; Y^{(1)}_3 | S^{(1)}, X_2 = 0) - \nu I(X^{(1)}_1; S^{(1)} | Y^{(1)}_3)$$
$$\left. + \bar{\nu} I(X^{(2)}_1, X_2; Y^{(2)}_3 | S^{(2)}) - \bar{\nu} I(X^{(2)}_1; S^{(2)} | Y^{(2)}_3) \right\}, \tag{G-58}$$

where the maximization is over all joint distributions of the form

$$Q_{S^{(1)}} P_{X^{(1)}_1} W_{Y_2, Y^{(1)}_3 | X^{(1)}_1, S^{(1)}} Q_{S^{(2)}} P_{X^{(2)}_1} P_{X_2 | X^{(2)}_1, S^{(2)}} W_{Y^{(2)}_3 | X^{(2)}_1, X_2, S^{(2)}}. \tag{G-59}$$

The bound in (G-58) is the counterpart, to the TD mode, of the upper bound (16) for the full-duplex case. By closely following the arguments and the algebra used in the proof of Theorem 4, it can be shown that this bound is maximized by choosing $S^{(1)}, S^{(2)}, X_1^{(1)}, X_1^{(2)}, X_2, Y_2, Y_3^{(1)}, Y_3^{(2)}$ that are jointly Gaussian, with $X_1^{(1)}$ with power $P_1^{(1)}$ is independent of $S^{(1)}$, and $X_1^{(2)}$ and X_2 with power $P_1^{(2)}$ and P_2, respectively, are such that

$$\mathbb{E}[X_1^{(2)}X_2] = \rho_{12}\sqrt{P_1^{(2)}P_2}, \quad \mathbb{E}[X_1^{(2)}S^{(2)}] = 0, \quad \mathbb{E}[X_2S^{(2)}] = \rho_{2s}\sqrt{P_2Q^{(2)}}.$$

Using this distribution, the evaluation of the RHS of (G-58) gives the RHS of (44).

2) Proof of Proposition 2: The proof follows by combining the technique of rate-splitting [53] and the Generalized DPC described in Section IV-A for the full-duplex mode. Rate splitting has the message W to be transmitted from the source node split into two independent parts: w_d transmitted directly to the destination at rate R_d, and w_r transmitted through the relay at rate R_r, with a total rate $R = R_r + R_d$.

The encoding and transmission scheme is as follows. During the relay-receive period, the source sends a Gaussian signal $X_{1,i}^{(1)}$ which carries w_r only and is independently drawn with a random variable $X_1^{(1)} \sim \mathcal{N}(0, P_1^{(1)})$ which is independent of the channel state $S^{(1)}$. During the relay-transmit period, the source transmits a Gaussian signal $X_{1,i}^{(2)}$ which carries both w_r and w_d and is independently drawn with $X_1^{(2)} \sim \mathcal{N}(0, P_1^{(2)})$. During the relay-transmit period, the relay sends a Gaussian signal $X_{2,i}$ which carries w_r only and is given by

$$X_{2,i} = U_{1,i} + \tilde{X}_{2,i}, \tag{G-60}$$

where $U_{1,i}$ is drawn with $U_1 \sim \mathcal{N}(0, \bar{\theta}P_2)$ and $\tilde{X}_{2,i}$ is obtained via a GDPC considering $S^{(2)}$ as non-causal channel state information during this period.
The random variables U_1 and $X_1^{(2)}$ are jointly Gaussian with $\mathbb{E}[X_1^{(2)}X_2] = \mathbb{E}[X_1^{(2)}U_1] = \rho'_{12}\sqrt{\bar{\theta}P_1^{(2)}P_2}$, and are both independent of the state $S^{(2)}$. For the GDPC, we use the following auxiliary random variable to generate the auxiliary codewords $U_{2,i}$,

$$U_2 = \tilde{X}_2 + \alpha' S^{(2)}, \tag{G-61}$$

where $\tilde{X}_2 \sim \mathcal{N}(0, \theta P_2)$ is jointly Gaussian with $S^{(2)}$, with $\mathbb{E}[X_2S^{(2)}] = \mathbb{E}[\tilde{X}_2S_2^{(2)}] = \rho'_{2s}\sqrt{\theta P_2Q^{(2)}}$; and α' is the GDPC scale parameter. Thus, using the above GDPC, $\tilde{X}_{2,i}$ is generated as

$$\tilde{X}_{2,i} = U_{2,i} - \alpha' S_i^{(2)} \tag{G-62}$$

where $U_{2,i}$ is independently drawn with U_2.

Furthermore, we let $X_{1,i}^{(2)} = \rho'_{12}\sqrt{P_1^{(2)}/\bar{\theta}P_2}U_{1,i} + \tilde{X}_{1,i}^{(2)}$, where $\tilde{X}_{1,i}^{(2)}$ is independently drawn with $\tilde{X}_1^{(2)} \sim \mathcal{N}(0, (1-\rho'^2_{12})P_1^{(2)})$, is independent of $U_1, X_2, S^{(2)}$, and carries w_d only.

For the decoding procedures at the source and the relay, we give simple arguments based on intuition (the rigorous decoding arguments use jointly typicality). Also, since all the random variables are i.i.d., we sometimes omit the time index. The relay subtracts out $S^{(1)}$ from the received Y_2 and then decodes w_r. Message w_r can be decoded correctly at the relay as long as

$$R_r < \frac{\nu}{2}\log\Big(1 + \frac{P_1^{(1)}}{N_2}\Big). \tag{G-63}$$

The destination decodes w_r from $(Y_3^{(1)}, Y_3^{(2)})$ by treating the part of $X_1^{(2)}$ that carries message w_d, i.e., $\tilde{X}_1^{(2)}$, as an unknown noise. Since w_r is carried over two parallel channels ($Y_3^{(1)}$ and $Y_3^{(2)}$) with $Y_3^{(2)}$ being a noisy version of channel inputs obtained by GDPC, the destination can decode w_r if

$$R_r < \nu I(X_1^{(1)}; Y_3^{(1)}) + \bar{\nu}[I(U_1, U_2; Y_3^{(2)}) - I(U_2; S^{(2)}|U_1)]. \tag{G-64}$$

Then the destination uses $Y_3^{(2)}$ and the decoded codewords U_1 and U_2 (which carry w_r only) to decode w_d. We note that, in decoding message w_r, the destination can decode U_2 *fully*, i.e., not only the bin index but also the correct sequence in the bin, if

$$I(U_2; Y_3^{(2)}|U_1) - I(U_2; S^{(2)}|U_1) > 0. \tag{G-65}$$

Under the condition (G-65), the destination can decode message w_d if

$$R_d < \bar{\nu} I(X_1^{(2)}; Y_3^{(2)}|U_1, U_2). \tag{G-66}$$

Adding (G-63) and (G-66), we obtain

$$R_r + R_d < \frac{\nu}{2}\log(1 + \frac{P_1^{(1)}}{N_2}) + \bar{\nu} I(X_1^{(2)}; Y_3^{(2)}|U_1, U_2), \tag{G-67}$$

and adding (G-64) and (G-66), we obtain

$$R_r + R_d < \nu I(X_1^{(1)}; Y_3^{(1)}) + \bar{\nu}[I(X_1^{(2)}, U_1, U_2; Y_3^{(2)}) - I(U_2; S^{(2)}|U_1)]. \tag{G-68}$$

We first compute the RHS of (G-68). The term $[I(X_1^{(2)}, U_1, U_2; Y_3^{(2)}) - I(U_2; S^{(2)}|U_1)]$ can be computed using simple algebra which is essentially similar to that in the evaluation of (D-24) in Appendix D, and which we omit here for brevity, to obtain

$$I(X_1^{(2)}, U_1, U_2; Y_3^{(2)}) - I(U_2; S^{(2)}|U_1) = \frac{1}{2}\log\Big(1 + \frac{P_1^{(2)} + \bar{\theta}P_2 + 2\rho'_{12}\sqrt{\bar{\theta}P_1^{(2)}P_2}}{\theta P_2 + Q^{(2)} + 2\rho'_{2s}\sqrt{\theta P_2 Q^{(2)}} + N_3}\Big) + \frac{1}{2}\log\Big(\frac{P'_2(P'_2 + Q^{(2)} + N_3)}{P'_2 Q^{(2)}(1-\alpha')^2 + N_3(P'_2 + \alpha'^2 Q^{(2)})}\Big), \quad \text{(G-69)}$$

with $P'_2 := \theta P_2(1-\rho'^2_{2s})$. Also, it is easy to show that

$$I(X_1^{(1)}; Y_3^{(1)}) = \frac{1}{2}\log\Big(1 + \frac{P_1^{(1)}}{N_3 + Q^{(1)}}\Big). \quad \text{(G-70)}$$

The mutual information on the RHS of (G-67) can be computed as follows. Let $\tilde{Y}_3^{(2)} = \tilde{X}_2 + S^{(2)} + Z_3$. We have

$$\begin{aligned} I(X_1^{(2)}; Y_3^{(2)}|U_1, U_2) &= I(\tilde{X}_1^{(2)}; \tilde{X}_1^{(2)} + \tilde{X}_2 + S^{(2)} + Z_3|U_1, U_2) \\ &\overset{(a)}{=} h(\tilde{X}_1^{(2)} + \tilde{Y}_3^{(2)}|U_2) - h(\tilde{Y}_3^{(2)}|U_2) \\ &\overset{(b)}{=} \frac{1}{2}\log\Big(\mathbb{E}[(\tilde{X}_1^{(2)})^2 + (\tilde{Y}_3^{(2)})^2] - \mathbb{E}[\mathbb{E}^2[\tilde{Y}_3^{(2)}|U_2]]\Big) \\ &\quad - \frac{1}{2}\log\Big(\mathbb{E}[(\tilde{Y}_3^{(2)})^2] - \mathbb{E}[\mathbb{E}^2[\tilde{Y}_3^{(2)}|U_2]]\Big) \\ &= \frac{1}{2}\log\Big(1 + \frac{\mathbb{E}[(\tilde{X}_1^{(2)})^2]}{\mathbb{E}[(\tilde{Y}_3^{(2)})^2] - \mathbb{E}[\mathbb{E}^2[\tilde{Y}_3^{(2)}|U_2]]}\Big) \\ &= \frac{1}{2}\log\Big(1 + \frac{(1-\rho'^2_{12})P_1^{(2)}}{N_3 + \Phi(\alpha', \theta, \rho'_{2s})}\Big), \end{aligned} \quad \text{(G-71)}$$

where (a) follows from the fact that U_1 is independent of U_2, $\tilde{X}_1^{(2)}$, $\tilde{Y}_3^{(2)}$; (b) follows from the fact that $\tilde{X}_1^{(2)}$ is independent of U_2; and $\Phi(\alpha', \theta, \rho'_{2s})$ is defined as in (48).

The conditional mutual information difference in (G-65) can be computed by subtracting the RHS of (G-71) and the term $I(U_1; Y_3^{(2)})$ from the RHS of (G-69), with

$$\begin{aligned} I(U_1; Y_3^{(2)}) &= h(Y_3^{(2)}) - h(\tilde{X}_1^{(2)} + \tilde{X}_2 + S^{(2)} + Z_3) \\ &= \frac{1}{2}\log\Big(\frac{P_1^{(2)} + P_2 + Q^{(2)} + 2\rho'_{12}\sqrt{\bar{\theta}P_1^{(2)}P_2} + 2\rho'_{2s}\sqrt{\theta P_2 Q^{(2)}} + N_3}{P_1^{(2)}(1-\rho'^2_{12}) + \theta P_2 + Q^{(2)} + N_3}\Big), \end{aligned} \quad \text{(G-72)}$$

to obtain

$$I(U_2; Y_3^{(2)}|U_1) - I(U_2; S^{(2)}|U_1) = \Theta(\alpha', \rho'_{12}, \theta, \rho'_{2s}), \quad \text{(G-73)}$$

where $\Theta(\alpha', \rho'_{12}, \theta, \rho'_{2s})$ is defined as in (49).

Finally, we obtain (47a) using (G-67) and (G-71); and we obtain (47b) using (G-68), (G-69) and (G-70). This completes the proof.

REFERENCES

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

[2] 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.

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

[4] 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.

[5] 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.

[6] U. Erez, S. Shamai (Shitz), and R. Zamir, "Capacity and lattice strategies for cancelling known interference," *IEEE Trans. Inf. Theory*, vol. IT-51, pp. 3820–3833, Nov. 2005.

[7] 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.

[8] S. Gel'fand and M. Pinsker, "On Gaussian channel with random parameters," in *Proc. IEEE Int. Symp. Information Theory*, Tashkent, 1984.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

[14] S. Kotagiri and J. N. Laneman, "Multiple access channel with state information known to some encoders and independent messages," *EURASIP Journal on Wireless Commnunications and Networking*, vol. 2008, Article ID 450680, 14 pages, 2008. doi:10.1155/2008/450680.

[15] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdù, "Cooperative multiple access encoding with states available at one transmitter," in *Proc. IEEE Int. Symp. Information Theory*, Nice, France, Jun. 2007, pp. 1556–1560.

[16] ——, "Cooperative multiple access encoding with states available at one transmitter," *IEEE Trans. Inf. Theory*, 2008 (accepted for publication).

[17] 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.

[18] 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.

[19] ——, "Bounds on the capacity of the relay channel with states at the source," in *Proc. IEEE International Symposium on Information Theory and its Applications*, Auckland, New Zealand, Dec. 2008 (accepted for publication).

[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] 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.

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

[23] A. S. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1639–1667, Jun. 2002.

[24] A. Somekh-Baruch and N. Merhav, "On the capacity game of public watermarking systems," *IEEE Trans. Inf. Theory*, vol. 50, pp. 511–524, Jun. 2004.

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

[26] A. Zaidi and L. Vandendorpe, "Coding schemes for relay-assisted information embedding," *IEEE Transactions on Information Security and Forensics*, 2008 (accepted for publication).

[27] N. Merhav, "On random coding error exponents of watermarking systems," *IEEE Trans. Inf. Theory*, vol. 46, pp. 420–430, Mar. 2000.

[28] G. Caire and S. Shamai (Shitz), "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. IT-48, pp. 2007–2019, Sep. 1999.

[29] ——, "On the throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-49, pp. 1691–1706, Jul. 2003.

[30] 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.

[31] 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.

[32] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-52, pp. 3936–3964, Jan. 2006.

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

[34] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1813–1827, May 2006.

[35] ——, "Limits on communications in a cognitive radio channel," *IEEE Communications Magazine*, vol. 44, pp. 44–49, Jun. 2006.

[36] A. Jovičić and P. Viswanath, "Cognitive radio: An information-theoretic perspective," in *Proc. IEEE Int. Symp. Information Theory*, Jul. 2006, pp. 2413–2417.

[37] I. Marić, A. Goldsmith, G. Kramer, and S. Shamai (Shitz), "On the capacity of interference channels with one cooperating transmitter," *European Trans. on Telecommunications*, pp. 405–420, 2008.

[38] S. Sridharan, S. Vishwanath, S. A. Jafar, and S. Shamai (Shitz), "On the capacity of cognitive relay assisted gaussian interference channel," in *Proc. IEEE Int. Symp. Information Theory*, Toronto, Canada, Jul. 2008, pp. 549–553.

[39] O. Sahin and E. Erkip, "On achievable rates for interference relay channel with interference cancellation," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Monterey, CA, Nov. 2007, pp. 805–809.

[40] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdù, "Cognitive interference channels with state information," in *Proc. IEEE Int. Symp. Information Theory*, Toronto, Canada, Jul. 2008, pp. 1353–1357.

[41] N. Merhav and S. Shamai (Shitz), "Information rates subjected to state masking," in *Proc. IEEE Int. Symp. Information Theory*, Seattle, USA, 2006, pp. 1184–1188.

[42] 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.

[43] F. M. J. Willems, *Informationtheoretical Results for the Discrete Memoryless Multiple Access Channel*. Leuven, Belgium: Doctor in de Wetenschappen Proefschrift dissertation, Oct. 1982.

[44] A. B. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Trans. Inf. Theory*, vol. IT-28, pp. 841–850, Nov. 1982.

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

[46] B. Schein and R. G. Gallager, "The Gaussian parallel relay network," in *Proc. IEEE Int. Symp. Information Theory*, Sorrento, Italy, Jun. 2000, p. 22.

[47] M. Gastpar and M. Vetterli, "On the capacity of large Gaussian relay networks," *IEEE Trans. Inf. Theory*, vol. IT-51, pp. 765–779–3063, Mar. 2005.

[48] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behaviour," *IEEE Trans. Inf. Theory*, vol. 50, pp. 3062–3080, Dec. 2004.

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

[50] A. Zaidi, S. Kotagiri, J. N. Laneman, and L. Vandendorpe, "Multiple access channels with states available at one transmitter and degraded message sets," *(to be submitted for publication)*.

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

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

[53] A. Høst-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Trans. Inf. Theory*, vol. IT-51, pp. 2020–2040, Jun. 2005.