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

Abdellatif Zaidi, Shlomo Shamai (Shitz), Pablo Piantanida, and Luc Vandendorpe, Fellow, IEEE

Abstract—We consider a three-terminal state-dependent relay channel with the channel state available noncausally at only the source. Such a model may be of interest for node cooperation in the framework of cognition, i.e., collaborative signal transmission involving cognitive and noncognitive radios. We study the capacity of this communication model. One principal problem is caused by the relay’s not knowing the channel state. For the discrete memoryless (DM) model, we establish two lower bounds and an upper bound on channel capacity. The first lower bound is obtained by a coding scheme in which the source describes the state of the channel to the relay and destination, which then exploit the gained description for a better communication of the source’s information message. The coding scheme for the second lower bound remedies the relay’s not knowing the states of the channel by first computing, at the source, the appropriate input that the relay would send had the relay known the states of the channel, and then transmitting this appropriate input to the relay. The relay simply guesses the sent input and sends it in the next block. The upper bound accounts for not knowing the state at the relay and destination. For the general Gaussian model, we derive lower bounds on the channel capacity by exploiting ideas in the spirit of those we use for the DM model; and we show that these bounds are optimal for small and large noise at the relay irrespective of the strength of the interference. Furthermore, we also consider a relay model with orthogonal channels from the source to the relay and from the source and relay to the destination in which the source input component that is heard by the relay does not depend on the channel states. We establish a better upper bound for both DM and Gaussian cases and we also characterize the capacity in a number of special cases.

Index Terms—Channel state information (CSI), cognitive radio, dirty paper coding (DPC), relay channel (RC), user cooperation.

I. INTRODUCTION

We consider a three-terminal state-dependent relay channel (RC) in which, as shown in Fig. 1, the source wants to communicate a message $W$ to the destination through the state-dependent RC in $n$ uses of the channel, with the help of the relay. The channel outputs $Y^n_2$ and $Y^n_3$ for the relay and the destination, respectively, are controlled by the channel input $X^n_1$ from the source, the relay input $X^n_R$ and the channel state $S^n$, through a given memoryless probability law $W_{Y_2,Y_3,X_1,X_2,S}$. The channel state $S^n$ is generated according to the $n$-product of a given memoryless probability law $Q_S$. It is assumed that the channel state is known, noncausally, to only the source. The destination estimates the message sent by the source from the received channel output. In this paper, we study the capacity of this communication system. We will refer to the model in Fig. 1 as general state-dependent RC with informed source.

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

$$W_{Y_2,Y_3,X_1,R,X_1,D,X_2,S} = W_{Y_2,X_1,R,S}W_{Y_3,X_1,D,X_2,S}. \quad (1)$$

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

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

A. Background and Related Work

Channels whose probabilistic input–output relation depends on random parameters, or channel states, have spurred much interest and can model a large variety of problems, each related to some physical situation of interest. Examples of applications include information embedding [2]–[4], interference imposed by adjacent users, certain storage applications such as computer memories [5], coding for certain broadcast channels [6]–[8], dispersive (ISI) channels [9], block fading in wireless environments [10], cooperation in the realm of cognition [11], and
Fig. 1. General state-dependent RC with state information $S^*$ available noncausally at only the source.

Fig. 2. State-dependent RC with the source input $X^*_n = (X^*_{1n}, X^*_{2n})$, and only the component $X^*_{1n}$ knowing the states of the channel noncausally.

The random state sequence may be known in a causal or noncausal manner. For single-user models, the concept of channel state available at only the transmitter dates back to Shannon [12] for the causal channel state case, and to Gel’fand and Pinsker [13] for the noncausal channel state case. In [14], Heegard and El Gamal study a model in which the state sequence is known noncausally to only the encoder or to only the decoder. They also derive achievable rates for the case in which partial channel state information (CSI) is given at varying rates to both the encoder and the decoder. In [15], Costa studies an additive Gaussian channel with additive Gaussian state known at only the encoder, and shows that Gel’fand–Pinsker coding with a specific auxiliary random variable, known as dirty paper coding (DPC), achieves the channel capacity. Interestingly, in this case, the DPC removes the effect of the additive channel state on the capacity as if there were no channel state present in the model or the channel state were known to the decoder as well. For a comprehensive review of state-dependent channels and related work, the reader may refer to [16].

A growing body of work studies multiuser state-dependent models. Recent advances in this regard can be found in [16]–[39], and many other works. Key to the investigation of a state-dependent model is whether the parameters controlling the channel are known to all or only some of the users in the communication model. If the parameters of the channel are known to only some of the users, the problem exhibits some asymmetry which makes its investigation more difficult in general. Also, in this case, one has to expect some rate penalty due to the lack of knowledge of the state at the uninformed encoders, relative to the case in which all encoders would be informed.

The state-dependent multiaccess channel (MAC) with only one informed encoder and degraded message sets is considered in [17], [18], and [40]–[43]; and the state-dependent RC with only informed relay is considered in [22] and [23]. For all these models, the authors develop nontrivial outer or upper bounds that permit to characterize the rate loss due to not knowing the state at the uninformed encoders. Key feature to the development of these outer or upper bounding techniques is that, in all these models, the uninformed encoder not only does not know the channel state but can learn no information about it.

The model for the RC with informed source that we study in this paper seemingly exhibits some similarities with the RC with informed relay considered in [22] and [23], and it also connects with the MAC with asymmetric channel state and degraded message sets considered in [17]–[19]. However, establishing a nontrivial upper bound for the present model is more involved, comparatively. Partly, this is because, here, one uninformed encoder (the relay) is also a receiver; and, so, it can potentially get some information about the channel states directly observing the past received sequence from the source. That is, at time $i$, the input $X^*_{2i}$ of the relay can potentially depend on the channel states through its past output $Y^*_{2i-1} = (Y^*_{2,1}, \ldots, Y^*_{2,i-1})$. For the general model in Fig. 1, the relay can even know the states noncausally, potentially. This is because $Y^*_{2i}$ may depend on future values of the state through past source inputs $X^*_{1,j}(W, S^n)$, $j = 1, \ldots, i - 1$. For the model of Fig. 2, the relay can know the states only strictly causally, but upper bounding the capacity seems still not easy.

In our recent work [44]–[46], we have shown that, in an MAC, strictly causal knowledge of the state at one encoder can be beneficial in general for the other encoder even if the latter is informed noncausally. In [45] and [46], we characterize the capacity region fully. Studying networks in which a subset of the nodes know the states noncausally and another subset know these states only strictly causally, i.e., networks with mixed—noncausal and strictly causal, states appears to be more challenging in general, and is likely to capture additional interest, especially after recent results on the utility of strictly causally known states in MACs [28], [29].

B. Main Contributions

For the general state-dependent RC with informed source shown in Fig. 1, we derive two lower bounds and an upper bound on the channel capacity. In the discrete memoryless (DM) case, the first lower bound is obtained by a block Markov coding
scheme in which the source describes the channel state to the relay and destination ahead of time. The source sends a two-layer description of the state consisting of two (possibly correlated) individual descriptions intended to be recovered at the relay and destination, respectively. The relay recovers the individual description intended to it and then utilizes the estimated state as noncausal state information at the transmitter to implement collaborative source–relay binning in subsequent blocks, through a combined decode-and-forward [47, Th. 5] and Gel’fand–Pinsker binning [13]. The destination guesses the source’s message sent cooperatively by the source and relay and the individual description which is intended to it from its output and the previously recovered state. The rationale for the coding scheme which we use for the first lower bound is that had the relay known the state with negligible distortion, then efficient cooperative source–relay binning in the spirit of [48] can be realized (recall that the model in [48] assumes availability of the state at both source and relay).

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

For the state-dependent general model, we also establish an upper bound on the capacity. This upper bound accounts for not knowing the state at the relay and the destination. Then, considering the relay model of Fig. 2, we derive a better upper bound that accounts also for the loss incurred by not knowing the state at one of the source encoder components. We show that this upper bound is strictly tighter than the max-flow min cut or cut-set upper bound. We point out the rate loss in the upper bound incurred by the availability of the channel state at only the one source encoder component. Using this upper bound, we characterize the channel capacity in a number of cases, including when the interference corrupts transmission to the destination but not to the relay.

C. Outline and Notation

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

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

II. System Model and Definitions

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

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

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

$$\phi^n_1 : \{1, \ldots , M\} \times S^n \to X^n_1$$

a sequence of encoding functions at the relay

$$\phi_{2,i} : Y^n_{2,i-1} \to X_{2,i}$$

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

$$\psi^n : Y^n_3 \to \{1, \ldots , M\}.$$

Let a $(M, n)$ code be given. The sequences $X^n_1$ and $X^n_2$ from the source and the relay, respectively, are transmitted across a state-dependent RC modeled as a memoryless conditional probability distribution $W_{Y_2,Y_3|X_1,X_2,S}$. The joint probability mass function on $W \times X^n_1 \times X^n_2 \times Y^n_2 \times Y^n_3$ is given by

$$P(w, x^n_1, x^n_2, y^n_2, y^n_3) = P(w) \prod_{i=1}^{n} Q_S(s_i) P(x_{1,i}|w,s^n_1) P(x_{2,i}|y_{2,i-1}) W_{Y_2,Y_3|X_1,X_2,S}(y_{2,i},y_{3,i}|x_{1,i},x_{2,i},s_i).$$

The destination estimates the message sent by the source from the channel output $Y^n_3$. The average probability of error is defined as $P^e_n = \mathbb{E}_S[P(\psi^n(Y^n_3) \neq W|S^n = s^n)]$.

An $(\epsilon, n, R)$ code for the state-dependent RC with informed source is an $(2^n R, n)$-code $(\phi^n_1, \phi^n_2, \psi^n)$ having average probability of error $P^e_n$ not exceeding $\epsilon$.

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

We shall also study the relay model shown in Fig. 2, in which the source alphabet $X_1 = X^n_{1R} \times X^n_{1D}$, $X^n_2 = (X^n_{2R}, X^n_{2D})$ with the input component $X^n_{1R}$ function of only the message $W$, and the input component $X^n_{2D}$ function of $(W, S^n)$, i.e., $X^n_{2R} = \phi^n_{2R}(W)$ and $X^n_{2D} = \phi^n_{2D}(W, S^n)$, and $\phi^n_{1R}$ and $\phi^n_{1D}$ are the source encoding functions, and the conditional distribution $W_{Y_2,Y_3|X_1,X_2,S}$ factorizing as (1). The encoding at the relay and the decoding at the destination remain as in the model of Fig. 1, i.e., given by (3) and (4), respectively.

III. DM RC WITH INFORMED SOURCE

In this section, we consider the general state-dependent RC model of Fig. 1. We assume that the alphabets $S, X_1, X_2, Y_2$, and $Y_3$ in the model are all discrete and finite.

A. Lower Bounds on Channel Capacity: State Description

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

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

$$R^u = \max_{R, S} \min \{I(U; Y_2|\hat{S}_R) - I(U; S, \hat{S}_D|V, \hat{S}_R) \}
\{I(U, V; Y_3|\hat{S}_D) - I(U, V, S, \hat{S}_R|\hat{S}_D)\}$$

subject to the constraints

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

$$I(S; \hat{S}_D) \leq I(U_D; Y_3, \hat{S}_D|V, U) - I(U_D; S, \hat{S}_R, \hat{S}_D|V)$$

where $|x| = \min\{x, 0\}$, and the maximization is over all joint measures on $S \times X^n_{1R} \times X^n_{1D} \times U^n_R \times U^n_D \times U \times X_2 \times Y^n_2 \times Y^n_3$ of the form (8) shown at the bottom of the page and satisfying

$$I(V; Y^n_3, \hat{S}_D) - I(V; \hat{S}_R) > 0.$$
the state $\hat{S}_R$. The allowed measure (8) implies the following Markov chains:
\[ V \leftrightarrow \hat{S}_R \leftrightarrow (S, \hat{S}_D), \quad (U, V, U_R, U_D) \leftrightarrow (X_1, X_2, S) \leftrightarrow (Y_2, Y_3). \]
(10)

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

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

**Remark 1:** The intuition for the coding scheme which we use to employ to establish the lower bound in Theorem 1 is as follows. Had the relay known the state, the source and the relay could implement collaborative binning against that state for transmission to the destination [48]. Since the source knows the state of the channel noncausally, it can transmit a description of it to the relay ahead of time. The relay recovers the state (with a certain distortion), and then utilizes it in the relevant subsequent block through a collaborative binning scheme. The hope is that the benefit that the source can get from being assisted by a more capable relay will compensate the loss caused by the source’s spending some of its resources to make the relay learn the state.

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

**Remark 2:** As can be seen from the proof in Appendix A, the source sends the descriptions intended to the relay and destination two blocks ahead of time. That is, at the beginning of block $i$, the source describes the state vector $s_{i+2}$ to the relay and destination. While one block delay is sufficient to describe the state to the relay, a minimum of two blocks is necessary for the state reconstruction at the destination because of the used window decoding technique.

**Outline of Proof of Theorem 1:**

A formal proof of Theorem 1 with complete error analysis is given in Appendix A. We now give a description of a random coding scheme which we use to obtain the lower bound given in Theorem 1. This scheme is based on an appropriate combination of block Markov encoding [47], Gel’fand–Pinsker binning [13], multiple descriptions [49], and Marton’s coding for general broadcast channels [50]–[52]. Next, we outline the encoding and decoding procedures.

We transmit in $B + 1$ blocks, each of length $n$. Let $s[i]$ denote the state sequence controlling the channel in block $i$, with $i = 1, \ldots, B + 1$. During each of the first $B$ blocks, the source encodes a message $w[i] \in \{1, 2^nR[i]\}$ and sends it over the channel. In addition, during each of the first $B - 1$ blocks, the source also sends two individual descriptions of $s[i+2]$ intended to be recovered at the relay and destination, respectively. We denote by $s_R[i], r_R \in \{1, 2^nR[i]\}$, the description of $s[i+2]$ intended to be recovered at the relay in block $i$, at rate $R_R$, and by $s_D[i], r_D \in \{1, 2^nR_D[i]\}$, the description of $s[i+2]$ intended to be recovered at the destination in block $i$, at rate $R_D$. For the last two blocks, for convenience, we set $w[B+1] = 1$, $(r_{R[B+1]}, r_{D[B+1]}) = (1, 1)$ and $(r_{R[B+1]}, r_{D[B+1]}) = (1, 1)$. For fixed $n$, the average (channel coding) rate $R(B/(B+1))$ of the information message over $B+1$ blocks approaches $R \to +\infty$, and the average (source coding) rates $R_R \to (B-1)/(B+1)$ and $R_D \to (B-1)/(B+1)$ approach $R_R$ and $R_D$, respectively, as $B \to +\infty$.

**Codebook generation:** Fix a measure $P_s\hat{s}_R, \hat{s}_D, U, U_R, U_D, V, U_V \in \mathbb{X}_1, X_2, Y_2, Y_3$ of the form (8). Calculate the marginals $P_{\hat{s}_R}$ and $P_{\hat{s}_D}$ induced by this measure. Fix $\epsilon > 0$, and let $M = 2^{n[R\epsilon]}$
\begin{align*}
J_V &= 2^n[I(V; \hat{s}_V)+\epsilon] \\
J_R &= 2^n[I(U_R; S, \hat{s}_R, \hat{s}_D, U, V)+\epsilon] \\
J_U &= 2^n[I(U; S, \hat{s}_R, \hat{s}_V, V)+\epsilon] \\
J_D &= 2^n[I(U; S, \hat{s}_R, \hat{s}_D, V, U)+\epsilon]
\end{align*}
with
\begin{align*}
R_R &= I(U; Y_2, \hat{s}_D[U, V]) - I(U; S, \hat{s}_R, \hat{s}_D, U, V) - \epsilon \\
R_D &= I(U_D; Y_3, \hat{s}_D[U, V]) - I(U_D; S, \hat{s}_R, \hat{s}_D, U, V) \\
&\quad + I(U_D; Y_3, \hat{s}_D[U, V]) - I(U; S, \hat{s}_R, \hat{s}_D, V)] - \epsilon
\end{align*}
(11)
(12)

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

We may assume that first term in (6) is nonnegative, i.e., $I(U; Y_2, \hat{s}_R V) - I(U; S, \hat{s}_R, \hat{s}_D V) \geq 0$.

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

1. Generate $2^{nR_R}$ $n$-vectors $\hat{s}_R[1], \ldots, \hat{s}_R[2^nR]$ independently according to a uniform distribution over the set $T^n(\hat{s}_R[n])$ of $n$-typical $\hat{s}_R$-vectors.
2. Generate $2^{nR_D}$ $n$-vectors $\hat{s}_D[1], \ldots, \hat{s}_D[2^nR_D]$ independently according to a uniform distribution over the set $T^n(\hat{s}_D[n])$ of $n$-typical $\hat{s}_D$-vectors.
3. Generate $J_V M$ independent and identically distributed (i.i.d.) codewords $\{w'[i, j_V]\}$ indexed by $w' = 1, \ldots, M$ and $J_V = 1, \ldots, J_V$. Each codeword $w'[i, j_V]$ is with i.i.d. components drawn according to $P_{w'}$.
4. For each codeword $w'[i, j_V]$, generate a collection of $J_U M$ codewords $\{w'[i, j_U, w, j_V]\}$ indexed by $w = 1, \ldots, M$ and $J_U = 1, \ldots, J_U$. Each codeword $w'[w, j_V, w, j_U]$ is with i.i.d. components drawn according to $P_{w' V}$. 

5) For each codeword $\mathbf{v}(w', j')$, for each codeword $\mathbf{u}_R(w', j', w, j)$, generate a collection of $J_R M_R$ codewords $\{\mathbf{u}_R(w', j', w, j, k, i, j, R)\}$ indexed by $k = 1, \ldots, M_R$ and $j = 1, \ldots, J_R$. Each codeword $\mathbf{u}_R(w', j', w, j, k, i, j, R)$ is with i.i.d. components drawn according to $P_{U_{R,w',j',w,j}}$.

6) For each codeword $\mathbf{v}(w', j')$, for each codeword $\mathbf{u}_R(w', j', w, j)$, generate a collection of $J_D M_D$ codewords $\{\mathbf{u}_D(w', j', w, j, l, j, D)\}$ indexed by $l = 1, \ldots, M_D$ and $j = 1, \ldots, J_D$. Each codeword $\mathbf{u}_D(w', j', w, j, l, j, D)$ is with i.i.d. components drawn according to $P_{U_{D,w',j',w,j}}$.

7) (Binning à-la Marton [50, 51]): For $t_R \in [1, 2^{n R_R}]$, define the cells

$$B_{t_R} = \left((t_R - 1)2^{n R_R - R_D - \epsilon} + 1, t_R2^{n R_R - R_D - \epsilon}\right].$$

Similarly, for $t_D \in [1, 2^{n R_D}]$, define the cells

$$C_{t_D} = \left((t_D - 1)2^{n R_D - R_D - \epsilon} + 1, t_D2^{n R_D - R_D - \epsilon}\right].$$

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

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

We pick up the story in block $i$. Let $w_i$ be the new message to be sent from the source node at the beginning of block $i$, and $w_{i-1}$ the message sent in the previous block $i-1$. The encoding at the beginning of block $i$ is as follows.

The source finds, if possible, a pair $\{t_{R_i}, t_{D_i}\}$ in $[1, 2^{n R_R}] \times [1, 2^{n R_D}]$ such that $(s[i] + 2, \hat{s}_R[t_{R_i}], \hat{s}_D[t_{D_i}])$ are jointly typical. If such a pair does not exist, simply set $(t_{R_i}, t_{D_i}) = (1, 1)$. We shall show that a successful encoding of $s[i] + 2$ at the source is accomplished with high probability provided $n$ is sufficiently large and

$$\hat{R}_R > I(S_i; \hat{s}_R),$$

$$\hat{R}_D > I(S_i; \hat{s}_D),$$

$$\hat{R}_R + \hat{R}_D > I(S_i; \hat{s}_R, \hat{s}_D) + I(\hat{s}_R; \hat{s}_D).$$

(13)

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

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

1) The relay knows $w_{i-1}$ and $t_{R_{i-2}}$ and finds an index $j' \in J_U$ such that $\mathbf{v}(w_{i-1}, j')$ is jointly typical with $\hat{s}_R[t_{R_{i-2}}]$. If there is more than one such index, it chooses the smallest. If there is no such index, it chooses an arbitrary index from $[1, J_U]$. Denote the chosen $j'$ by $j'_{R_i} = j' \{s[i], \hat{s}_R[t_{R_{i-2}}], w_{i-1}\}$. (Note that since $V \leftrightarrow \hat{s}_R \rightarrow \{S_i, \hat{s}_D\}$ forms a Markov chain and $\hat{s}_R[t_{R_{i-2}}]$ is jointly typical with $\{s[i], \hat{s}_R[t_{R_{i-2}}], \hat{s}_D[t_{D_{i-2}}]\}$, chosen as such, $\mathbf{v}(w_{i-1}, j'_{R_i})$ is jointly typical with $\{s[i], \hat{s}_R[t_{R_{i-2}}], \hat{s}_D[t_{D_{i-2}}]\}$, by the Markov Lemma [53, Lemma 12.1].) Then, the relay sends a vector $\mathbf{x}_2[i]$ with i.i.d. components drawn according to $\mathbf{v}(w_{i-1}, j'_{R_i})$ and $\hat{s}_R[t_{R_{i-2}}]$, drawn according to the marginal $P_{X_2 \{s[i], \hat{s}_R[t_{R_{i-2}}], \hat{s}_D[t_{D_{i-2}}]\}}$ induced by the distribution (8). (For $i = 1, 2$, the relay does not know an estimate of the channel state and so it sends some default codeword.)

2) The source first finds an index $j' \in J_U$ such that $\mathbf{u}_D(w_{i-1}, j'_{R_i}, w_i, j'_{D_i})$ is jointly typical with the vector $s[i], \hat{s}_R[t_{R_{i-2}}], \hat{s}_D[t_{D_{i-2}}]$, chosen as such, $\mathbf{v}(w_{i-1}, j'_{R_i})$ is jointly typical with $\{s[i], \hat{s}_R[t_{R_{i-2}}], \hat{s}_D[t_{D_{i-2}}]\}$, by the Markov Lemma [53, Lemma 12.1]. Then, the relay sends a vector $\mathbf{x}_2[i]$ with i.i.d. components drawn according to $\mathbf{v}(w_{i-1}, j'_{R_i})$ and $\hat{s}_R[t_{R_{i-2}}]$. Next, the source searches for one pair

$$\left(\mathbf{u}_R(w_{i-1}, j'_{R_i}, w_i, j'_{R_i}, k_i, j_{R_i}), \mathbf{u}_D(w_{i-1}, j'_{R_i}, w_i, j'_{R_i}, k_i, j_{R_i})\right)$$

where the set $D_{t_{R_i}, t_{D_i}}$ is defined by (14) at the bottom of the next page.

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

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

(15)

Denote the found pair as $\mathbf{u}_R(w_{i-1}, j'_{R_i}, w_i, j'_{R_i}, k_i, j_{R_i})$, $\mathbf{u}_D(w_{i-1}, j'_{R_i}, w_i, j'_{R_i}, k_i, j_{R_i})$.

4) The source then sends a vector $\mathbf{x}_2[i]$ with i.i.d. components given the vectors $\mathbf{v}(w_{i-1}, j'_{R_i})$, $\mathbf{u}_R(w_{i-1}, j'_{R_i}, w_i, j'_{R_i}, k_i, j_{R_i})$, $\mathbf{u}_D(w_{i-1}, j'_{R_i}, w_i, j'_{R_i}, k_i, j_{R_i})$, and $\{s[i], \hat{s}_R[t_{R_{i-2}}], \hat{s}_D[t_{D_{i-2}}]\}$, drawn according to the marginal $P_{X_2 \{s[i], \hat{s}_R[t_{R_{i-2}}], \hat{s}_D[t_{D_{i-2}}]\}}$ induced by the distribution (8).

**Decoding:** Decoding and state reconstruction at the relay are based on classical joint typicality. Decoding and state reconstruction at the destination are based on joint typicality and

\footnotetext[1]{Note that, strictly speaking, the encoder is not allowed to randomize at this stage. A more rigorous analysis consists in generating the desired input distribution at the codebook generation stage.}
The decoding and reconstruction procedures at the end of block $i$ are as follows.

1) The relay knows $w_{i,1}$ and $u_{R,i,2}$ (in fact, the relay knows also $i_{R,i-1}$ but does not use it for decoding in this step). It declares that $(\hat{w}_i, i_{R,i})$ are sent if there exists a unique triple $(\hat{w}_i, \hat{j}_{U,i}, \hat{k}_i)$, $\hat{w}_i \in [1, M'_i]$, $\hat{j}_{U,i} \in J_{U,i}$, $\hat{k}_i \in [1, M_R]$, such that $u_{U,i-1, j_{U,i-1}, w_{i-1}, i_{R,i}}$ and $u_{D}(w_{i-1, j_{D,i-1}, w_{i-1}, i_{R,i-1}})$ are jointly typical with $(y_2[i], \hat{S}_R[t_{R,i-2}], w_{i-1})$ given $v(w_{i-1, j_{D,i-1}}, \hat{w}_i, \hat{j}_{U,i}, \hat{k}_i, j_{R,i})$ are jointly typical with $(y_2[i], \hat{S}_R[t_{R,i-2}], w_{i-1})$, for some $j_{R,i} \in J_{R,i}$, where $j_{U,i} = j_{U}[s_{R}[t_{R,i-2}], w_{i-1}]$. One can show that, with the choice (12), the decoding error in this step is small for sufficiently large $n$ if

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

If (16) is satisfied, the estimate $\hat{i}_{R,i}$ of $i_{R,i}$ at the relay is the index of $\hat{S}_R$ containing the found $\hat{k}_i$, i.e., $\hat{k}_i \in B_{R,i}$. If (16) is satisfied, the estimate $\hat{i}_{R,i}$ of $i_{R,i}$ at the relay is the index of $\hat{S}_R$ containing the found $\hat{k}_i$, i.e., $\hat{k}_i \in B_{R,i}$.

2) The destination knows the pair $(w_{i-1, j_{D,i-1}})$ and the index $\hat{j}_{V,i-1} = j_V[s_{R}[t_{R,i-2}], w_{i-1}]$ and decodes the pair $(w_{i-1, j_{D,i-1}})$ based on the information received in block $i-1$ and block $i$. It declares that $(\hat{w}_{i-1, j_{D,i-1}})$ is sent if there is a unique triple $(\hat{w}_{i-1, j_{D,i-1}}, \hat{k}_{i-1}, j_{D,i-1})$, $\hat{w}_{i-1} \in [1, M'_i], \hat{j}_{D,i-1} \in J_{D,i-1}, \hat{k}_{i-1} \in [1, M_D]$, and a unique $j_{D,i-1}$ such that $u_{U,i-1, j_{U,i-1}, l_{i-1}, \hat{k}_{i-1}}$, $\hat{w}_{i-1} \in [1, M'_i]$, $\hat{j}_{U,i-1} \in J_{U,i-1}, \hat{k}_{i-1} \in [1, M_D]$, and a unique $\hat{j}_{U,i-1}$ are jointly typical with $(y_3[i-1], \hat{S}_D[t_{D,i-1}], w_{i-1})$ and $v(w_{i-1, j_{D,i-1}}, \hat{w}_{i-1}, \hat{j}_{U,i-1})$ are jointly typical with $(y_3[i-1], \hat{S}_D[t_{D,i-1}], w_{i-1})$. One can show that, with the choice (12), the decoding error in this step is small for sufficiently large $n$ if

$$R < I(V; U, Y_3, \hat{S}_D) - I(V; U, S, \hat{S}_R, \hat{S}_D). \quad (17)$$

If (17) is satisfied, the estimate $\hat{i}_{D,i-1}$ of $i_{D,i-1}$ at the destination is the index of the $\hat{C}_{i_{D,i-1}}$ containing the found $l_{i-1}$, i.e., $\hat{l}_{i-1} \in \hat{C}_{i_{D,i-1}}$. Also, the destination obtains the correct index $j_{V,i} = j_V[s_{R}[t_{R,i-2}], w_{i-1}]$.

The achievable rate in Theorem 1 requires the relay to decode the message sent by the source fully, and this can be rather a severe constraint. We can generalize Theorem 1 by allowing the relay to decode the message sent by the source only partially [54]. This can be done by splitting the information sent by the source into two independent parts: one part is sent through the relay and the other part is sent directly to the destination. In the following theorem, the random variables $V$, $U$, $U_R$, and $U_D$ play the same roles as in Theorem 1 and $U_1$ is a new random variable that represents the information directly sent to the destination.

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

$$R^b = \max \min \{I(U; Y_2, \hat{S}_R V) - I(U; S, \hat{S}_R, \hat{S}_D V), I(U; V, Y_3) - I(U; V, S, \hat{S}_R V), I(U_1; Y_3) - I(U_1; S, \hat{S}_R V)\} - I(U_1; S, \hat{S}_R V). \quad (18)$$

subject to the constraints

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

$$I(S; \hat{S}_D) \leq I(U; Y_3, \hat{S}_D U, V) - I(U; S, \hat{S}_R, \hat{S}_D U, V) + I(U_1; Y_3, \hat{S}_D V) - I(U_1; S, \hat{S}_R, \hat{S}_D V). \quad (19b)$$

$$I(S; \hat{S}_R, \hat{S}_D) + I(\hat{S}_R; \hat{S}_D) \leq I(U; Y_2, \hat{S}_R U, V) - I(U; S, \hat{S}_R, \hat{S}_D U, V) + I(U_1; Y_3, \hat{S}_D V) - I(U_1; S, \hat{S}_R, \hat{S}_D V) - I(U; U_D) + I(U_1; U_D) + I(U; V, S, \hat{S}_R, \hat{S}_D V) \quad (19c)$$

where $[x] = \max(x, 0)$, and the maximization is over all joint measures on $S \times \hat{S}_R \times \hat{S}_D \times U_R \times U_D \times U \times X_1 \times X_2 \times Y_2 \times Y_3$ of the form (20) shown at the bottom of the next page and satisfying $U_1 \leftrightarrow (V, U, S, \hat{S}_R, \hat{S}_D) \leftrightarrow U_R$ is a Markov chain and

$$0 < I(V; Y_3, \hat{S}_D) - I(V; \hat{S}_R)$$

$$0 \leq I(U; Y_2, \hat{S}_R) - I(U; S, \hat{S}_R V)$$

$$0 \leq I(U_1; Y_3, U, \hat{S}_D) - I(U_1; S, \hat{S}_R U, \hat{S}_D). \quad (20)$$

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

$$ R = \max \{ \sum \{ I(U;Y_2) + I(U;Y_3) - I(U;S|X_2) \} - I(U;U_1;S|X_2) \} $$

with the maximization over joint measures of the form

$$ P_{S,U,V,X_1,X_2,Y_2} = Q_{S} P_{X_2} P_{U} S_{X_2} P_{U_1} X_1 U_S X_2 W_{Y_2} Y_3 |X_1 X_2 S $$

and satisfying

$$ 0 \leq I(U;Y_2) - I(U;S) \leq I(U;Y_2) - I(U;S) $$

$$ 0 \leq I(U;Y_3 U) - I(U;U_1 S) \leq I(U;Y_3 U) - I(U;U_1 S) $$

B. Lower Bound on Channel Capacity: Analog Input Description

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

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

$$ R^u = \max \{ I(U;Y_3) - I(U;S) \} $$

subject to the constraint

$$ I(X;\bar{X}) < I(U;Y_2) - I(U;S) $$

where maximization is over all joint measures on $S \times U \times U_R \times X_1 \times X_2 \times X_3 \times Y_2 \times Y_3$ of the form

$$ P_{S,U,U_R,X_1,X_2,X_3,Y_2,Y_3} = Q_{S} P_{X_2} P_{U_R} U S_{X_2} P_{U_1} U S_{X_2} P_{X_3} X_2 Y_2 X_1 Y_3 S $$

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

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

$$(X_1,U_R) \leftrightarrow (U,S) \leftrightarrow (X_1, X_2, S) \leftrightarrow (Y_2,Y_3).$$

**Remark 4:** The rationale for the coding scheme which we use to obtain the lower bound in Theorem 3 is as follows. Had the relay known the message to be sent in each block and the state that corrupts the transmission in that block, then the relay generates its input using a collaborative Gel’fand–Pinsker scheme as in [48].

For our model, the source knows the message that the relay should optimally send in each block (if the relay gets the message correctly). It also knows the state sequence that corrupts the transmission in that block. It can then generate the appropriate relay input vector that the relay would send had the relay known the message and the state. The source can send this vector to the relay ahead of time, and if the relay can estimate it to high accuracy, then collaborative source–relay binning in the sense of [48] is readily realized for transmission from the source and relay to the destination.

More precisely, a block Markov encoding is used to establish Theorem 3. Let us consider transmission in two adjacent blocks $i$ and $i + 1$. In the beginning of block $i$, the source sends the information $w_i$ of the current block, and, in addition, describes to the relay the input $x[i+1]$ that the relay should send in the next block $i + 1$ had the relay known the message $w_{i+1}$ and the state $s[i+1]$. Let $\hat{x}[m]$ be a description of $x[i+1]$. The message $w_i$ and the index $m_i$ which the source sends in block $i$ are precoded using binning against the state that controls transmission in the current block $i$. The vector $x[i+1]$, however, is the input that the relay would send in the next block $i + 1$ had the relay known the state $s[i+1]$, and so is generated at the source using binning against the state $s[i+1]$. The vector $x[i+1]$ and its description which is sent to the relay during block $i$ are intended to combine coherently with the source transmission in block $i + 1$.

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

C. Upper Bound on Channel Capacity

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

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

**Theorem 4:** The capacity of the state-dependent DM RC with informed source is upper bounded by

$$R_{\text{triv}}^{\text{up}} = \max_{P(z,x)} \min \left\{ I(X_1; Y_3, S, X_2), I(X_1, X_2; Y_3|S) \right\}$$

(29)

not easy.

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

**Theorem 4:** The capacity of the state-dependent DM RC with informed source is upper bounded by

$$R_{\text{triv}}^{\text{up}} = \max_{P(z,x)} \min \left\{ I(V; Y_3, U, X_2) - I(V; S) \right\}$$

(30)

where the maximization is over measures of the form

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

(31)

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

$$|\mathcal{U}| \leq |S| |X_1| |X_2|$$

(32a)

$$|\mathcal{V}| \leq \left( |S| |X_1| |X_2| \right)^2$$

(32b)

respectively.

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

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

**Remark 6:** In the case in which $Y_2 = S$, the relay in the model of Fig. 1 has no message of its own to transmit and only acts as a helper who knows the state strictly causally. The capacity of this model can be obtained as a special case of that of the multiaccess model solved in [46]. In particular, in [46], it is shown that even though it only knows the states strictly causally, the relay can still be of some utility for the source, which knows the states fully. This special case model also has connections with the model studied in [55].

IV. DM MODEL WITH ORTHOGONAL COMPONENTS

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

A. Bounds on Channel Capacity

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

**Proposition 1:** The capacity of the state-dependent DM RC with orthogonal components of Fig. 2 is lower bounded by

$$R_{\text{orth}}^{\text{lo}} = \max \left\{ I(X_1; Y_2, X_2), I(X_1; X_2; Y_3) \right\}$$

(33)

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

$$P_{S,U_1,X_1R,X_1D,X_2,Y_3} = Q_S P_{X_1R} P_{X_1D|S} P_{X_2|S} W_{Y_2,Y_3,X_1D,X_2}$$

(34)

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

In the rate (33), the variable $U_1$ represents the Gel’fand–Pinsker auxiliary random variable associated with the information sent directly to the destination. More specifically, the message $W$ from the source is split into two independent parts: one of which is transmitted through the relay at rate $R_r$ and the other is transmitted directly to the destination without the help of the relay at rate $R_d$. The total rate is $R = R_r + R_d$. The message that is transmitted through the relay can be decoded correctly if the rate $R_r$ satisfies [47, Th. 1]

$$R_r < \min \left\{ I(X_1R; Y_2, X_2), I(X_1, X_2; Y_3) \right\}$$

(35)

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

$$R_d < I(U_1; Y_3, X_1R, X_2) - I(U_1; S; X_1R, X_2)$$

(36)
This shows that message $W$ can be sent reliably at the rate (33).

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

**Theorem 5:** The capacity of the state-dependent DM RC with orthogonal components of Fig. 2 is upper bounded by

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

(37)

where the maximization is over all joint measures of the form

$$P_{S,X_{1R},X_{1D},X_2,Y_2,Y_3} = Q_{S} P_{X_{1R}} P_{X_{1D}} P_{X_2} S W_{Y_2} X_{1R} S W_{Y_3} X_{1D} X_2 S.$$  

(38)

**Proof:** The proof of Theorem 5 appears in Appendix D. Observe that the second term of the minimization in (37) upper bounds the information that the source and the relay can send to the destination by

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

(39)

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

**B. Comments and Digression**

There is a connection between the state-dependent relay model of Fig. 2 and a state-dependent two-user multiaccess model with degraded message sets that we treated recently in [44]–[46]. In particular, setting in the multiaccess model of [44]–[46], the channel state are known noncausally to one of the encoders and only strictly causally to the other encoder. Also, both encoders transmit a common message and, in addition, the encoder that knows the states noncausally transmits an individual message. In [44], we derive bounds on the capacity region; and in [45] and [46], we characterize the full capacity region of this multiaccess model. In [44]–[46], we show that the knowledge of the states only strictly causally at the encoder that sends only the common message can increase the capacity region in general. We also observe that the capacity region is increased even in the extreme case in which the encoder that knows the states only strictly causally has no message to transmit (i.e., common-message rate equal to zero). This suggests that in the relay model of Fig. 2, although it can only know the states strictly causally, the relay can potentially help the source combat the effect of the state (in addition to its classic role of relaying the information message). Although it is not clear yet how the relay could exploit optimally the information about the strictly causal part of the state sequence that it can get by observing its output, the upper bound in Theorem 5 makes one step ahead toward this end; by bounding the information that the source and relay can transmit cooperatively; and so, in a sense, the capacity increase that the source can get through the relay’s help.

**V. MEMORYLESS GAUSSIAN RC WITH INFORMED SOURCE**

**A. System Model**

In this section, we consider a full-duplex state-dependent RC informed source in which the channel state and the noise are additive and Gaussian. In this model, the channel state can model an additive Gaussian interference which is assumed to be known (noncausally) to only the source. The channel outputs $Y_{z,i}$ 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 $S_{z,i}$ by

$$Y_{z,i} = X_{1,i} + S_{z,i}$$

(40a)

$$Y_{z,i} = X_{1,i} + X_{2,i} + S_{z,i}.$$  

(40b)

The channel state $S_{z,i}$ is zero-mean Gaussian random variable with variance $Q_{z,i}$ and only the source knows the state sequence $S^n$ (noncausally). The noises $Z_{z,i}$ are zero-mean Gaussian random variables with variances $N_{z,i}$ and $N_{z,i}$, respectively; and they are mutually independent and independent from the state sequence $S^n$ and the channel inputs $\{X_1^n, X_2^n\}$. Also, we consider the following individual power constraints on the average transmitted power at the source and the relay

$$\sum_{i=1}^{n} X_{1,i}^2 \leq n P_1,$$

$$\sum_{i=1}^{n} X_{2,i}^2 \leq n P_2.$$  

(41)

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

**B. Bounds on Channel Capacity**

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

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

$$R_{\text{GI}}^{\text{lo}} = \max \left\{ \frac{1}{2} \log \left( 1 + \frac{\sqrt{P_1} + \sqrt{P_2} - D^{1/2}}{N_2 + D + P_1} \right) \right\}$$

(42)

where

$$D := P_2 \frac{N_2}{N_2 + P_1}.$$  

(43)

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

**Remark 7:** It is insightful to observe that the rate in Theorem 6 does not depend on the strength of the state $S$. This makes the coding scheme appreciable, particularly for the case of arbitrary strong interference in which classical coding schemes suffer greatly from the strong interference unknown at the relay.

**Outline of Proof of Theorem 6:** The result in Theorem 3 for the DM case can be extended to memoryless channels with discrete time and continuous alphabets using standard techniques [56, Ch. 7]. The proof of Theorem 6 follows through evaluation of the lower bound of Theorem 3 using the following jointly Gaussian input distribution. For $0 \leq \gamma \leq 1$, we let $X \sim
$N(0, P_2)$ and $X_{1R} \sim N(0, \gamma P_1)$, with $X$ jointly Gaussian with $S$ with $E[XS] = 0$; and $X_{1R}$ jointly Gaussian with $(S, X)$, with $E[X_{1R}S] = E[X_{1R} X] = 0$. Also, for $0 \leq t \leq D \leq P_2$ given, we consider the test channel $\tilde{X} = aX + \tilde{X}$, where $a := 1 - D/P_2$ and $\tilde{X}$ is a Gaussian random variable with zero mean and variance $\tilde{P}_2 = D(1 - D/P_2)$, independent from $X$ and $S$. Using this test channel, we calculate $E[|X - \tilde{X}|^2] = D$ and $E[|\tilde{X}|^2] = P_2 - D$.

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

\[
U = \left(\sqrt{\frac{\gamma P_1}{P_2}}, \sqrt{\frac{P_2 - D}{P_2}}\right) X + \alpha S \tag{44}
\]

\[
U_R = X_{1R} + \alpha_R \left(S + \sqrt{\frac{\gamma P_1}{P_2}} X\right) \tag{45}
\]

where

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

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

**Alternative Proof of Theorem 6:** The encoding and transmission scheme is as follows. For $0 \leq \gamma \leq 1$, let $X \sim N(0, P_2)$ and $X_{1R} \sim N(0, \gamma P_1)$, with $X$ jointly Gaussian with $S$ with $E[XS] = 0$; and $X_{1R}$ jointly Gaussian with $(S, X)$, with $E[X_{1R}S] = E[X_{1R} X] = 0$. Also, let $0 \leq t \leq D \leq P_2$ be given, and consider the test channel $\tilde{X} = aX + \tilde{X}$, where $a := 1 - D/P_2$ and $\tilde{X}$ is a Gaussian random variable with zero mean and variance $\tilde{P}_2 = D(1 - D/P_2)$, independent from $X$ and $S$. Using this test channel, we calculate $E[|X - \tilde{X}|^2] = D$ and $E[|\tilde{X}|^2] = P_2 - D$. We use the two random variables $U$ and $U_R$ given by (45) to generate the auxiliary codewords $U_i$ and $U_{R,i}$, which will be used in the sequel.

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

\[
\left(\sqrt{\frac{\gamma P_1}{P_2}} + \frac{P_2 - D}{P_2}\right) x[i] = u[i] - \alpha s[i]. \tag{47}
\]

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

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

\[
\hat{R} = I(X; \hat{X}) = \frac{1}{2} \log\left(\frac{P_2}{D}\right). \tag{48}
\]

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

\[
x_1[i] = x_{1R}[m_i] + \sqrt{\frac{\gamma P_1}{P_2}} x[w_i]. \tag{49}
\]

In (49), the signal $x_{1R}[m_i]$ carries message $m_i$ and is obtained via a DPC considering $(s[i], x[w_i])$ as noncausal CSI, as

\[
x_{1R}[m_i] = u_R[i] - \alpha_R(s[i] + \sqrt{\frac{\gamma P_1}{P_2}} x[w_i]) \tag{50}
\]

where the components of $u_R[i]$ are generated i.i.d. using the auxiliary random variable $U_R$.

In the beginning of block $i$, the relay has decoded message $m_{i-1}$ correctly (this will be justified below) and sends

\[
x_2[i] = \frac{\sqrt{P_2}}{\sqrt{P_2 - D}} \hat{x}[m_{i-1}]. \tag{51}
\]

For the decoding arguments at the source and the relay, we give simple arguments based on intuition (the rigorous decoding uses joint typicality). Also, since all the random variables are i.i.d., we sometimes omit the time index. The relay decodes the index $m_i$ from the received $y_3[i]$ at the end of block $i$. Since signal $x_{1R}[m_i]$ is precoded at the source against the interference caused by the information message $w_i$, decoding at the relay can be done reliably as long as $n$ is large and

\[
\hat{R} \leq \frac{1}{2} \log \left(1 + \frac{\gamma P_1}{N_0}\right). \tag{52}
\]

The destination decodes message $w_i$ from the received $y_3[i]$ at the end of block $i$, considering signal $x_{1R}[m_i]$ as unknown noise, with

\[
y_3[i] = x_1[i] + x_2[i] + s[i] + z_3[i] = \left(\sqrt{\frac{\gamma P_1}{P_2}} x[w_i] + \sqrt{\frac{P_2}{P_2 - D}} \hat{x}[m_{i-1}]\right) + s[i] + (z_3[i] + x_{1R}[m_i]). \tag{53}
\]

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

\[
x'[i] = E\left[\sqrt{\frac{\gamma P_1}{P_2}} x[w_i] + \sqrt{\frac{P_2}{P_2 - D}} \hat{x}[m_{i-1}] \mid x[i]\right] \tag{54}
\]

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

We can alternatively write the output $y_3[i]$ in (53) as

\[
y_3[i] = \xi x[w_i] + s[i] + (z_3[i] + e_x[i] + x_{1R}[m_i]) \tag{56}
\]

where

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

and $e_x[i]$ is Gaussian with variance $D$ and is independent of $x[w_i]$ and $s[i]$. 

Now, considering the equivalent form (56) of the output $y_3[i]$, it is easy to see that the destination can decode message $w_i$ correctly at the end of block $i$ as long as $n$ is large and

$$R \leq I(U; Y_3) - I(U; S) = \frac{1}{2} \log \left( 1 + \frac{\sqrt{\rho_{12}^2 + \sqrt{\rho_{12}^2}} \frac{N_3}{N_3 + D + \gamma P_1} }{D} \right). \quad (58)$$

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

$$D \geq P_2 \frac{N_2}{N_2 + \gamma P_1}. \quad (59)$$

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

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

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

**Definition 1:**

$$\tilde{Q}_S(t, Q, D) := (1 - t)^2 Q - t(t - 2)D,$$

$$R(\alpha, P, Q, N) := \frac{1}{2} \log \left( \frac{P(P + Q + N)}{PQ(1 - \alpha)^2 + N(P + \alpha^2 Q)} \right)$$

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

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

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

$$R_{III}^U = \max \min \left\{ R \left( \alpha, (1 - \rho_{12}^2 - \rho_{12}^2), \beta P_{1r}, \epsilon^2 \tilde{Q}, N_2 + \beta P_{1r}, P_{1d} \right) \right\}$$

$$= \max \min \left\{ R \left( \alpha, (1 - \rho_{12}^2 - \rho_{12}^2), \beta P_{1r}, \epsilon^2 \tilde{Q}, N_3 + \beta P_{1r}, P_{1d} \right) + \frac{1}{2} \log \left( 1 + \frac{(\rho_{12}^2 \sqrt{\beta P_{1r}} + \sqrt{\rho_{12}^2})^2}{N_3 + \epsilon^2 D + \rho_{12}^2 \beta P_{1r} + (1 - \rho_{12}^2 - \rho_{12}^2) \beta P_{1r} + P_{1d}} \right) \right\} \right\}$$

$$\neq \frac{1}{2} \log \left( 1 + \frac{P_{1d}}{N_3 + \beta P_{1r}} \right) \quad (60)$$

where

$$D = Q \frac{N_2 + P_{1d}}{N_2 + \beta P_{1r} + P_{1d}}, \quad \xi = 1 + \rho_{12} \sqrt{\beta P_{1r}} / Q, \quad (61)$$

$$\alpha_2 = \frac{(\rho_{12}^2 \sqrt{\beta P_{1r}} + \sqrt{\rho_{12}^2})^2}{N_3 + P_{1r} + P_{1d} + 2 \rho_{12} \sqrt{\beta P_{1r}} P_2 - \rho_{12}^2 \beta P_{1r} + \epsilon^2 D}, \quad (62)$$

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

**Proof:** A formal proof of Theorem 7 appears in Appendix E. An outline of proof of Theorem 7 is as follows. The result in Theorem 1 for the DM case can be extended to memoryless channels with discrete time and continuous alphabets using standard techniques [56, Ch. 7]. For the state-dependent Gaussian RC (40), we evaluate the rate (6) with the following choice of input distribution. We choose $\tilde{S}_D = \emptyset$ and $U_D = \emptyset$. Furthermore, we consider the test channel $\tilde{S}_R = \alpha S + \tilde{S}_R$, where $\alpha := 1 - D/Q$ and $\tilde{S}_R$ is a Gaussian random variable with zero mean and variance $\sigma_{\tilde{S}_R}^2 = D(1 - D/Q)$, independent from $S$. The random variable $X_2$ is Gaussian with zero mean and variance $P_2$, independent of $S$ and $\tilde{S}_R$. The random variable $X_1$ is composed of three parts: $X_1 = X_{SR} + X_{WR} + X_{WD}$, where $X_{SR}$ is Gaussian with zero mean and variance $\theta P_{1r}$, for some $\theta \in [0, 1]$, is independent of $S$, $\tilde{S}_R$, $X_2$; and $X_{WR} = (\rho_{12} \sqrt{\beta P_{1r}} / Q) S + \rho_{12} \sqrt{\beta P_{1r}} P_{1r} X_2 + X_{WR}^H$, where $X_{WR}$ is Gaussian with zero mean and variance $(1 - \rho_{12}^2) \beta P_{1r}$, for some $\rho_{12} \in [0, 1]$ and $\rho_{12} \times [-1, 0]$ and is independent of $X_{SR}$, $X_2$, and $(S, \tilde{S}_R)$; and $X_{WD}$ is a Gaussian with zero mean and variance $P_{1d}$, chosen independently from all the other variables. The auxiliary random variables are chosen as

$$V = (\rho_{12} \sqrt{\beta P_{1r}} / P_2 + 1) X_2 + \alpha_2 \left( \rho_{12} \sqrt{\beta P_{1r}} / Q + 1 \right) \tilde{S}_R \quad (64a)$$

$$U = X_{WR} + \alpha \xi (S - \alpha_2 \tilde{S}_R) \quad (64b)$$

$$U_1 = X_{WD} + \frac{P_{1d}}{P_{1d} + N_2 + \beta P_{1r}} \xi (1 - \alpha)(S - \alpha_2 \tilde{S}_R) \quad (64c)$$

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

with

$$D = Q \frac{N_2 + P_{1d}}{N_2 + \beta P_{1r} + P_{1d}}, \quad \xi = 1 + \rho_{12} \sqrt{\beta P_{1r}} / Q \quad (65)$$

and $\alpha_2$ is given by (66) shown at the bottom of the next page.

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

**Remark 8:** The parameter $\alpha$ in Theorem 7 stands for DPC’s scale factor in precoding the information message against the interference on its way to the relay and to the destination. Because the model (40) has the links to the relay and to the destination corrupted by noise terms with distinct variances, one cannot remove the effect of the interference on the two links simultaneously via one single DPC as in [20]. This explains why the parameter $\alpha$ is left to be optimized over in (60). However, in the spirit of [20], one can improve the rate of Theorem 7 by...
time-sharing coding schemes that are similar to the one we employed for Theorem 7 but with different inflation parameters tailored, respectively, for the link to the relay and the link to the destination, as in [25].

Similar to the general DM model of Section III, in the general Gaussian model (40), the relay does not know the states of the channel directly but can potentially get information about \( S^n \) from the observed output sequence \( Y_2^{n-1} \). Also, \( Y_2^{n-1} \) may even contain information about future values of the state, and this makes establishing upper bounds on the capacity that are strictly better than the cut-set upper bound more difficult. Note that the cut-set upper bound is in general non-tight essentially because both \( X_1 \) and \( X_2 \) know the state \( S \) in (67).

C. Analysis of Some Extreme Cases

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

1) If \( N_2 \to 0 \), e.g., the relay is located spatially very close to the source, the lower bound of Theorem 6 and the cut-set upper bound (67) tend asymptotically to the same value

\[
C_G = \frac{1}{2} \log \left( 1 + \frac{(\sqrt{P_1} + \sqrt{P_2})^2}{N_3} \right) - o(1)
\]

(68)

where \( o(1) \to 0 \) as \( N_2 \to 0 \). Equation (68) reflects the rationale for our coding scheme for the lower bound in Theorem 6 which is tailored to be asymptotically optimal whenever the relay can learn with negligible distortion the input that it should send. In this case, the rate (68) can be interpreted as the information between two transmit antennas which both know the channel state and one receive antenna. (For comparison, note that the coding scheme of Theorem 7 achieves rate smaller than that of Theorem 6 if \( N_2 \to 0 \), because even though with the coding scheme of Theorem 7 as well the relay obtains the state estimate at almost no expense if \( N_2 \) is arbitrarily small, it also needs to know the information message to perform binning, however.)

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

\[
R_G^{\text{up}} = \max_{P_1, P_2} \min I(X_1; Y_2; Y_3|S, X_2), I(X_1, X_2; Y_3|S) \}
\]

(67)

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

3) If \( N_2 \to \infty \), i.e., the link to the relay is broken or too noisy, the cut-set upper bound (67) and the lower of Theorem 7 agree and give the channel capacity

\[
C_G = \frac{1}{2} \log \left( 1 + \frac{P_1}{N_3} \right),
\]

(70)

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

\[
R_G^{\text{up}} = \frac{1}{2} \log \left( 1 + \frac{P_1}{N_3 + P_2} \right).
\]

(71)

This is because the distortion in Theorem 6 is equal to its maximum value \( P_2 \) in this case. Equation (71) reflects a limitation of our coding scheme for the lower bound in Theorem 6 if the relay fails to reconstruct the input described by the source. In this case, the input from the relay acts as additional noise at the destination, thus causing the cooperative transmission to perform worse than simple direct transmission. The achievable rate (71) is, however, still better than had the state been merely treated as unknown noise if \( P_1 \leq Q \). (For comparison, note that the lower bound of Theorem 7 vanishes if \( N_2 \to \infty \).)

VI. MEMORYLESS GAUSSIAN MODEL WITH ORTHOGONAL COMPONENTS

In this section, we study an important class of state-dependent Gaussian RCs with orthogonal components. In this model, the source input \( X_1, i \) is independent of the channel state \( S^n \), and the channel outputs \( Y_2, i \) and \( Y_3, i \) at time instant \( t \) for the relay and the destination, respectively, are related to the channel inputs from the source and relay and the channel state \( S \) by

\[
Y_{2, i} = X_{1, i} + S_i + Z_{2, i}
\]

(72a)

\[
Y_{3, i} = X_{1, i} + X_{2, i} + S_i + Z_{3, i}
\]

(72b)

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

\[
\sum_{i=1}^{n} X_{1, i}^2 \leq nP_1, \sum_{i=1}^{n} X_{2, i}^2 \leq nP_2,
\]

(73)

The definition of a code for this Gaussian model follows that for the discrete case of Section IV, with the additional constraint that the channel inputs should satisfy the power constraint (73).

A. Bounds on Channel Capacity

The following proposition provides a lower bound on the capacity of the state-dependent Gaussian relay model (72).

\[
\alpha_2 = \frac{(\rho_{12} \sqrt{P_1} + \sqrt{P_2})^2}{(\rho_{12} \sqrt{P_1} + \sqrt{P_2})^2 + (1 - \rho_{12}^2 - \rho_{13}^2) \sqrt{P_1} + (N_3 + \theta D + \theta P_1) + \theta P_2)}
\]

(66)
Proposition 2: The capacity of the state-dependent Gaussian relay model (72) is lower bounded by

\[ R_{G-orth}^{lo} = \max \min \left\{ \frac{1}{2} \log \left(1 + \frac{P_{HR}}{N_3 + Q}\right), \frac{1}{2} \log \left(1 + \frac{P_{ID}(1 - \rho_{12}^2 - \rho_{1s}^2)}{N_3}\right) \right\} \]

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

\[ \rho_{12}^2 + \rho_{1s}^2 \leq 1. \] (75)

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

We now turn to establish an upper bound on the capacity of the Gaussian model (72). It is easy to show that the cut-set-upper bound (67) can be written as

\[ R_{G-orth}^{up} = \max_{p(x_2,s)p(x_1|s,x_2)p(x_1,n,s_2)} \min \left\{ I(X_1H_1; Y_2|S, X_2), I(X_2; Y_3|S) \right\} \]

in this case. In what follows, we establish an upper bound that is strictly better than (76) by accounting for that the source input component \( X_1H_1 \) at time \( i \) does not know the state \( S^n \) at all and that the relay output \( Y_2^{-1} \) is function of only the strictly causal part of the state in this case. The following theorem states the corresponding result.

Theorem 8: The capacity of the state-dependent Gaussian relay model (72) is upper bounded by

\[ R_{G-orth}^{up} = \max \min \left\{ \frac{1}{2} \log \left(1 + \frac{P_{HR}}{N_2}\right), \frac{1}{2} \log \left(1 + \frac{P_{ID}(1 - \rho_{12}^2 - \rho_{1s}^2)}{N_3}\right) \right\} \]

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

\[ \rho_{12}^2 + \rho_{1s}^2 \leq 1. \] (78)

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

Remark 9: Similar to the DM case, the upper bound in Theorem 8 improves upon the cut-set upper bound through the second term of the minimization. The second term of the minimization is strictly tighter than that of the cut-set upper bound because it accounts for the rate loss incurred by not knowing the state \( S^n \) at all at the source encoder component \( X_1H_1 \), that is heard at the relay and that the relay output \( Y_2^{-1} \) can depend on the state only strictly causally in this case. Further, investigating closely the proof in Appendix G, it can be seen that, by opposition to the corresponding DM case, the relay ignores completely any information about the state in the multiaccess part of (77).

B. Capacity for Some Special Cases

In this section, we characterize the capacity for some special Gaussian models. The achievable rate of Proposition 2 differs from the upper bound of Theorem 8 only through the logarithm term in (74) in which the state is taken as unknown noise in the lower bound. Substituting \( \rho := \rho_{1s} \) and \( \zeta := 1 - \rho_{12}^2 - \rho_{1s}^2 \) in (77) and (74), it is easy to see that if \( P_{HR}, P_{ID}, P_2, Q, N_2, \) and \( N_3 \) satisfy

\[ N_2 < \max_{\zeta \in [0, 1]} \frac{P_{HR}P_{ID}\zeta + (\sqrt{Q} + \rho\sqrt{P_{ID}})^2 + N_3}{\sqrt{P_2} + \sqrt{1 - \zeta - \rho^2\sqrt{P_{ID}}}^2} - Q \] (79)

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

\[ C_{G-orth} \]

\[ \frac{1}{2} \log \left(1 + \frac{\sqrt{P_2} + \sqrt{1 - \zeta - \rho^2\sqrt{P_{ID}}}^2 + N_3}{\sqrt{P_2} + \sqrt{1 - \zeta - \rho^2\sqrt{P_{ID}}}^2} + \frac{1}{N_3} \right). \] (80)

Let us now consider an important special case of (72) in which the interference affects only the channel to the destination, i.e.,

\[ Y_{2,i} = X_{1H,i} + Z_{2,i}, \quad Y_{3,i} = X_{1D,i} + Y_{2,i} + S_i + Z_{3,i}. \] (81a, 81b)

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

Theorem 9: The capacity of the state-dependent Gaussian relay model (81) is given by

\[ C_{G-orth} = \max \min \left\{ \frac{1}{2} \log \left(1 + \frac{P_{HR}}{N_2}\right), \frac{1}{2} \log \left(1 + \frac{P_{ID}(1 - \rho_{12}^2 - \rho_{1s}^2)}{N_3}\right) \right\} \]

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

\[ \rho_{12}^2 + \rho_{1s}^2 \leq 1. \] (83)

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

Another important special case of the state-dependent Gaussian relay model of Fig. 2 is one such that \( Y_3 = (Y_3^{(1)}, Y_3^{(2)}) \) and the conditional distribution \( W_{Y_3^{(1)}|X_1H_1,S,x_2} \) factorizes as \( W_{Y_3^{(1)}|x_2} W_{Y_3^{(2)}|X_1H_1,S} \)

\[ Y_{3,1}^{(1)} = X_{1H,1} + S_1 + Z_{3,1}, \quad Y_{3,2}^{(2)} = X_{1H,2} + Z_{3,2}. \] (84a, 84b, 84c)
where the noises $Z_{31}^{(1)}$ and $Z_{32}^{(2)}$ are zero-mean Gaussian random variables with variances $N_3$, and are mutually independent and independent from the state sequence $S^n$, the source input $X_n^r = (X_{1R}^n, X_{1U}^n)$, and the relay input $X_n^g$. Considering average power constraint $\sum_{i=1}^n X_{1i}^2 \leq nP_1$ on $X_n^r$ and $\sum_{i=1}^n X_{2i}^2 \leq nP_2$ on $X_n^g$, the following corollary states the capacity of this model.

**Corollary 1:** The capacity of the state-dependent Gaussian relay model (84) is given by

$$C_{G, \text{Orth}} = \max \min \left\{ \frac{1}{2} \log \left( 1 + \frac{P_1}{N_1} \right), \frac{1}{2} \log \left( 1 + \frac{P_2}{N_2} \right) \right\} + \frac{1}{2} \log \left( 1 + \frac{(1-\gamma)P_1}{N_3} \right)$$

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

The proof of Corollary 1 follows by specializing the cut-set upper bound to the model (84) and then observing that this upper bound can actually be attained using a combination of binning and generalized block Markov scheme where we let $X_{1R}$ and $X_{1U}$ to be zero-mean Gaussian with variances $\gamma P_1$ and $(1-\gamma)P_1$, respectively, for some $0 \leq \gamma \leq 1$, independent of $S$ and $X_2$; $X_2$ is zero-mean Gaussian with variance $P_2$, independent of $S$; and $X_{1R}$ and $X_{1U}$ obtained with standard DPCs for the links to the relay and to the receiver component $Y_2^{(3)}$, respectively. The source sends information to the receiver via the relay through the dirty paper coded $X_{1R}$, and independent information via the direct link through the dirty paper coded $X_{1U}$.

**Extreme cases:**

1) *Arbitrarily strong channel state:* In the asymptotic case $Q \to \infty$, the capacity of the model (72) is given by

$$C_{G, \text{Orth}} = \frac{1}{2} \log \left( 1 + \frac{P_1}{N_1} \right).$$

This can be easily seen since both the upper bound of Theorem 9 and the lower bound (74) tend to the RHS of (86) in this case. The RHS of (86) is also clearly achievable by turning the relay OFF and applying standard DPC at the source.

2) *If $N_2 \to \infty,*$ i.e., the link to the relay is broken or too noisy, the lower and upper bounds on the capacity of the model (72) agree and give the channel capacity as the RHS of (86).

**VII. NUMERICAL EXAMPLES AND DISCUSSION**

In this section, we discuss some numerical examples, for the general Gaussian RC with informed source (40), the model (72), and the special case (81). We illustrate the results of Theorems 5–8 and, for the model (40), we also include comparisons with previously known achievable rates for this model such as that obtained using compress-and-forward (CF) and binning in [34, Th. 4] and that with partial decode-and-forward and binning in [25, Th. 3].

Fig. 3 illustrates the lower bound of Theorem 6 and the lower bound of Theorem 7 for the model (40), as functions of the signal-to-noise ratio (SNR) at the relay, i.e., $\text{SNR} = P_1/N_2$, for $P_1 = 10$ dB and $Q = 15$ dB. The figure shows that the lower bound of Theorem 6 is asymptotically optimal at small SNR, and the lower bound of Theorem 7 is asymptotically optimal at small SNR. This shows the relevance of transmitting to the relay only a description of the appropriate input that it should send upon sending to it a description of the state itself at large SNR. At moderate SNR, however, sending a description of the state to the relay may improve upon sending to it a description of the appropriate Gel’fand–Pinsker binned codeword that it should send. (How the two bounds compare depends essentially on the strength of the state.) For example, at large SNR, the stronger the state, the larger the advantage of the lower bound of Theorem 7 upon that of Theorem 7.) Furthermore, the figure also shows that the lower bound of Theorem 7 is better than that of [25, Th. 3], thereby reflecting the utility of describing the state to the relay (recall that the coding scheme that we employed for the lower bound of Theorem 7 involves also a partial cancellation of the state by the source to the relay so that the relay benefits from it and the source benefits in turn). Fig. 4 shows similar bounds computed for an example degraded Gaussian RC.

**Remark 10:** The lower bound of Theorem 6 is asymptotically close to optimal in $\text{SNR}$ as we mentioned in Section V-C and is visible from Fig. 3. This is because the appropriate relay input, which is precoded at the source against the state and is encoded in a manner that it should combine coherently with the source transmission in next block, can be sent by the source to the relay at almost no expense in power and can be learned by the relay with negligible distortion in this case. One can be tempted to
expect a similar behavior for the lower bound of Theorem 7 since, for the latter as well, the relay can learn a “good” estimate of the state at almost no expense in source’s power and with negligible distortion. This should not be, however, since our coding scheme for Theorem 7 requires the relay to also decode the source’s information message. Related to this aspect, the effect of the limitation which we mentioned in Remark 8 is visible at large $\text{SNR}$ for this lower bound.

Fig. 5 illustrates the upper bound (77) of Theorem 8 and the lower bound (74) for the model (72). For comparison, the figure shows also the cut-set upper bound had the state been known also at the relay and the destination, i.e., (76), and the trivial lower bound obtained by considering the channel state as unknown noise and using a generalized block Markov coding scheme as in [1]. The curves are plotted against the SNR at the relay, i.e., $\text{SNR} = P_{1R}/N_2$ (in decibels). Observe that the upper bound (77) is strictly better than the cut-set upper bound. The improvement is due to that the upper bound (77) accounts for some inevitable rate loss which is caused by not knowing the state at the relay, as we mentioned previously. Also, the improvement is visible mainly at small to relatively large values of $\text{SNR}$.

Fig. 6 illustrates the capacity result of (81) as given by Theorem 9, as a function of the SNR in the link source to relay of $P_{1R}/N_2$ (in decibels). Also shown for comparison are the cut-set upper bound and the trivial lower bound obtained by considering the channel state as unknown noise and using a generalized block Markov coding scheme as in [1].

VIII. CONCLUSION AND DISCUSSION

In this paper, we consider a state-dependent RC with the channel states available noncausally at only the source, i.e., neither at the relay nor at the destination. We refer to this communication model as state-dependent RC with informed source. This setup may model some scenarios of node cooperation over wireless networks with some of the terminals equipped with cognition capabilities that enable estimating to high accuracy the states of the channel.
We investigate this problem in the DM case and in the Gaussian case. For both cases, we derive lower and upper bounds on the channel capacity. A key feature of the model we study is that, assuming decode-and-forward relaying, the input of the relay should be generated using binning against the state that controls the channel in order to combat its effect and, at the same time, combine coherently with the source transmission. We develop two lower bounds on the capacity by using coding schemes which achieve this goal differently. In the first coding scheme, the source describes the channel state to the relay and to the destination, through a combined coding for multiple descriptions, binning, and decode-and-forward scheme. The relay guesses an estimate of the transmitted information message and of the channel state and then utilizes the state estimate to perform cooperative binning with the source for sending the information message. The destination utilizes its output and the already recovered state to guess an estimate of the currently transmitted message and state description. In the second coding scheme, the source describes to the relay the appropriate input that the relay would send had the relay known the channel state. The relay then simply guesses this input and sends it in the appropriate subsequent block. The lower bound obtained with this scheme achieves close to optimal for some special cases.

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

We close this paper with a discussion on related aspects. Our coding scheme of Theorem 1 is, in essence, of decode-and-forward relaying type (though the relay also sends a compression version of the state on top of the decoded information message). Our coding scheme of Theorem 3 can be seen as being more of a nonstandard CF relaying type, since the relay sends a compressed version of the input produced at the source. Although not optimal in general, these schemes are tailored specifically to deal (at least partially) with the presence of the channel state in our model. The relay can of course employ other relaying schemes to assist the source, such as estimate-and-forward, amplify-and-forward, or combinations of these. However, while these schemes may outperform the schemes that we described in this paper for certain channel parameters, in general they do not really offer inherently better mechanisms of dealing with the presence of the channel state and exploiting its full knowledge at the source. In the case of states known causally or only strictly causally, the new noisy network coding by Lim et al. [57] and quantize-map-and-forward by Avestimeher et al. [58], which implement standard compression without Wyner–Ziv binning, have been proved to in general offer better rates for certain related relay [32] and multiaccess [32], [44]–[46] models. For the model at hand, however, like for the standard state-independent three-terminal RC, noisy network coding offers exactly the same rate as classic CF at the relay, but no better, as observed recently in [59].

Appendix A

Proof of Theorem 1

Throughout this appendix, we denote the set of strongly jointly \( \epsilon \)-typical sequences [60, Ch. 14.2] with respect to the distribution \( P^n_{X,Y} \) as \( T^n_{\epsilon}(P^n_{X,Y}) \). Sometimes, when the considered probability distribution is clear from the context, we shall denote this simply as \( T^n_{\epsilon} \).

Consider the random coding scheme that we outlined in Section III. We now analyze the average probability of error. Analysis of Probability of Error: The average probability of error is given by

\[
\Pr\{\text{Error}\} = \sum_{s \in S^n} \Pr(s) \Pr(\text{error}|s) \\
\leq \sum_{s \notin T^n_{\epsilon}(Q_s)} \Pr(s) + \sum_{s \in T^n_{\epsilon}(Q_s)} \Pr(s) \Pr(\text{error}|s).
\]

The first term, \( \Pr(s \notin T^n_{\epsilon}(Q_s)) \), on the RHS of (A-1) goes to zero as \( n \to +\infty \), by the strong asymptotic equipartition property (AEP) [60, p. 384]. Thus, it is sufficient to upper bound the second term on the RHS of (A-1).

We now examine the probabilities of the error events associated with the encoding and decoding procedures. The error event is contained in the union of the following error events, where the events \( E_{i,j} \) and \( E_{i,i} \) correspond to encoding errors at block \( i \); the events \( E_{k,i} \), \( k = 3, \ldots, 6 \), correspond to decoding errors at the relay at block \( i \); and the events \( E_{k,i} \), \( k = 7, \ldots, 13 \), correspond to decoding errors at the destination at block \( i \).

We note that the indices \( j_{i}^{v_1} \) and \( j_{i}^{n_1} \) are random. The decoding procedure at the relay involves computing the index \( j_{i}^{v_1} \) and decoding explicitly the index \( j_{i}^{n_1} \); and the decoding procedure at the destination involves decoding explicitly both indices. The analysis of error events that involve explicit decoding of random binning indices in the context of state-dependent channels needs some care. This is addressed explicitly in [61, pp. 854–855] for an example network. The approach of [61] relies essentially on the two lemmas, Lemma 1 and Lemma 2, therein as well as their proofs. A particular key element in the proof of [61, Lemma 1] is an upper bound on the probability that the random index (which is a message in [61]) takes a specific value given a specific state vector and a specific choice of the codebook \( C \). In what follows, the analysis of the error events \( E_{k,i} \), \( k = 5, 9, 10, 11 \), follows in a way that is essentially similar to the analysis of the event \( E_3 \) in [61, pp. 854–855], with minor modifications, as well as standard arguments on jointly typical sequences. For the sake of brevity, in the analysis of each of the error events that will follow, we will outline the steps that differ from [61] and refer to [61] each time the analysis is
For convenience, let us denote, with a slight abuse of notation, \( \tilde{S}[i] := (S[i], \tilde{S}_R[i], \tilde{S}_D[i]) \).

1) Let \( E^{(1)}_{1i} = E^{(1)}_{1i} \cup E^{(2)}_{1i} \cup E^{(3)}_{1i} \), with

\[
E^{(1)}_{1i} = \left\{ (s[i+2], \tilde{S}_R[i], \tilde{S}_D[i]) \neq T^n_P(S_i, S_i) \right\}
\]
for all \( i \in [1, 2^n R_n] \).

\[
E^{(2)}_{1i} = \left\{ (s[i+2], \tilde{S}_D[i]) \neq T^n_P(S_i, S_i) \right\}
\]
for all \( i \in [1, 2^n R_n] \).

\[
E^{(3)}_{1i} = \left\{ (s[i+2], \tilde{S}_R[i], \tilde{S}_D[i]) \neq T^n_P(S_i, S_i) \right\}
\]
for all \( (i, j, j_D) \in [1, 2^n R_n] \times [1, 2^n R_n] \).

From known results in rate distortion theory [60, p. 336], it follows that

\[
\text{Exponentially with } \frac{1}{n} \text{ if } \tilde{R}_R \geq I(S; \tilde{S}_R).
\]

Similarly, \( \tilde{R}_D \) is exponentially with \( \frac{1}{n} \) if \( \tilde{D}_D \geq I(S; \tilde{S}_D) \). It remains to show that \( \tilde{R}_R + \tilde{R}_D > I(S; \tilde{S}_R, \tilde{S}_D) + I(\tilde{S}_R; \tilde{S}_D) \), and this can be proved by following straightforwardly the arguments and algebra in [49].

2) Let \( E_{2i} \) be the event that there is no pair \( (u_R^{(i)}, u_D^{(i)}, l, j, j_D) \) satisfying (14), i.e., the set \( D_{i,n+1} \) is empty. Using Chebyshev’s inequality, it is easy to see that

\[
P\left( \left\| D_{i,n+1} \right\| > 0 \right) \leq \frac{\text{Var}\left( \left\| D_{i,n+1} \right\| \right)}{\epsilon^2 \left( \text{Ex} D_{i,n+1} \right)^2}.
\]

\[\text{(A-2)}\]

We obtain bounds on \( \text{Ex} D_{i+1,n+1} \) and \( \text{Var}\left( \left\| D_{i+1,n+1} \right\| \right) \) by proceeding in a way similar to [51]. We define the indicator functions as given by (A-4), shown at the bottom of the page.

The cardinality of the set \( D_{i,n+1} \) is given by (A-5), shown at the bottom of the page. Thus, \( \text{Ex} D_{i,n+1} \) can be bounded as given by (A-6) at the bottom of the page.

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

\[
\text{Var}\left( \left\| D_{i,n+1} \right\| \right) \leq 2^{2(R_R + R_D - R_R - R_D - I(U, V, S; \tilde{S}_R, \tilde{S}_D) + o(1))}.
\]

\[\text{(A-7)}\]

Therefore, for sufficiently large \( n \)

\[
P\left( \left\| D_{i,n+1} \right\| = 0 \right) \leq \epsilon
\]

\[\text{(A-8)}\]

provided that (15) is true.

**Lemma 1:**

\[\text{Var}\left( \left\| D_{i,n+1} \right\| \right) \leq 2^{2(R_R + R_D - R_R - R_D - I(U, V, S; \tilde{S}_R, \tilde{S}_D) + o(1))}.
\]

\[\text{(A-9)}\]

**Proof:** For notational convenience, let us use temporarily in the proof of this lemma the shorthand notation \( u_R^{(i,j)} := u_R^{(i,j)}(w_{i-1}, j_{i-1}, w_{i+1}, j_{i+1}, j_D) \) and \( u_D^{(i,j)} := u_D^{(i,j)}(w_{i-1}, j_{i-1}, w_{i+1}, j_{i+1}, j_D) \). Then, \( \left\| D_{i,n+1} \right\| \) can be expressed as given by (A-10), shown at the bottom of the next page.

Taking the expectation and dividing by \( \left\| B_{i,n} \times C_{i,n} \times \mathcal{J}_R \times \mathcal{J}_D \right\| \) in both sides of (A-10), we get (A-11) at the bottom of the next page.
Let \( \Delta := I(U_R; S, \hat{S}_R, \hat{S}_D) + I(U_D; S, \hat{S}_R, \hat{S}_D) - I(U_R; U_D, V, S, \hat{S}_R, \hat{S}_D) \). It can be shown easily that

i) For \( (k_i, j_{Ri}) = (k'_i, j'_{Ri}) \) and \( (l_i, j_{Di}) = (l'_i, j'_{Di}) \),
\[
\Pr\left\{ \left( u_R(k_i, j_{Ri}), u_D(l_i, j_{Di}) \right) \in \mathcal{D}_{k_i,j_{Di}} \right\} \leq 2^{-n(\Delta - \delta)}.
\]  
(A-12)

ii) For \( (k_i, j_{Ri}) = (k'_i, j'_{Ri}) \) and \( (l_i, j_{Di}) \neq (l'_i, j'_{Di}) \),
\[
\Pr\left\{ \left( u_R(k_i, j_{Ri}), u_D(l_i, j_{Di}) \right) \in \mathcal{D}_{k_i,j_{Di}} \right\} \leq 2^{-n(\Delta - \delta)}.
\]  
(A-13)

iii) For \( (k_i, j_{Ri}) \neq (k'_i, j'_{Ri}) \) and \( (l_i, j_{Di}) = (l'_i, j'_{Di}) \),
\[
\Pr\left\{ \left( u_R(k_i, j_{Ri}), u_D(l_i, j_{Di}) \right) \in \mathcal{D}_{k_i,j_{Di}} \right\} \leq 2^{-n(\Delta - \delta)}.
\]  
(A-14) This completes the proof of Lemma 1.

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

\[
\text{var}\left(\|\mathcal{D}_{k_i,j_{Di}}\|\right)
= E[\|\mathcal{D}_{k_i,j_{Di}}\|^2] - E^2[\|\mathcal{D}_{k_i,j_{Di}}\|],
\]
\[
\leq |\mathcal{B}_{R_i}||\mathcal{C}_{n_i}|J_RJ_D2^{-n(\Delta - \delta)}
+ 2^nR_R + R_D - R_R - i(U_RU_DV, S, \hat{S}_R, \hat{S}_D + \epsilon)|.
\]  
(A-16)
1) Let $E_{3i}$ be the event that $u(w_{i-1}, j_{V_i}, w_i, j_{V_i})$ and $u_R(w_{i-1}, j_{V_i}, w_i, j_{V_i}, k_i, j_{R_i})$ are not jointly typical with $(y_2[i], s_R[i | R_{i-2}])$ given $v(w_i, j_{V_i})$. That is

$$E_{3i} = \{ \left( u(w_{i-1}, j_{V_i}, w_i, j_{V_i}), u_R(w_{i-1}, j_{V_i}, w_i, j_{V_i}, k_i, j_{R_i}) \right)$$

$$\notin T^n_n(P_{V,U,R,V_2,S_R}) \}. \quad (A-17)$$

For $v(w_{i-1}, j_{V_i}), u(w_{i-1}, j_{V_i}, w_i, j_{V_i})$, $u_R(w_{i-1}, j_{V_i}, w_i, j_{V_i}, k_i, j_{R_i})$, $u_D(w_{i-1}, j_{V_i}, w_i, j_{V_i}, k_i, j_{R_i})$, jointly typical with $s_i, s_R[i | R_{i-2}], s_D[i | D_{i-2}]$, and with the source input $x_1[i]$, the relay input $x_2[i]$, we have

$$\Pr(E_{3i} | E_{3i}, E_{3i}) \to 0$$

as $n \to \infty$ by the Markov Lemma [60, p. 436].

2) Let $E_{4i}$ be the event that $u(w_{i-1}, j_{V_i}, w_i, j_{V_i})$ and $u_R(w_{i-1}, j_{V_i}, w_i, j_{V_i}, k_i, j_{R_i})$ are jointly typical with $(y_2[i], s_R[i | R_{i-2}])$ given $v(w_{i-1}, j_{V_i})$, for some $w_i \in [1, M], j_{V_i} \in J_V, k_i \in [1, M_R]$, and $j_{R_i} \in J_R$, with $w_i \neq w_i$. That is

$$E_{4i} = \{ (j_{V_i} \in [1, M], j_{V_i} \in J_V, k_i \in [1, M_R])$$

$$J_{R_i} \in J_R \; s.t. \; w_i \neq w_i)$$

$$u(w_{i-1}, j_{V_i}, w_i, j_{V_i})$$

$$u_R(w_{i-1}, j_{V_i}, w_i, j_{V_i}, k_i, j_{R_i})$$

$$v(w_{i-1}, j_{V_i}, y_{2[i]}, s_R[i | R_{i-2}])$$

$$\in T_n(P_{V,U,R,V_2,S_R}) \}. \quad (A-18)$$

Conditioned on $E_{3i}, E_{3i}, E_{3i}$, the probability of the event $E_{4i}$ can be bounded as

$$\Pr(E_{4i} | E_{3i}, E_{3i}, E_{3i}) \leq M_J M_R J_{R_2}^{-n \cdot | V | (U, V_2, S_R) | V | - n} \cdot e$$

$$= 2^{-n | I(U; Y_2 | V, S_R) - I(U; S, S_R | V, S_R) - \frac{e}{2} + \frac{e}{2}}. \quad (A-19)$$

Conditioned on the events $E_{3i}, E_{3i}, E_{3i}$, and $E_{4i}$, the probability of the event $E_{4i}$ can be bounded as

$$\Pr(E_{4i} | E_{3i}, E_{3i}, E_{3i}) \leq 2^{-n | I(U; Y_2 | V, S_R) - I(U; S, S_R | V, S_R) - \frac{e}{2} + \frac{e}{2}}. \quad (A-20)$$

(Conditioned on the events $E_{3i}, E_{3i}, E_{3i}$, the probability of the event $E_{4i}$ can be bounded as

$$\Pr(E_{4i} | E_{3i}, E_{3i}, E_{3i}) \leq 2^{-n | I(U; Y_2 | V, S_R) - I(U; S, S_R | V, S_R) - \frac{e}{2} + \frac{e}{2}}. \quad (A-21)$$

(Note the multiplicative term $4 = 2^n(\frac{e}{2})$ in the RHS of (A-21).) The proof of (A-21) follows by proceeding in a way that is essentially similar to the analysis of the event $E_{3i}$ in [61, pp. 854-855], with minor modifications. More specifically, let, for given $j_{V_i}$,

$$C(j_{V_i}) = \{ u(w_{i-1}, j_{V_i}, w_i, j_{V_i}) : j_{V_i} \neq j_{V_i} \}.$$ First, following the lines of [61, eq. (5), p. 855], one can easily show that

$$P[\mathbb{E}(j_{V_i} : 1 \mathbb{S}[i] = \mathbb{S}[i] | \mathbb{S}(j_{V_i}) = \mathbb{E}(V(w_{i-1}, j_{V_i}) = v)] \leq \frac{1}{2}$$

(22)

Thus, using (A-22) and the approach in [61], it can be shown easily (A-23) shown at the bottom of the page holds, where (a) follows since for given $j_{V_i}$, the event $U(w_{i-1}, j_{V_i}, w_i, 1, 1) = u, u_R(w_{i-1}, j_{V_i}, w_i, 1, 1, 1) = u_R$ is independent of $j_{V_i} = j_{V_i}$ conditioned given

$$\{ V(w_{i-1}, j_{V_i}) = v, j_{V_i} \neq 1, \mathbb{C}(j_{V_i}) = \mathbb{E}, \mathbb{S}[i] = \mathbb{S}[i] \};$$ and (b) follows using (A-22) and an approach similar to [61, Lemma 1].

$$\text{P} \{ U(w_{i-1}, j_{V_i}, w_i, 1, 1) = u, u_R(w_{i-1}, j_{V_i}, w_i, 1, 1, 1) = u_R \}$$

$$= 2 \text{P} \{ U(w_{i-1}, j_{V_i}, w_i, 1, 1) = u, u_R(w_{i-1}, j_{V_i}, w_i, 1, 1, 1) = u_R \}$$

$$\leq 2 \text{P} \{ U(w_{i-1}, j_{V_i}, w_i, 1, 1) = u, u_R(w_{i-1}, j_{V_i}, w_i, 1, 1, 1) = u_R \} - \mathbb{E}(V(w_{i-1}, j_{V_i}) = v)$$

(A-23)
Similarly

$$\Pr\{Y_2[i] = y_2, \bar{S}_R[u_{R1-2}] = \bar{s}_R, J_{U1} \neq 1, J_{V1} = j_{V1}
\leq 2P\{Y_2[i] = y_2, \bar{S}_R[u_{R1-2}] = \bar{s}_R, \bar{S}[i] = \bar{s}[i]
$$

where (c) follows since for given \(j_{V1}\), the event \(\{Y_2[i] = y_2, \bar{S}_R[u_{R1-2}] = \bar{s}_R\}\) is independent of \(\{J_{V1} = j_{V1}\}\) conditionally given \(\{V(w_{U1-1}, j_{V1}) = v, \bar{S}_U[i] = \bar{s}[i]\}\); and (d) follows using (A-22) and an approach similar to [61, Lemma 2].

Finally, using (A-23) and (A-24), and following straightforwardly the approach in [61, pp. 854-855], we obtain (A-21). Thus, summarizing, \(\Pr(E_{6i}|E_{1i}, E_{2i}, E_{5i}, E_{6i}) \rightarrow 0\) as \(n \rightarrow \infty\).

4) Let \(E_{6i}\) be the event that \(u_R(w_{U1-1}, j_{V1}, w_i, j_{Ri}, k_i, j_{Ri})\) is jointly typical with \(\{y_2[i], \bar{S}_R[u_{R1-2}]\}\) for some \(k_i \in [1, M_R]\), \(j_{Ri} \in J_R\) with \(k_i' \neq k_i\). That is

$$E_{6i} = \left\{ \exists k_i' \in [1, M_R], j_{Ri} \in J_R s.t.: k_i' \neq k_i
\leq 2P\{Y_2[i] = y_2, \bar{S}_R[u_{R1-2}] = \bar{s}_R, \bar{S}[i] = \bar{s}[i]
$$

Proceeding in a way similar to the event \(E_{4i}\), it can be shown easily that, conditioned on the events \(E_{1i}, E_{2i}, E_{5i}, E_{6i}, E_{5i}, E_{6i}\), the probability of the event \(E_{6i}\) can be bounded as

$$\Pr(E_{6i}|E_{1i}, E_{2i}, E_{5i}, E_{6i}) \leq M_R J_R 2^{-n[I(U, Y, \bar{S}_R, \bar{S}_U, U V] - \epsilon]} - 2^n(\epsilon).$$

Thus, \(\Pr(E_{6i}|E_{1i}, E_{2i}, E_{5i}, E_{6i}) \rightarrow 0\) as \(n \rightarrow \infty\).

5) For decoding the triple \(\{w_{U1-1}, j_{U1-1}, j_{V1}\}\) and the index \(j_{V1}\) at the destination, let \(E_{7i}\) be the union of the following two events:

$$E_{7i}^{(1)} = \left\{ \left( v(w_{U1-1}, j_{V1}), u(w_{U1-1}, j_{U1-1}, j_{V1}) \right)
$$

6) For decoding the triple \(\{w_{U1-1}, j_{U1-1}, j_{V1}\}\) and the index \(j_{V1}\) at the destination, let \(E_{8i}\) be the event

$$E_{8i} = \left\{ w_{U1-1}' \in [1, M_U], j_{U1-1} \in J_U, k_{i-1} \in [1, M_D],
$$

Proceeding in a way similar to for the events \(E_{4i}, E_{5i}, E_{6i}\), and noticing that, for given \(w_{i-1}' \in [1, M]\), the two subevents in \(E_{5i}\) are independent because the codebooks used for blocks \(i - 1\) and \(i\) are different, it can be shown easily that, conditioned on \(\cap_{k=1}^{i-1} E_{ki}\), the probability of the event \(E_{8i}\) can be bounded as given by (A-27) shown at the bottom of this page. Thus, \(\Pr(E_{8i}|\cap_{k=1}^{i-1} E_{ki}) \rightarrow 0\) as \(n \rightarrow \infty\) if \(R < I(U, V; Y, \bar{S}_D) - I(U, V; S, \bar{S}_R, \bar{S}_U, U V) - 2(\epsilon)\).
7) For decoding the triple \((\hat{w}_{i-1}, \hat{J}_{U_{i-1}}, \hat{t}_{i-1})\) and the index \(\hat{j}_{Vi}\) at the destination, let \(E_{Bi}\) be the event
\[
E_{Bi} = \left\{ j_{Vi} \in J_V \right\} \text{s.t.} : j_{Vi} \neq j_{Vi}^{*}\n\]
\[
= \left\{ u(w_{i-2}, j_{Vi}^{*}-1, w_{i-1}; 1, l_i, j_{D_i} - 1), v(w_{i-1}, j_{Vi}^{*}-1), y_3[l_i - 1], s_D[l_{D_i} - 1] \right\} \in T_n(P_{U,V,Y,S})\n\]
\[
= \left\{ v(w_{i-1}, j_{Vi}), y_3[l_i], s_D[l_{D_i} - 1] \right\} \in T_n(P_{U,V,Y,S,D})\n\]

It is clear that
\[
\text{Pr}(E_{Bi}) \leq \text{Pr}\left\{ \left\{ v(w_{i-1}, j_{Vi}), y_3[l_i], s_D[l_{D_i} - 1] \right\} \in T_n(P_{U,V,Y,S,D}) \right\} \text{ for some } j_{Vi} \neq j_{Vi}^{*}\right\}. \quad (A-28)
\]

Now, proceeding in a way similar to [61, pp. 854-855], with the rather minor modifications outlined in the following, we can analyze the error event in the RHS of (A-28), and get the following bound on the probability of the error event \(E_{Bi}\) conditioned on \(\gamma_{k-1}^E_{k_i}\)
\[
\text{Pr}(E_{Bi} | \gamma_{k-1}^E_{k_i}) \leq J_{Vi} 2^{-n I(V;Y,S,h_n) - \frac{1}{2} - \epsilon} = 2^{-n I(V;Y,S,h_n) - l(V;S,S,h_n, S,D) - \frac{1}{2} - 2\epsilon}. \quad (A-29)
\]

An outline of the proof of (A-29) is as follows. Let \(C = \left\{ v(w_{i-1}, j_{Vi}) : j_{Vi} \neq j_{Vi}^{*} \right\}\). First, we show that for \(n\) sufficiently large
\[
\text{Pr}\left\{ \left\{ v(w_{i-1}, j_{Vi}) : j_{Vi} \neq j_{Vi}^{*} \right\} \in T_n \right\} \leq \frac{1}{2}. \quad (A-30)
\]
Let \(\xi := |\{ v(w_{i-1}, j_{Vi}) : j_{Vi} \neq j_{Vi}^{*} \} \in T_n |\). Then, if \(\xi \geq 1\), for \(n\) sufficiently large
\[
\text{Pr}\left\{ \left\{ v(w_{i-1}, j_{Vi}) : j_{Vi} \neq j_{Vi}^{*} \right\} \in T_n \right\} \leq \frac{1}{2}. \quad (A-31)
\]
If \(\xi = 0\), for \(n\) sufficiently large
\[
\text{Pr}\left\{ \left\{ v(w_{i-1}, j_{Vi}) : j_{Vi} \neq j_{Vi}^{*} \right\} \in T_n \right\} \leq \frac{1}{2}. \quad (A-32)
\]
where the first inequality follows by the union of events bound.

Next, using (A-30) and following the approach in [61, Lemmas 1 and 2], it can be shown easily that for sufficiently large \(n\)
\[
\text{Pr}\left\{ v(w_{i-1}, j_{Vi}) : j_{Vi} \neq j_{Vi}^{*} \right\} \leq 2^n 2^{-n I(V;Y,S,h_n) - l(V;S,S,h_n, S,D) - \frac{1}{2} - 2\epsilon}. \quad (A-29)
\]

8) For decoding the triple \((\hat{w}_{i-1}, \hat{J}_{U_{i-1}}, \hat{t}_{i-1})\) and the index \(\hat{j}_{Vi}\) at the destination, let \(E_{1i_0}\) be the event
\[
E_{1i_0} = \left\{ j_{U_{i-1}} \in J_D \right\} \text{s.t.} : j_{U_{i-1}} \neq j_{U_{i-1}}^{*}, j_{D_{i-1}} \in J_D
\]
\[
\text{Pr}(v(w_{i-1}, j_{Vi}) : j_{Vi} \neq j_{Vi}^{*} \right\}) \leq 2^n 2^{-n I(V;Y,S,h_n) - l(V;S,S,h_n, S,D) - \frac{1}{2} - 2\epsilon}. \quad (A-33)
\]

Finally, using (A-33) and following essentially straightforwardly the approach of [61, pp. 854-855], we obtain
\[
E_{1i_0} = \left\{ j_{U_{i-1}} \in J_D \right\} \text{s.t.} : j_{U_{i-1}} \neq j_{U_{i-1}}^{*}, j_{D_{i-1}} \in J_D
\]
\[
\text{Pr}(v(w_{i-1}, j_{Vi}) : j_{Vi} \neq j_{Vi}^{*} \right\}) \leq 2^n 2^{-n I(V;Y,S,h_n) - l(V;S,S,h_n, S,D) - \frac{1}{2} - 2\epsilon}. \quad (A-33)
\]

Note that the first event in \(E_{1i_0}\) (i.e., the one relative to block \(i - 1\)) and the second event in \(E_{1i_0}\) (i.e., the one relative to block \(i\)) are independent since the codebooks used for successive blocks \(i - 1\) and \(i\) are different. Then, proceeding in a way similar to the event \(E_{Bi}\) to analyze the first event in \(E_{1i_0}\) and in a way similar to the event \(E_{Bi}\) to analyze the second event in \(E_{1i_0}\), it can be shown easily that, conditioned on the events \(\gamma_{k-1}^E_{k_i}\), the probability of the event \(E_{1i_0}\) can be bounded as (A-34), shown at the bottom of the next page. Thus, \(\text{Pr}(E_{1i_0} | \gamma_{k-1}^E_{k_i}) \rightarrow 0\)

9) For decoding the triple \((\hat{w}_{i-1}, \hat{J}_{U_{i-1}}, \hat{t}_{i-1})\) and the index \(\hat{j}_{Vi}\) at the destination, let \(E_{1i+1}\) be the event
\[
E_{1i+1} = \left\{ j_{U_{i+1}} \in J_D \right\} \text{s.t.} : j_{U_{i+1}} \neq j_{U_{i+1}}^{*}, j_{D_{i+1}} \in J_D
\]
Proceeding similarly to the event $E_5$, it can be shown easily that, conditioned on $\cap_{k=1}^{n} E_{k+i}^c$, the probability of the event $E_{11i}$ can be bounded as $(A-35)$, shown at the bottom of the page. Thus, $\Pr(E_{11i} \cap \cap_{k=1}^{n} E_{k+i}^c) \to 0$ as $n \to \infty$.

10. For decoding the triple $(\hat{x}_{i-1}, \hat{y}_{i-1}, \hat{z}_{i-1})$ and the index $\hat{j}_{1}$ at the destination, let $E_{12i}$ be the event

$$
E_{12i} = \left\{ i \in \mathbb{N} : \hat{j}_{k+1} = 1 \right\}
$$

11. For decoding the triple $(\hat{x}_{i-1}, \hat{y}_{i-1}, \hat{z}_{i-1})$ and the index $\hat{j}_{1}$ at the destination, let $E_{13i}$ be the event

$$
E_{13i} = \left\{ i \in \mathbb{N} : \hat{j}_{k+1} = 1 \right\}
$$

This concludes the proof of Theorem 1.

**APPENDIX B**

**PROOF OF THEOREM 3**

First we generate a random codebook that we use to obtain the lower bound in Theorem 3. This scheme is based on a combination of block Markov coding [47], Gel’fand–Pinsker binning [13], and classic rate distortion theory [60, Ch. 13]. Next, we outline the encoding and decoding procedures.

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

**Codebook generation:** Fix a measure

$$P_{S, U, V, X, X, X, y, v, x, y, z, w}$$

of the form (27). Calculate the marginal $P_X$ induced by this measure. Fix $\epsilon > 0$ and let

$$J = 2^n [I(U; V) + 2\epsilon] \quad J_R = 2^n [I(U; V) + 2\epsilon]$$

$$M = 2^n [R - 4\epsilon] \quad R_M = 2^n [R - 4\epsilon]$$

1) We generate $JM$ independent and identically distributed (i.i.d.) codewords $\{u(w; j)\}$ indexed by $w = 1, \ldots, M$, $j = 1, \ldots, J$, each with i.i.d. components drawn according to $P_U$.
2) We generate $J_R M_R$ i.i.d. codewords $\{u_{R}(m, j_R)\}$ indexed by $m = 1, \ldots, M_R$, $j_R = 1, \ldots, J_R$, each with i.i.d. components drawn according to $P_{U_R}$.

3) Independently, we randomly generate a rate distortion codebook consisting of $M_R$ sequences $\hat{x}$ drawn i.i.d. according to the $n$-product of the marginal $P_{\hat{x}}$. We index these sequences as $\hat{x}[m, m = 1, \ldots, M_R]$.

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

The encoding at the beginning of block $i$ is as follows.

1) The relay searches for the smallest $j \in \{1, \ldots, J\}$ such that $u_j[w_i, j]$ is jointly typical with $s[i]$. (The properties of strongly typical sequences guarantee that there exists one such $j$.) Denote this $j$ by $j^* = j(s[i], w_i)$.

2) Similarly, the source finds $j^*_R = j^*(s[i]+1, w_{i+1})$ such that $u_{j^*_R}[w_{i+1}]$ is jointly typical with $s[i]+1$ and then generates a vector $x[w_{i+1}]$ with i.i.d. components drawn according to $u_{j^*_R}[w_{i+1}]$ and $s[i]+1$, drawn according to the marginal $P_X U_{U,S}$.

iii) Then, the source indices $x[w_{i+1}]$ by $m_i$, if there exists an $m_i \in \{1, \ldots, M_R\}$ such that $x[w_{i+1}]$ and $\hat{x}[m_i]$ are jointly strongly typical. If there is more than one such $m_i$, the source selects the first in lexicographic order. If there is no such $m_i$, let $m_i = 1$. Shannon’s rate-distortion theory [60, Ch. 13] ensures that the encoding of $x[w_{i+1}]$ is accomplished successfully with high probability provided that $n$ is sufficiently large and $P_R > I(X; \hat{X})$.

2) The relay node transmits $x[w_{i+1}]$ to the relay and the decoding procedure at destination at the end of block $i$ is as follows.

1) The relay node transmits $x[w_{i+1}]$. It sends a vector $x_1[i]$ with i.i.d. components given the vectors $u_i[w_i, j^*_i]$, $u_{j^*_R}[m_i, j^*_R]$, and $s[i]$, drawn according to the marginal $P_{X, U_{U_R,S}}$ induced by the distribution (27).

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

1) The relay knows $m_{i-1}$ (this will be justified below), and sends $x_2[i] = \hat{x}[m_{i-1}]$.

2) The source searches for the smallest $j \in \{1, \ldots, J\}$ such that $u_j[w_i, j]$ is jointly typical with $s[i], u_j[w_i, j^*_i]$. (Again, the properties of strongly typical sequences guarantee that there exists one such $j$.) Denote this $j$ by $j^* = j(s[i], w_i)$.

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

1) The relay knows $m_{i-1}$ (this will be justified below), and sends $x_2[i] = \hat{x}[m_{i-1}]$.

2) The source searches for the smallest $j \in \{1, \ldots, J\}$ such that $u_j[w_i, j]$ is jointly typical with $s[i], u_j[w_i, j^*_i]$. (Again, the properties of strongly typical sequences guarantee that there exists one such $j$.) Denote this $j$ by $j^* = j(s[i], w_i)$.

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

1) The relay knows $m_{i-1}$ (this will be justified below), and sends $x_2[i] = \hat{x}[m_{i-1}]$.

2) The source searches for the smallest $j \in \{1, \ldots, J\}$ such that $u_j[w_i, j]$ is jointly typical with $s[i], u_j[w_i, j^*_i]$. (Again, the properties of strongly typical sequences guarantee that there exists one such $j$.) Denote this $j$ by $j^* = j(s[i], w_i)$.

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

1) The relay knows $m_{i-1}$ (this will be justified below), and sends $x_2[i] = \hat{x}[m_{i-1}]$.

2) The source searches for the smallest $j \in \{1, \ldots, J\}$ such that $u_j[w_i, j]$ is jointly typical with $s[i], u_j[w_i, j^*_i]$. (Again, the properties of strongly typical sequences guarantee that there exists one such $j$.) Denote this $j$ by $j^* = j(s[i], w_i)$.

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

1) The relay knows $m_{i-1}$ (this will be justified below), and sends $x_2[i] = \hat{x}[m_{i-1}]$.

2) The source searches for the smallest $j \in \{1, \ldots, J\}$ such that $u_j[w_i, j]$ is jointly typical with $s[i], u_j[w_i, j^*_i]$. (Again, the properties of strongly typical sequences guarantee that there exists one such $j$.) Denote this $j$ by $j^* = j(s[i], w_i)$.
For $u(w_i, j_i^*)$, $u_R(m_i, j_R^i)$ jointly typical with $s[i]$, and with the source input $x[i]$ and the relay input $x_2[i]$, we have $Pr(E_4, E_5^i) \rightarrow 0$ as $n \rightarrow \infty$ by the Markov Lemma [60, p. 436].

4) For the decoding at the relay, let $E_5$ be the event that $u_R(m_i, j_R^i)$ is jointly typical with $y_2[i]$ for some $m'_i \in [1, M_R]$ and $j_R^i \in J_R$, with $m'_i \neq m_i$. That is

$$E_5 = \left\{ \exists m'_i \in [1, M_R], j_R^i \in J_R \text{ s.t. } m'_i \neq m_i \right\}$$

$$\cup \{u_R(m'_i, j_R^i), y_2[i] \in T^n_R(P_{U_R}y_2, x^i) \} \} \}$$

(B-8)

Conditioned on the events $E_4^i, E_5^i, E_6^i, E_7^i$, and $E_8^i$, the probability of the event $E_5$ can be bounded using the union bound,

$$Pr(E_5|E_4^i, E_5^i, E_6^i, E_7^i, E_8^i) \leq M_R J R^{-n[I(U_R ; Y_2) - \epsilon]} - n[I(U_R ; Y_2) - I(U_R ; U_S)]$$

$$= 2^{-n[I(U_R ; Y_2) - I(U_R ; U_S)] - \epsilon}$$

(B-9)

Thus, $Pr(E_5|E_4^i, E_5^i, E_6^i, E_7^i, E_8^i) \rightarrow 0$ as $n \rightarrow \infty$ if $R < I(U_R ; Y_2) - I(U_R ; U_S) - I(U_R ; U_S)$.

5) For the decoding at the destination, let $E_6$ be the event that $u(w_i, j_i^*)$ is not jointly typical with $y_3[i]$. That is

$$E_6 = \left\{ \exists w'_i \in [1, M], j_i \in J \text{ s.t. } w'_i \neq k_i \right\}$$

$$\cup \{u(w'_i, j_i), y_3[i] \in T^n(U, y_3) \} \} \}$$

(B-10)

For $u(w_i, j_i^*)$, $u(R(m_i, j_R^i)$ jointly typical with $s[i]$, and with the source input $x[i]$ and the relay input $x_2[i]$, we have $Pr(E_6|E_4^i, E_5^i, E_6^i, E_7^i, E_8^i) \rightarrow 0$ as $n \rightarrow \infty$ by the Markov Lemma [60, p. 436].

6) For the decoding at the destination, let $E_7$ be the event that $u(w'_i, j_i)$ is jointly typical with $y_3[i]$ for some $w'_i \in [1, M]$ and $j_i \in J$, with $w'_i \neq k_i$. That is

$$E_7 = \left\{ \exists w'_i \in [1, M], j_i \in J \text{ s.t. } w'_i \neq k_i \right\}$$

$$\cup \{u(w'_i, j_i), y_3[i] \in T^n(U, y_3) \} \} \}$$

(B-11)

Conditioned on the events $E_4^i, E_5^i, E_6^i, E_7^i$, and $E_8^i$, the probability of the event $E_7$ can be bounded using the union bound,

$$Pr(E_7|E_4^i, E_5^i, E_6^i, E_7^i, E_8^i) \leq M J R^{-n[I(U ; Y_3) - \epsilon]}$$

$$= 2^{-n[I(U ; Y_3) - I(U ; U_S) - \epsilon]}$$

(B-12)

Thus, $Pr(E_7|E_4^i, E_5^i, E_6^i, E_7^i, E_8^i) \rightarrow 0$ as $n \rightarrow +\infty$ if $R < I(U ; Y_3) - I(U ; U_S)$.

This concludes the proof of Theorem 3.

APPENDIX C
PROOFS OF THEOREM 4

Let an $(\epsilon_n, n, R)$ code be given. By Fano’s inequality, we have

$$nR = H(W) \leq I(W; Y_3^n) + 1 + n \epsilon_n.$$  

(C-1)

Let us define $\bar{d}_i = (S_{i+1}^n, Y_{i+1}^{i-1}, Y_i^{i-1})$ and $V_i = (W, S_{i+1}^n, Y_i^{i-1})$, $i = 1, \ldots, n$.

We have

$$I(W; Y_3^n) \leq I(W; Y_2^n, Y_3^n)$$

$$\leq I(W; Y_2^n, Y_3^n) - I(W; S^n)$$

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

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

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

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

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

(C-2)

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

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

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

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

(C-3)

where; (a) follows since message $W$ is independent of the state $S^n$; (b) follows from Csiszar and Korner’s “summation by parts” lemma [62]

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

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

(C-4)

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

(C-5)

(c) follows similarly, from Csiszar and Korner’s “summation by parts”

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

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

(C-6)

(C-7)
and (d) follows from the fact that $X_{2i}$ is a deterministic function of $Y_{2}^{i-1}$.

Similarly

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

$$= \sum_{i=1}^{n} I(\bar{V}_{i};Y_{3,i}) - I(\bar{V}_{i};S_{i})$$

(C-6)

where (e) follows exactly as in the converse part of the proof of the capacity of Gel’fand–Pinsker channel [13] by replacing $Y_{3}^{n}$ with $Y_{3,n}$.

From the above, we have

$$R \leq \frac{1}{n} \sum_{i=1}^{n} I(\bar{V}_{i};Y_{3,i}) - I(\bar{V}_{i};S_{i}) + 1 + nR_{e,n}$$

$$R \leq \frac{1}{n} \sum_{i=1}^{n} I(\bar{V}_{i};Y_{3,i}) - I(\bar{V}_{i};S_{i}) + 1 + nR_{e,n}$$

(C-7)

We introduce a random variable $T$ which is uniformly distributed over $\{1, \ldots, n\}$. Set $S_i = S_i^{n}, \bar{U} = \bar{U}_i, \bar{V} = \bar{V}_i, X_1 = X_{1,T}, X_2 = Y_{2,T}, Y_1 = Y_{1,T}$, and $Y_3 = Y_{3,T}$. We substitute $T$ into the aforementioned bounds. Considering the first bound in (C-7), we have

$$\frac{1}{n} \sum_{i=1}^{n} I(\bar{V}_{i};Y_{3,i}) - I(\bar{V}_{i};S_{i}) + 1 + nR_{e,n}$$

(C-8)

where in the last equality we used the fact that $T$ is independent of all the other variables.

Similarly, considering the second bound in (C-7), we obtain

$$\frac{1}{n} \sum_{i=1}^{n} I(\bar{V}_{i};Y_{3,i}) - I(\bar{V}_{i};S_{i})$$

(C-9)

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

$$R \leq I(V;Y_{2,n}U|X_{2}) - I(V;S|U,X_{2}) + 1 + nR_{e,n}$$

(C-10)

So far, we have shown that, for a given sequence of $(r,n,R)$-codes consisting of a mapping $\phi_{1}^{n} : X_{1}^{n} \rightarrow X_{2}^{n}$ and $\phi_{2}^{n} : W \times S^{n} \rightarrow X_{2}^{n}$, a sequence of mappings $\phi_{2,i} : Y_{2,i}^{n-1} \rightarrow X_{2,i}, i = 1, \ldots, n$, at the relay, and a mapping $\psi_{2}^{n} : Y_{3}^{n} \rightarrow W$ at the decoder with average error probability $P_{e}^{n} \rightarrow 0$ as $n \rightarrow 0$, the rate $R$ must satisfy (37).

By Fano’s inequality, we have

$$H(W|Y_{3}^{n}) \leq nr_{e,n} + \frac{1}{n} \delta_{n}.$$  

(D-1)

Thus

$$nR - H(W) \leq I(W;Y_{3}^{n}) + n\delta_{n}.$$  

(D-2)

We now upper bound $I(W;Y_{3}^{n})$ as in the following lemma, the proof of which follows.
Lemma 2:

\[ I(W; Y^n_3) \leq \sum_{i=1}^{n} I(X_i; Y_i) + I(X_i; S_i) \]

Proof: To simplify the notation, we use \( S^i = (S_1, S_2, \ldots, S_i) \), \( Y^i = (Y_1, Y_2, \ldots, Y_i) \), \( X^i_j = (X_{1,j}, X_{2,j}, \ldots, X_{j,t}, j = 1, 2, 3, \ldots, n) \), and \( X^i_{j-1} = (X_{1,j-1}, X_{2,j-1}, \ldots, X_{j-1,t}, j = 1, 2, 3, \ldots, n) \).

1) The proof of the bound on \( I(W; Y^n_3) \) given in i) follows straightforwardly by revealing the state to the destination and using the channel structure given by (1).

\[ I(W; Y^n_3) \leq \sum_{i=1}^{n} I(X_i; Y_i) + I(X_i; S_i) \]

2) The proof of the bound on \( I(W; Y^n_3) \) given in ii) follows as follows:

\[ I(W; Y^n_3) = I(W, S^n; Y^n_3) - I(S^n; Y^n_3 | S) \]

where \([a]\) follows trivially by revealing the state to the destination; \([b]\) follows since \( X_i \) is i.i.d. and is independent of the message \( W \); \([f]\) follows from \( (W, S^n, Y^n_i) \) is a Markov chain; \([g]\) follows from \( Y_i \) is a deterministic function of \( S_i \); \([h]\) follows from the fact that \( X_2, i \) is independent of \( S_i \).

We introduce a random variable \( T \) which is uniformly distributed over \( \{1, \ldots, n\} \). Set \( S = S_T, X_1R = X_{1R,T}, X_{1D} = X_{1D,T}, X_2 = X_{2,T}, Y_2 = Y_{2,T}, \) and \( Y_3 = Y_{3,T} \). We substitute \( T \) into the aforementioned bounds. Considering the bound (D-12), we obtain

\[ \frac{1}{n} \sum_{i=1}^{n} I(X_i; Y_i) = I(X_1D; Y_3 | S, T) + I(X_2; Y_3 | T) \]

where

\[ \frac{1}{n} \leq \sum_{i=1}^{n} I(X_i; Y_i) \]
and similarly
\[
\frac{1}{n} \sum_{i=1}^{n} I(X_{1R,i}; Y_{2,i}, X_{2,i}, S_i) + I(X_{1D,i}; Y_{3,i}, X_{2,i}, S_i)
= I(X_{1R}; vY_2 S, X_2, T) + I(X_{1D}; Y_3 S, X_2, T)
\]
\hspace{1cm} (D-14)
where the distribution on \(\{T, S, X_{1R}, X_{1D}, X_2, Y_2, Y_3\}\) from a given code is of the form
\[
P_{T,S,X_{1R},X_{1D},X_2,Y_2,Y_3} = QS P_{X_{1R}} P_{X_2} \tau P_{X_{1D}|X_2} \tau P_{X_2|S} S X_2 T
\times W_{Y_2 S X_{1R} W_Y S X_{1D} X T}.
\]
\hspace{1cm} (D-15)

We now eliminate the variable \(T\) from (D-13) and (D-14) as follows. The RHS of (D-13) can be bounded as
\[
I(X_{1D}; X_2, S; Y_3[T]) - I(S; X_2, Y_3[T])
\leq H(Y_3) - H(Y_3|X_{1D}, X_2, S) - H(S|T)
+ H(S|X_2, Y_3, T)
= I(X_{1D}, X_2, S; Y_3) - H(S|T) + H(S|X_2, Y_3, T)
\leq I(X_{1D}, X_2, S; Y_3) - I(S; X_2, Y_3)
= I(X_{1D}; Y_3 S, X_2) + I(X_2; Y_3)
\]
\hspace{1cm} (D-16)
where
\((k)\) holds since \(H(Y_3|T) \leq H(Y_3)\) and \(H(Y_3|X_{1D}, X_2, S) = H(Y_3|X_{1D}, X_2, S)\) (by the Markovian relation \(T \leftrightarrow (X_{1D}, X_2, S) \leftrightarrow Y_3\)); and
\((l)\) holds since \(S\) is independent of \(T\) and \(H(S|X_2, Y_3, T) \leq H(S|X_2, Y_3)\).

Similarly, the RHS of (D-13) can be bounded as
\[
I(X_{1R}; Y_2 S, X_2, T) + I(X_{1D}; Y_3 S, X_2, T)
\leq I(X_{1R}; Y_2 S X_2) + I(X_{1D}; Y_3 S X_2).
\]
\hspace{1cm} (D-17)

Finally, combining (D-2), (D-12), (D-16) at one hand, and (D-2), (D-11), (D-17) at the other hand, we get
\[
R \leq I(X_{1D}; Y_3 S X_2) + I(X_{1D}; Y_3 S X_2)
\]
\hspace{1cm} (D-18a)
\[
R \leq I(X_{1R}; Y_2 S X_2) + I(X_{1D}; Y_3 S X_2)
\]
\hspace{1cm} (D-18b)
where the distribution on \(\{S, X_{1R}, X_{1D}, X_2, Y_2, Y_3\}\), obtained by marginalizing (D-15) over the variable \(T\), has the form given in (38).

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

APPENDIX E
PROOF OF THEOREM 7

The encoding and transmission scheme is as follows. Let \(P_{t_r} \geq 0, P_{t_d} \geq 0,\) and \(D \geq 0\) be given such that \(P_{t_r} + P_{t_d} \leq P_t\) and \(0 \leq D \leq Q\). Also, consider the test channel \(\tilde{S}_R = aS + \tilde{S}_R\), where \(a := 1 - D/Q\) and \(\tilde{S}_R\) is a Gaussian random variable with zero mean and variance \(\sigma_{\tilde{S}_R}^2 = D(1 - D/Q)\), independent from \(S\). Using this test channel, we calculate \(E[(S - \tilde{S}_R)^2] = D \) and \(E[\tilde{S}_R^2] = Q - D\). Let \(X_2 \sim \mathcal{N}(0, P_2)\) be jointly Gaussian with \(S\) with \(E[X_2\tilde{S}_R] = 0\) and independent from \(S\), and \(X_{SR} \sim \mathcal{N}(0, \theta P_{t_r})\) jointly Gaussian with \(\{S, \tilde{S}_R\}\) with \(E[X_{SR} S] = 0\) and \(E[X_{SR}\tilde{S}_R] = 0\), where \(0 \leq \theta \leq 1\). Also, let \(X_{wR} \sim \mathcal{N}(0, \tilde{P}_{t_r})\) be jointly Gaussian with \(\{X_2, S\}\) and independent of \(X_{SR}\), with \(E[X_{wR} S] = \sigma_{s_1}\) and \(E[X_{wR}\tilde{S}_R] = \sigma_{s_2}\) and \(X_{wD} \sim \mathcal{N}(0, P_{t_d})\) jointly Gaussian with and independent of \(\{X_{wR}, X_{SR}, X_2, S, \tilde{S}_R\}\). In what follows, we use the random variables \(V, U, U_{t_r}, U_{t_d}\) which we will use in the sequel. Also, recall the definition of \(Q, \xi, a_2\) in (62) and (63), respectively, which we will use in the rest of this proof.

We decompose the message \(W\) to be sent from the source into two parts \(W_r\) and \(W_d\). The input \(X_n^r\) from the source is divided into three independent parts, i.e., \(X_n^r = X_n^{SR} + X_n^{wR} + X_n^{wD}\), where \(X_n^{SR}\) carries a description \(\hat{S}_R\) of the state \(S\) that is intended to be recovered only at the relay and has power constraint \(n\tilde{P}_{t_r}\), \(X_n^{wR}\) carries message \(W_r\) and has power constraint \(n\tilde{P}_{t_r}\), and \(X_n^{wD}\) carries message \(W_d\) and has power constraint \(nP_{t_d}\), with \(P_t = P_{t_r} + P_{t_d}\). The message \(W_r\) is sent through the relay at rate \(R\), and the message \(W_d\) is sent directly to the destination at rate \(R_d\). The total rate is \(R = R_t + R_d\).

As in the discrete case, a block Markov encoding is used. Let \(w_{r,i} = (w_{t_r,i}, w_{t_d,i}) \in [1, 2^{nR_r}] \times [1, 2^{nR_d}]\) denote the message to be transmitted in block \(i\) and \(s_{[i]}\) denote the state controlling the channel in block \(i\). The source quantizes \(s_{[i]}\) into \(\hat{s}_R \in \tilde{S}_R\), where \(\tilde{S}_R = \{1, \ldots, \alpha S\}\). Using the aforementioned test channel, the source can encode \(s_{[i]}\) successfully at the quantization rate \(R_{t_r} = I(S; \hat{S}_R)\).

\begin{equation}
R_{t_r} = I(S; \hat{S}_R) - \frac{1}{2} \log(Q/D).
\end{equation}

In the beginning of block \(i\), the relay has decoded correctly message \(w_{r,i-1}\) and the index \(t_{r,i-1}\) of the description \(\hat{s}_R\) sent by the source in the previous block \(i-1\) (this will be justified below) and sends a Gaussian signal \(x_{t_r,i} = x_{t_r,i}^{SR} + x_{t_r,i}^{wR} + x_{t_r,i}^{wD}\) as noncausal CSI at the transmitter, as \(x_{t_r,i}^{wR} = x_{t_r,i}^{wR} + x_{t_r,i}^{wD}\) and obtained via a DPC considering \(s_{[i]}\). The components of \(v[i]\) are generated i.i.d. using the auxiliary random variable \(V\).

Let \(t_{r,i}\) be the index associated with the state \(s_{[i+1]}\) of the next block \(i+1\). In the beginning of block \(i\), the source sends a superposition of three Gaussian vectors
\[
\begin{align*}
x_{t_r,i}^{SR} &= \sum_{i=1}^{\alpha S} \sqrt{\tilde{P}_{t_r}} \hat{s}_R[i] + \sqrt{\tilde{P}_{t_r}} x_{t_r,i}^{wR} + \sqrt{\tilde{P}_{t_r}} x_{t_r,i}^{wD} \\
x_{t_r,i}^{wR} &= \sqrt{\tilde{P}_{t_r}} x_{t_r,i}^{wR} + \sqrt{\tilde{P}_{t_r}} x_{t_r,i}^{wD}
\end{align*}
\]
\hspace{1cm} (E-2)
where the components of \(v[i]\) are generated i.i.d. using the auxiliary random variable \(V\).

Let \(t_{r,i}\) be the index associated with the state \(s_{[i]}\) of the next block \(i+1\). In the beginning of block \(i\), the source sends a superposition of three Gaussian vectors
\[
\begin{align*}
x_{t_r,i}^{SR} &= \sum_{i=1}^{\alpha S} \sqrt{\tilde{P}_{t_r}} \hat{s}_R[i] + \sqrt{\tilde{P}_{t_r}} x_{t_r,i}^{wR} + \sqrt{\tilde{P}_{t_r}} x_{t_r,i}^{wD} \\
x_{t_r,i}^{wR} &= \sqrt{\tilde{P}_{t_r}} x_{t_r,i}^{wR} + \sqrt{\tilde{P}_{t_r}} x_{t_r,i}^{wD}
\end{align*}
\]
\hspace{1cm} (E-3)
In (E-3), the vectors $\mathbf{x}_{SR}[t_{Ri}]$ and $\mathbf{x}_{wd}[w_{di}]$ are generated i.i.d. using the auxiliary random variables $X_{SR}$ and $X_{WD}$, respectively, and the vector $\mathbf{x}_{wR}[w_{ri}]$ has power $n(1 - \rho_2^2 - \rho_3^2)\bar{B}P_{tr}$ and is independent of $\mathbf{s}[i], \mathbf{x}_2[w_{ri-1}], \mathbf{x}_{SR}[t_{Ri}], \text{ and } \mathbf{x}_{wd}[w_{di}]$. Furthermore, the vector $\mathbf{x}_{SR}[t_{Ri}]$ carries a description $\hat{s}_{R}[t_{Ri}]$ of the state $t_{Ri+1}$ that affects transmission in the next block $i + 1$, intended to be recovered only at the relay; the vector $\mathbf{x}_2[w_{ri-1}]$ carries cooperative information $w_{ri-1}$, and the vector $\mathbf{x}_{wR}[w_{ri}]$ carries new information $w_{ri}$. The vectors $\mathbf{x}_{SR}[t_{Ri}], \mathbf{x}_{wd}[w_{di}], \text{ and } \mathbf{x}_{wR}[w_{ri}]$ are obtained via DPCs considering (E-4a) as noncausal CSI at the transmitter, as

$$\mathbf{x}_{SR}[t_{Ri}] = \mathbf{u}_R[i] - \frac{\theta P_{tr}}{\theta P_{tr} + N_2 + P_{id}}(1 - \alpha)s[i] \quad (E-4a)$$

$$\mathbf{x}_{wd}[w_{di}] = \mathbf{u}_1[i] - \frac{P_{id}}{P_{id} + N_2 + \theta P_{tr}Q(1 - \alpha)}(\mathbf{s}[i] - \alpha\hat{s}_{R}[t_{Ri-1}]) \quad (E-4b)$$

$$\mathbf{x}_{wR}[w_{ri}] = \mathbf{u}_2[i] - \alpha\mathbf{x}_2[i] - \alpha\hat{s}_{R}[t_{Ri-1}] \quad (E-4c)$$

where the components of $\mathbf{u}_R[i], \mathbf{u}_1[i], \text{ and } \mathbf{u}_2[i]$ are generated i.i.d. using the auxiliary random variables $U_R, U_1, \text{ and } U$, respectively.

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

$$\mathbf{y}_2[i] = \mathbf{x}_{SR}[t_{Ri}] + \rho_{12}\sqrt{\frac{\bar{B}P_{tr}}{P_2}}\mathbf{x}_2[w_{ri-1}] + \mathbf{x}_{wR}[w_{ri}]$$

$$+ \left(1 + \rho_{12}\sqrt{\frac{\bar{B}P_{tr}}{Q}}\right)\mathbf{s}[i] + (\mathbf{z}_2[i] + \mathbf{x}_{wd}[w_{di}]). \quad (E-5)$$

The relay knows $w_{ri-1}$ and $t_{Ri-1}$ and decodes the pair $(w_{ri-1}, t_{Ri-1})$ from $\mathbf{y}_2[i]$. The relay decodes $w_{ri-1}$ and $t_{Ri}$ successively, starting by $w_{ri-1}$. To decode $w_{ri}$, the relay subtracts out the quantity $\rho_{12}\sqrt{\bar{B}P_{tr}/P_2}\mathbf{x}_2[w_{ri-1}] + \alpha\hat{s}_{R}[t_{Ri-1}]$ from $\mathbf{y}_2[i]$ to make the channel equivalent to

$$\hat{\mathbf{y}}_2[i] = \mathbf{x}_{wR}[w_{ri}] + \mathbf{z}_2[i] - \alpha\hat{s}_{R}[t_{Ri-1}] + \mathbf{z}_2[i] + \mathbf{x}_{wd}[w_{di}]$$

$$+ \left(\mathbf{z}_2[i] + \mathbf{x}_{SR}[t_{Ri}] + \mathbf{x}_{wd}[w_{di}]. \quad (E-6)$$

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

$$R_c \leq I(U; \hat{Y}_2) - I(U; S - \alpha_2\hat{s}_{R})$$

$$= R\left(\alpha, (1 - \rho_2^2 - \rho_3^2)\bar{B}P_{tr}, \xi^2\bar{Q}, N_2 + \theta P_{tr} + P_{id}\right) \quad (E-7)$$

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

$$\hat{\mathbf{y}}_2[i] = \mathbf{x}_{wR}[w_{ri}] + \mathbf{z}_2[i] + (\mathbf{z}_2[i] + \mathbf{x}_{wd}[w_{di}]). \quad (E-8)$$

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

$$\hat{R}_R \leq I(U_R; \hat{Y}_2) - I(U_R; S)$$

$$= \frac{1}{2}\log\left(1 + \frac{\theta P_{tr}}{N_2 + P_{id}}\right) \quad (E-9)$$

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

$$\mathbf{y}_3[i] = \mathbf{x}_2[i] + \mathbf{x}_3[w_{ri-1}] + \mathbf{s}[i] + \mathbf{z}_3[i]$$

$$= (\rho_{12}\sqrt{\frac{\bar{B}P_{tr}}{P_2}} + 1)\mathbf{x}_2[w_{ri-1}] + \mathbf{x}_{wR}[w_{ri}] + \mathbf{x}_{wd}[w_{di}]$$

$$+ (\rho_{12}\sqrt{\frac{\bar{B}P_{tr}}{Q}} + 1)\mathbf{s}[i] + \mathbf{z}_3[i] + \mathbf{x}_{SR}[t_{Ri}]). \quad (E-10)$$

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

$$R_r \leq I(V, U; Y_3) - I(V, U; S, \hat{s}_{R})$$

$$- [I(V; Y_3) - I(V; \hat{s}_{R})] + [I(U; Y_3|V) - I(U; S, \hat{s}_{R} V)] \quad (E-11)$$

where the equality follows since the choice of $(V, \hat{s}_{R})$ in (64) satisfying $V \rightarrow \hat{s}_{R} \rightarrow S$ is a Markov chain.

We first compute the term $I(V; Y_3) - I(V; \hat{s}_{R})$. Let $\hat{s}[i]$ be the estimation error of $s[i]$ given $s_{R}[t_{Ri-1}]$ under MMSE criterion. Since $s[i]$ and $s_{R}[t_{Ri-1}]$ are jointly Gaussian, $\hat{s}[i]$ is i.i.d. Gaussian with variance $E[(s - \hat{s}_{R})^2] = \xi^2D$ per element and is independent from $s_{R}[t_{Ri-1}]$. Thus, we can alternatively write the output $\mathbf{y}_3[i]$ as

$$\mathbf{y}_3[i] = (\rho_{12}\sqrt{\frac{\bar{B}P_{tr}}{P_2}} + 1)\mathbf{x}_2[w_{ri-1}] + \mathbf{x}_{wR}[w_{ri}] + \mathbf{x}_{wd}[w_{di}]$$

$$+ \xi\hat{s}_{R}[t_{Ri-1}] + (\mathbf{z}_3[i] + \mathbf{x}_{SR}[t_{Ri}] + \hat{s}[i]). \quad (E-12)$$

With the choice of the auxiliary random variable $V$ as in (64) and that of the associated Costa’s scale factor $\alpha_2$ set to its optimal value as in (63), the destination decodes the vector $\mathbf{v}[i]$ correctly from $\mathbf{y}_3[i]$ at rate

$$I(V; Y_3) - I(V; \hat{s}_{R})$$

$$= \frac{1}{2}\log\left(1 + \frac{\rho_{12}\sqrt{\bar{B}P_{tr}} + \sqrt{P_2}}{N_3 + \xi^2D + 6P_{tr} + (1 - \rho_2^2 - \rho_3^2)\bar{B}P_{tr} + P_{id}}\right) \quad (E-13)$$
where the equality follows through straightforward algebra. Let us now compute the term \([I(U; Y_3 | V) - I(U; S, \mathcal{S}_R | V)]\). Observing that the destination can peel off \(v[i-1]\) from \(y_3[i-1]\) to make the channel equivalent to

\[
\hat{y}_3[i-1] = y_3[i-1] - \left( \rho_{12} \sqrt{\frac{P_{tr}}{P_2}} \mathbf{x}_2[w_{tr-2}] + \alpha_2 \xi \mathcal{S}_R [\mathcal{R}i-2] \right)
\]

\[
- \mathbf{x}_w[w_{tr-1}] + \xi (S[i-1] - \alpha_2 \xi \mathcal{S}_R [\mathcal{R}i-2])
\]

\[
+ (z_3[i-1] + \mathbf{x}_{SR}[\mathcal{R}i-1] + \mathbf{x}_{wd}[w_{di-1}])
\]

(E-14)

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

\[
I(U; Y_3 | V) - I(U; S, \mathcal{S}_R | V)
\]

\[
= I(U; \hat{Y}_3) - I(U; S, \xi (S - \alpha_2 \xi \mathcal{S}_R))
\]

\[
=R(\alpha_1 (1 - \rho_{12}^2 - \rho_{13}^2 \xi^2 P_{tr}^2 Q N_3 + \eta P_{tr} + P_{tr}^a))
\]

(E-15)

where the last equality follows through straightforward algebra. Finally, the destination can peel off \(u[i-1]\) from \(\hat{y}_3[i-1]\) to make the channel equivalent to

\[
\hat{y}_3[i-1] = y_3[i-1] - \left( \rho_{12} \sqrt{\frac{P_{tr}}{P_2}} \mathbf{x}_2[w_{tr-2}] + \alpha_2 \xi \mathcal{S}_R [\mathcal{R}i-2] \right)
\]

\[
- \mathbf{x}_w[w_{tr-1}] + \xi (1 - \alpha_2) (S[i-1] - \alpha_2 \xi \mathcal{S}_R [\mathcal{R}i-2])
\]

\[
+ (z_3[i-1] + \mathbf{x}_{SR}[\mathcal{R}i-1] + \mathbf{x}_{wd}[w_{di-1}])
\]

(E-16)

From (E-16), it is easy to see that if \(n\) is large, and with the choice of the auxiliary random variable \(U\) as in (64), the destination obtains the vector \(u[i-1]\) correctly from \(\hat{y}_3[i-1]\) which carries message \(w_{di-1}\) at rate

\[
R_d \leq I(U_1; \hat{Y}_3) - I(U_1; \xi (S - \alpha_2 \xi \mathcal{S}_R))
\]

\[
= \frac{1}{2} \log(1 + \frac{P_{tr}^a}{N_3 + \eta P_{tr}}).
\]

(E-17)

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

**APPENDIX F**

**PROOF OF PROPOSITION 2**

In the proof, we compute the rate (33) of Proposition 1 using an appropriate jointly Gaussian distribution on \((S, U_1, X_{1R}, X_{1D}, X_2)\). The algebra in this section is similar to that in the proof of [23, Th. 3] and [17, Th. 6]. We first compute the term \([I(U_1; Y_3 | X_{1R}, X_2) - I(U_1; S, X_{1R}, X_2)]\) in the RHS of (33) because this gives insights about the distribution that we should use to compute the lower bound. We assume that \(X_{1R}, X_{1D}, X_2\) are jointly Gaussian random variables with zero mean and variance \(P_{1R}, P_{1D}, P_2\), respectively. The random variables \(X_{1R}\) and \(X_2\) are independent and independent of the state \(S\). The random variable \(X_{1D}\) is independent of \(X_{1R}\) and jointly Gaussian with \((S, X_2)\), with \(E[X_{1D} X_2] = \rho_{12} \sqrt{P_{1D} P_2}\) and \(E[X_{1D} S] = \rho_{13} \sqrt{P_{1D} Q}\), for some correlation coefficients \(\rho_{12} \in [-1, 1]\) and \(\rho_{13} \in [-1, 1]\).

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

\[
\hat{X}_{1D} = \rho_{12} \sqrt{\frac{P_{1D}}{P_2}} X_2 + \rho_{13} \sqrt{\frac{P_{1D}}{Q}} S + Z_3.
\]

(F-1)

\[
X'_{1D} = X_{1D} - \hat{X}_{1D}.
\]

(F-2)

We can then write \(Y_3\) as shown in (81) alternatively as

\[
Y_3 = X'_{1D} + \left(1 + \rho_{12} \sqrt{\frac{P_{1D}}{P_2}}\right) X_2 + \left(1 + \rho_{13} \sqrt{\frac{P_{1D}}{Q}}\right) S + Z_3.
\]

(F-3)

Let now

\[
Y'_3 : = Y_3 - E[Y_3 | X_{1R}, X_2] = X'_{1D} + \left(1 + \rho_{13} \sqrt{\frac{P_{1D}}{Q}}\right) S + Z_3.
\]

(F-4)

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

\[
U_1 = X'_{1D} + \alpha (1 + \rho_{13} \sqrt{\frac{P_{1D}}{Q}}) S
\]

(F-5)

where \(\alpha\) is Costa's parameter given by

\[
\alpha = \frac{E[X'_{1D}^2]}{E[X'_{1D}^2] + E[Z_3^2]} = \frac{P_{1D}(1 - \rho_{12}^2 - \rho_{13}^2)}{P_{1D}(1 - \rho_{12}^2 - \rho_{13}^2) + N_3}
\]

(F-6)

Then, we can easily show that

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

(F-7)

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

\[
U_1 = X_{1D} - \rho_{12} \sqrt{\frac{P_{1D}}{P_2}} X_2 + \alpha_{opt} S
\]

(F-8)

with

\[
\alpha_{opt} = \alpha (1 + \rho_{13} \sqrt{\frac{P_{1D}}{Q}}) - \rho_{13} \sqrt{\frac{P_{1D}}{Q}}
\]

\[
\left(1 - \frac{P_{1D}}{P_2} \right) \left(1 - \frac{P_{1D}}{Q} \right) + N_3 \left(1 + \frac{P_{1D}}{Q} \right) - \rho_{13} \sqrt{\frac{P_{1D}}{Q}}
\]

(F-9)
Now, it is easy to see that, with the choice (F-8), we have
\[
I(U_1; Y_2, X_{1R}, X_2) - I(U_1; S | X_{1R}, X_2) = I(U_1; Y'_2) - I(U_1; S) = \frac{1}{2} \log \left( 1 + \frac{E[X_2^2]}{N_2} \right) - \frac{1}{2} \log \left( 1 + \frac{P_{1D}(1 - \rho_{12}^2 - \rho_{1s}^2)}{N_3} \right).
\]  
(F-10)

We now compute the terms \(I(X_{1R}; Y_2 | X_2)\) and \(I(X_2; Y_3)\). It is easy to see that with the aforementioned jointly Gaussian input distribution
\[
I(X_{1R}; Y_2 | X_2) = I(X_{1R}; Y_2) = \frac{1}{2} \log(1 + \frac{P_{1R}}{N_2 + Q}).
\]  
(F-11)

Also, we have
\[
I(X_{1R}, X_2, Y_3) \overset{(a)}{=} I(X_2; Y_3) = h(Y_3) - h(Y_3 | X_2) = h(Y_3) - h(X_{1D} + E|X_{1D} |X_2| + E|X_{1D}| S + Z_3 | X_2) = \frac{1}{2} \log \left( 1 + \frac{E[X_{1D}^2] + E|S + E|X_{1D}|S|^2 + E[Z_3^2]}{P_{1D}(1 - \rho_{12}^2 - \rho_{1s}^2)} \right) \overset{(b)}{=} \frac{1}{2} \log \left( 1 + \frac{E[X_{1D}^2] + E|S + E|X_{1D}|S|^2 + E[Z_3^2]}{P_{1D}(1 - \rho_{12}^2 - \rho_{1s}^2)} \right) \overset{(c)}{=} \frac{1}{2} \log \left( 1 + \frac{E[X_{1D}^2] + E|S + E|X_{1D}|S|^2 + E[Z_3^2]}{P_{1D}(1 - \rho_{12}^2 - \rho_{1s}^2)} \right)
\]  
(F-12)

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

Adding (F-10) and (F-11), we obtain the first term of the minimization in (82); and adding (F-10) and (F-12), we obtain the second term of the minimization in (82).

Finally, we obtain the capacity in Theorem 9 by maximizing the RHS of (82) over all possible values of \(\rho_{12} \in [-1, 1]\) and \(\rho_{1s} \in [-1, 0]\). Investigating the two terms of the minimization, we can easily see that it suffices to consider \(\rho_{12} \in [0, 1]\) and \(\rho_{1s} \in [-1, 0]\). This concludes the proof of Theorem 9.

**APPENDIX G**

**PROOF OF THEOREM 8**

In this section, we first use the upper bound for the DM case in Theorem 5 to obtain a new upper bound on the capacity of the state-dependent additive Gaussian model (72). Then, we show that this new upper bound is maximized by jointly Gaussian \((S, X_{1R}, X_{1D}, X_2, Z_2, Z_3)\).

From Theorem 5, we have that, given any \((\epsilon_n, n, R)\) sequence of codes with average error probability \(P_e^m \longrightarrow 0\) as \(n \longrightarrow +\infty\), the transmission rate \(R\) satisfies
\[
R \leq \min \left\{ I(X_{1R}; Y_2, X_2; S), I(X_{1D}; Y_3 | X_2, S) \right\} + I(X_{1D}; Y_3 | X_2, S)
\]  
(G-1)

for some joint measure of the form
\[
P_{S, X_{1R}, X_{1D}, X_2, Y_2, Y_3} = Q_{S} P_{X_{1R}} X_{1D} P_{X_{1D}|X_2} P_{X_2} Y_2 X_{1R}, S W_{Y_2} x_{1R}, s W_{Y_3} | x_{1D}, x_2 s.
\]  
(G-2)

Since the channel structure (72) satisfies \(W_{Y_2} x_{1R}, x_2 s = W_{Y_2} | x_{1R} s\), it follows that
\[
I(X_{1R}; Y_2 | X_2, S, X_2) = H(Y_2 | S, X_2) - H(Y_2 | S, X_2, X_{1R}) = H(Y_2 | S, X_2) - H(Y_2 | S, X_{1R}) \leq H(Y_2 | S) - H(Y_2, S, X_{1R}) = I(X_{1R}; Y_2 | S).
\]  
(G-3)

An upper bound on the capacity of the channel (72) is then given by
\[
R \leq \min \left\{ I(X_{1R}; Y_2; S), I(X_2; Y_3) \right\} + I(X_{1D}; Y_3 | X_2, S)
\]  
(G-4)

for some joint measure of the form
\[
P_{S, X_{1R}, X_{1D}, X_2, Y_2, Y_3} = Q_{S} P_{X_{1R}} X_{1D} P_{X_{1D}|X_2} P_{X_2} Y_2 X_{1R}, S W_{Y_2} x_{1R}, s W_{Y_3} | x_{1D}, x_2 s.
\]  
(G-5)

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

Fix a joint distribution on \((S, X_{1R}, X_{1D}, X_2, Y_2, Y_3)\) of the form (G-5) satisfying
\[
E[X_{1R}^2] = \tilde{P}_{1R} \leq P_{1R}, E[X_{1D}^2] = \tilde{P}_{1D} \leq P_{1D}, E[X_{1D} X_2] = \sigma_{12}, E[X_{1D} S] = \sigma_{1s}.
\]  
(G-6)

We shall also use the correlation coefficients \(\rho_{12} \in [-1, 1]\), \(\rho_{1s} \in [-1, 0]\) defined as
\[
\rho_{12} = \frac{\sigma_{12}}{\sqrt{\tilde{P}_{1D} \tilde{P}_{2}}}, \ \rho_{1s} = \frac{\sigma_{1s}}{\sqrt{\tilde{P}_{1D} Q}}.
\]  
(G-7)

We first compute the first term in the minimization on the RHS of (G-4). We have
\[
R \leq I(X_{1R}; Y_2 | S) + I(X_{1D}; Y_3 | X_2, S) = h(X_{1R} + Z_2 | S) - h(Z_2) + h(X_{1D} + Z_3 | X_2, S) - h(Z_3)
\]  
(G-8)

\[
\overset{(u)}{\leq} h(X_{1R} + Z_2) - h(Z_2) + h(X_{1D} + Z_3 | X_2, S) - h(Z_3)
\]  
(G-9)

\[
\overset{(v)}{\leq} \frac{1}{2} \log \left( 1 + \frac{\tilde{P}_{1R}}{N_2} \right) + \frac{1}{2} \log \left( 1 + \frac{\tilde{P}_{1D}(1 - \rho_{12}^2 - \rho_{1s}^2)}{N_3} \right)
\]  
(G-10)

where \(u\) holds since conditioning reduces entropy; and \(h\) holds since the conditional differential entropy \(h(X_{1R} + Z_2)\) is maximized if \((X_{1R}, Z_2)\) are jointly Gaussian and, by the Maximum Conditional Differential Entropy Lemma [53, Part I], the conditional differential entropy \(h(X_{1D} + Z_3 | X_2, S)\) is maximized if \((S, X_{1D}, X_2, Z_3)\) are jointly Gaussian.
We now compute the term $[I(X_2; Y_3) + I(X_{1D}; Y_3, X_2, S)]$. We have

$$I(X_2; Y_3) + I(X_{1D}; Y_3 | X_2, S)$$

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

$$= h(Y_3 | S, X_{1D}, X_2) - h(Y_3 | S) - h(S | X_{1D}, X_2)$$

$$+ h(S | X_1, Y_3)$$

$$= h(Y_3) - h(S) + h(S | X_2, Y_3) - h(Z_2) \quad (G-12)$$

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

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

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

where equality is attained if $Y_2$ is Gaussian. Similarly, the term $h(S | X_2, Y_3)$ is maximized if $(S, X_2, Y_3)$ are jointly Gaussian.

Let $\hat{S}(X_2, Y_3) = E[S | X_2, Y_3]$ be the MMSE estimator of $S$ given $(X_2, Y_3)$, i.e.,

$$\hat{S}(X_2, Y_3) = E[S | X_2, X_{1D} + S + Z_3]$$

$$= \gamma_1 X_2 + \gamma_2 (X_{1D} + S + Z_3) \quad (G-14)$$

with

$$\gamma_1 = -\frac{\sigma_{1s}(Q + \sigma_{1s})}{P_{1D} + 2\sigma_{1s} + Q + N_3} - \sigma_{12}$$

$$\gamma_2 = \frac{1}{P_{1D} + 2\sigma_{1s} + Q + N_3} \quad (G-15)$$

$$h(S | X_2, Y_3) = h(S - \hat{S}(X_2, Y_3) | X_2, Y_3)$$

$$\leq h(S - \gamma_1 X_2 - \gamma_2 (X_{1D} + S + Z_3))$$

$$= \frac{1}{2} \log(2\pi e) \left( \left( S - \gamma_1 X_2 - \gamma_2 (X_{1D} + S + Z_3) \right)^2 \right)$$

$$= \frac{1}{2} \log (2\pi e) \left( \frac{Q P_{1D} + \tilde{P}_2 + 2\sigma_{1s} + Q - \sigma_{12}^2 + \sigma_{1s}^2}{P_{1D} + 2\sigma_{1s} + Q + N_3} \right)$$

$$= \frac{1}{2} \log \left( \frac{Q P_{1D} + \tilde{P}_2 + 2\sigma_{1s} + Q - \sigma_{12}^2 + \sigma_{1s}^2}{N_3} \right) \quad (G-16)$$

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

$$I(X_2; Y_3) + I(X_{1D}; Y_3 | S, X_2)$$

$$= \frac{1}{2} \log \left( \frac{Q P_{1D} + \tilde{P}_2 + 2\sigma_{1s} + Q - \sigma_{12}^2 + \sigma_{1s}^2}{N_3} \right)$$

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

$$A_{X_{1R}, X_{1D}, X_2, S} = \begin{pmatrix} P_{1R} & 0 & 0 & 0 \\ 0 & \tilde{P}_{1R} & \rho_{12} \sqrt{P_{1D} P_2} & \rho_{1s} \sqrt{P_{1D} Q} \\ 0 & \rho_{12} \sqrt{P_{1D} P_2} & \tilde{P}_2 & 0 \\ 0 & \rho_{1s} \sqrt{P_{1D} Q} & 0 & Q \end{pmatrix} \quad (G-19)$$

that satisfy

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

and have nonnegative discriminant

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

i.e., for $Q > 0$

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

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

Also, it is easy to see that, for fixed $\tilde{P}_{1D}$, the functions $\Theta_1(\tilde{P}_{1R}, \tilde{P}_{1D}, \rho_{12}, \rho_{1s})$ and $\Theta_2(\tilde{P}_{1R}, \tilde{P}_{1D}, \rho_{12}, \rho_{1s})$ increase monotonically with $\tilde{P}_{1R}$ and $\tilde{P}_2$. So, for fixed $\tilde{P}_{1D}$, they are maximized at $\tilde{P}_{1R} = P_{1R}$ and $\tilde{P}_2 = P_2$. To complete the proof, we should show that $\Theta_1(P_{1R}, \tilde{P}_{1D}, \rho_{12}, \rho_{1s})$ and $\Theta_2(\tilde{P}_{1R}, \tilde{P}_{1D}, \rho_{12}, \rho_{1s})$ are also maximized at $\tilde{P}_{1D} = P_{1D}$.

It is clear that the function $\Theta_1(P_{1R}, \tilde{P}_{1D}, \rho_{12}, \rho_{1s})$ increases with $\tilde{P}_{1D}$. The term $\Theta_2(\tilde{P}_{1R}, \tilde{P}_{1D}, \rho_{12}, \rho_{1s})$ can be seen as the sum rate of a two-user state-dependent MAC with state information $S^* \text{ known to one encoder, both encoders sending a common message, and the informed encoder sending, in addition, an individual message}$ [17]. As argued in [17], this sum rate increases with the power of the informed encoder [17, Appendix E], i.e., $\tilde{P}_{1D}$ here. This concludes the proof of Theorem 5.

APPENDIX H

PROOF OF THEOREM 9

1) Converse Part: the proof of the converse part of Theorem 9 follows by noticing that the computation of the upper bound (G-4) in the proof of Theorem 8 for the special case (81), and using the same jointly Gaussian distribution as in Appendix G, gives the RHS of (82).

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

[1] A. El Gamal and S. Zahedi, “Capacity of a class of relay channels with orthogonal components,” IEEE Trans. Inf. Theory, vol. IT-51, no. 5, pp. 1803–1817, May 2005.

[2] P. Moulin and J. A. O’Sullivan, “Information-theoretic analysis of information hiding,” IEEE Trans. Inf. Theory, vol. 49, no. 3, pp. 563–593, Mar. 2003.

[3] A. Zaidi, P. Piantanida, and P. Duhamel, “Broadcast- and MAC-aware coding strategies for multiple user information embedding,” IEEE Trans. Signal Process., vol. 55, no. 6, pp. 2974–2992, Jun. 2007.

[4] A. Zaidi and L. Vandendorpe, “Coding schemes for relay-assisted information embedding,” IEEE Trans. Inf. Forensics Security, vol. 4, no. 1, pp. 70–85, Jan. 2009.

[5] A. V. Kusnetsov and B. S. Tsybakov, “Coding in a memory with defective cells,” Probl. Predach. Inf., vol. 10, pp. 52–60, 1974.

[6] G. Caire and S. Shamai (Shitz), “On the achievable throughput of a multi-antenna Gaussian broadcast channel,” IEEE Trans. Inf. Theory, vol. IT-49, no. 7, pp. 1691–1706, Jul. 2003.

[7] 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, no. 10, pp. 2658–2668, Oct. 2003.

[8] P. Viswanath and D. N. C. Tse, “Sum capacity of the vector Gaussian MIMO broadcast channel,” IEEE Trans. Inf. Theory, vol. IT-49, no. 8, pp. 1912–1921, Aug. 2003.

[9] R. Zamir, S. Shamai (Shitz), and U. Erez, “Nested linear/lattice codes for structured multi-terminal binning,” IEEE Trans. Inf. Theory, vol. IT-48, no. 6, pp. 1250–1276, Jun. 2002.

[10] E. Biglieri, J. Proakis, and S. Shamai (Shitz), “Fading channels: Information-theoretic and communication aspects,” IEEE Trans. Inf. Theory, vol. 44, no. 6, pp. 2610–2622, Oct. 1998.

[11] J. Mitola, Cognitive radio: An integrated agent architecture for software defined radio Ph.D. dissertation, KTH Royal Inst. Technol., Dept. Elect. Eng., Stockholm, Sweden, 2000.

[12] C. E. Shannon, “Channels with side information at the transmitter,” IBM J. Res. Develop., vol. 2, pp. 289–293, Oct. 1958.

[13] S. I. Gel’fand and M. S. Pinsker, “Coding for channel with random parameters,” Probl. Control Inf. Theory, vol. 9, pp. 19–31, 1980.

[14] C. D. Heegard and A. El Gamal, “On the capacity of computer memory with defects,” IEEE Trans. Inf. Theory, vol. IT-29, no. 5, pp. 731–739, Sep. 1983.

[15] M. H. M. Costa, “Writing on dirty paper,” IEEE Trans. Inf. Theory, vol. 29, no. 3, pp. 439–441, May 1983.

[16] G. Keshtev, Y. Steinberg, and N. Merhav, Channel Coding in the Presence of Side Information: Subject Review, ser. Foundations and Trends in Communications and Information Theory. Nowell, USA: Now Publishers, 2008.

[17] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdú, “Cooperative multiple access encoding with states available at one transmitter,” IEEE Trans. Inf. Theory, vol. 55, no. 6, pp. 4448–4469, Oct. 2009.

[18] S. Kotagiri and J. N. Laneman, “Multiaccess channels with state known to some encoders and independent messages,” EURASIP J. Wireless Commun. Netw., vol. 2008, pp. 450680-1–450680-14, 2008.

[19] A. Zaidi, S. Kotagiri, J. N. Laneman, and L. Vandendorpe, “Multiaccess channels with state known to one encoder: Another case of degraded message sets,” in Proc. IEEE Int. Symp. Inf. Theory, Seoul, Korea, Jun.–Jul. 2009, pp. 2376–2380.

[20] A. Khisti, U. Erez, A. Lapidoth, and G. Wornell, “Carbon copying onto independent states each known causally at one encoder,” in Proc. IEEE Int. Symp. Inf. Theory, Austin, TX, USA, Jun. 2010, pp. 480–484.

[21] A. Zaidi and L. Vandendorpe, “Lower bounds on the capacity of the relay channel with states at the source,” EURASIP J. Wireless Commun. Netw., vol. 2009, pp. 634296-1–634296-22, 2009.

[22] Y. Cemal and Y. Steinberg, “The multiple-access channel with partial state information at the encoders,” IEEE Trans. Inf. Theory, vol. IT-51, no. 11, pp. 3992–4003, Nov. 2005.

[23] Y. Steinberg, “Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information,” IEEE Trans. Inf. Theory, vol. IT-51, no. 8, pp. 2867–2877, Aug. 2005.

[24] A. Lapidoth and Y. Steinberg, “The multiple access channel with causal and strictly causal side information at the encoders,” in Proc. Int. Zurich. Semin. Commun., Zurich, Switzerland, Mar. 2010, pp. 13–16.

[25] A. Zaidi and L. Vandendorpe, “Rate regions for the partially-cooperative relay broadcast channel with non-causal side information,” in Proc. IEEE Int. Symp. Inf. Theory, vol. 57, no. 10, pp. 6379–6396, Oct. 2011.

[26] M. Li, I. Simeone, and A. Yener, “Multiple access channels with states causally known at transmitters,” IEEE Trans. Inf. Theory vol. 59, no. 3, pp. 1394–1404, Mar. 2013.

[27] M. Li, I. Simeone, and A. Yener, “Message and state cooperation in a relay channel when only the relay knows the state,” IEEE Trans. Inf. Theory 2011, to be published [Online]. Available: http://arxiv.org/abs/1102.0768.

[28] I. Bross and A. Lapidoth, “The state-dependent multiple-access channel with states available at a cribbing encoder,” IEEE 26th Conv. Elect. Electron. Eng.. Eilat, Israel, 2010.

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

[30] M. N. Khormuji and M. Skoglund, “On cooperative downlink transmission with frequency reuse,” in Proc. IEEE Int. Symp. Inf. Theory, Seoul, Korea, Jun.–Jul. 2009, pp. 849–853.

[31] G. Como and S. Yüksel, “On the capacity of memoryless finite state multiple-access channels with asymmetric state information at the encoders,” IEEE Trans. Inf. Theory vol. 57, no. 3, pp. 1267–1273, Mar. 2011.

[32] N. Sen, G. Como, S. Yüksel, and F. Alajaji, “On the capacity of memoryless finite-state multiple-access channels with asymmetric noisy state information at the encoders,” in Proc. 49th Annu. Allerton Conf. Commun., Control, Comput., Monticello, IL, USA, Sep. 2011, pp. 1210–1215.

[33] R. Khosravi-Farsani and F. Marvasti, “Capacity bounds for multiuser channels with non-causal channel state information at the transmitters,” in Proc. IEEE Inf. Theory Workshop, Paraty, Brazil, Oct. 2011, pp. 195–199.

[34] S. I. Bross, A. Lapidoth, and M. Wigger, “Dirty-paper coding for the Gaussian multiaccess channel with conferencing,” IEEE Trans. Inf. Theory, vol. 58, no. 9, pp. 5640–5668, Sep. 2012.

[35] S. Kotagiri and J. N. Laneman, “Achievable rates for multiple access channels with state information known at one encoder,” presented at the Allerton Conf. Commun., Control, Comput., Monticello, IL, USA, 2004.

[36] S. Kotagiri and J. Laneman, “Multiaccess channels with state known to one encoder: A case of degraded message sets,” in Proc. IEEE Int. Symp. Inf. Theory, Nice, France, Jun. 2007, pp. 1556–1559.

[37] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdú, “Cooperative encoding with asymmetric state information at the transmitters,” presented at the Allerton Conf. Commun., Control, Comput., Monticello, IL, USA, Sep. 2006.

[38] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdú, “Cooperative multiple access decoding with states available at one transmitter,” in Proc. IEEE Int. Symp. Inf. Theory, St. Petersburg, Russia, 2011, pp. 2801–2805.

[39] A. Zaidi, P. Piantanida, and S. Shamai (Shitz), “Wyner-Ziv type versus noisy network coding for a state-dependent MAC,” in Proc. IEEE Int. Symp. Inf. Theory, Cambridge, MA, USA, Jul. 2012, pp. 1682–1686.

[40] A. Zaidi, P. Piantanida, and S. Shamai (Shitz), “Capacity region of multiple access channel with states known noncausally at one encoder and only strictly causally at the other encoder,” IEEE Trans. Inf. Theory 2012, to be published [Online]. Available: http://arxiv.org/abs/1201.3278.
[47] T. M. Cover and A. E. Gamal, “Capacity theorems for the relay channel,” IEEE Trans. Inf. Theory, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.
[48] Y.-H. Kim, A. Sutivong, and S. Sigurjonnson, “Multiple user writing on dirty paper,” in Proc. IEEE Int. Symp. Inf. Theory, Chicago, IL, USA, Jun. 2004, pp. 534–538.
[49] A. E. Gamal and T. M. Cover, “Achievable rates for multiple descriptions,” IEEE Trans. Inf. Theory, vol. IT-28, no. 6, pp. 851–857, Nov. 1982.
[50] K. Marton, “A coding theorem for the discrete memoryless broadcast channel,” IEEE Trans. Inf. Theory, vol. IT-25, no. 3, pp. 306–311, May 1979.
[51] A. E. Gamal and E. C. van der Meulen, “A proof of Marton’s coding theorem for the discrete memoryless broadcast channel,” IEEE Trans. Inf. Theory, vol. IT-27, pp. 120–122, Jan. 1981.
[52] Y. Steinberg and S. Shamai (Shitz), “Achievable rates for the broadcast channel with states known at the transmitter,” in Proc. IEEE Int. Symp. Inf. Theory, Adelaide, Australia, Sep. 2005, pp. 2184–2188.
[53] A. E. Gamal and Y.-H. Kim, Network Information Theory. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[54] A. E. Gamal and M. Aref, “The capacity of the semideterministic relay channel,” IEEE Trans. Inf. Theory, vol. IT-28, no. 3, p. 536, May 1982.
[55] Y.-K. Chia, R. Soundararajan, and T. Weissman, “Estimation with a helper who knows the interference,” IEEE Int. Symp. Inf. Theory 2012, pp. 706–710.
[56] R. G. Gallager, Information Theory and Reliable Communication. New York, NY, USA: Wiley, 1968.
[57] S. H. Lim, Y.-H. Kim, A. E. Gamal, and S.-Y. Chung, “Noisy network coding,” IEEE Trans. Inf. Theory, vol. 57, no. 5, pp. 3132–3152, May 2011.
[58] A. Avestimehr, S. Diggavi, and D. Tse, “Wireless network information flow: A deterministic approach,” IEEE Trans. Inf. Theory, vol. 57, no. 4, pp. 1872–1905, Apr. 2011.
[59] M. N. Khormuji and M. Skoglund, “Noisy network coding approach to the relay channel with a random state,” in Proc. 45th Annu. Conf. Inf. Sci. Syst., Mar. 2011, pp. 1–5.
[60] T. M. Cover and J. A. Thomas, Elements of Information Theory. New York, NY, USA: Wiley, 1991.
[61] S. H. Lim, P. Minero, and Y.-H. Kim, “Lossy communication of correlated sources over multiple access channels,” in 48th Annu. Allerton Conf. Commn., Control Comput., Monticello, IL, USA, Sep.–Oct. 2010.
[62] I. Csiszár and J. Körner, “Broadcast channels with confidential messages,” IEEE Trans. Inf. Theory, vol. 24, no. 3, pp. 339–348, May 1978.
[63] I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems. London, U.K.: Academic, 1981.

Abdellatif Zaidi received the B.S. degree in Electrical Engineering from École Nationale Supérieure de Techniques Avancées, ENSTA ParisTech, France in 2002 and the M. Sc. and Ph.D. degrees in Electrical Engineering from École Nationale Supérieure des Télécommunications, TELECOM ParisTech, Paris, France in 2002 and 2005, respectively. From December 2002 to December 2005, he was with the Communications an Electronics Dept., TELECOM ParisTech, Paris, France and the Signals and Systems Lab, CNRS/Supélec, France pursuing his PhD degree. From May 2006 to September 2010, he was at École Polytechnique de louvain, Université Catholique de Louvain, Belgium, working as a research assistant. Dr. Zaidi was “Research Visitor” at the University of Notre Dame, Indiana, USA, during fall 2007 and Spring 2008. He is now, an assistant professor at Université Paris-Est Marne-La-Vallee, France. He is a member of Laboratoire d’Informatique Gaspard Monge (LIGM).

Dr. Zaidi’s research interests cover a broad range of topics from network information theory and communication theory. Of particular interest are the problems of multi-terminal information theory, Shannon theory, relaying and cooperation, network coding, physical layer security, source coding and interference mitigation in multi-user channels.

Shlomo Shamai (Shitz) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, in 1975, 1981 and 1986 respectively.

During 1975–1985 he was with the Communications Research Labs, in the capacity of a Senior Research Engineer. Since 1986 he is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, where he is now a Technion Distinguished Professor, and holds the William Fondiller Chair of Telecommunications. His research interests encompasses a wide spectrum of topics in information theory and statistical communications.

Dr. Shamai (Shitz) is an IEEE Fellow and a member of the Israeli Academy of Sciences and Humanities. He is the recipient of the 2011 Claude E. Shannon Award. He has also been awarded the 1999 van der Pol Gold Medal of the Union RadioScientifique Internationale (URSI), and is a co-recipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003, and the 2004 joint IT/COM societies paper award, the 2007 IEEE Information Theory Society Paper Award, the 2009 European Commission FP7, Network of Excellence in Wireless Communications (NEWCOM++) Best Paper Award, and the 2010 Thomson Reuters Award for International Excellence in Scientific Research. He is also the recipient of 1985 Alon Grant for distinguished young scientists and the 2000 Technion Henry Taub Prize for Excellence in Research. He has served as Associate Editor for the Shannon Theory of the IEEE Transactions on Information Theory, and has also served twice on the Board of Governors of the Information Theory Society. He is a member of the Executive Editorial Board of the IEEE Transactions on Information Theory.

Pablo Piantanida received the B.Sc. and M.Sc degrees (with honors) in Electrical Engineering from the University of Buenos Aires (Argentina), in 2003, and the Ph.D. from the Paris-Sud University (France) in 2007. In 2008, he has been with the Department of Communications and Radio-Frequency Engineering at Vienna University of Technology (Austria). Since October 2007 he has joined in 2007 the Department of Telecommunications, SUPELEC, as an Assistant Professor in network information theory. His research interests include multi-terminal information theory, Shannon theory, cooperative communications, physical-layer security and coding theory for wireless applications.

Luc Vandendorpe (M’93–SM’99–F’06) was born in Monsour, Belgium, in 1962. He received the electrical engineering degree (summa cum laude) and the Ph.D. degree from the Université Catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in 1985 and 1991, respectively.

Since 1985, he has been with the Communications and Remote Sensing Laboratory of UCL, where he first worked in the field of bit rate reduction techniques for video coding. In 1992, he was a Visiting Scientist and Research Fellow at the Telecommunications and Traffic Control Systems Group of the Delft Technical University, The Netherlands, where he worked on spread spectrum techniques for personal communications systems. From October 1992 to August 1997, he was Senior Research Associate of the Belgian NSF at UCL, and invited Assistant Professor. He is currently a Professor and head of the Institute for Information and Communication Technologies, Electronics and Applied Mathematics. His current interest include digital communication systems and, more precisely, resource allocation for OFDMA-based multicell systems, MIMO and distributed MIMO, sensor networks, turbo-based communications systems, physical layer security, and UWB based positioning.