Bounds and Capacity Theorems for Cognitive Interference Channels with State [1] [2]

Ruchen Duan, Yingbin Liang [3]

Abstract

A class of cognitive interference channels with state are investigated, in which a primary transmitter sends a message to two receivers (receivers 1 and 2) with assistance of a cognitive transmitter (that knows the primary transmitter's message), and the cognitive transmitter also sends a separate message to receiver 2. The channel is corrupted by an independent and identically distributed (i.i.d.) state sequence. The scenario, in which the state sequence is noncausally known at both the cognitive transmitter and receiver 2, is first studied. The capacity region is obtained for both the discrete memoryless and Gaussian channels. The second scenario, in which the state sequence is noncausally known only at the cognitive transmitter, is further studied. Inner and outer bounds on the capacity region are obtained for the discrete memoryless channel and its degraded version. The capacity region is characterized for the degraded semideterministic channel and for channels that satisfy a less noisy condition. The Gaussian channels are further studied, which are partitioned into two cases based on how the interference compares with the signal at receiver 1. For each case, inner and outer bounds on the capacity region are derived, and partial boundaries of the capacity region are characterized. The full capacity region is also characterized for channels that satisfy certain conditions. It is shown that certain Gaussian channels achieve the capacity of the same channels with state noncausally known at both the cognitive transmitter and receiver 2.

1 Introduction

Interference channels with state can model many communication scenarios in practical wireless systems such as cellular networks, sensor networks, and cognitive radio networks. In these networks, communication between one transmitter-receiver pair may be interfered by signals from other communicating pairs which share the same spectrum resource with them. The state may be caused by many reasons such as channel uncertainty and transmitter-side signal interference. Consequently, transmission rates of these users (or in general, the throughput of a system), are affected by strength of interference and state, and by how interference and state are handled in designing transmission schemes. Therefore, it is important to understand the fundamental communication limits (i.e., the capacity region) of interference channels with state.

---

[1]The material in this paper was presented in part at the 49th Allerton Conference on Communication, Control, and Computing, Monticello, Illinois, USA, September 2011 and in part at the IEEE International Symposium on Information Theory, Cambridge, Massachusetts, USA, July 2012.

[2]The work of R. Duan and Y. Liang was supported by a National Science Foundation under Grant CCF-10-26566 and by the National Science Foundation CAREER Award under Grant CCF-10-26565.

[3]The authors are with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 USA (email: {yliang06,rduan}@syr.edu).

A few interference channel models with state noncausally known at transmitters have been studied, which are generalizations of the Gel'fand-Pinsker model [1] for the point-to-point channel with state. In [2] and [3], the interference channel model with two transmitters sending two messages respectively to two receivers was studied. The channel is corrupted by an independent and identically distributed (i.i.d.) state sequence, which is noncausally known at both transmitters. A number of achievable schemes were proposed and their corresponding rate regions were compared. In [4], the interference channel is corrupted by two independent states, each of which interferes one receiver. The states are available at their corresponding transmitters. The capacity region is obtained for the strong interference regime with the state power going to infinity. In [5] and [6], a model of the cognitive interference channel with state was studied, in which both transmitters (i.e., the primary and cognitive transmitters) jointly send one message to receiver 1, and the cognitive transmitter sends an additional message separately to receiver 2. The i.i.d. state sequence is noncausally known at the cognitive transmitter only. Inner and outer bounds on the capacity region were provided.

In this paper, we investigate a different cognitive interference channel model with state (see Fig. 1), in which a primary transmitter sends a message to both receivers (receivers 1 and 2) with assistance of a cognitive transmitter that knows the primary transmitter's message, and the cognitive transmitter also sends an additional message to receiver 2. The channel is corrupted by an i.i.d. state sequence. The difference of our model from the model studied in [5] and [6] lies in that the common message jointly sent by both transmitters needs to be decoded at both receivers instead of only at receiver 1 as in [5] and [6]. Although the two models appear similar to each other, their capacity regions can have different forms, and the transmission schemes achieving these regions can also be different. This fact is already demonstrated by the two corresponding models without state studied respectively in [7–13] and [14]. The capacity bounds in [7–13] and the capacity region in [14] have different forms, and are achieved by different achievable schemes. Therefore, our study can lead to new information theoretic insights.

We further note that compared to the basic Gel'fand-Pinsker model, the cognitive interference channel model we study here and in [5] and [6] capture more communication features such as the transmitter-side signal cognition and receiver-side signal interference in addition to random state corruption of the channel. This type of models are well motivated in practical networks. For example, it is often the case in cognitive radio networks that a primary transmitter wishes to send a common message to a number of primary receivers, and a cognitive transmitter (which often knows the primary transmitter's message via its necessary coordination with the primary transmitter) can cooperatively send the common message to the primary receivers. This cognitive transmitter may also have its own message intended to one of the primary receivers. State corruption of the channel may arise because the cognitive transmitter can communicate to some secondary receivers simultaneously, and its signals to these receivers then interfere with the primary receivers as state, which is clearly known by the cognitive transmitter. A similar scenario can also occur in cellular networks. For example, two base stations may cooperatively send certain common information to many receivers which are near the edge between the two cells that the two base stations serve. In addition, one of the base stations may transmit additional information to receivers in its own cell.

We investigate two scenarios of cognitive interference channels with state. The first scenario assumes that the state sequence is noncausally known at both the cognitive transmitter and receiver 2, and is referred to as the CIC-DM-STR (which stands for the cognitive interference channel with degraded message sets and with state information noncausally known at both the cognitive transmitter and receiver 2). The second scenario assumes that the state sequence is noncausally known only at the cognitive transmitter, and is referred to as the CIC-DM-ST (which stands for the cognitive interference channel with degraded message sets and with state information noncausally known at only the cognitive transmitter). Our goal in this paper is to study the performance (i.e., the capacity region) of both scenarios and correspondingly design communication schemes to exploit the noncausal state information in the context of signal cognition and interference.

In the following, we summarize the main results of this paper. We note that due to the channel properties of cognition, interference, random channel state, and asymmetry of the state knowledge, it is natural that an achievable scheme employs coding techniques of superposition, rate splitting, and Gel'fand-Pinsker coding. The novelty of this paper lies in finding optimality of such types of achievable schemes (i.e., achievement of the capacity region) by properly integrating these coding techniques for various channel parameters. The new ingredients that we develop in the converse arguments are also mentioned below.

For the CIC-DM-STR model, we characterize the full capacity region for both the discrete memoryless and Gaussian channels. More specifically, for the discrete memoryless channel, we first obtain an inner bound which is based on an achievable scheme as a special case and a much simpler version of the scheme designed for the CIC-DM-ST. This is because receiver 2 knows the state and can hence remove state interference from its output. In this way, an achievable scheme needs to deal with only state interference at receiver 1. Whereas for the CIC-DM-ST, in which the state information is known at neither receiver, an achievable scheme needs to deal with state interference for both receivers. Such a channel with compound states is much more challenging to analyze. We then derive an outer bound on the capacity region, which takes a different form from the inner bound. Standard techniques do not provide an easy argument of the equivalence of the two bounds. We apply the technique recently developed by Lapidoth and Wang in [15] for proving equivalence of two rate regions characterized by different sets of auxiliary random variables, and show that our inner and outer bounds match. We note that the CIC-DM-STR model was also studied in [16], in which the capacity region of the channel is not fully obtained.

We further study the Gaussian CIC-DM-STR model. Although for the Gaussian channel, it is natural to obtain an achievable region by applying the general jointly Gaussian input distribution to the inner bound derived for the discrete memoryless channel, the resulting region is too complex to analyze. It is then very difficult to develop a converse proof for capacity characterization. Our approach is to partition the Gaussian channel parameters into two classes, and to characterize the full capacity region for each class. More specifically, for each case, we develop a simpler inner bound that exploits the conditions that the channel satisfies, and we are then able to derive outer bounds that match the inner bounds and hence establish the capacity region. We further note that outer bounds we derive for the CIC-DM-STR are useful for providing outer bounds (sometimes tight outer bounds as demonstrated in Section 5.2) on the capacity region for the CIC-DM-ST.

We then study the CIC-DM-ST model. For the discrete memoryless channel, we derive inner and outer bounds on the capacity region. In particular, due to asymmetry of the state knowledge (i.e., the primary transmitter does not know the channel state but the cognitive transmitter does), the cognitive transmitter not only helps the primary transmitter in the conventional way of superposition, but also helps to correlate the input with the state sequence via Gel'fand-Pinsker scheme. Thus, we employ the Gel'fand-Pinsker scheme for these two cooperative transmitters in the way that the primary transmitter generates signals with only the message index, superposing on which the cognitive transmitter generates auxiliary random variables with the bin index. We show by special cases that such auxiliary random variables are all necessary to achieve the capacity.

We then study the degraded channel, and obtain bounds on the capacity region. It is not surprising that the capacity region for the degraded channel is not obtained because it is difficult to obtain the capacity region even for the degraded broadcast channel with state [17]. Nevertheless, we establish the capacity region for degraded channels, which further satisfy the semideterminsitic condition. This example channel demonstrates that both superposition and Gel'fand-Pinsker coding for state treatment in the cognitive transmitter's cooperation are necessary for achieving the capacity. Besides the semideterministic degraded channel, we also identify a less noisy condition under which we obtain the capacity region.

We further study the Gaussian channel. As for the CIC-DM-STR model, we also partition the channel parameters into two cases. For the first case, in which the channel gain of interference is stronger than the channel gain of the signal at receiver 1, it is reasonable to let receiver 1 decode full information intended for receiver 2. We derive inner bound based on such a scheme. We also provide an outer bound and further identify rate points, at which inner and outer bounds match at the boundary. These points hence characterize partial boundary of the capacity region. We further identify a condition, under which the outer bound fully characterizes the capacity region.

For the second Gaussian case, in which the channel gain of interference is weaker than the channel gain of the signal at receiver 1, we obtain two inner bounds with the Gel'fand-Pinsker scheme canceling the state respectively at receivers 1 and 2. Similarly to the first Gaussian case, for each inner bound, we provide an outer bound and identify rate points that the inner and outer bounds match at the boundary. We further show that respectively under two channel conditions, each outer bound characterizes the full capacity region. In particular, one of these conditions leads to the case that the Gaussian CIC-DM-ST achieves the capacity region of the Gaussian CIC-DM-STR. This is similar to the case that dirty paper coding achieves the capacity of the Gaussian channel when the state is also known at the receiver [18]. Here, the channel does not achieve the capacity with both receivers knowing the channel state due to asymmetry of the state knowledge at the transmitter side.

The rest of the paper is organized as follows. In Section 2, we describe the channel model and explain the notations used in this paper. In Sections 3, we present our results for the CIC-DM-STR for both the discrete memoryless and Gaussian channels. In Sections 4 and 5, we present our results for the CIC-DM-ST for the discrete memoryless channel and the Gaussian channel, respectively. Finally, in Section 6, we conclude with a few remarks.

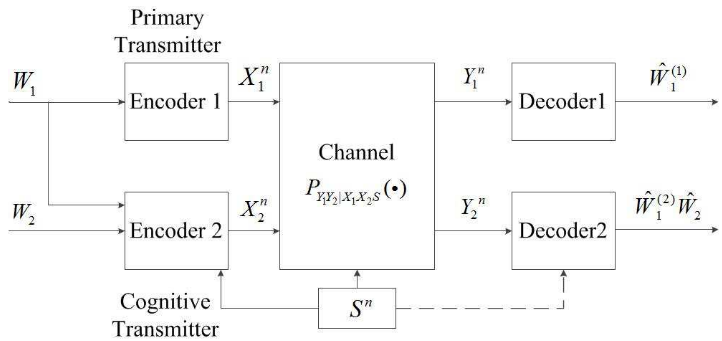

Figure 1: An illustration of the CIC-DM-STR (including the dashed line) and CIC-DM-ST (without the dashed line) models

2 Channel Model

We consider a class of cognitive interference channels with state (see Fig. 1), in which two transmitters (referred to as the primary transmitter and the cognitive transmitter) jointly send a message W_1 to two receivers (say receivers 1 and 2), and the cognitive transmitter sends a message W_2 to receiver 2. The channel is also corrupted by an i.i.d. state sequence S^n. We investigate two scenarios. The first scenario assumes that the state sequence is noncausally known at both the cognitive transmitter and receiver 2, and is referred to as the CIC-DM-STR. The second scenario assumes that the state sequence is known only at the cognitive transmitter, and is referred to as the CIC-DM-ST. In this paper, we use s^n to denote the vector $(s_1,\cdots,s_n)$, and use s_i^n to denote the vector $(s_i,\cdots,s_n)$. We formally define the channel model as follows.

Definition 1. *The discrete memoryless CIC-DM-ST and CIC-DM-STR consist of two finite channel input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, a finite state alphabet $\mathcal{S}$, two finite channel output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, and a transition probability distribution $P_{Y_1Y_2|X_1X_2S}$, where $X_1 \in \mathcal{X}_1$ and $X_2 \in \mathcal{X}_2$ are the channel inputs from the primary and cognitive transmitters, respectively, $S \in \mathcal{S}$ is the state variable, and $Y_1 \in \mathcal{Y}_1$ and $Y_2 \in \mathcal{Y}_2$ are the channel outputs at receivers 1 and 2, respectively.*

Definition 2. *A $(2^{nR_1}, 2^{nR_2}, n)$ code for the CIC-DM-ST consists of the following:*

- *two message sets: $\mathcal{W}_k = 1, 2,\cdots, 2^{nR_k}$ for $k = 1, 2$;*
- *two messages: W_1 and W_2 are independent random variables and are uniformly distributed over $\mathcal{W}_1$ and $\mathcal{W}_2$, respectively;*
- *two encoders: an encoder $f_1 : \mathcal{W}_1 \to \mathcal{X}_1^n$, which maps a message $w_1 \in \mathcal{W}_1$ to a codeword $x_1^n \in \mathcal{X}_1^n$; and an encoder $f_2 : \mathcal{W}_1 \times \mathcal{W}_2 \times S^n \to \mathcal{X}_2^n$, which maps a message pair $(w_1, w_2) \in \mathcal{W}_1 \times \mathcal{W}_2$ and a state sequence $s^n \in S^n$ to a codeword $x_2^n \in \mathcal{X}_2^n$;*
- *two decoders: $g_1 : \mathcal{Y}_1^n \to \mathcal{W}_1$, which maps a received sequence y_1^n into a message $\hat{w}_1^{(1)} \in \mathcal{W}_1$; and $g_2 : \mathcal{Y}_2^n \to \mathcal{W}_1 \times \mathcal{W}_2$, which maps a received sequence y_2^n into a message pair $\left(\hat{w}_1^{(2)}, \hat{w}_2\right) \in \mathcal{W}_1 \times \mathcal{W}_2$.*

We note that the above definition is also applicable to the CIC-DM-STR, if the second decoder is changed to $g_2 : (\mathcal{Y}_2^n, S^n) \to \mathcal{W}_1 \times \mathcal{W}_2$.

For a given code, we define the probability of error as

$$P_e^{(n)} = \frac{1}{2^{n(R_1+R_2)}} \sum_{w_1=1}^{2^{nR_1}} \sum_{w_2=1}^{2^{nR_2}} Pr\left\{ \left(\hat{w}_1^{(1)}, \hat{w}_1^{(2)}, \hat{w}_2 \right) \neq (w_1, w_1, w_2) \right\}. \tag{1}$$

A rate pair (R_1, R_2) is said to be *achievable* if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes such that

$$\lim_{n\to\infty} P_e^{(n)} = 0. \tag{2}$$

Definition 3. *The capacity region is defined to be the closure of the set of all achievable rate pairs* (R_1, R_2).

In the following, we define a number of channel conditions for classifying the channels in our study:

- $P_{Y_1Y_2|X_1X_2S} = P_{Y_2|X_1X_2S}P_{Y_1|Y_2}$ (3)
- $P_{Y_1Y_2|X_1X_2S} = P_{Y_2|X_1X_2S}P_{Y_1|Y_2X_1S}$ (4)
- $P_{Y_1Y_2|X_1X_2S} = P_{Y_1|X_1X_2S}P_{Y_2|Y_1X_1S}$ (5)
- $I(X_1; Y_1) \leq I(X_1; Y_2)$ *and* $I(U; Y_1|X_1) \leq I(U; Y_2|X_1)$
  for all $P_{UX_1X_2S}$ s.t. $P_{X_1SUX_2Y_1Y_2} = P_{X_1}P_SP_{UX_2|SX_1}P_{Y_1Y_2|SX_1X_2}$ (6)
- $I(X_1U; Y_1) \geq I(X_1U; Y_2)$
  for all $P_{UX_1X_2S}$ s.t. $P_{X_1SUX_2Y_1Y_2} = P_{X_1}P_SP_{UX_2|SX_1}P_{Y_1Y_2|SX_1X_2}$ (7)

We describe the intuitive meaning of the above conditions as follows. The concrete meaning will be clearer in the contexts when these conditions are applied in later sections.

- Condition (3) is a degradedness condition, which implies that Y_1 is a degraded version of Y_2, i.e., receiver 2 receives stronger signals than receiver 1, and hence, receiver 2 is able to decode all information that receiver 1 can decode.

- Condition (4) is a conditional degradedness condition, which is weaker than condition (3). For this case, only given X_1 and S, Y_1 is a degraded version of Y_2. Intuitively, this implies that receiver 2 is stronger in decoding X_2 (and hence W_2) than receiver 1 if X_1 and S are given.

- Condition (5) is opposite to condition (4) with roles of Y_1 and Y_2 being switched. Here, receiver 1 is stronger in decoding X_2 (and hence W_2) than receiver 2 if X_1 and S are given. Such a condition suggests an achievable scheme, in which receiver 1 decodes W_2 (although not required by the system) in order to perform successive interference cancellation for decoding W_1 at a higher rate.

- Condition (6) is a less noisy condition implying that Y_2 is a less noisy version of Y_1, similar to the less noisy condition defined in [19]. Compared to condition (3), it is a weaker condition.

- Condition (7) is a less noisy condition indicating that Y_1 is a less noisy version of Y_2.

We also study the Gaussian CIC-DM-ST and CIC-DM-STR models defined as follows.

Definition 4. *The Gaussian CIC-DM-ST and CIC-DM-STR have outputs at receivers 1 and 2 for one symbol time given by*

$$Y_1 = X_1 + aX_2 + S + N_1 \tag{8a}$$
$$Y_2 = bX_1 + X_2 + cS + N_2 \tag{8b}$$

where the noise variables $N_1 \sim \mathcal{N}(0,1)$ and $N_2 \sim \mathcal{N}(0,1)$, and the state variable $S \sim \mathcal{N}(0,Q)$. Both the noise variables and the state variable are i.i.d. over channel uses. The channel inputs are subject to the average power constraints

$$\frac{1}{n}\sum_{i=1}^{n} X_{1i}^2 \leqslant P_1 \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} X_{2i}^2 \leqslant P_2. \tag{9}$$

We note that although some channel gain coefficients in (8a) and (8b) are set to be one and the noise variables have unit variance, the above definition for Gaussian channels is without loss of generality. This is because the power constraints for X_1, X_2 and S can be arbitrary, and hence any linear relationship between (X_1, X_2, S) and (Y_1, Y_2) can be reduced to the above model with certain power constraints on X_1, X_2 and S. It is also due to the same reason that the noise variables can be normalized without loss of generality. More specifically, suppose we have a general Gaussian model described as follows for one symbol time:

$$Y_1 = a_{11}X_1 + a_{12}X_2 + a_{13}S + N_1$$
$$Y_2 = b_{21}X_1 + b_{22}X_2 + b_{23}S + N_2$$

with power constraints on X_1, X_2 and S being respectively P_1', P_2' and Q', and with noise variances being σ_1^2 and σ_2^2. It can be seen that this model is equivalent to the model defined in (8a) and (8b) with $|a| = \left|\frac{a_{12}\sigma_2}{b_{22}\sigma_1}\right|$, $|b| = \left|\frac{b_{21}\sigma_1}{a_{11}\sigma_2}\right|$, $|c| = \left|\frac{b_{23}\sigma_1}{a_{13}\sigma_2}\right|$, $P_1 = \frac{a_{11}^2}{\sigma_1^2}P_1'$, $P_2 = \frac{b_{22}^2}{\sigma_2^2}P_2'$, and $Q = \frac{a_{13}^2}{\sigma_1^2}Q'$.

3 The CIC-DM-STR Model

In this section, we study the CIC-DM-STR model, in which the state information is known at both the cognitive transmitter and receiver 2. This channel is of interest by its own, and the capacity of this channel also provides a useful outer bound for characterizing the capacity for the CIC-DM-ST, in which the state information is known only at the cognitive transmitter.

3.1 The CIC-DM-STR Model: Discrete Memoryless Channels

We first design an achievable scheme that includes superposition coding, rate-splitting, and Gel'fand-Pinsker binning scheme. We outline the achievable scheme as follows. The primary transmitter first encodes W_1. Then the cognitive transmitter cooperatively encodes and transmits W_1 using superposition. Moreover, the cognitive transmitter employs rate splitting for transmitting W_2, i.e., splits W_2 into two components W_{21} and W_{22} with W_{21} intended for both receivers to decode and W_{22} intended only for receiver 2 to decode. The cognitive transmitter encodes W_{21} and W_{22} by superposing them on W_1. Furthermore, since the cognitive transmitter knows the channel state information, it employs Gel'fand-Pinsker scheme via an auxiliary random variable U (in the following capacity region) to reduce state interference for receiver 1 to decode W_1 and W_{21}. Hence, U contains information of both W_1 and W_{21}, and plays dual roles: helping to cancel state interference and serving as a rate splitting random variable for carrying the message W_{21}. We also note that since receiver 2 has the knowledge of the state, no additional auxiliary random variable is needed for cancelling state interference for receiver 2.

As we commented in Section 1, the CIC-DM-STR is easier to analyze than the CIC-DM-ST, because receiver 2 knows the state and can hence remove the state interference from its output. In this way, the design of achievable schemes needs to deal with only the state interference at receiver 1. Whereas for the CIC-DM-ST, in which the state information is known at neither receiver, the achievable scheme needs to deal with state interference at both receivers. This involves the design for compound states, and hence results in a more challenging problem.

We characterize the full capacity region for the CIC-DM-STR in the following theorem.

Theorem 1. *(Capacity) The capacity region for the CIC-DM-STR consists of rate pairs* (R_1, R_2) *satisfying:*

$$R_1 \leqslant I(X_1U;Y_1) - I(U;S|X_1) \tag{10a}$$
$$R_2 \leqslant I(X_2;Y_2|SX_1) \tag{10b}$$
$$R_1 + R_2 \leqslant I(X_1X_2;Y_2|S) \tag{10c}$$
$$R_1 + R_2 \leqslant I(X_1U;Y_1) + I(X_2;Y_2|X_1US) - I(U;S|X_1) \tag{10d}$$

for some distribution $P_{X_1SUX_2Y_1Y_2} = P_{X_1}P_SP_{UX_2|X_1S}P_{Y_1Y_2|SX_1X_2}$, *where* U *is an auxiliary random variable and its cardinality is bounded by* $|\mathcal{U}| \leqslant |\mathcal{X}_1||\mathcal{X}_2||\mathcal{S}| + 1$.

Proof. Since the CIC-DM-STR can be viewed as a special case of the CIC-DM-ST with $Y_2 = (Y_2, S)$, the achievability proof follows directly from the achievable region for the CIC-DM-ST given in (22a)-(22e) by setting $T = X_1$, $V = X_2$ and $Y_2 = Y_2S$.

For the converse, we first obtain the following outer bound consisting of rate pairs (R_1, R_2) satisfying

$$R_1 \leqslant I(KX_1;Y_1) - I(K;S|X_1) \tag{11a}$$
$$R_2 \leqslant I(X_2;Y_2|SX_1) \tag{11b}$$
$$R_1 + R_2 \leqslant I(X_1X_2;Y_2|S) \tag{11c}$$
$$R_1 + R_2 \leqslant I(TKX_1;Y_1) - I(TK;S|X_1) + I(X_2;Y_2|X_1TKS) \tag{11d}$$

for some distribution $P_{X_1STKX_2Y_1Y_2} = P_{X_1}P_SP_{KT|X_1S}P_{X_2|X_1SKT}P_{Y_1Y_2|SX_1X_2}$, where K and T are auxiliary random variables. The proof is detailed in Appendix A.

In order to show that the region (10a)-(10d) is the capacity region, it is sufficient to show that the above outer bound (11a)-(11d) is a subset of the region (10a)-(10d). Towards this end, we apply the technique in [15] and analyze the outer bound (11a)-(11d) by considering the following two cases.

If $I(T;Y_1|KX_1) - I(T;S|KX_1) \leqslant 0$, the outer bound (11a)-(11d) can be further bounded as:

$$\begin{aligned}
R_1 &\leqslant I(KX_1;Y_1) - I(K;S|X_1) && \text{(12a)}\\
R_2 &\leqslant I(X_2;Y_2|SX_1) && \text{(12b)}\\
R_1 + R_2 &\leqslant I(X_1X_2;Y_2|S) && \text{(12c)}\\
R_1 + R_2 &\leqslant I(KX_1;Y_1) - I(K;S|X_1) + [I(T;Y_1|KX_1) - I(T;S|KX_1)] + I(X_2;Y_2|X_1TKS) \\
&\leqslant I(KX_1;Y_1) - I(K;S|X_1) + I(X_2;Y_2|X_1KS). && \text{(12d)}
\end{aligned}$$

which implies that the outer bound (11a)-(11d) is contained in (10a)-(10d) by setting $U = K$ in (10a)-(10d).

If $I(T;Y_1|KX_1) - I(T;S|KX_1) \geqslant 0$, the outer bound (11a)-(11d) can be further bounded as:

$$\begin{aligned}
R_1 &\leqslant I(KX_1;Y_1) - I(K;S|X_1) \\
&= I(KTX_1;Y_1) - I(KT;S|X_1) - [I(T;Y_1|KX_1) - I(T;S|KX_1)] \\
&\leqslant I(KTX_1;Y_1) - I(KT;S|X_1) && \text{(13a)}\\
R_2 &\leqslant I(X_2;Y_2|SX_1) && \text{(13b)}\\
R_1 + R_2 &\leqslant I(X_1X_2;Y_2|S) && \text{(13c)}\\
R_1 + R_2 &\leqslant I(TKX_1;Y_1) - I(TK;S|X_1) + I(X_2;Y_2|X_1KTS) && \text{(13d)}
\end{aligned}$$

which also implies that the outer bound (11a)-(11d) is contained in (10a)-(10d) by setting $U = KT$ in (10a)-(10d). □

Remark 1. *By setting X_1 to be deterministic, Theorem 1 reduces to the capacity region for the broadcast channel with degraded message sets and with the state information noncausally known at both the transmitter and receiver 2, which consists of rate pairs satisfying*

$$\begin{aligned}
R_1 &\leqslant I(U;Y_1) - I(U;S) && \text{(14a)}\\
R_1 + R_2 &\leqslant I(X;Y_2|S) && \text{(14b)}\\
R_1 + R_2 &\leqslant I(U;Y_1) + I(X;Y_2|US) - I(U;S) && \text{(14c)}
\end{aligned}$$

for some distribution $P_{SUXY_1Y_2} = P_SP_{UX|S}P_{Y_1Y_2|SX}$, where X is the channel input, and Y_1 and Y_2 are channel outputs respectively at two receivers.

3.2 The CIC-DM-STR Model: Gaussian Channels

In this section, we characterize the capacity region for the Gaussian CIC-DM-STR as described in Definition 4. We partition Gaussian channels into two classes based on the value of the channel parameter

a, and characterize the capacity region for each class. We note that our results for Gaussian channels exploit the fact that for both $|a| \leqslant 1$ and $|a| > 1$, the Gaussian channel is stochastically degraded given X_1 and S, i.e., its marginal distributions at the two receivers are the same as a physically degraded Gaussian channel that satisfies the conditions (4) and (5), respectively. Because the capacities of the two Gaussian channels are the same, our results below are applicable to both stochastically degraded and physically degraded channels with the proofs exploiting the physical degradedness conditions (4) and (5).

We first provide the capacity region for the Gaussian channel with $|a| \leqslant 1$. As later shown in Theorems 11 and 13, this capacity region serves as the tight converse for characterizing the partial or full boundary of the capacity region for the Gaussian CIC-DM-ST.

Theorem 2. *(Capacity) For the Gaussian CIC-DM-STR, if $|a| \leqslant 1$, the capacity region consists of rate pairs (R_1, R_2) satisfying:*

$$R_1 \leqslant \frac{1}{2}\log\left(1+\frac{P_1+2a\rho_{21}\sqrt{P_1P_2}+a^2\rho_{21}^2P_2}{a^2(1-\rho_{21}^2)P_2+2a\rho_{2s}\sqrt{P_2Q}+Q+1}\right)+\frac{1}{2}\log\left(1+\frac{a^2P_2'}{a^2P_2''+1}\right) \tag{15a}$$

$$R_2 \leqslant \frac{1}{2}\log(1+P_2'') \tag{15b}$$

$$R_1+R_2 \leqslant \frac{1}{2}\log\left(1+b^2P_1+2b\rho_{21}\sqrt{P_1P_2}+(1-\rho_{2s}^2)P_2\right) \tag{15c}$$

where $P_2'+P_2''=(1-\rho_{21}^2-\rho_{2s}^2)P_2$, $P_2' \geqslant 0$, $P_2'' \geqslant 0$, and $\rho_{21}^2+\rho_{2s}^2 \leqslant 1$.

We explain the achievable scheme used for obtaining the above capacity region as follows. Here, the cognitive transmitter's power P_2 is split into three parts (as in (17)): $\rho_{21}^2P_2$ is for cooperatively transmitting W_1 via beamforming (as reflected in the first term in (15a)); $P_2'+\rho_{2s}^2P_2$ is for transmitting additional W_1 via an auxiliary random variable U to deal with the state at receiver 1 using dirty paper coding (as reflected in the second term in (15a)); and P_2'' is for transmitting W_2 (as reflected in (15b)). Here, rate splitting is not used, i.e., $W_{21}=\phi$, because for the case $|a| \leqslant 1$, forcing receiver 1 to decode certain W_{21} may reduce the achievable region.

Proof. Consider the following rate region, which consists of rate pairs (R_1, R_2) satisfying

$$R_1 \leqslant I(X_1U;Y_1)-I(U;S|X_1) \tag{16a}$$

$$R_2 \leqslant I(X_2;Y_2|UX_1S) \tag{16b}$$

$$R_1+R_2 \leqslant I(X_1X_2;Y_2|S) \tag{16c}$$

for some distribution $P_{SX_1UX_2Y_1Y_2}=P_{X_1}P_SP_{UX_2|X_1S}P_{Y_2|X_1X_2S}P_{Y_1|Y_2X_1S}$. This region is contained in (10a)-(10d), and is hence achievable. This can be seen by observing that $I(X_2;Y_2|UX_1S) \leqslant I(X_2U;Y_2|X_1S)$ and the sum rate bound (10d) is equal to the sum of the two bounds on the individual rates in (16a) and (16b).

The achievability of (15a)-(15c) is then obtained by choosing the following jointly Gaussian distribution for the random variables:

$$
\begin{aligned}
&X_1 \sim \mathcal{N}(0, P_1), \quad X_2' \sim \mathcal{N}(0, P_2'), \quad X_2'' \sim \mathcal{N}(0, P_2''), \quad P_2' + P_2'' = (1-\rho_{21}^2-\rho_{2s}^2)P_2 \\
&X_2 = \rho_{21}\sqrt{\frac{P_2}{P_1}}X_1 + X_2' + X_2'' + \rho_{2s}\sqrt{\frac{P_2}{Q}}S \\
&U = X_2' + \alpha\left(1 + a\rho_{2s}\sqrt{\frac{P_2}{Q}}\right)S
\end{aligned}
\tag{17}
$$

where X_1, X_2' , X_2'' and S are independent, and $\alpha = \frac{a^2P_2'}{a^2P_2'+a^2P_2''+1}$.

The converse proof is detailed in Appendix B. □

We next characterize the capacity region for the Gaussian channel with $|a| > 1$.

Theorem 3. *(Capacity) For the Gaussian CIC-DM-STR, if $|a| > 1$, the capacity region consists of rate pairs (R_1, R_2) satisfying:*

$$R_2 \leqslant \frac{1}{2}\log(1 + (1-\rho_{21}^2-\rho_{2s}^2)P_2) \tag{18a}$$

$$R_1 + R_2 \leqslant \frac{1}{2}\log(1 + b^2P_1 + 2b\rho_{21}\sqrt{P_1P_2} + (1-\rho_{2s}^2)P_2) \tag{18b}$$

$$R_1 + R_2 \leqslant \frac{1}{2}\log\left(1 + \frac{P_1 + 2a\rho_{21}\sqrt{P_1P_2} + a^2\rho_{21}^2P_2}{a^2(1-\rho_{21}^2)P_2 + 2a\rho_{2s}\sqrt{P_2Q} + Q + 1}\right) + \frac{1}{2}\log(1 + a^2(1-\rho_{2s}^2-\rho_{21}^2)P_2) \tag{18c}$$

where $\rho_{21}^2 + \rho_{2s}^2 \leqslant 1$.

Differently from Theorem 2, due to the fact that $|\alpha| > 1$, receiver 1 is stronger in decoding W_2. Hence, the achievable scheme sets $W_{21} = W_2$, i.e., requires receiver 1 to decode the full message W_2. The cognitive transmitter's power P_2 is split into two parts (as in (19)): $\rho_{21}^2P_2$ is for cooperatively transmitting W_1 via beamforming (as reflected in the first term in (18c)); and $(1-\rho_{21}^2)P_2$ is for transmitting additional W_1 and $W_{21} = W_2$ via an auxiliary random variable U to deal with the state at receiver 1 using dirty paper coding (as reflected in the second term in (18c)).

Proof. The achievability follows from (10a)-(10d) by choosing jointly Gaussian distribution for random variables as follows:

$$
\begin{aligned}
&X_1 \sim \mathcal{N}(0, P_1), \quad X_2' \sim \mathcal{N}(0, (1-\rho_{21}^2-\rho_{2s}^2)P_2) \\
&X_2 = \rho_{21}\sqrt{\frac{P_2}{P_1}}X_1 + X_2' + \rho_{2s}\sqrt{\frac{P_2}{Q}}S \\
&U = X_2' + \alpha\left(1 + a\rho_{2s}\sqrt{\frac{P_2}{Q}}\right)S
\end{aligned}
\tag{19}
$$

where X_1, X_2' and S are independent, and $\alpha = \frac{a^2(1-\rho_{21}^2-\rho_{2s}^2)P_2}{a^2(1-\rho_{21}^2-\rho_{2s}^2)P_2+1}$. We note that with this choice of the random variables, the first bound in (10a)-(10d) is redundant.

In order to prove the converse for Theorem 3, we first prove the following outer bound.

Lemma 1. *For the CIC-DM-STR, if it satisfies the condition* (5)*, an outer bound on the capacity region consists of rate pairs* (R_1, R_2) *satisfying*

$$R_2 \leqslant I(X_2; Y_2|SX_1) \tag{20a}$$
$$R_1 + R_2 \leqslant I(X_1X_2; Y_2|S) \tag{20b}$$
$$R_1 + R_2 \leqslant I(X_1; Y_1) + I(X_2; Y_1|SX_1) \tag{20c}$$

for some distribution $P_{SX_1UX_2Y_1Y_2} = P_{X_1}P_SP_{UX_2|X_1S}P_{Y_1|X_1X_2S}P_{Y_2|Y_1X_1S}$.

The proof for the above lemma is detailed in Appendix C. For the Gaussian channel with $|a| > 1$, it satisfies the condition (5). We then use the above lemma for developing the converse proof, which is detailed in Appendix D. □

4 The CIC-DM-ST Model: Discrete Memoryless Channels

In this section, we investigate the discrete memoryless CIC-DM-ST model. We first provide inner and outer bounds on the capacity region, and then identify a few special cases, for which we establish the capacity region.

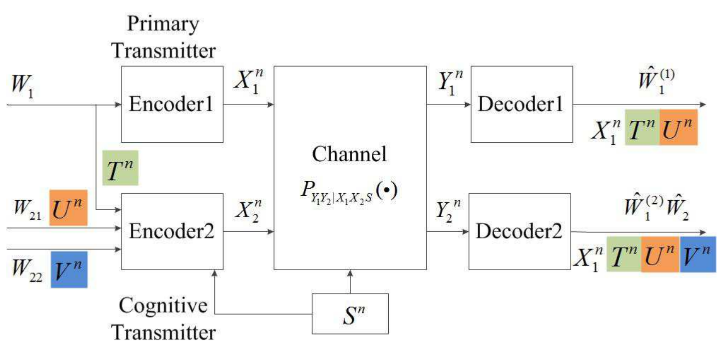

Figure 2: An illustration of an achievable scheme for the CIC-DM-ST

In order to derive an inner bound on the capacity region, we design an achievable scheme, the main idea of which is illustrated in Fig. 2. As for the CIC-DM-STR, this achievable scheme includes superposition coding, rate-splitting, and Gel'fand-Pinsker binning scheme. The primary transmitter and the cognitive transmitter cooperatively transmit W_1. The cognitive transmitter splits W_2 into two components W_{21} and W_{22} with W_{21} intended for both receivers and W_{22} intended only for receiver 2. Differently from the scheme for the CIC-DM-STR, here the cognitive transmitter employs Gel'fand-Pinsker scheme via three auxiliary random variables T, U and V (as in Lemma 2) to reduce state interference respectively for W_1, W_{21} and W_{22}. In particular, T deals with state interference for either receiver 1 or receiver 2 to decode W_1, U deals with state interference for either receiver 1 or receiver 2 to decode W_{21}, and V deals with state interference for receiver 2 to decode W_{22}. In particular, T and U cannot be combined because it is possible that U deals with the state at receiver 2 whereas T deals with the state at receiver 1. This also explains the reason that only one auxiliary random variable U is needed for obtaining the capacity region for the CIC-DM-STR model, in which only state interference at receiver 1 needs to be handled, and hence a single auxiliary random variable U (combining T and U) is sufficient for receiver 1 to decode both W_1 and W_{21}. At the receiver end, since receiver 1 can decode W_{21}, it can eliminate the interference caused by this message when it decodes W_1. In summary, we list the one-to-one correspondence of the codewords, the random variables that represent the codewords, and the messages carried by the codewords as follows:

1) A codeword from the primary transmitter to receivers 1 and 2: represented by X_1^n, associated with W_1;

2) A codeword from the cognitive transmitter to receivers 1 and 2: represented by T^n, associated with W_1 that the cognitive transmitter helps to send to receivers 1 and 2, coded using Gel'fand-Pinsker binning;

3) A codeword from the cognitive transmitter to receivers 1 and 2: represented by U^n, associated with W_{21} intended for both receivers 1 and 2 to decode, coded using Gel'fand-Pinsker binning;

4) A codeword from the cognitive transmitter to receiver 2: represented by V, associated with W_{22} intended for receiver 2 to decode, coded using Gel'fand-Pinsker binning.

We now provide an achievable region based on the above achievable scheme, which is useful in establishing our main inner bound.

Lemma 2. *An achievable region for the CIC-DM-ST consists of rate pairs (R_1, R_2) satisfying:*

$$R_2 = R_{21} + R_{22}, \quad R_{21} \geqslant 0, \quad R_{22} \geqslant 0 \tag{21a}$$
$$R_1 + R_{21} \leqslant I(TUX_1;Y_1) - I(TU;S|X_1) \tag{21b}$$
$$R_{22} \leqslant I(V;Y_2|UTX_1) - I(V;S|UTX_1) \tag{21c}$$
$$R_{21} + R_{22} \leqslant I(UV;Y_2|X_1T) - I(UV;S|X_1T) \tag{21d}$$
$$R_{21} + R_{22} \leqslant I(TUV;Y_2|X_1) - I(TUV;S|X_1) \tag{21e}$$
$$R_1 + R_{21} + R_{22} \leqslant I(TUVX_1;Y_2) - I(TUV;S|X_1) \tag{21f}$$

for some distribution $P_{X_1STUVX_2Y_1Y_2} = P_{X_1}P_SP_{TUVX_2|SX_1}P_{Y_1Y_2|SX_1X_2}$, where T, U and V are auxiliary random variables.

Proof. The detailed proof is relegated to Appendix E. □

Based on Lemma 2, our main inner bound on the capacity region is given in the following theorem.

Theorem 4. *(Inner Bound) For the CIC-DM-ST, an achievable region consists of rate pairs* (R_1, R_2) *satisfying:*

$$
\begin{aligned}
R_1 &\leqslant I(X_1TU;Y_1) - I(TU;S|X_1) && \text{(22a)}\\
R_2 &\leqslant I(UV;Y_2|X_1T) - I(UV;S|X_1T) && \text{(22b)}\\
R_2 &\leqslant I(TUV;Y_2|X_1) - I(TUV;S|X_1) && \text{(22c)}\\
R_1 + R_2 &\leqslant I(X_1TUV;Y_2) - I(TUV;S|X_1) && \text{(22d)}\\
R_1 + R_2 &\leqslant I(X_1TU;Y_1) + I(V;Y_2|X_1TU) - I(TUV;S|X_1) && \text{(22e)}
\end{aligned}
$$

for some distribution $P_{X_1STUVX_2Y_1Y_2} = P_{X_1}P_SP_{TUVX_2|SX_1}P_{Y_1Y_2|SX_1X_2}$ *that satisfies*

$$I(V;Y_2|UTX_1) - I(V;S|UTX_1) \geqslant 0. \tag{23}$$

Proof. By applying Fourier-Motzkin elimination [20], we eliminate R_{21} and R_{22} from the bounds in Lemma 2 and obtain the bounds in Theorem 4. □

We note that the condition (23) follows from Fourier-Motzkin elimination to guarantee validness of the region in Lemma 2.

Remark 2. *The achievable region in Theorem 4 reduces to the capacity region of the multiple-access channel with state known noncausally at one transmitter in [21] by setting* $Y_1 = Y_2$, $T = \phi$ *and* $V = U$.

Remark 3. *The achievable region in Theorem 4 reduces to the capacity region of the same cognitive interference channel without state in [14, Theorem 4] by setting* $S = \phi$, $T = \phi$ *and* $V = X_2$.

We next derive the following inner bound, which is achieved by a simpler scheme that combines T and U together as one auxiliary random variable. This inner bound is useful for studying Gaussian channels in Section 5.2.

Corollary 1. *(Inner Bound) For the CIC-DM-ST, an achievable region consists of rate pairs* (R_1, R_2) *satisfying:*

$$
\begin{aligned}
R_1 &\leqslant I(X_1T;Y_1) - I(T;S|X_1) && \text{(24a)}\\
R_2 &\leqslant I(V;Y_2|X_1T) - I(V;S|X_1T) && \text{(24b)}\\
R_2 &\leqslant I(TV;Y_2|X_1) - I(TV;S|X_1) && \text{(24c)}\\
R_1 + R_2 &\leqslant I(X_1TV;Y_2) - I(TV;S|X_1) && \text{(24d)}
\end{aligned}
$$

for some distribution $P_{X_1STVX_2Y_1Y_2} = P_{X_1}P_SP_{TVX_2|X_1S}P_{Y_1Y_2|SX_1X_2}$ *that satisfies*

$$I(V;Y_2|TX_1) - I(V;S|TX_1) \geq 0. \tag{25}$$

Proof. The achievable region in Corollary 1 follows directly from Theorem 4 by setting $U = T$. □

We next provide an outer bound on the capacity region for the CIC-DM-ST.

Theorem 5. *(Outer Bound) An outer bound for the the CIC-DM-ST consists of the rate pairs* (R_1, R_2) *satisfying:*

$$
\begin{aligned}
R_1 &\leqslant I(X_1TU;Y_1) - I(TU;S|X_1) && (26a)\\
R_2 &\leqslant I(TV;Y_2|X_1) - I(TV;S|X_1) && (26b)\\
R_1 + R_2 &\leqslant I(X_1TV;Y_2) - I(TV;S|X_1) && (26c)
\end{aligned}
$$

for some distribution $P_{X_1STUVX_2Y_1Y_2} = P_{X_1}P_S P_{TUVX_2|X_1S} P_{Y_1Y_2|SX_1X_2}$*, which satisfies the Markov chain conditions* $T \leftrightarrow UV \leftrightarrow X_1X_2S \leftrightarrow Y_1Y_2$.

Proof. The proof employs the techniques in [1] for the Gel'fand-Pinsker model, and exploits independence properties among variables in our model. In particular, the auxiliary random variables are carefully constructed. The detailed proof is relegated to Appendix F. □

We note that although we use the same set of auxiliary random variables in inner and outer bounds for the CIC-DM-ST (as in Theorem 4 and Theorem 5) to make it easy for comparing the two bounds, in general, these random variables may have different joint distributions in different bounds as we specify for each bound.

We now provide inner and outer bounds for the degraded channel, which are useful for further identifying the cases for which we obtain the capacity region.

Theorem 6. *(Inner and Outer Bounds) If the CIC-DM-ST satisfies the degradedness condition* (3) *(i.e., receiver 1 is degraded with regard to receiver 2), then an achievable region consists of the rate pairs* (R_1, R_2) *satisfying:*

$$
\begin{aligned}
R_1 &\leqslant I(X_1T;Y_1) - I(T;S|X_1) && (27a)\\
R_2 &\leqslant I(V;Y_2|X_1T) - I(V;S|X_1T) && (27b)\\
R_2 &\leqslant I(TV;Y_2|X_1) - I(TV;S|X_1) && (27c)
\end{aligned}
$$

for some distribution $P_{X_1STVX_2Y_1Y_2} = P_{X_1}P_S P_{TVX_2|X_1S} P_{Y_1Y_2|SX_1X_2}$ *that satisfies*

$$
I(V;Y_2|TX_1) - I(V;S|TX_1) \geq 0. \quad (28)
$$

An outer bound on the capacity region for such a channel consists of the rate pairs (R_1, R_2) *satisfying:*

$$
\begin{aligned}
R_1 &\leqslant I(X_1T;Y_1) - I(T;S|X_1) && (29a)\\
R_2 &\leqslant I(TV;Y_2|X_1) - I(TV;S|X_1) && (29b)
\end{aligned}
$$

for some distribution $P_{X_1STVX_2Y_1Y_2} = P_{X_1}P_S P_{TVX_2|X_1S} P_{Y_1Y_2|SX_1X_2}$*, which satisfies the Markov chain conditions* $T \leftrightarrow V \leftrightarrow X_1X_2S \leftrightarrow Y_1Y_2$.

Proof. The achievability follows from the achievable region given in Corollary 1 by removing the bound (24d) due to the degradedness condition. The proof of the outer bound is detailed in Appendix G. □

Remark 4. *By setting $X_1 = \phi$,* (27c) *is redundant, and the achievable region in Theorem 6 coincides with the achievable region for the degraded broadcast channel with state noncausally known at the transmitter in [17]. This is reasonable because although the model in [17] does not require receiver 2 to decode W_1 as in our model, receiver 2 is able to do so due to the degradedness condition.*

The inner and outer bounds given in Theorems 4 and 5 do not match in general. We next identify two classes of channels, for which we obtain the capacity region. We first provide the capacity region for the degraded semideterministic channel in the following theorem.

Theorem 7. *(Capacity) If the CIC-DM-ST model satisfies the degradedness condition* (3) *and the semideterministic condition such that Y_2 is a deterministic function of X_1, X_2 and S, then the capacity region of the channel consists of rate pairs (R_1, R_2) satisfying:*

$$R_1 \leqslant I(X_1T;Y_1) - I(T;S|X_1) \tag{30a}$$
$$R_2 \leqslant H(Y_2|X_1TS) \tag{30b}$$
$$R_2 \leqslant H(Y_2|X_1) - I(TY_2;S|X_1) \tag{30c}$$

for some distribution $P_{X_1STX_2Y_1Y_2} = P_{X_1}P_SP_{TX_2|SX_1}P_{Y_2|X_1X_2S}P_{Y_1|Y_2}$, where T is an auxiliary random variable and its cardinality is bounded by $|\mathcal{T}| \leqslant |\mathcal{X}_1||\mathcal{X}_2||\mathcal{S}| + 1$.

Proof. The achievability follows from (27a)-(27c) by setting $V = Y_2$. The proof of the converse is detailed in Appendix H. ☐

We note that due to the fact that receiver 2's output Y_2 is a deterministic function of the inputs, no auxiliary random variable is needed to deal with the state at receiver 2 any more. Hence, similarly to the CIC-DM-STR, only one auxiliary random variable T is needed to deal with the state at receiver 1 via the Gelfand-Pinsker binning scheme.

Following Theorem 7, we also obtain the capacity region for the semideterministic degraded broadcast channel with the noncausal state information known at the transmitter by setting $X_1 = \phi$ in Theorem 7, which consists of rate pairs (R_1, R_2) satisfying:

$$R_1 \leqslant I(T;Y_1) - I(T;S) \tag{31a}$$
$$R_2 \leqslant H(Y_2|TS) \tag{31b}$$

for some distribution that $P_{STXY_1Y_2} = P_SP_{TX|S}P_{Y_2|XS}P_{Y_1|Y_2}$, where X is the channel input, and Y_1 and Y_2 are the channel outputs. We note that the bound (30c) becomes redundant when setting $X_1 = \phi$, because

$$H(Y_2) - I(TY_2;S) = H(Y_2|TS) + (I(T;Y_2) - I(T;S)) \tag{32a}$$
$$\geq H(Y_2|TS) + (I(T;Y_1) - I(T;S)) \tag{32b}$$
$$\geq H(Y_2|TS) \tag{32c}$$

where $I(T;Y_1) - I(T;S) \geq 0$ is necessary to guarantee $R_1 \geq 0$ in (31a).

We next obtain the following capacity region when receiver 1 is less noisy than receiver 2, i.e, the channel satisfies the condition (7).

Theorem 8. *(Capacity) For the CIC-DM-ST, if it satisfies the condition (7), the capacity region consists of rate pairs (R_1, R_2) satisfying:*

$$R_2 \leqslant I(U;Y_2|X_1) - I(U;S|X_1) \tag{33a}$$

$$R_1 + R_2 \leqslant I(X_1U;Y_2) - I(U;S|X_1) \tag{33b}$$

for some distribution $P_{X_1SUX_2Y_1Y_2} = P_{X_1}P_SP_{UX_2|X_1S}P_{Y_1Y_2|SX_1X_2}$, where U is an auxiliary random variable and its cardinality is bounded by $|\mathcal{U}| \leqslant |\mathcal{X}_1||\mathcal{X}_2||\mathcal{S}|$.

We note that if condition (7) is satisfied, receiver 1 is less noisy than receiver 2. Thus, bounds on achievable rates are dominated by receiver 2, and only one auxiliary random variable U is needed for dealing with state interference for receiver 2 to decode all messages.

Proof. The achievability follows from Theorem 1 by setting $T = \phi$, $V = U$ and using (7) to remove the redundant bounds. The converse follows from the capacity region of the multiple access channel (with its receiver being receiver 2 in our model) with state available at one transmitter given in [21], which clearly is an outer bound for our model. □

5 The CIC-DM-ST Model: Gaussian Channels

In this section, we consider the Gaussian CIC-DM-ST model. Similarly to Section 3, we partition the Gaussian CIC-DM-ST into two classes corresponding to $|a| \leqslant 1$ and $|a| > 1$, and study these two classes separately in this and next subsections. In each subsection, we first provide inner and outer bounds on the capacity region, and then characterize partial boundaries of the capacity region based on these bounds. We also obtain the full capacity region for channels that satisfy certain conditions.

5.1 Gaussian Channel: $|a| > 1$

If $|a| > 1$, the Gaussian channel satisfies the condition (5). We first provide an inner bound for this class of channels.

Proposition 1. *(Inner Bound) For the Gaussian CIC-DM-ST, if $|a| > 1$, an inner bound consists of rate pairs (R_1, R_2) satisfying:*

$$R_2 \leqslant \frac{1}{2}\log(1 + P_2') \tag{34a}$$

$$R_1 + R_2 \leqslant \frac{1}{2}\log\left(1 + \frac{b^2P_1 + 2b\rho_{21}\sqrt{P_1P_2} + \rho_{21}^2P_2}{(1-\rho_{21}^2)P_2 + 2c\rho_{2s}\sqrt{P_2Q} + c^2Q + 1}\right) + \frac{1}{2}\log(1 + P_2') \tag{34b}$$

$$\begin{aligned} R_1 + R_2 \leqslant &\frac{1}{2}\log\left(1 + \frac{P_1 + 2a\rho_{21}\sqrt{P_1P_2} + a^2\rho_{21}^2P_2}{a^2(1-\rho_{21}^2)P_2 + 2a\rho_{2s}\sqrt{P_2Q} + Q + 1}\right) \\ &+ \frac{1}{2}\log\left(1 + \frac{a^2P_2'^2 + 2a\rho_{2s1}\rho_{2s2}P_2' - a^2\rho_{2s1}^2P_2' - \rho_{2s1}^2}{a^2\rho_{2s1}^2P_2' + \rho_{2s2}^2P_2' + P_2' + \rho_{2s1}^2 - 2a\rho_{2s1}\rho_{2s2}P_2'}\right) \end{aligned} \tag{34c}$$

where $P_2' = (1-\rho_{21}^2 - \rho_{2s}^2)P_2$ *and* $\rho_{21}^2 + \rho_{2s}^2 \leqslant 1$, $\rho_{2s1} = \alpha(c\sqrt{Q} + \rho_{2s}\sqrt{P_2})$, $\rho_{2s2} = (\sqrt{Q} + a\rho_{2s}\sqrt{P_2})$, $\alpha = \frac{P_2'}{P_2'+1}$.

Similarly to the Gaussian CIC-DM-STR, due to the fact that $|a| > 1$, i.e., receiver 1 is stronger in decoding W_2, the achievable scheme sets $W_{21} = W_2$, i.e., requires receiver 1 to decode full message W_2. The cognitive transmitter's power P_2 is split into two parts (as in (36)): $\rho_{21}^2 P_2$ is for cooperatively transmitting W_1 via beamforming (as reflected in the first term in (34b) and (34c)); and $(1-\rho_{21}^2)P_2$ is for transmitting additional W_1 and $W_{21} = W_2$ via dirty paper coding (as reflected in the second term in (34b)). Differently from the CIC-DM-STR, the auxiliary random variable U is used here to deal with the state interference at receiver 2 (instead of receiver 1 for the CIC-DM-STR). This is also due to the fact that $|a| > 1$ so that receiver 2 is weaker in decoding information from the cognitive transmitter and hence needs additional help in state cancellation via dirty paper coding than receiver 1. Therefore, in the above achievable region, (34a) reflects the fact that receiver 2 decodes $W_{21} = W_2$, and (34b) and (34c) respectively reflect the facts that receiver 2 and receiver 1 decode both W_1 and $W_{21} = W_2$.

Proof. By setting $T = X_1$ and $U = V$ in the inner bound given in Theorem 4, we obtain an inner bound that includes the following bounds:

$$R_2 \leqslant I(U;Y_2|X_1) - I(U;S|X_1) \tag{35a}$$
$$R_1 + R_2 \leqslant I(X_1U;Y_2) - I(U;S|X_1) \tag{35b}$$
$$R_1 + R_2 \leqslant I(X_1U;Y_1) - I(U;S|X_1)\ . \tag{35c}$$

Based on the above bounds, we choose the jointly Gaussian input distribution and employ dirty paper coding for U to deal with the state in Y_2. More specifically, we set the random variables as follows and obtain the desired inner bound:

$$\begin{aligned} & X_1 \sim \mathcal{N}(0, P_1), \quad X_2' \sim \mathcal{N}(0, P_2') \\ & X_2 = \rho_{21}\sqrt{\frac{P_2}{P_1}}X_1 + X_2' + \rho_{2s}\sqrt{\frac{P_2}{Q}}S \\ & U = X_2' + \alpha\left(c + \rho_{2s}\sqrt{\frac{P_2}{Q}}\right)S \end{aligned} \tag{36}$$

where X_1, X_2' and S are independent random variables, and $\alpha = \frac{P_2'}{P_2'+1}$. □

We next provide an outer bound on the capacity region based on the following idea.

Since both W_1 and W_2 must be decoded at receiver 2, the two transmitters and receiver 2 form a cognitive MAC with state known at the cognitive transmitter. Hence, the capacity region for such a MAC serves as an outer bound for the Gaussian CIC-DM-ST.

Proposition 2. *(Outer Bound) For the Gaussian CIC-DM-ST, if* $|a| > 1$*, an outer bound consists of rate pairs* (R_1, R_2) *satisfying:*

$$R_2 \leqslant \frac{1}{2}\log(1 + P_2') \tag{37a}$$

$$R_1 + R_2 \leqslant \frac{1}{2}\log\left(1 + \frac{b^2P_1 + 2b\rho_{21}\sqrt{P_1P_2} + \rho_{21}^2P_2}{(1-\rho_{21}^2)P_2 + 2c\rho_{2s}\sqrt{P_2Q} + c^2Q + 1}\right) + \frac{1}{2}\log(1 + (1-\rho_{21}^2-\rho_{2s}^2)P_2) \tag{37b}$$

where $P_2' \leq (1-\rho_{21}^2-\rho_{2s}^2)P_2$ *and* $\rho_{21}^2 + \rho_{2s}^2 \leqslant 1$*.*

Proof. It is clear that the outer bound in Proposition 2 is equivalent to the region that consists of rate pairs (R_1, R_2) satisfying:

$$R_2 \leqslant \frac{1}{2}\log(1 + (1-\rho_{21}^2-\rho_{2s}^2)P_2) \tag{38a}$$

$$R_1 + R_2 \leqslant \frac{1}{2}\log\left(1 + \frac{b^2P_1 + 2b\rho_{21}\sqrt{P_1P_2} + \rho_{21}^2P_2}{(1-\rho_{21}^2)P_2 + 2c\rho_{2s}\sqrt{P_2Q} + c^2Q + 1}\right) + \frac{1}{2}\log(1 + (1-\rho_{21}^2-\rho_{2s}^2)P_2) \tag{38b}$$

where $\rho_{21}^2 + \rho_{2s}^2 \leqslant 1$. This region is the capacity region of the multiple access channel with state (with its receiver being receiver 2 in our model) given in [21], and hence serves as an outer bound for our model. □

We note that although the region (37a)-(37b) is equivalent to the region (38a)-(38b), the form of (37a)-(37b) is more convenient for characterizing the boundary points of the capacity region as we demonstrate below.

Although the inner bound (34a)-(34c) and the outer bound (37a)-(37b) do not match in general, we show that these bounds characterize some boundary points of the capacity region. Such characterization of individual points on the capacity boundary is meaningful and useful in practice, because each point on the boundary of the capacity region has an operational meaning, and represents one maximum weighted sum rate $R_1 + \lambda R_2$ for some $\lambda > 0$. In practice, depending on fairness requirements of the system and priority of sending either message, certain values of λ are of particular interest. Hence, the corresponding points on the boundary of the capacity region are regarded as the best weighted sum rates (with reasonable balance between the two rates). One special case of such boundary points is the sum capacity, which has been studied intensively in the literature.

In order to characterize boundary points of the capacity region, we first change the inner bound (34a)-(34c) into a more convenient form, which consists of rate pairs (R_1, R_2) satisfying:

$$R_2 \leqslant \frac{1}{2}\log(1 + P_2') \tag{39a}$$

$$R_1 + R_2 \leqslant \frac{1}{2}\log\left(1 + \frac{b^2P_1 + 2b\rho_{21}\sqrt{P_1P_2} + \rho_{21}^2P_2}{(1-\rho_{21}^2)P_2 + 2c\rho_{2s}\sqrt{P_2Q} + c^2Q + 1}\right) + \frac{1}{2}\log(1 + (1-\rho_{21}^2-\rho_{2s}^2)P_2) \tag{39b}$$

$$\begin{aligned} R_1 + R_2 \leqslant &\frac{1}{2}\log\left(1 + \frac{P_1 + 2a\rho_{21}\sqrt{P_1P_2} + a^2\rho_{21}^2P_2}{a^2(1-\rho_{21}^2)P_2 + 2a\rho_{2s}\sqrt{P_2Q} + Q + 1}\right) \\ &+ \frac{1}{2}\log\left(1 + \frac{\left(a^2(1-\rho_{21}^2-\rho_{2s}^2)P_2 + 2a\rho_{2s1}\rho_{2s2} - a^2\rho_{2s1}^2\right)(1-\rho_{21}^2-\rho_{2s}^2)P_2 - \rho_{2s1}^2}{\left(a^2\rho_{2s1}^2 + \rho_{2s2}^2 + 1 - 2a\rho_{2s1}\rho_{2s2}\right)(1-\rho_{21}^2-\rho_{2s}^2)P_2 + \rho_{2s1}^2}\right) \end{aligned} \tag{39c}$$

where $P_2' \leq (1-\rho_{21}^2-\rho_{2s}^2)P_2$ and $\rho_{21}^2+\rho_{2s}^2 \leqslant 1$, $\rho_{2s1} = \alpha(c\sqrt{Q}+\rho_{2s}\sqrt{P_2})$, $\rho_{2s2} = (\sqrt{Q}+a\rho_{2s}\sqrt{P_2})$, $\alpha = \frac{(1-\rho_{21}^2-\rho_{2s}^2)P_2}{(1-\rho_{21}^2-\rho_{2s}^2)P_2+1}$. The above region is equivalent to (34a)-(34c), because it is obtained by substituting the equality constraint $P_2' = (1-\rho_{21}^2-\rho_{2s}^2)P_2$ into (34a) and (34b) (which does not change the bounds), and relaxing the constraint on P_2' to be $P_2' \leq (1-\rho_{21}^2-\rho_{2s}^2)P_2$, that affects only (34a) and clearly does not enlarge the region. We now denote the bounds in (39a)-(39c) by $r_2(P_2')$, $r_{12}(\rho_{21},\rho_{2s})$, and $\tilde{r}_{12}(\rho_{21},\rho_{2s})$. For $0 \leq P_2' \leq P_2$, let

$$(\rho_{21}^*(P_2'), \rho_{2s}^*(P_2')) = \underset{(\rho_{21},\rho_{2s}):P_2' \leq (1-\rho_{21}^2-\rho_{2s}^2)P_2}{\operatorname{argmax}} r_{12}(\rho_{21},\rho_{2s}). \tag{40}$$

Based on these notations, we characterize partial boundary of the capacity region for the Gaussian channel as follows.

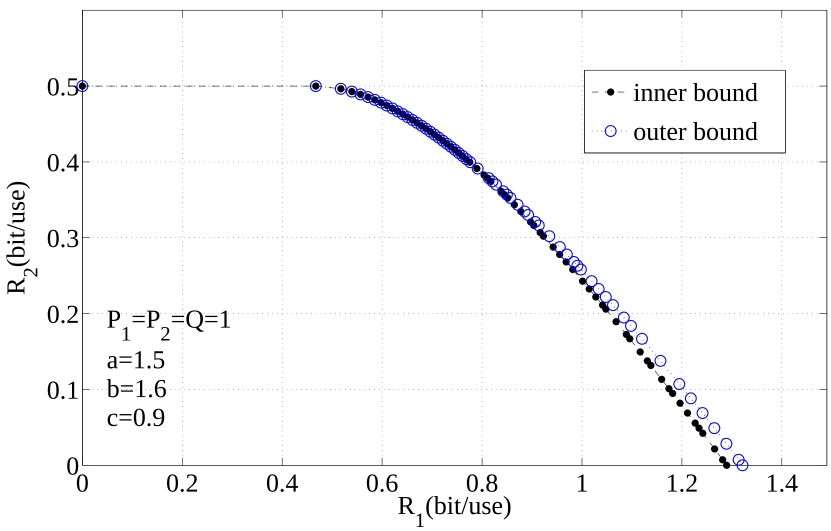

Figure 3: An illustration of the partial boundary of the capacity region for a Gaussian CIC-DM-ST with $|a| > 1$.

Theorem 9. *(Partial Boundary of Capacity Region) Consider the Gaussian CIC-DM-ST with* $|a| > 1$*. For* $0 \leq P_2' \leq P_2$*, the rate pairs* $\Big(r_{12}\big(\rho_{21}^*(P_2'), \rho_{2s}^*(P_2')\big) - r_2(P_2'),\ r_2(P_2')\Big)$ *are on the boundary of the capacity region if* $r_{12}(\rho_{21}^*(P_2'), \rho_{2s}^*(P_2')) \leq \tilde{r}_{12}(\rho_{21}^*(P_2'), \rho_{2s}^*(P_2'))$*. The rate pairs* $(R_1, r_2(P_2))$ *are also on the boundary of the capacity region if* $R_1 \leq \min\{r_{12}, \tilde{r}_{12}\}|_{\rho_{21}=0,\rho_{2s}=0} - r_2(P_2)$*.*

Proof. The rate pairs given in the theorem are achievable due to the condition given in the theorem. They are also on the boundary of the outer bound given in Proposition 2, because r_2 and r_{12} are the same as the bounds on R_1 and on R_1+R_2, respectively, and the chosen parameters $(\rho_{21}^*(P_2'), \rho_{2s}^*(P_2'))$ for each value of P_2' guarantees that the rate pairs are on the boundary. The second statement is clear because when $P_2' = P_2$, R_2 achieves the maximum value, and hence any such rate pair is on the boundary if it is achievable. □

In Fig. 3, we demonstrate the partial boundary of the capacity region characterized in Theorem 9. We consider the channel defined by the parameters $P_1 = P_2 = Q = 1$, $a = 1.5$, $b = 1.6$ and $c = 0.9$. We plot the boundaries of the inner bound given in Proposition 1 and the outer bound given in Proposition 2, respectively. It is clear that the two boundaries match when R_2 is above a certain threshold, and this matching part thus characterizes some boundary points of the capacity region as studied in Theorem 9.

We next show that under certain channel conditions, the outer bound given in Proposition 2 fully characterizes the capacity region.

Theorem 10. *(Capacity) For the Gaussian CIC-DM-ST, if $|a| > 1$ and the channel satisfies the condition (7), the capacity region consists of rate pairs (R_1, R_2) satisfying:*

$$R_2 \leqslant \frac{1}{2}\log(1 + P_2') \tag{41a}$$

$$R_1 + R_2 \leqslant \frac{1}{2}\log\left(1 + \frac{b^2P_1 + 2b\rho_{21}\sqrt{P_1P_2} + \rho_{21}^2P_2}{(1-\rho_{21}^2)P_2 + 2c\rho_{2s}\sqrt{P_2Q} + c^2Q + 1}\right) + \frac{1}{2}\log(1 + P_2')\ . \tag{41b}$$

where $P_2' = (1 - \rho_{21}^2 - \rho_{2s}^2)P_2$ and $\rho_{21}^2 + \rho_{2s}^2 \leqslant 1$.

As explained after Theorem 8, if the less noisy condition (7) is satisfied, receiver 2 dominates the performance of the channel. Thus, the achievable scheme that uses the auxiliary random variable for dealing with the state at receiver 2 via dirty paper coding turns out to be optimal.

Proof. Following from the region in Theorem 8, we set the random variables as in (36) and obtain an achievable region as given in (41a)-(41b). Such an achievable region is equivalent to the outer bound given in Proposition 2 as we comment in the proof of Proposition 2. □

5.2 Gaussian Channel: $|a| \leqslant 1$

We first note that the inner bound given in Proposition 1 for the case when $|a| > 1$ also serves as an inner bound for the case when $|a| \leqslant 1$. However, the choice of auxiliary random variables ($T = \phi$ and $U = V$) for obtaining this inner bound requires receiver 1 to decode all information for receiver 2. As such, this bound works well only when receiver 1 is stronger than receiver 2, and does not serve as a good inner bound for the case when $|a| \leqslant 1$. Thus, in this subsection, we develop two new inner bounds and one new outer bound on the capacity region for the case when $|a| \leqslant 1$. We also note that the outer bound in Proposition 2 is applicable and useful here as demonstrated in the sequel.

We derive the two inner bounds based on the same achievable region for the discrete memoryless channel with different choices of the distributions for the auxiliary random variables.

Proposition 3. *(Inner Bound 1) For the Gaussian CIC-DM-ST, if $|a| \leqslant 1$, then an inner bound on* *the capacity region consists of rate pairs* (R_1, R_2) *satisfying*

$$R_1 \leqslant \frac{1}{2}\log\left(1+\frac{P_1+2a\rho_{21}\sqrt{P_1P_2}+a^2\rho_{21}^2P_2}{a^2(1-\rho_{21}^2)P_2+2a\rho_{2s}\sqrt{P_2Q}+Q+1}\right)+\frac{1}{2}\log\left(1+\frac{a^2P_2'}{a^2P_2''+1}\right) \tag{42a}$$

$$R_2 \leqslant \frac{1}{2}\log(1+P_2'') \tag{42b}$$

$$\begin{aligned} R_2 \leqslant &\frac{1}{2}\log\left(1+\frac{a^2P_2'^2+2a\rho_{2s1}\rho_{2s2}P_2'-\rho_{2s1}^2(P_2'+P_2''+1)}{a^2P_2'P_2''+\rho_{2s1}^2(P_2'+P_2''+1)+a^2\rho_{2s2}P_2'+a^2P_2'-2a\rho_{2s1}\rho_{2s2}P_2'}\right)\\ &+\frac{1}{2}\log(1+P_2'') \end{aligned} \tag{42c}$$

$$\begin{aligned} R_1+R_2 \leqslant &\frac{1}{2}\log\left(1+\frac{b^2P_1+2b\rho_{21}\sqrt{P_1P_2}+\rho_{21}^2P_2}{(1-\rho_{21}^2)P_2+2c\rho_{2s}\sqrt{P_2Q}+c^2Q+1}\right)\\ &+\frac{1}{2}\log\left(1+\frac{a^2P_2'^2+2a\rho_{2s1}\rho_{2s2}P_2'-\rho_{2s1}^2(P_2'+P_2''+1)}{a^2P_2'P_2''+\rho_{2s1}^2(P_2'+P_2''+1)+a^2\rho_{2s2}P_2'+a^2P_2'-2a\rho_{2s1}\rho_{2s2}P_2'}\right)\\ &+\frac{1}{2}\log(1+P_2'') \end{aligned} \tag{42d}$$

where $\rho_{2s1}=\alpha\left(1+a\rho_{2s}\sqrt{\frac{P_2}{Q}}\right)\sqrt{Q}$, $\rho_{2s2}=\left(c+\rho_{2s}\sqrt{\frac{P_2}{Q}}\right)\sqrt{Q}$, $\alpha=\frac{a^2P_2'}{a^2P_2'+a^2P_2''+1}$, $|\rho_{21}|\leqslant 1$, $|\rho_{2s}|\leqslant 1$, $P_2'\geqslant 0$, $P_2''\geqslant 0$, *and* $P_2'+P_2''=(1-\rho_{21}^2-\rho_{2s}^2)P_2$.

Similarly to the CIC-DM-STR, if $|a|\leqslant 1$, the cognitive transmitter's power P_2 is split into three parts (as in (43)): $\rho_{21}^2P_2$ is for cooperatively transmitting W_1 via beamforming (as reflected in the first term in (42a)); $P_2'+\rho_{2s}^2P_2$ are for either transmitting additional W_1 or transmitting W_2 using dirty paper coding (via T) to deal with the state at receiver 1 (as reflected in the second term in (42a) or the first term in (42c)); and the remaining P_2 is for transmitting W_2 using dirty paper coding (via V) to deal with the state at receiver 2 (as reflected in (42b) and the second term in (42c)). Therefore, in the above achievable region, (42a) reflects the fact that receiver 1 decodes W_1 contained in both X_1 and T, (42b) reflects the fact that receiver 2 decodes W_2 contained in V, (42c) reflects the fact that receiver 2 decodes W_2 contained in both T and V, and (42d) reflects the fact that receiver 2 decodes W_1 contained in X_1, and W_2 contained in both T and V. We note that T plays two roles: either transmitting W_1 or transmitting W_2.

Proof. The above theorem is based on Corollary 1 by choosing (T, V, X_1, X_2) to be jointly Gaussian and employing dirty paper coding with T chosen for dealing with the state for Y_1 and V chosen for dealing with the state for Y_2. More specifically, We set the random variables as follows:

$$X_1 \sim \mathcal{N}(0, P_1), \quad X_2' \sim \mathcal{N}(0, P_2'), \quad X_2'' \sim \mathcal{N}(0, P_2''), \quad P_2' + P_2'' = (1-\rho_{21}^2 - \rho_{2s}^2)P_2$$
$$X_2 = \rho_{21}\sqrt{\frac{P_2}{P_1}}X_1 + X_2' + X_2'' + \rho_{2s}\sqrt{\frac{P_2}{Q}}S$$
$$T = X_2' + \alpha\left(1 + a\rho_{2s}\sqrt{\frac{P_2}{Q}}\right)S$$
$$V = X_2'' + \beta\left(c - \alpha + (1-a\alpha)\rho_{2s}\sqrt{\frac{P_2}{Q}}\right)S \tag{43}$$

where X_1, X_2', X_2'' and S are independent random variables, $\alpha = \frac{a^2P_2'}{a^2P_2' + a^2P_2'' + 1}$, and $\beta = \frac{P_2''}{P_2''+1}$. □

Proposition 4. *(Inner Bound 2) For the Gaussian CIC-DM-ST, if $|a| \leqslant 1$, then an inner bound on the capacity region consists of rate pairs (R_1, R_2) satisfying*

$$R_1 \leqslant \frac{1}{2}\log\left(1 + \frac{P_1 + 2a\rho_{21}\sqrt{P_1P_2} + a^2\rho_{21}^2P_2}{a^2(1-\rho_{21}^2)P_2 + 2a\rho_{2s}\sqrt{P_2Q} + Q + 1}\right)$$
$$+ \frac{1}{2}\log\left(1 + \frac{a^2P_2'^2 + 2a\rho_{2s1}\rho_{2s2}P_2' - a^2\rho_{2s1}^2(P_2' + P_2'') - \rho_{2s1}^2}{a^2\rho_{2s1}^2P_2' + \rho_{2s2}^2P_2' + a^2\rho_{2s1}^2P_2'' + a^2P_2'P_2'' + P_2' + \rho_{2s1}^2 - 2a\rho_{2s1}\rho_{2s2}P_2'}\right) \tag{44a}$$
$$R_2 \leqslant \frac{1}{2}\log(1 + P_2'') \tag{44b}$$
$$R_1 + R_2 \leqslant \frac{1}{2}\log\left(1 + \frac{b^2P_1 + 2b\rho_{21}\sqrt{P_1P_2} + \rho_{21}^2P_2}{(1-\rho_{21}^2)P_2 + 2c\rho_{2s}\sqrt{P_2Q} + c^2Q + 1}\right)$$
$$+ \frac{1}{2}\log(1 + (1-\rho_{21}^2 - \rho_{2s}^2)P_2) \tag{44c}$$

where $\rho_{2s1} = \alpha(c\sqrt{Q} + \rho_{2s}\sqrt{P_2})$, $\rho_{2s2} = (\sqrt{Q} + a\rho_{2s}\sqrt{P_2})$, $\alpha = \frac{P_2'}{P_2' + P_2'' + 1}$, $|\rho_{21}| \leqslant 1$, $|\rho_{2s}| \leqslant 1$, $P_2' \geqslant 0$, $P_2'' \geqslant 0$, and $P_2' + P_2'' = (1-\rho_{21}^2 - \rho_{2s}^2)P_2$.

We note that the inner bounds in Proposition 3 and 4 are based on the same achievable region for the discrete memoryless channel, i..e., Corollary 1, except that the auxiliary random variable T is designed to deal with the state at receiver 2 in Proposition 4 .

Proof. The above theorem is based on Corollary 1 by choosing (T, V, X_1, X_2) to be jointly Gaussian and employing dirty paper coding by choosing T and V as follows:

$$\begin{aligned}
& X_1 \sim \mathcal{N}(0, P_1), \quad X_2' \sim \mathcal{N}(0, P_2'), \quad X_2'' \sim \mathcal{N}(0, P_2''), \quad P_2' + P_2'' = (1 - \rho_{21}^2 - \rho_{2S}^2)P_2 \\
& X_2 = \rho_{21}\sqrt{\frac{P_2}{P_1}}X_1 + X_2' + X_2'' + \rho_{2s}\sqrt{\frac{P_2}{Q}}S \\
& T = X_2' + \alpha\left(c + \rho_{2s}\sqrt{\frac{P_2}{Q}}\right)S \\
& V = X_2'' + \beta(1-\alpha)\left(c + \rho_{2s}\sqrt{\frac{P_2}{Q}}\right)S
\end{aligned} \tag{45}$$

where X_1, X_2', X_2'' and S are independent random variables, $\alpha = \frac{P_2'}{P_2'+P_2''+1}$, and $\beta = \frac{P_2''}{P_2''+1}$. Here, T is chosen for dealing with the state for Y_2 (differently from the proof for Proposition 3) based on dirty paper coding where X_2'' is taken as noise. We then subtract T from Y_2 and design V for dealing with the state for $Y_2' = Y_2 - T$ based on dirty paper coding. For this choice of the auxiliary random variables, the second bound on R_2 in Corollary 1 is redundant because $I(T; Y_2|X_1) - I(T; S|X_1) > 0$. □

We next provide two outer bounds, both of which are useful for characterizing the capacity results. The first outer bound is given by the capacity region of the Gaussian CIC-DM-STR that we present in Theorem 2 in Section 3. For convenience, we rewrite this bound below.

Corollary 2. *(Outer Bound 1) For the Gaussian CIC-DM-ST, if $|a| \leqslant 1$, then the capacity region of CIC-DM-STR serves as an outer bound on the capacity region, which consists of rate pairs (R_1, R_2) satisfying*

$$\begin{aligned}
R_1 &\leqslant \frac{1}{2}\log\left(1 + \frac{P_1 + 2a\rho_{21}\sqrt{P_1P_2} + a^2\rho_{21}^2P_2}{a^2(1-\rho_{21}^2)P_2 + 2a\rho_{2s}\sqrt{P_2Q} + Q + 1}\right) + \frac{1}{2}\log\left(1 + \frac{a^2P_2'}{a^2P_2'' + 1}\right) \\
R_2 &\leqslant \frac{1}{2}\log(1 + P_2'') \\
R_1 + R_2 &\leqslant \frac{1}{2}\log(1 + b^2P_1 + 2b\rho_{21}\sqrt{P_1P_2} + (1-\rho_{2s}^2)P_2)
\end{aligned}$$

where $P_2' + P_2'' = (1 - \rho_{21}^2 - \rho_{2s}^2)P_2$, $P_2' \geqslant 0$, $P_2'' \geqslant 0$, and $\rho_{21}^2 + \rho_{2s}^2 \leqslant 1$.

As we comment at the beginning of this subsection, the outer bound in Proposition 2 is also applicable and useful for the case with $|a| \leqslant 1$. For convenience, we rewrite it below as a corollary.

Corollary 3. *(Outer Bound 2) For the Gaussian CIC-DM-ST, if $|a| \leqslant 1$, an outer bound on the capacity region consists of rate pairs (R_1, R_2) satisfying:*

$$R_2 \leqslant \frac{1}{2}\log(1 + P_2'') \tag{47a}$$

$$R_1 + R_2 \leqslant \frac{1}{2}\log\left(1 + \frac{b^2P_1 + 2b\rho_{21}\sqrt{P_1P_2} + \rho_{21}^2P_2}{(1-\rho_{21}^2)P_2 + 2c\rho_{2s}\sqrt{P_2Q} + c^2Q + 1}\right) + \frac{1}{2}\log(1 + (1 - \rho_{21}^2 - \rho_{2s}^2)P_2) \tag{47b}$$

where $P_2'' \leqslant (1 - \rho_{21}^2 - \rho_{2s}^2)P_2$, $P_2'' \geq 0$, and $\rho_{21}^2 + \rho_{2s}^2 \leqslant 1$.

For Gaussian channels with $|a| \leqslant 1$, we characterize partial boundaries of the capacity region based on the inner and outer bounds respectively given in Proposition 3 and 4, and Corollaries 2 and 3. Although the forms of inner bounds are complicated, we show that some boundary points on the capacity region are determined only by a subset of there bounds, and can hence be characterized via the given outer bounds.

We let $\Delta = (\rho_{21}, \rho_{2s}, P_2')$ and use $r_1'(\Delta, P_2'')$, $r_2'(P_2'')$, $\tilde{r}_2'(\Delta, P_2'')$, $r_{12}'(\Delta, P_2'')$ to denote the bounds (42a)-(42d) given in Proposition 3. For $0 \leq P_2'' \leq P_2$, let

$$\Delta^*(P_2'') = \underset{\Delta: P_2' + P_2'' = (1-\rho_{21}^2 - \rho_{2s}^2)P_2}{\operatorname{argmax}} \; r_1'(\Delta, P_2''). \tag{48}$$

Based on these notations, we characterize partial boundary of the capacity region for the Gaussian channel as follows.

Theorem 11. *(Partial Boundary of Capacity Region) Consider the Gaussian CIC-DM-ST with* $|a| \leqslant 1$*. For* $0 \leq P_2'' \leq P_2$*, the rate pairs* $(r_1'(\Delta^*(P_2''), P_2''), r_2'(P_2''))$ *are on the boundary of the capacity region if* $r_2'(P_2'') \leq \tilde{r}_2'(\Delta^*(P_2''), P_2'')$ *and* $r_1'(\Delta^*(P_2''), P_2'') + r_2'(P_2'') \leq r_{12}'(\Delta^*(P_2''), P_2'')$*.*

Proof. The rate pairs given in the theorem are contained in inner bound 1 given in Proposition 3 due to the conditions given in the theorem. We next show that these rate pairs are also on the boundary of an outer bound. Following from outer bound 1 in Corollary 2, $R_1 \leqslant r_1'(\Delta, P_2'')$ and $R_2 \leqslant r_2'(P_2'')$ also determine an outer bound with (Δ, P_2'') taking the same values as in inner bound 1 given in Proposition 3. Then the chosen parameters $\Delta^*(P_2'')$ for each value of P_2'' guarantee that the rate pairs are on the boundary of this outer bound. ☐

Remark 5. *The rate pairs characterized in Theorem 11 are on the boundary of the capacity region CIC-DM-STR, which is the outer bound 1 in Corollary 2.*

We next characterize additional boundary points of the capacity region based on inner bound 2 given in Proposition 4 and outer bound 2 given in Corollary 3. We use $r_1''(\rho_{21}, \rho_{2s}, P_2', P_2'')$, $r_2''(P_2'')$, and $r_{12}''(\rho_{21}, \rho_{2s})$ to denote the bounds (44a)-(44c) given in Proposition 4. For $0 \leq P_2'' \leq P_2$, let

$$(\rho_{21}^*(P_2''), \rho_{2s}^*(P_2'')) = \underset{(\rho_{21}, \rho_{2s}): P_2'' \leq (1-\rho_{21}^2 - \rho_{2s}^2)P_2}{\operatorname{argmax}} \; r_{12}''(\rho_{21}, \rho_{2s}), \tag{49}$$

and let $P_2'^*(P_2'') = (1 - \rho_{21}^*(P_2'')^2 - \rho_{2s}^*(P_2'')^2)P_2 - P_2''$. Based on these notations, we characterize partial boundary of the capacity region as follows.

Theorem 12. *(Partial Boundary of Capacity Region) Consider the Gaussian CIC-DM-ST with* $|a| \leqslant 1$*. For* $0 \leq P_2'' \leq P_2$*, the rate pairs* $(r_{12}''(\rho_{21}^*(P_2''), \rho_{2s}^*(P_2'')) - r_2''(P_2''),\ r_2''(P_2''))$ *are on the boundary of the capacity region if* $r_{12}''(\rho_{21}^*(P_2''), \rho_{2s}^*(P_2'')) - r_2''(P_2'') \leq r_1''(\rho_{21}^*(P_2''), \rho_{2s}^*(P_2''), P_2'^*(P_2''), P_2'')$*. The rate pairs* $(R_1, r_2''(P_2))$ *are also on the boundary of the capacity region if* $R_1 \leq \min\{r_1'', r_{12}'' - r_2''(P_2)\}|_{\rho_{21}=0, \rho_{2s}=0, P_2'=0}$*.*

Proof. The rate pairs given in the theorem are clearly contained in inner bound 2 given in Proposition 4. These rate pairs are also on the boundary of outer bound 2 given in Corollary 3, because r_2''

and r''_{12} are the same as the bounds on R_2 and on $R_1 + R_2$, respectively, and the chosen parameters $(\rho^*_{21}(P''_2), \rho^*_{2s}(P''_2))$ for each value of P''_2 guarantee that the rate pairs are on the boundary. The second statement is clear because when $P''_2 = P_2$, R_2 achieves the maximum value, and hence any rate pair with such R_2 is on the boundary if it is achievable. □

Theorems 11 and 12 collectively characterize partial boundary of the capacity region for the Gaussian channel with $|a| \leqslant 1$. In Fig. 4, we demonstrate these boundary points of the capacity region for an example channel with the parameters $P_1 = P_2 = Q = 1$, $b = 0.85$, $c = 0.9$ and $a = 0.8$. We plot the boundaries of the two inner bounds given in Proposition 3 and Proposition 4, and the boundaries of the two outer bounds given in Corollary 2 and Corollary 3, respectively. It can be seen that the two inner bounds are very close to each other. The boundary of inner bound 1 matches the boundary of outer bound 1 when R_1 is above a certain value, and this part is thus on the boundary of the capacity region. We also note that this part of the boundary achieves the capacity region of the CIC-DM-STR. It can further be seen that the boundary of inner bound 2 matches the boundary of outer bound 2 when R_2 is above a certain threshold, and this part is hence also on the boundary of the capacity region.

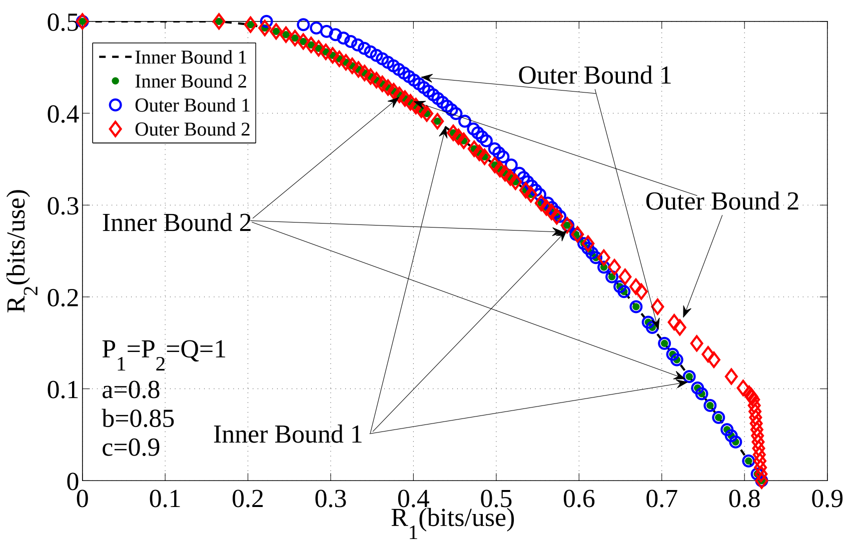

Figure 4: An illustration of inner and outer bounds and the partial boundary of the capacity region for a Gaussian CIC-DM-ST with $|a| \leqslant 1$

It can be seen that outer bounds 1 and 2 separately characterize certain parts of the boundary of the capacity region for Gaussian channels with $|a| \leqslant 1$. We further show that each of these two outer bounds can characterize the full capacity region for channels that satisfy certain conditions.

Theorem 13. *(Capacity) For the Gaussian CIC-DM-ST, if $|a| \leqslant 1$ and the channel satisfies the condition (6), the capacity region consists of rate pairs (R_1, R_2) satisfying:*

$$R_1 \leqslant \frac{1}{2}\log\left(1 + \frac{P_1 + 2a\rho_{21}\sqrt{P_1P_2} + a^2\rho_{21}^2P_2}{a^2(1-\rho_{21}^2)P_2 + 2a\rho_{2s}\sqrt{P_2Q} + Q + 1}\right) + \frac{1}{2}\log\left(1 + \frac{a^2P'_2}{a^2P''_2 + 1}\right) \tag{50a}$$

$$R_2 \leqslant \frac{1}{2}\log(P''_2 + 1) \tag{50b}$$

where $P_2' + P_2'' = (1 - \rho_{21}^2 - \rho_{2s}^2)P_2$, $P_2' \geqslant 0$, $P_2'' \geqslant 0$ *and* $\rho_{21}^2 + \rho_{2s}^2 \leqslant 1$, $|\rho_{21}| \leqslant 1$, $|\rho_{2s}| \leqslant 1$.

We note that the above capacity region matches the capacity region in [5] of a cognitive interference model with state, in which W_1 is intended only for receiver 1. This is reasonable because under the condition (6), receiver 1 is weaker in decoding W_1 than receiver 2, and receiver 2 can hence always decode W_1, which satisfies the additional requirement in the model of this paper. Consequently, in the designation of auxiliary random variables, more resources are used to help receiver 1 to cancel signal and state interference. This is why only part of P_2 is used to transmit W_2, and there is a tradeoff between the rates R_1 and R_2.

Proof. Under the condition (6), the bounds in the achievable region in Corollary 1 reduce to:

$$R_1 \leqslant I(X_1T;Y_1) - I(T;S|X_1) \tag{51a}$$

$$R_2 \leqslant I(V;Y_2|X_1T) - I(V;S|X_1T) \tag{51b}$$

$$R_2 \leqslant I(TV;Y_2|X_1) - I(TV;S|X_1) \tag{51c}$$

Based on the above bounds, we choose the same jointly Gaussian input distribution as in (43). In particular, since the auxiliary random variable T is chosen to employ dirty paper coding to deal with the state in Y_1, it guarantees that $I(T;Y_1|X_1) - I(T;S|X_1) \geqslant 0$, which implies that $I(T;Y_2|X_1) - I(T;S|X_1) \geqslant 0$ due to the condition (6). Hence, (51c) is redundant. Thus, we obtain an achievable region that matches the first two bounds of outer bound 1 in Corollary 2 and is hence tight. ☐

Remark 6. *The above theorem implies that the Gaussian CIC-DM-ST achieves the capacity of the CIC-DM-STR if the channel satisfies* $|a| \leqslant 1$ *and the condition (6). This is similar to the result that dirty paper coding achieves the capacity of the Gaussian channel with state also known at the receiver [18]. Here, the channel cannot achieve the capacity with both receivers knowing the channel state due to the fact that the primary transmitter does not know the channel state.*

The following theorem identifies the channels for which outer bound 2 given in Corollary 3 characterizes the full capacity region.

Theorem 14. *(Capacity) For the Gaussian CIC-DM-ST, if* $|a| \leqslant 1$ *and the channel satisfies the condition (7), the capacity region consists of rate pairs* (R_1, R_2) *satisfying:*

$$R_1 \leqslant \frac{1}{2}\log\left(1 + \frac{P_2'}{P_2''+1}\right) + \frac{1}{2}\log\left(1 + \frac{b^2P_1 + 2b\rho_{21}\sqrt{P_1P_2} + \rho_{21}^2P_2}{(1-\rho_{21}^2)P_2 + 2c\rho_{2s}\sqrt{P_2Q} + c^2Q + 1}\right) \tag{52a}$$

$$R_2 \leqslant \frac{1}{2}\log(1 + P_2'') \tag{52b}$$

where $P_2' + P_2'' = (1 - \rho_{21}^2 - \rho_{2s}^2)P_2$, $P_2' \geqslant 0$, $P_2'' \geq 0$ *and* $\rho_{21}^2 + \rho_{2s}^2 \leqslant 1$, $|\rho_{21}| \leqslant 1$, $|\rho_{2s}| \leqslant 1$.

We note that Theorems 10 and 14 imply that under the condition (7), the Gaussian CIC-DM-ST has the same capacity region as the multiple access channel with state given in [21]. This is reasonable because the condition (7) implies that receiver 2 is weaker than receiver 1 in decoding W_1, and hence dominates the rate region.

Proof. With the condition (7), it can be seen that an achievable region determined by the following bounds is contained in the inner bound given in Corollary 1, and is hence achievable.

$$R_1 \leqslant I(X_1T;Y_2) - I(T;S|X_1) \tag{53a}$$
$$R_2 \leqslant I(V;Y_2|X_1T) - I(V;S|X_1T) \tag{53b}$$
$$R_2 \leqslant I(TV;Y_2|X_1) - I(TV;S|X_1)\ . \tag{53c}$$

The achievability follows from the above region by choosing the jointly Gaussian distribution and employing dirty paper coding for T to deal with the state for Y_2 and for V to deal with the remaining state for Y_2 after subtracting $\frac{1}{a}T$. More specifically, we set the auxiliary random variable as follows:

$$\begin{aligned}
&X_1 \sim \mathcal{N}(0,P_1), \quad X_2' \sim \mathcal{N}(0,P_2'), \quad X_2'' \sim \mathcal{N}(0,P_2''), \quad P_2' + P_2'' = (1-\rho_{21}^2-\rho_{2S}^2)\\
&X_2 = \rho_{21}\sqrt{\frac{P_2}{P_1}}X_1 + X_2' + X_2'' + \rho_{2s}\sqrt{\frac{P_2}{Q}}S\\
&T = X_2' + \alpha\left(c + \rho_{2s}\sqrt{\frac{P_2}{Q}}\right)S\\
&V = X_2'' + \beta(1-\alpha)\left(c + \rho_{2s}\sqrt{\frac{P_2}{Q}}\right)S
\end{aligned} \tag{54}$$

where X_1, X_2', X_2'' and S are independent random variables, $\alpha = \frac{P_2'}{P_2'+P_2''+1}$, and $\beta = \frac{P_2''}{P_2''+1}$. Such a choice of the input distribution also implies that $I(T;Y_2|X_1) - I(T;S|X_1) \geqslant 0$, and the bound (53c) is hence redundant. The proof for the converse follows by observing that the region (52a)-(52b) has the same boundary points as outer bound 2 given in Corollary 3, and hence the two regions are equivalent. □

6 Conclusion

In this paper, we performed a comprehensive study of a class of cognitive interference channels with degraded message sets, which are corrupted by i.i.d. state sequences. For the CIC-DM-STR, in which the state sequence is noncausally known at both the cognitive transmitter and receiver 2, we obtained the capacity region for both the discrete memoryless and Gaussian channels. In order to characterize the capacity region for the Gaussian channel, we partitioned the channel parameters into two classes depending on the interference of X_2 to receiver 1, and the achievable schemes that achieve the capacity region for these two classes represent two extreme cases of rate splitting: either no W_2 is decoded at receiver 1 or entire W_2 is decoded at receiver 1. For both cases, we employed dirty paper coding to deal with state interference at receiver 1, and showed that such a scheme achieves the capacity region.

For the CIC-DM-ST, in which the state sequence is noncausally known only at the cognitive transmitter, we obtained inner and outer bounds on the capacity region for the discrete memoryless channel and its degraded version. We characterized the capacity region for the degraded semideterministic channel and channels that satisfy a less noisy condition. For both cases, the channel performance is dominated by one receiver, and hence it is optimal (i.e., capacity achievable) to deal with the state interference only at such a receiver. We further studied the Gaussian channels, and also partitioned these channels into two cases as for the CIC-DM-STR. For each case, we derived inner and outer bounds on the capacity region, and characterized partial boundary of the capacity region. We also characterized the full capacity region for channels that satisfy certain conditions.

We note that our characterization of the capacity region for the CIC-DM-STR applied the technique developed recently in [15] for proving equivalence of inner and outer bounds characterized by different sets of auxiliary random variables. Such a technique may also be useful for other network models. Furthermore, we anticipate that our technique of characterizing partial boundary of the capacity region for the Gaussian channel may be applicable to other Gaussian network models.

We further note that this paper focused on the cases, in which the state is noncausally known at the cognitive transmitter, or at both the cognitive transmitter and receiver 2. For the case in which the state is causally known to these transceivers, the analysis can be very different and is hence out of scope of this paper. However, it is very interesting to study such a topic in the future, and compare the rate regions with those provided in this paper. Such comparison will imply the impact of causality of the state information on rate regions.

Appendix

A Proof of the Outer Bound (11a)-(11d)

Consider a $(2^{nR_1}, 2^{nR_2}, n)$ code with an average error probability $P_e^{(n)}$. The probability distribution on $\mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{S}^n \times \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}_1^n \times \mathcal{Y}_2^n$ is given by

$$P_{W_1W_2S^nX_1^nX_2^nY_1^nY_2^n} = P_{W_1}P_{W_2}\left[\prod_{i=1}^{n} P_{S_i}\right] P_{X_1^n|W_1}P_{X_2^n|W_1W_2S^n}\prod_{i=1}^{n} P_{Y_{1i}Y_{2i}|X_{1i}X_{2i}S_i}. \tag{55}$$

By Fano's inequality, we have

$$\begin{aligned} H(W_1|Y_1^n) &\leqslant nR_1P_e^{(n)} + 1 = n\delta_{1n} \\ H(W_1W_2|S^nY_2^n) &\leqslant n(R_1+R_2)P_e^{(n)} + 1 = n\delta_{2n} \end{aligned} \tag{56}$$

where δ_{1n}, $\delta_{2n} \to 0$ as $n \to +\infty$. Let $\delta_n = \delta_{1n} + \delta_{2n}$, which also satisfies that $\delta_n \to 0$ as $n \to +\infty$.

We define the following auxiliary random variables:

$$\begin{aligned} K_i &= (W_1, S_{i+1}^n, X_1^n, Y_1^{i-1}) \\ T_i &= Y_{2(i+1)}^n \end{aligned} \tag{57}$$

which satisfies the Markov chain condition:

$$K_iT_i \leftrightarrow X_{1i}X_{2i}S_i \leftrightarrow Y_{1i}Y_{2i} \tag{58}$$

for $i = 1,\cdots, n$.

We first bound R_1 based on the Fano's inequality as follows:

$$\begin{aligned}
nR_1 &\leqslant I(W_1; Y_1^n) + n\delta_n \\
&\overset{(a)}{=} \sum_{i=1}^{n} [I(W_1S_{i+1}^n; Y_1^i) - I(W_1S_i^n; Y_1^{i-1})] + n\delta_n \\
&\overset{(b)}{=} \sum_{i=1}^{n} [I(W_1S_{i+1}^n; Y_1^{i-1}) + I(W_1S_{i+1}^n; Y_{1i}|Y_1^{i-1}) \\
&\qquad - I(W_1S_{i+1}^n; Y_1^{i-1}) - I(S_i; Y_1^{i-1}|W_1S_{i+1}^n)] + n\delta_n \\
&= \sum_{i=1}^{n} [I(W_1S_{i+1}^n; Y_{1i}|Y_1^{i-1}) - I(S_i; Y_1^{i-1}|W_1S_{i+1}^n)] + n\delta_n \\
&= \sum_{i=1}^{n} [H(Y_{1i}|Y_1^{i-1}) - H(Y_{1i}|W_1S_{i+1}^nY_1^{i-1}) - H(S_i|W_1S_{i+1}^n) + H(S_i|W_1S_{i+1}^nY_1^{i-1})] + n\delta_n \\
&\overset{(c)}{\leqslant} \sum_{i=1}^{n} [H(Y_{1i}) - H(Y_{1i}|W_1S_{i+1}^nY_1^{i-1}X_1^n) - (H(S_i|X_{1i}) + H(S_i|W_1S_{i+1}^nY_1^{i-1}X_1^n))] + n\delta_n \\
&\overset{(d)}{\leqslant} \sum_{i=1}^{n} [I(K_iX_{1i}; Y_{1i}) - I(K_i; S_i|X_{1i})] + n\delta_n
\end{aligned} \tag{59}$$

where (a) follows due to cancellation of the terms in the sum, and the fact that $Y_1^0 = \phi$, (b) follows from the chain rule of mutual information, (c) follows because X_1^n is a function of W_1, and (d) follows from the definition of K_i. The single letter characterization follows standard steps and is hence omitted.

We next bound R_2 as follows:

$$\begin{aligned}
nR_2 &\overset{(a)}{\leqslant} I(W_2; Y_2^nS^n) + n\delta_n \leqslant I(W_2; Y_2^nS^nW_1) + n\delta_n \\
&\overset{(b}{\leqslant} I(W_2; Y_2^n|W_1S^n) + n\delta_n \\
&= \sum_{i=1}^{n} I(W_2; Y_{2i}|Y_{2(i+1)}^nS^nW_1X_1^n) + n\delta_n \\
&\overset{(c)}{\leqslant} \sum_{i=1}^{n} [H(Y_{2i}|S_iX_{1i}) - H(Y_{2i}|W_2Y_{2(i+1)}^nS^nW_1X_1^nX_{2i})] + n\delta_n \\
&\overset{(d)}{\leqslant} \sum_{i=1}^{n} [H(Y_{2i}|S_iX_{1i}) - H(Y_{2i}|S_iX_{1i}X_{2i})] + n\delta_n \\
&= \sum_{i=1}^{n} I(X_{2i}; Y_{2i}|S_iX_{1i}) + n\delta_n.
\end{aligned} \tag{60}$$

where (a) follows from Fano's inequality (56), (b) follows from chain rule and the fact that W_2 and (W_1, S^n) are independent, (c) follows because conditioning does not increase entropy, and (d) follows from the Markov chain $K_iT_i \leftrightarrow X_{1i}X_{2i}S_i \leftrightarrow Y_{1i}Y_{2i}$.

We further bound $R_1 + R_2$ based on Fano's inequality as follows:

$$
\begin{aligned}
n(R_1+R_2) &\leqslant I(W_1W_2;Y_2^nS^n)+n\delta_n \\
&= \sum_{i=1}^{n} I(W_2W_1;Y_{2i}|Y_{2(i+1)}^nS^n)+n\delta_n \\
&\overset{(a)}{\leqslant} \sum_{i=1}^{n}[H(Y_{2i}|S_i)-H(Y_{2i}|W_2Y_{2(i+1)}^nS^nW_1X_{1i}X_{2i})]+n\delta_n \\
&\overset{(b)}{\leqslant} \sum_{i=1}^{n}[H(Y_{2i}|S_i)-H(Y_{2i}|S_iX_{1i}X_{2i})]+n\delta_n \\
&= \sum_{i=1}^{n} I(X_{1i}X_{2i};Y_{2i}|S_i)+n\delta_n.
\end{aligned} \tag{61}
$$

where (a) follows because conditioning does not increase entropy, and (b) follows because Y_{2i} is independent of other variables given X_{1i}, X_{2i} and S_i.

We introduce a lemma which is useful in the proof.

Lemma 3. *: [22, Lemma 7] For a set of random variables* $(T, Y_1 \dots, Y_n, Z_1, \dots, Z_n)$,

$$
\sum_{i=1}^{n} I(Y_i;Z^{i-1}|TY_{i+1}^n) = \sum_{i=1}^{n} I(Y_{i+1}^n;Z_i|TZ^{i-1}). \tag{62}
$$

We proceed to derive an alternative bound on $R_1 + R_2$ as follows:

$$
\begin{aligned}
n(R_1+R_2) &\leqslant I(W_1;Y_1^n)+I(W_2;Y_2^nS^n)+n\delta_n \\
&\overset{(a)}{\leqslant} I(W_1;Y_1^n)+I(W_2;Y_2^nS^n|W_1)+n\delta_n
\end{aligned} \tag{63}
$$

where (a) follows because W_1 and W_2 are independent.

The first term in (63) can be bounded as follows:

$$
\begin{aligned}
& I(W_1;Y_1^n) \\
=& \sum_{i=1}^{n} I(W_1;Y_{1i}|Y_1^{i-1}) \\
\overset{(a)}{\leqslant}& \sum_{i=1}^{n} I(W_1Y_1^{i-1};Y_{1i}) \\
\overset{(b)}{=}& \sum_{i=1}^{n} \left[I(W_1Y_1^{i-1}S_{i+1}^nY_{2(i+1)}^n;Y_{1i}) - I(S_{i+1}^nY_{2(i+1)}^n;Y_{1i}|W_1Y_1^{i-1})\right] \\
\overset{(c)}{=}& \sum_{i=1}^{n} \left[I(W_1Y_1^{i-1}S_{i+1}^nY_{2(i+1)}^n;Y_{1i}) - I(S_iY_{2i};Y_1^{i-1}|W_1S_{i+1}^nY_{2(i+1)}^n)\right] \\
\overset{(d)}{=}& \sum_{i=1}^{n} \left[I(W_1Y_1^{i-1}S_{i+1}^nY_{2(i+1)}^n;Y_{1i}) - I(S_iY_{2i};Y_1^{i-1}W_1S_{i+1}^nY_{2(i+1)}^n) + I(W_1S_{i+1}^nY_{2(i+1)}^n;S_iY_{2i})\right] \\
\overset{(e)}{=}& \sum_{i=1}^{n} [I(W_1Y_1^{i-1}S_{i+1}^nY_{2(i+1)}^n;Y_{1i}) - I(S_i;Y_1^{i-1}W_1S_{i+1}^nY_{2(i+1)}^n) \\
& \qquad + I(W_1S_{i+1}^nY_{2(i+1)}^n;S_iY_{2i}) - I(Y_{2i};Y_1^{i-1}W_1S_{i+1}^nY_{2(i+1)}^n|S_i)] \\
\overset{(f)}{=}& \sum_{i=1}^{n} [I(T_iK_iX_{1i};Y_{1i}) - I(T_iK_iX_{1i};S_i) \\
& \qquad + I(W_1S_{i+1}^nY_{2(i+1)}^n;S_iY_{2i}) - I(Y_{2i};Y_1^{i-1}W_1S_{i+1}^nY_{2(i+1)}^n|S_i)] \qquad (64)
\end{aligned}
$$

where (a) follows from chain rule and the fact that mutual information is nonnegative, (b) follows from chain rule, (c) follows from Lemma 3, (d) and (e) follows from chain rule, and (f) follows from the definition for T_i and K_i.

We next consider the last two terms in (64) together with the second term in (63) as follows:

$$
\begin{aligned}
& I(W_2;Y_2^nS^n|W_1)+\sum_{i=1}^n\left[I(W_1S_{i+1}^nY_{2(i+1)}^n;S_iY_{2i})-I(Y_{2i};Y_1^{i-1}W_1S_{i+1}^nY_{2(i+1)}^n|S_i)\right] \\
&\overset{(a)}{=}\sum_{i=1}^n\left[I(W_2;Y_{2i}S_i|W_1S_{i+1}^nY_{2(i+1)}^n)+I(W_1S_{i+1}^nY_{2(i+1)}^n;S_iY_{2i})-I(Y_{2i};Y_1^{i-1}W_1S_{i+1}^nY_{2(i+1)}^n|S_i)\right] \\
&\overset{(b)}{=}\sum_{i=1}^n[I(W_1W_2S_{i+1}^nY_{2(i+1)}^n;S_iY_{2i})+I(S^{i-1};S_iY_{2i}|W_1W_2S_{i+1}^nY_{2(i+1)}^n)-I(S_{i+1}^nY_{2(i+1)}^n;S_i|W_1W_2S^{i-1}) \\
&\qquad\qquad -I(Y_{2i};Y_1^{i-1}W_1S_{i+1}^nY_{2(i+1)}^n|S_i)] \\
&\overset{(c)}{=}\sum_{i=1}^n[I(W_1W_2S_{i+1}^nS^{i-1}Y_{2(i+1)}^n;S_iY_{2i})-I(S_{i+1}^nY_{2(i+1)}^n;S_i|W_1W_2S^{i-1}) \\
&\qquad\qquad -I(Y_{2i};Y_1^{i-1}W_1S_{i+1}^nY_{2(i+1)}^n|S_i)] \\
&\overset{(d)}{=}\sum_{i=1}^n[I(W_1W_2S_{i+1}^nS^{i-1}Y_{2(i+1)}^n;S_iY_{2i})-I(S_{i+1}^nY_{2(i+1)}^nW_1W_2S^{i-1};S_i)-I(Y_{2i};Y_1^{i-1}W_1S_{i+1}^nY_{2(i+1)}^n|S_i)] \\
&\overset{(e)}{=}\sum_{i=1}^n\left[I(W_1W_2S_{i+1}^nS^{i-1}Y_{2(i+1)}^n;Y_{2i}|S_i)-I(Y_{2i};Y_1^{i-1}W_1S_{i+1}^nY_{2(i+1)}^n|S_i)\right] \\
&\overset{(f)}{\leqslant}\sum_{i=1}^n\left[H(Y_{2i}|S_iY_1^{i-1}W_1X_1^nS_{i+1}^nY_{2(i+1)}^n)-H(Y_{2i}|S_iY_1^{i-1}W_1X_1^nW_2S_{i+1}^nS^{i-1}Y_{2(i+1)}^nX_{2i})\right] \\
&\overset{(g)}{\leqslant}\sum_{i=1}^n\left[H(Y_{2i}|S_iY_1^{i-1}W_1X_1^nS_{i+1}^nY_{2(i+1)}^n)-H(Y_{2i}|S_iY_1^{i-1}W_1X_1^nS_{i+1}^nY_{2(i+1)}^nX_{2i})\right] \\
&\overset{(h)}{=}\sum_{i=1}^n I(X_{2i};Y_{2i}|X_{1i}T_iK_iS_i)
\end{aligned}
\tag{65}
$$

where (a) follows from chain rule, (b) follows from chain rule to combine the first two terms in the previous step and Lemma 3, (c) follows from chain rule, (d) follows from chain rule and because W_1, W_2 and S^{i-1} are independent from S_i, (e) follows from chain rule, (f) follows because X_1^n is a function of W_1 and conditioning does not increase entropy, (g) follows because Y_{2i} is independent of other variables given X_{1i}, X_{2i} and S_i, and (h) follows from the definition of T_i and K_i.

Therefore, substituting (64) and (65) into (63), we obtain

$$
n(R_1+R_2)\leqslant\sum_{i=1}^n[I(T_iK_iX_{1i};Y_{1i})-I(T_iK_i;S_i|X_{1i})+I(X_{2i};Y_{2i}|X_{1i}T_iK_iS_i)]+n\delta_n. \tag{66}
$$

B Proof of the Converse for Theorem 2

For the Gaussian channel, if $|a| \leqslant 1$, it satisfies the condition (4). For these channels, we first prove the following bounds.

$$nR_1 \leqslant \sum_{i=1}^{n}[I(U_iX_{1i};Y_{1i}) - I(U_i;S_i|X_{1i})] + n\delta_n \tag{67a}$$

$$nR_2 \leqslant \sum_{i=1}^{n} I(X_{2i};Y_{2i}|U_iX_{1i}S_i) + n\delta_n \tag{67b}$$

$$n(R_1+R_2) \leqslant \sum_{i=1}^{n} I(X_{1i}X_{2i};Y_{2i}|S_i) + n\delta_n \tag{67c}$$

The bound (67a) follows from (59) by setting $U_i = K_i = (W_1S_{i+1}^nX_1^nY_1^{i-1})$ for $i = 1, \ldots, n$. The bound (67c) follows from (61).

We then bound R_2 as follows and obtain (67b):

$$\begin{aligned}
nR_2 =& I(W_2;Y_2^nS^n) + n\delta_n \\
\overset{(a)}{\leqslant}& I(W_2;Y_2^nS^n|W_1) + n\delta_n \\
\overset{(b)}{=}& I(W_2;Y_2^n|W_1S^n) + n\delta_n \\
=& \sum_{i=1}^{n} I(W_2;Y_{2i}|W_1S^nY_2^{i-1}) + n\delta_n \\
=& \sum_{i=1}^{n} [H(Y_{2i}|W_1S^nY_2^{i-1}) - H(Y_{2i}|W_1W_2S^nY_2^{i-1})] + n\delta_n \\
\overset{(c)}{=}& \sum_{i=1}^{n} [H(Y_{2i}|W_1S^nX_1^nY_1^{i-1}Y_2^{i-1}) - H(Y_{2i}|W_1W_2S^nY_2^{i-1}X_1^nY_1^{i-1})] + n\delta_n \\
\overset{(d)}{\leqslant}& \sum_{i=1}^{n}[H(Y_{2i}|W_1S_{i+1}^nX_1^nY_1^{i-1}S_i) - H(Y_{2i}|W_1S_{i+1}^nS_iX_1^nY_1^{i-1}X_{2i})] + n\delta_n \\
\overset{(e)}{\leqslant}& \sum_{i=1}^{n} [H(Y_{2i}|S_iX_{1i}U_i) - H(Y_{2i}|S_iX_{1i}U_iX_{2i})] + n\delta_n \\
\leqslant& \sum_{i=1}^{n} I(X_{2i};Y_{2i}|U_iX_{1i}S_i) + n\delta_n
\end{aligned} \tag{68}$$

where (a) follows because W_1 and W_2 are independent, (b) follows because W_2 and S are independent, (c) follows from the degradedness condition (4) so that X_1^n and Y_1^{i-1} can be added into the conditioning, (d) follows from the fact that given X_{1i}, X_{2i}, and S_i, Y_{2i} is independent of all other variables, and (e) follows from the definition of U_i.

We further derive the bounds (67a)-(67c) for Gaussian channels. We first consider the bound on R_1 as follows:

$$\begin{aligned}
R_1 &\leqslant \frac{1}{n}\sum_{i=1}^{n}[I(X_{1i}U_i;Y_{1i}) - I(U_i;S_i|X_{1i})] \\
&= \frac{1}{n}\sum_{i=1}^{n}[h(Y_{1i}) - h(Y_{1i}|X_{1i}U_i) - h(S_i|X_{1i}) + h(S_i|X_{1i}U_i)] \\
&\overset{(a)}{=} \frac{1}{n}\sum_{i=1}^{n}[h(Y_{1i}) - h(Y_{1i}|X_{1i}U_iS_i) - I(S_i;Y_{1i}|X_{1i}U_i) - h(S_i|X_{1i}) + h(S_i|X_{1i}U_i)] \\
&= \frac{1}{n}\sum_{i=1}^{n}[h(Y_{1i}) - h(Y_{1i}|X_{1i}U_iS_i) - h(S_i|X_{1i}) + h(S_i|X_{1i}U_iY_{1i})] \\
&\overset{(b)}{\leqslant} \frac{1}{n}\sum_{i=1}^{n}[h(Y_{1i}) - h(Y_{1i}|X_{1i}U_iS_i) - h(S_i) + h(S_i|X_{1i}Y_{1i})] \qquad (69)
\end{aligned}$$

where (a) follows because addition of the second and third terms equals the second term in the previous step, and (b) follows because S_i and X_{1i} are independent and conditioning does not increase entropy.

We then derive bound for each term in (69) respectively as follows. The first term in (69) can be derived as:

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}h(Y_{1i}) \\
\overset{(a)}{\leqslant}&\frac{1}{2n}\sum_{i=1}^{n}\log 2\pi e(E(X_{1i}+aX_{2i}+S_i+N_i)^2) \\
\leqslant&\frac{1}{2n}\sum_{i=1}^{n}\log 2\pi e\Big(E[X_{1i}^2]+2aE(X_{1i}X_{2i})+a^2E[X_{2i}^2]+2aE(X_{2i}S_i)+E[S_i^2]+E[N_i^2]\Big) \\
\overset{(b)}{\leqslant}&\frac{1}{2}\log 2\pi e\Bigg(\frac{1}{n}\sum_{i=1}^{n}E[X_{1i}^2]+\frac{2a}{n}\sum_{i=1}^{n}E(X_{1i}X_{2i})+\frac{a^2}{n}\sum_{i=1}^{n}E[X_{2i}^2]+\frac{2a}{n}\sum_{i=1}^{n}E(X_{2i}S_i) \\
&+\frac{1}{n}\sum_{i=1}^{n}E[S_i^2]+\frac{1}{n}\sum_{i=1}^{n}E[N_i^2]\Bigg) \\
\overset{(c)}{\leqslant}&\frac{1}{2}\log 2\pi e\left(P_1+a^2P_2+Q+1+\frac{2a}{n}\sum_{i=1}^{n}E(X_{1i}X_{2i})+\frac{2a}{n}\sum_{i=1}^{n}E(X_{2i}S_i)\right) \\
\leqslant&\frac{1}{2}\log 2\pi e\left(P_1+a^2P_2+Q+1+2a\rho_{21}\sqrt{P_1P_2}+2a\rho_{2s}\sqrt{P_2Q}\right) \qquad (70)
\end{aligned}$$

where $\rho_{21}=\frac{\frac{1}{n}\sum_{i=1}^{n}E(X_{1i}X_{2i})}{\sqrt{P_1P_2}}$ and $\rho_{2s}=\frac{\frac{1}{n}\sum_{i=1}^{n}E(X_{2i}S_i)}{\sqrt{P_2Q}}$. In the above derivation, (a) follows from the fact that the Gaussian distribution maximizes the entropy given the variance of the random variable, (b) follows from the concavity of the logarithm function and Jensen's inequality, and (c) follows from the power constraints.

We next quantify the term $\frac{1}{n}\sum_{i=1}^{n} h(Y_{1i}|X_{1i}U_iS_i)$ via its upper and lower bounds. We first have

$$\frac{1}{n}\sum_{i=1}^{n} h(Y_{1i}|X_{1i}X_{2i}S_i) \overset{(a)}{\leqslant} \frac{1}{n}\sum_{i=1}^{n} h(Y_{1i}|X_{1i}U_iS_i) \leqslant \frac{1}{n}\sum_{i=1}^{n} h(Y_{1i}|X_{1i}S_i) \tag{71}$$

where (a) follows because conditioning does not increase entropy and given X_{1i}, X_{2i}, and S_i, Y_{1i} is independent of all other variables.

For the left-hand side, we have

$$\frac{1}{n}\sum_{i=1}^{n} h(Y_{1i}|X_{1i}X_{2i}S_i) = \frac{1}{2}\log 2\pi e. \tag{72}$$

For the right-hand side, by setting $\alpha = a\rho_{21}\sqrt{\frac{P_2}{P_1}}$ and $\beta = a\rho_{2S}\sqrt{\frac{P_2}{Q}}$, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} h(Y_{1i}|X_{1i}S_i)\\
=&\frac{1}{n}\sum_{i=1}^{n} h(X_{1i}+aX_{2i}+S_i+N_{1i}|S_iX_{1i})\\
=&\frac{1}{n}\sum_{i=1}^{n} h(aX_{2i}+N_{1i}-\alpha X_{1i}-\beta S_i|S_iX_{1i})\\
\overset{(a)}{\leqslant}&\frac{1}{n}\sum_{i=1}^{n} h(aX_{2i}+N_{1i}-\alpha X_{1i}-\beta S_i)\\
\overset{(b)}{\leqslant}&\frac{1}{2n}\sum_{i=1}^{n} \log(2\pi e E[(aX_{2i}+N_{1i}-\alpha X_{1i}-\beta S_i)^2])\\
\overset{(c)}{\leqslant}&\frac{1}{2}\log 2\pi e\left(a^2P_2+1+\alpha^2P_1+\beta^2Q-2a\alpha\frac{1}{n}\sum_{i=1}^{n}E[X_{1i}X_{2i}]-2a\beta\frac{1}{n}\sum_{i=1}^{n}E[X_{2i}S_i]\right)\\
=&\frac{1}{2}\log 2\pi e\left(1+a^2(1-\rho_{2S}^2-\rho_{21}^2)P_2\right).
\end{aligned} \tag{73}$$

where (a) follows because conditioning does not increase entropy, (b) follows because the Gaussian distribution maximizes the entropy for variables with certain variance, and (c) follows from the concavity of the log function and Jensen's inequality.

Therefore, combining (72) and (73), we conclude that there exists $0 \leqslant P_2'' \leqslant (1-\rho_{2S}^2-\rho_{21}^2)P_2$ such that

$$\frac{1}{n}\sum_{i=1}^{n} h(Y_{1i}|X_{1i}U_iS_i) = \frac{1}{2}\log 2\pi e(1+a^2P_2'')\,. \tag{74}$$

The third term in (69) is given by

$$\frac{1}{n}\sum_{i=1}^{n} h(S_i) = \frac{1}{2}\log 2\pi eQ\,. \tag{75}$$

Finally, for the fourth term in (69), we first define $\alpha' = \frac{-a\rho_{21}\sqrt{P_2P_1}(a\rho_{2s}\sqrt{P_2Q}+Q)}{\left(a^2(1-\rho_{21}^2)P_2+Q+2a\rho_{2s}\sqrt{P_2Q}+1\right)P_1}$ and $\beta' = -\frac{P_1}{a\rho_{21}\sqrt{P_1P_2}}\alpha'$, and then have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} h(S_i|X_{1i}Y_{1i}) \\
=&\frac{1}{n}\sum_{i=1}^{n} h(S_i|X_{1i},\, X_{1i}+aX_{2i}+S_i+N_{1i}) \\
=&\frac{1}{n}\sum_{i=1}^{n} h(S_i-\alpha' X_{1i}-\beta'(aX_{2i}+S_i+N_{1i})|X_{1i},\, X_{1i}+aX_{2i}+S_i+N_{1i}) \\
\overset{(a)}{\leqslant}&\frac{1}{n}\sum_{i=1}^{n} h(S_i-\alpha' X_{1i}-\beta'(aX_{2i}+S_i+N_{1i})) \\
\overset{(b)}{\leqslant}&\frac{1}{n}\sum_{i=1}^{n} \log\left(2\pi e E(S_i-\alpha' X_{1i}-\beta'(aX_{2i}+S_i+N_{1i}))^2\right) \\
\overset{(c)}{\leqslant}&\frac{1}{2}\log 2\pi e\Bigg(Q+\alpha'^2P_1+a^2\beta'^2P_2+\beta'^2Q+2a\beta'^2\frac{1}{n}\sum_{i=1}^{n}E(X_{2i}S_i)+\beta'^2 \\
&\quad+2\alpha'\beta' a\frac{1}{n}\sum_{i=1}^{n}E(X_{1i}X_{2i})-2\beta' a\frac{1}{n}\sum_{i=1}^{n}E(X_{2i}S_i)-2\beta' Q\Bigg) \\
=&\frac{1}{2}\log 2\pi e\frac{(a^2(1-\rho_{21}^2-\rho_{2s}^2)P_2+1)Q}{a^2(1-\rho_{21}^2)P_2+2a\rho_{2s}\sqrt{P_2Q}+Q+1}
\end{aligned} \tag{76}
$$

where (a) follows because conditioning does not increase entropy, (b) follows because the Gaussian distribution maximizes the entropy for variables with certain variance, and (c) follows from the concavity of the log function and Jensen's inequality.

Substituting (70), (72), (75) and (76) into (69), we obtain

$$
R_1 \leqslant \frac{1}{2}\log\left(1+\frac{P_1+2a\rho_{21}\sqrt{P_1P_2}+a^2\rho_{21}^2P_2}{a^2(1-\rho_{21}^2)P_2+2a\rho_{2s}\sqrt{P_2Q}+Q+1}\right)+\frac{1}{2}\log\left(1+\frac{a^2P_2'}{a^2P_2''+1}\right)
$$

where $P_2' = (1-\rho_{21}^2-\rho_{2s}^2)P_2-P_2''$.

We then bound R_2 by further deriving (67b). When $a \leqslant 1$, we have $Y_{1i} = aY_{2i}+(1-ab)X_{1i}+(1-ac)S_i+N_i'$, where $N_i' \sim \mathcal{N}(0,1-a^2)$ and is independent from Y_2^n, X_1^n and S^n. By applying the conditional entropy power inequality [23], we have

$$
\begin{aligned}
2^{2h(Y_{1i}|U_iS_iX_{1i})} =&2^{2h(aY_{2i}+(1-ab)X_{1i}+(1-ac)S_i+N_i'|U_iS_iX_{1i})} \\
=&2^{2h(aY_{2i}+N_i'|U_iS_iX_{1i})} \\
\geqslant&2^{2h(aY_{2i}|U_iS_iX_{1i})}+2^{2h(N_i'|U_iS_iX_{1i})} \\
=&2^{2h(Y_{2i}|U_iS_iX_{1i})+\log(a^2)}+2\pi e(1-a^2).
\end{aligned} \tag{77}
$$

Thus,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} h(Y_{2i}|U_iS_iX_{1i}) \leqslant& \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\log\left(\frac{2^{2h(Y_{1i}|U_iS_iX_{1i})}-2\pi e(1-a^2)}{a^2}\right)\\ \overset{(a)}{\leqslant}& \frac{1}{2}\log\left(\frac{2^{2\frac{1}{n}\sum_{i=1}^{n} h(Y_{1i}|U_iS_iX_{1i})}-2\pi e(1-a^2)}{a^2}\right)\\ \overset{(b)}{=}& \frac{1}{2}\log(2\pi e(1+P_2'')) \end{aligned} \tag{78}$$

where (a) follows from the concavity of the function $\log\left(2^x-b\right)$ for $b\geqslant 0$, and (b) follows from (74).

Therefore, we have

$$\begin{aligned}R_2 &\leqslant \frac{1}{n}\sum_{i=1}^{n} I(X_{2i};Y_{2i}|X_{1i}S_iU_i)\\ &= \frac{1}{n}\sum_{i=1}^{n}\left[h(Y_{2i}|X_{1i}S_iU_i)-h(Y_{2i}|X_{1i}S_iX_{2i})\right]\\ &\overset{(a)}{=} \frac{1}{2}\log(2\pi e(1+P_2''))-\frac{1}{2}\log(2\pi e)\\ &= \frac{1}{2}\log(1+P_2'')\ . \end{aligned} \tag{79}$$

where (a) follows from (74).

We finally bound $R_1 + R_2$ by further deriving (67c). We set $\alpha'' = \rho_{2s}\sqrt{\frac{P_2}{Q}}$, and have

$$\begin{aligned}
R_1 + R_2 \leqslant & \frac{1}{n}\sum_{i=1}^{n} I(X_{1i}X_{2i};Y_{2i}|S_i) \\
= & \frac{1}{n}\sum_{i=1}^{n}[h(Y_{2i}|S_i) - h(Y_{2i}|X_{1i}S_iX_{2i})] \\
= & \frac{1}{n}\sum_{i=1}^{n} h(bX_{1i} + X_{2i} + cS_i + N_{1i}|S_i) - \frac{1}{2}\log 2\pi e \\
= & \frac{1}{n}\sum_{i=1}^{n} h(bX_{1i} + X_{2i} + N_{1i} - \alpha'' S_i|S_i) - \frac{1}{2}\log 2\pi e \\
\overset{(a)}{\leqslant} & \frac{1}{n}\sum_{i=1}^{n} h(bX_{1i} + X_{2i} + N_{1i} - \alpha'' S_i) - \frac{1}{2}\log 2\pi e \\
\overset{(b)}{\leqslant} & \frac{1}{n}\sum_{i=1}^{n} \log(2\pi e E(bX_{1i} + X_{2i} + N_{1i} - \alpha'' S_i)^2) - \frac{1}{2}\log 2\pi e \\
\overset{(c)}{\leqslant} & \frac{1}{2}\log 2\pi e\left(b^2P_1 + P_2 + 1 + \alpha''^2 Q + 2b\frac{1}{n}\sum_{i=1}^{n}E[X_{1i}X_{2i}] - 2\alpha''\frac{1}{n}\sum_{i=1}^{n}E[X_{2i}S_i]\right) \\
& - \frac{1}{2}\log 2\pi e \\
= & \frac{1}{2}\log(b^2P_1 + P_2 + 1 + 2b\rho_{21}\sqrt{P_1P_2} - \rho_{2s}^2 P_2).
\end{aligned} \tag{80}$$

where (a) follows because conditioning does not increase entropy, (b) follows because the Gaussian distribution maximizes the entropy for variables with certain variance, and (c) follows from the concavity of the log function and Jensen's inequality.

C Proof of Lemma 1

Following (60) and (61), we obtain

$$nR_2 \leqslant \sum_{i=1}^{n} I(X_{2i};Y_{2i}|S_iX_i) + n\delta_n \tag{81a}$$

$$n(R_1 + R_2) \leqslant \sum_{i=1}^{n} I(X_{1i}X_{2i};Y_{2i}|S_i) + n\delta_n. \tag{81b}$$

We then prove an alternative bound on $R_1 + R_2$ as follows:

$$
\begin{aligned}
& n(R_1+R_2) \\
\leqslant & I(W_1;Y_1^n) + I(W_2;Y_2^n|S^n) + n\delta_n \\
\overset{(a)}{\leqslant} & I(W_1;Y_1^n) + I(W_2;Y_2^n|S^nW_1) + n\delta_n \\
= & I(W_1;Y_1^n) + H(W_2|S^nW_1) - H(W_2|S^nW_1Y_2^n) + n\delta_n \\
\overset{(b)}{\leqslant} & I(W_1;Y_1^n) + H(W_2|S^nW_1) - H(W_2|S^nW_1Y_2^nY_1^nX_1^n) + n\delta_n \\
\overset{(c)}{=} & I(W_1;Y_1^n) + H(W_2|S^nW_1) - H(W_2|S^nW_1Y_1^n) + n\delta_n \\
= & I(W_1;Y_1^n) + I(W_2;Y_1^n|S^nW_1) + n\delta_n \\
= & \sum_{i=1}^n [H(Y_{1i}|Y_1^{i-1}) - H(Y_{1i}|W_1Y_1^{i-1}) + H(Y_{1i}|S^nW_1Y_1^{i-1}) - H(Y_{1i}|S^nW_1W_2Y_1^{i-1})] + n\delta_n \\
\overset{(d)}{=} & \sum_{i=1}^n [H(Y_{1i}|Y_1^{i-1}) - H(Y_{1i}|X_{1i}) + H(Y_{1i}|X_{1i}) - H(Y_{1i}|W_1Y_1^{i-1}) \\
& \qquad + H(Y_{1i}|S^nW_1Y_1^{i-1}) - H(Y_{1i}|S^nW_1W_2Y_1^{i-1})] + n\delta_n \\
\overset{(e)}{\leqslant} & \sum_{i=1}^n [H(Y_{1i}) - H(Y_{1i}|X_{1i}) + H(Y_{1i}|X_{1i}) \\
& \qquad - I(S^n;Y_{1i}|W_1Y_1^{i-1}) - H(Y_{1i}|S^nX_{1i}X_{2i}W_1W_2Y_1^{i-1})] + n\delta_n \\
\overset{(f)}{=} & \sum_{i=1}^n [I(X_{1i};Y_{1i}) + H(Y_{1i}|X_{1i}) - H(Y_{1i}|S_iX_{1i}X_{2i})] - I(S^n;Y_1^n|W_1) + n\delta_n \\
= & \sum_{i=1}^n [I(X_{1i};Y_{1i}) + H(Y_{1i}|X_{1i}) - H(Y_{1i}|S_iX_{1i}X_{2i})] - H(S^n) + H(S^n|Y_1^nW_1) + n\delta_n \\
= & \sum_{i=1}^n [I(X_{1i};Y_{1i}) + H(Y_{1i}|X_{1i}) - H(Y_{1i}|S_iX_{1i}X_{2i}) - H(S_i) + H(S_i|Y_1^nW_1S_{i+1}^n)] + n\delta_n \\
\overset{(g)}{\leqslant} & \sum_{i=1}^n [I(X_{1i};Y_{1i}) + H(Y_{1i}|X_{1i}) - H(Y_{1i}|S_iX_{1i}X_{2i}) - H(S_i) + H(S_i|Y_{1i}X_{1i})] + n\delta_n \\
= & \sum_{i=1}^n [I(X_{1i};Y_{1i}) + H(Y_{1i}|X_{1i}) - H(Y_{1i}|S_iX_{1i}X_{2i}) - I(S_i;Y_{1i}|X_{1i})] + n\delta_n \\
= & \sum_{i=1}^n [I(X_{1i};Y_{1i}) - H(Y_{1i}|S_iX_{1i}X_{2i}) + H(Y_{1i}|S_iX_{1i})] + n\delta_n \\
= & \sum_{i=1}^n [I(X_{1i};Y_{1i}) + I(X_{2i};Y_{1i}|S_iX_{1i})] + n\delta_n
\end{aligned}
\tag{82}
$$

where (a) follows due to the chain rule and the fact that W_1 and W_2 are independent, (b) follows because conditioning does not increase entropy, (c) follows from degradedness condition (5), (d) follows because the term $H(Y_{1i}|X_{1i})$ is added and subtracted, (e) follows because conditioning does not increase entropy, (f) follows because given X_{1i}, X_{2i}, and S_i, Y_{2i} is independent of all other variables, and (g) follows because X_1^n is a function of W_1 and conditioning does not increase entropy.

D Proof of the Converse for Theorem 3

Based on the outer bound derived in Appendix C, we further derive an outer bound for the Gaussian channel. We first derive a bound on R_2 based on (81a). We set $\alpha = \rho_{21}\sqrt{\frac{P_2}{P_1}}$ and $\beta = \rho_{2s}\sqrt{\frac{P_2}{Q}}$, where $\rho_{21} = \frac{\frac{1}{n}\sum_{i=1}^n E(X_{1i}X_{2i})}{\sqrt{P_1P_2}}$ and $\rho_{2s} = \frac{\frac{1}{n}\sum_{i=1}^n E(X_{2i}S_i)}{\sqrt{P_2Q}}$. We then obtain:

$$
\begin{aligned}
R_2 \leqslant & \frac{1}{n}\sum_{i=1}^n h(Y_{2i}|X_{1i}S_i) - h(Y_{2i}|X_{1i}X_{2i}S_i) \\
= & \frac{1}{n}\sum_{i=1}^n h(bX_{1i} + X_{2i} + cS_i + N_{1i}|S_iX_{1i}) - \frac{1}{2}\log 2\pi e \\
= & \frac{1}{n}\sum_{i=1}^n h(X_{2i} + N_{1i} - \alpha X_{1i} - \beta S_i|S_iX_{1i}) - \frac{1}{2}\log 2\pi e \\
\overset{(a)}{\leqslant} & \frac{1}{n}\sum_{i=1}^n h(X_{2i} + N_{1i} - \alpha X_{1i} - \beta S_i) - \frac{1}{2}\log 2\pi e \\
\overset{(b)}{\leqslant} & \frac{1}{2n}\sum_{i=1}^n \log\left(2\pi e E(X_{2i} + N_{1i} - \alpha X_{1i} - \beta S_i)^2\right) - \frac{1}{2}\log 2\pi e \\
\overset{(c)}{\leqslant} & \frac{1}{2}\log\left(P_2 + 1 + \alpha^2 P_1 + \beta^2 Q - 2\alpha\frac{1}{n}\sum_{i=1}^n E[X_{1i}X_{2i}] - 2\beta\frac{1}{n}\sum_{i=1}^n E[X_{2i}S_i]\right) \\
= & \frac{1}{2}\log(1 + (1 - \rho_{2s}^2 - \rho_{21}^2)P_2)
\end{aligned}
\tag{83}
$$

where (a) follows because conditioning does not increase entropy, (b) follows because the Gaussian distribution maximizes the entropy for variables with certain variance, and (c) follows from the concavity of the log function and Jensen's inequality.

Following (80), we obtain the following bound on $R_1 + R_2$ based on (81b)

$$
R_1 + R_2 \leqslant \frac{1}{2}\log\left(b^2P_1 + P_2 + 1 + 2b\rho_{21}\sqrt{P_1P_2} - \rho_{2s}^2P_2\right). \tag{84}
$$

We further derive (82) for the Gaussian channel as follows:

$$
\begin{aligned}
R_1 + R_2 &\leqslant \frac{1}{n}\sum_{i=1}^{n}[I(X_{1i};Y_{1i}) + I(X_{2i};Y_{1i}|X_{1i}S_i)] \\
&= \frac{1}{n}\sum_{i=1}^{n}[h(Y_{1i}) - h(Y_{1i}|X_{1i}) + h(Y_{1i}|X_{1i}S_i) - h(Y_{1i}|S_iX_{1i}X_{2i})] \\
&= \frac{1}{n}\sum_{i=1}^{n}[h(Y_{1i}) - I(S_i;Y_{1i}|X_{1i}) - h(Y_{1i}|S_iX_{1i}X_{2i})] \\
&\overset{(a)}{=} \frac{1}{n}\sum_{i=1}^{n}[h(Y_{1i}) - h(S_i) + h(S_i|X_{1i}Y_{1i}) - h(Y_{1i}|S_iX_{1i}X_{2i})]
\end{aligned} \tag{85}
$$

where (a) follows because S_i and X_{1i} are independent.

Following (70), (72), (75), and (76) in Appendix B, we obtain

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} h(Y_{1i}) \leqslant \frac{1}{2}\log 2\pi e(P_1 + a^2P_2 + Q + 1 + 2a\rho_{21}\sqrt{P_1P_2} + 2a\rho_{2s}\sqrt{P_2Q}) \\
&\frac{1}{n}\sum_{i=1}^{n} h(Y_{1i}|X_{1i}X_{2i}S_i) = \frac{1}{2}\log 2\pi e \\
&\frac{1}{n}\sum_{i=1}^{n} h(S_i) = \frac{1}{2}\log 2\pi e Q \\
&\frac{1}{n}\sum_{i=1}^{n} h(S_i|X_{1i}Y_{1i}) \leqslant \frac{1}{2}\log 2\pi e \frac{(a^2(1-\rho_{21}^2-\rho_{2s}^2)P_2+1)Q}{a^2(1-\rho_{21}^2)P_2 + 2a\rho_{2s}\sqrt{P_2Q} + Q + 1}
\end{aligned}
$$

Substituting the above bounds into (85), we obtain

$$
\begin{aligned}
R_1 + R_2 \leqslant &\frac{1}{2}\log\left(1 + \frac{P_1 + 2a\rho_{21}\sqrt{P_1P_2} + a^2\rho_{21}^2P_2}{a^2(1-\rho_{21}^2)P_2 + 2a\rho_{2s}\sqrt{P_2Q} + Q + 1}\right) \\
&+ \frac{1}{2}\log\left(1 + a^2(1-\rho_{2s}^2-\rho_{21}^2)P_2\right)
\end{aligned} \tag{86}
$$

which concludes the proof.

E Proof of Lemma 2

The achievable scheme applies rate splitting, superposition coding and Gel'fand-Pinsker binning scheme. We use random codes and fix the following joint distribution:

$$
P_{SX_1TUVX_2Y_1Y_2} = P_{X_1}P_SP_{T|X_1S}P_{U|X_1TS}P_{V|TUX_1S}P_{X_2|TUVX_1S}P_{Y_1Y_2|X_1X_2S}.
$$

Let $T_\epsilon^n(P_{SX_1TUVX_2Y_1Y_2})$ denote the strongly joint ϵ-typical set based on the above distribution. For a given sequence x^n, let $T_\epsilon^n(P_{U|X}|x^n)$ denote the set of sequences u^n such that (u^n, x^n) is jointly typical based on the distribution P_{XU}.

Code Construction:

1. Generate 2^{nR_1} codewords $x_1^n(w_1)$ with i.i.d. components based on P_{X_1}. Index these codewords by $w_1 = 1,\cdots,2^{nR_1}$.
2. For each $x_1^n(w_1)$, generate $t^n(w_1, v_1)$ with i.i.d. components based on $P_{T|X_1}$. Index these codewords by $v_1 = 1,\cdots,2^{n\tilde{R}_1}$.
3. For each $x_1^n(w_1)$ and $t^n(w_1, v_1)$, generate $u^n(w_1, v_1, w_{21}, v_{21})$ with i.i.d. components based on $P_{U|X_1T}$. Index these codewords by $w_{21} = 1,\cdots,2^{nR_{21}}$ and $v_{21} = 1,\cdots,2^{n\tilde{R}_{21}}$.
4. For each $x_1^n(w_1)$, $t^n(w_1, v_1)$, and $u^n(w_1, v_1, w_{21}, v_{21})$, generate $v^n(w_1, v_1, w_{21}, v_{21}, w_{22}, v_{22})$ with i.i.d. components based on $P_{V|X_1TU}$. Index these codewords by $w_{22} = 1,\cdots,2^{nR_{22}}$ and $v_{22} = 1,\cdots,2^{n\tilde{R}_{22}}$.

Encoding:

1. Encoder 1: Given w_1, map w_1 into $x_1^n(w_1)$ for transmission.
2. Encoder 2:
   - Given w_1, $x_1^n(w_1)$ and s^n, select $t^n(w_1, \tilde{v}_1)$ such that
   $$(t^n(w_1, \tilde{v}_1), s^n, x_1^n(w_1)) \in T_\epsilon^n(P_{X_1}P_SP_{T|X_1S}).$$
   Otherwise, set $\tilde{v}_1 = 1$. It can be shown that for large n, such t^n exists with high probability if
   $$\tilde{R}_1 > I(T; S|X_1). \tag{87}$$
   - Given w_{21} and selected $t^n(w_1, \tilde{v}_1)$, select $u^n(w_1, \tilde{v}_1, w_{21}, \tilde{v}_{21})$ such that
   $$(u^n(w_1, \tilde{v}_1, w_{21}, \tilde{v}_{21}), t^n(w_1, \tilde{v}_1), s^n, x_1^n(w_1)) \in T_\epsilon^n(P_{X_1}P_SP_{T|X_1S}P_{U|X_1ST}).$$
   Otherwise, set $\tilde{v}_{21} = 1$. It can be shown that for large n, such u^n exists with high probability if
   $$\tilde{R}_{21} > I(U; S|X_1T). \tag{88}$$
   - Given w_{22} and selected $u^n(w_1, \tilde{v}_1, w_{21}, \tilde{v}_{21})$, select $v^n(w_1, \tilde{v}_1, w_{21}, \tilde{v}_{21}, w_{22}, \tilde{v}_{22})$ such that
   $$\begin{aligned}&(v^n(w_1, \tilde{v}_1, w_{21}, \tilde{v}_{21}, w_{22}, \tilde{v}_{22}), u^n(w_1, \tilde{v}_1, w_{21}, \tilde{v}_{21}), t^n(w_1, \tilde{v}_1), s^n, x_1^n(w_1))\\ &\in T_\epsilon^n(P_{X_1}P_SP_{T|X_1S}P_{U|X_1ST}P_{V|UX_1ST}).\end{aligned}$$
   Otherwise, set $\tilde{v}_{22} = 1$. It can be shown that for large n, such v^n exists with high probability if
   $$\tilde{R}_{22} > I(V; S|UX_1T). \tag{89}$$
   - Given selected $x_1^n(w_1)$, $t^n(w_1, \tilde{v}_1)$, $u^n(w_1, \tilde{v}_1, w_{21}, \tilde{v}_{21})$, $v^n(w_1, \tilde{v}_1, w_{21}, \tilde{v}_{21}, w_{22}, \tilde{v}_{22})$ and s^n, generate x_2^n with i.i.d. components based on $P_{X_2|TUVX_1S}$ for transmission.

Decoding:

1 Decoder 1: Given y_1^n, find the unique tuple $(\hat{w}_1, \hat{v}_1, \hat{w}_{21}, \hat{v}_{21})$ such that

$$(x_1^n(\hat{w}_1), t^n(\hat{w}_1, \hat{v}_1), u^n(\hat{w}_1, \hat{v}_1, \hat{w}_{21}, \hat{v}_{21}), y_1^n) \in T_\epsilon^n(P_{X_1TUY_1}).$$

If no or more than one such tuples with different w_1 can be found, then declare error. One can show that for sufficiently large n, decoding is correct with high probability if

$$R_1 + \tilde{R}_1 + R_{21} + \tilde{R}_{21} \leqslant I(TUX_1; Y_1) \tag{90}$$

We note that since receiver 1 is not required to decode W_{21} correctly by the channel model, the corresponding error events do not need to be analyzed.

2. Decoder 2: Given y_2^n, find a tuple $(\hat{w}_1, \hat{v}_1, \hat{w}_{21}, \hat{v}_{21}, \hat{w}_{22}, \hat{v}_{22})$ such that

$$\begin{aligned}(x_1^n(\hat{w}_1), t^n(\hat{w}_1, \hat{v}_1), u^n(\hat{w}_1, \hat{v}_1, \hat{w}_{21}, \hat{v}_{21}), v^n(\hat{w}_1, \hat{v}_1, \hat{w}_{21}, \hat{v}_{21}, \hat{w}_{22}, \hat{v}_{22}), y_2^n)\\ \in T_\epsilon^n(P_{X_1TUVY_2}).\end{aligned}$$

If no or more than one such tuples can be found, then declare error. It can be shown that for sufficiently large n, decoding is correct with high probability if

$$R_{22} + \tilde{R}_{22} \leqslant I(V; Y_2|UX_1T) \tag{91a}$$

$$R_{21} + \tilde{R}_{21} + R_{22} + \tilde{R}_{22} \leqslant I(UV; Y_2|X_1T) \tag{91b}$$

$$\tilde{R}_1 + R_{21} + \tilde{R}_{21} + R_{22} + \tilde{R}_{22} \leqslant I(TUV; Y_2|X_1) \tag{91c}$$

$$R_1 + \tilde{R}_1 + R_{21} + \tilde{R}_{21} + R_{22} + \tilde{R}_{22} \leqslant I(TUVX_1; Y_2) \tag{91d}$$

Lemma 2 is thus proved by combining (87)-(91d).

F Proof of Theorem 5

Consider a $(2^{nR_1}, 2^{nR_2}, n)$ code with an average error probability $P_e^{(n)}$. The probability distribution on $\mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{S}^n \times \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}_1^n \times \mathcal{Y}_2^n$ is given by

$$P_{W_1W_2S^nX_1^nX_2^nY_1^nY_2^n} = P_{W_1}P_{W_2}\left[\prod_{i=1}^n P_{S_i}\right] P_{X_1^n|W_1}P_{X_2^n|W_1W_2S^n}\prod_{i=1}^n P_{Y_{1i}Y_{2i}|X_{1i}X_{2i}S_i}. \tag{92}$$

By Fano's inequality, we have

$$H(W_1|Y_1^n) \leqslant nR_1P_e^{(n)} + 1 = n\delta_{1n} \tag{93a}$$

$$H(W_1W_2|Y_2^n) \leqslant n(R_1 + R_2)P_e^{(n)} + 1 = n\delta_{2n} \tag{93b}$$

where δ_{1n}, $\delta_{2n} \to 0$ as $n \to +\infty$. Let $\delta_n = \delta_{1n} + \delta_{2n}$, which also satisfies that $\delta_n \to 0$ as $n \to +\infty$.

We define the following auxiliary random variables:

$$\begin{aligned} T_i &= (W_1, S_{i+1}^n, X_1^n) \\ U_i &= (T_i, Y_1^{i-1}) \\ V_i &= (T_i, W_2, Y_2^{i-1}) \end{aligned} \tag{94}$$

which satisfy the Markov chain conditions:

$$T_i \longleftrightarrow U_i V_i \longleftrightarrow X_{1i} X_{2i} S_i \longleftrightarrow Y_{1i} Y_{2i} \tag{95}$$

for $i = 1,\cdots, n$.

The following bound on R_1 follows the same steps as in (59) in Appendix A, and we have

$$nR_1 \leqslant \sum_{i=1}^{n} [I(T_i U_i X_{1i}; Y_{1i}) - I(T_i U_i; S_i | X_{1i})] + n\delta_n. \tag{96}$$

I.e.we define $T_i = (W_1, S_{i+1}^n, X_1^n)$ and $U_i = (T_i, Y_1^{i-1})$

We next bound R_2 and obtain

$$\begin{aligned} nR_2 =& I(W_2; Y_2^n) + n\delta_n \leqslant I(W_2; Y_2^n | W_1) + n\delta_n \\ \overset{(a)}{=}& \sum_{i=1}^{n} [I(W_2 S_{i+1}^n; Y_2^i | W_1) - I(W_2 S_i^n; Y_2^{i-1} | W_1)] + n\delta_n \\ \overset{(b)}{=}& \sum_{i=1}^{n} [I(W_2 S_{i+1}^n; Y_2^{i-1} | W_1) + I(W_2 S_{i+1}^n; Y_{2i} | W_1 Y_2^{i-1}) \\ &- I(W_2 S_{i+1}^n; Y_2^{i-1} | W_1) - I(S_i; Y_2^{i-1} | W_1 W_2 S_{i+1}^n)] + n\delta_n \\ =& \sum_{i=1}^{n} [I(W_2 S_{i+1}^n; Y_{2i} | W_1 Y_2^{i-1}) - I(S_i; Y_2^{i-1} | W_1 W_2 S_{i+1}^n)] + n\delta_n \\ =& \sum_{i=1}^{n} [H(Y_{2i} | W_1 Y_2^{i-1}) - H(Y_{2i} | W_1 W_2 S_{i+1}^n Y_2^{i-1}) \\ &- H(S_i | W_1 W_2 S_{i+1}^n) + H(S_i | W_1 W_2 S_{i+1}^n Y_2^{i-1})] + n\delta_n \\ \overset{(c)}{=}& \sum_{i=1}^{n} [H(Y_{2i} | W_1 Y_2^{i-1} X_{1i}) - H(Y_{2i} | W_1 W_2 S_{i+1}^n X_1^n Y_2^{i-1}) \\ &- H(S_i | W_1 W_2 S_{i+1}^n X_{1i}) + H(S_i | W_1 W_2 S_{i+1}^n X_1^n Y_2^{i-1})] + n\delta_n \\ \overset{(d)}{\leqslant}& \sum_{i=1}^{n} [H(Y_{2i} | X_{1i}) - H(Y_{2i} | X_{1i} T_i V_i) - H(S_i | X_{1i}) + H(S_i | X_{1i} T_i V_i)] + n\delta_n \\ =& \sum_{i=1}^{n} [I(T_i V_i; Y_{2i} | X_i) - I(T_i V_i; S_i | X_{1i})] + n\delta_n. \end{aligned} \tag{97, 98}$$

where (a) follows due to cancellation of the terms in the sum and because $Y_1^0 = \phi$, (b) follows from chain rule, (c) follows because X_1^n is a function of W_1, and (d) follows because conditionning does not increase entropy, and from the definition of T_i and V_i.

We then bound the sum rate $R_1 + R_2$ as follows.

$$\begin{aligned}
&n(R_1+R_2)\\
=&I(W_1W_2;Y_2^n)+n\delta_n \qquad (99)\\
\overset{(a)}{=}&\sum_{i=1}^{n}[I(W_1W_2S_{i+1}^n;Y_2^i)-I(W_1W_2S_i^n;Y_2^{i-1})]+n\delta_n\\
\overset{(b)}{=}&\sum_{i=1}^{n}[I(W_1W_2S_{i+1}^n;Y_2^{i-1})+I(W_1W_2S_{i+1}^n;Y_{2i}|Y_2^{i-1})\\
&\quad -I(W_1W_2S_{i+1}^n;Y_2^{i-1})-I(S_i;Y_2^{i-1}|W_1W_2S_{i+1}^n)]+n\delta_n\\
=&\sum_{i=1}^{n}[I(W_1W_2S_{i+1}^n;Y_{2i}|Y_2^{i-1})-I(S_i;Y_2^{i-1}|S_{i+1}^nW_1W_2)]+n\delta_n\\
=&\sum_{i=1}^{n}[H(Y_{2i}|Y_2^{i-1})-H(Y_{2i}|W_1W_2S_{i+1}^nY_2^{i-1})\\
&\quad -H(S_i|S_{i+1}^nW_1W_2)+H(S_i|S_{i+1}^nW_1W_2Y_2^{i-1})]+n\delta_n\\
\overset{(c)}{\leqslant}&\sum_{i=1}^{n}[H(Y_{2i})-H(Y_{2i}|W_1W_2S_{i+1}^nX_1^nY_2^{i-1})-H(S_i|X_{1i})+H(S_i|W_1W_2S_{i+1}^nX_1^nY_2^{i-1})]+n\delta_n\\
\overset{(d)}{=}&\sum_{i=1}^{n}[I(X_{1i}T_iV_i;Y_{2i})-I(T_iV_i;S_i|X_{1i})]+n\delta_n \qquad (100)
\end{aligned}$$

where (a) follows due to cancellation of the terms in the sum and because $Y_1^0 = \phi$, (b) follows due to chain rule, (c) follows because S^n is independent of (X_1^n, W_1, W_2), S^n is i.i.d. and because X_1^n is a function of W_1, and (d) follows from the definition of T_i and V_i.

G Proof of the Outer Bound for Theorem 6

We define the following auxiliary random variables:

$$\begin{aligned}
T_i &= (W_1, S_{i+1}^n, X_1^n, Y_1^{i-1})\\
V_i &= (T_i, W_2, Y_2^{i-1}) \qquad (101)
\end{aligned}$$

which satisfy the Markov chain conditions:

$$T_i \longleftrightarrow V_i \longleftrightarrow X_{1i}X_{2i}S_i \longleftrightarrow Y_{1i} \longleftrightarrow Y_{2i} \qquad (102)$$

for $i = 1,\cdots,n$.

By following the step similar to those in (59), we obtain the following bound on R_1:

$$nR_1 \leqslant \sum_{i=1}^{n}[I(T_iX_{1i};Y_{1i}) - I(T_i;S_i|X_{1i})] + n\delta_n. \tag{103}$$

We next derive a bound on R_2 by continuing to derive the bound (97) as follows:

$$\begin{aligned} nR_2 &\leqslant \sum_{i=1}^{n}[H(Y_{2i}|W_1Y_2^{i-1}) - H(Y_{2i}|W_1W_2S_{i+1}^nY_2^{i-1}) \\ &\quad - H(S_i|W_1W_2S_{i+1}^n) + H(S_i|W_1W_2S_{i+1}^nY_2^{i-1})] + n\delta_n \\ &\overset{(a)}{=} \sum_{i=1}^{n}[H(Y_{2i}|W_1Y_2^{i-1}X_{1i}) - H(Y_{2i}|W_1W_2S_{i+1}^nX_1^nY_1^{i-1}Y_2^{i-1}) \\ &\quad - H(S_i|W_1W_2S_{i+1}^nX_{1i}) + H(S_i|W_1W_2S_{i+1}^nX_1^nY_1^{i-1}Y_2^{i-1})] + n\delta_n \\ &\leqslant \sum_{i=1}^{n}[H(Y_{2i}|X_{1i}) - H(Y_{2i}|X_{1i}T_iV_i) - H(S_i|X_{1i}) + H(S_i|X_{1i}T_iV_i)] + n\delta_n \\ &= \sum_{i=1}^{n}[I(T_iV_i;Y_{2i}|X_i) - I(T_iV_i;S_i|X_{1i})] + n\delta_n. \end{aligned} \tag{104}$$

where (a) follows due to the degradedness condition (4), and because X_{1i} is a function of W_1.

H Proof of the Converse for Theorem 7

We define the auxiliary random variable $T_i = (W_1S_{i+1}^nX_1^nY_1^{i-1})$, which satisfies the Markov chain:

$$T_i \leftrightarrow X_{1i}X_{2i}S_i \leftrightarrow Y_{1i}Y_{2i}, \quad \text{for } i = 1, \cdots, n. \tag{105}$$

Following (103), we obtain

$$nR_1 \leqslant \sum_{i=1}^{n}[I(T_iX_{1i};Y_{1i}) - I(T_i;S_i|X_{1i})] + n\delta_n.$$

We next bound R_2 as follows.

$$\begin{aligned}
nR_2 &= I(W_2; Y_2^n) + n\delta_n \\
&\overset{(a)}{\leqslant} I(W_2; Y_2^n | W_1 S^n X_1^n) + n\delta_n \\
&= \sum_{i=1}^{n} [I(W_2; Y_{2i} | W_1 S^n X_1^n Y_2^{i-1})] + n\delta_n \\
&\leqslant \sum_{i=1}^{n} H(Y_{2i} | W_1 S^n X_1^n Y_2^{i-1}) + n\delta_n \\
&\overset{(b)}{=} \sum_{i=1}^{n} H(Y_{2i} | W_1 S^n X_1^n Y_1^{i-1} Y_2^{i-1}) + n\delta_n \\
&\overset{(c)}{\leqslant} \sum_{i=1}^{n} H(Y_{2i} | W_1 S_{i+1}^n X_1^n Y_1^{i-1} S_i) + n\delta_n \\
&= \sum_{i=1}^{n} H(Y_{2i} | X_{1i} T_i S_i) + n\delta_n
\end{aligned} \tag{106}$$

where (a) follows because W_2 is independent of (W_1, S^n, X_1^n), (b) follows due to the degradedness condition (3), and (c) follows because conditioning does not increase entropy.

We then derive another bound on R_2 by continuing to derive the bound (97) as follows:

$$\begin{aligned}
nR_2 &\leqslant \sum_{i=1}^{n} [H(Y_{2i} | W_1 Y_2^{i-1}) - H(Y_{2i} | W_1 W_2 S_{i+1}^n Y_2^{i-1}) \\
&\quad - H(S_i | W_1 W_2 S_{i+1}^n) + H(S_i | W_1 W_2 S_{i+1}^n Y_2^{i-1})] + n\delta_n \\
&\overset{(a)}{=} \sum_{i=1}^{n} [H(Y_{2i} | W_1 X_1^n Y_2^{i-1}) - H(Y_{2i} | W_1 W_2 S_{i+1}^n Y_2^{i-1}) - H(S_i | W_1 W_2 X_1^n S_{i+1}^n) \\
&\quad + H(S_i | W_1 W_2 X_1^n S_{i+1}^n Y_1^{i-1} Y_2^{i-1} Y_{2i}) + I(Y_{2i}; S_i | W_1 W_2 S_{i+1}^n Y_2^{i-1})] + n\delta_n \\
&\overset{(b)}{\leqslant} \sum_{i=1}^{n} [H(Y_{2i} | X_{1i}) - H(S_i | X_{1i}) + H(S_i | X_{1i} T_i Y_{2i})] + n\delta_n \\
&= \sum_{i=1}^{n} [H(Y_{2i} | X_{1i}) - I(T_i Y_{2i}; S_i | X_{1i})] + n\delta_n.
\end{aligned} \tag{107}$$

where (a) follows because X_1^n is a function of W_1 and from the degradedness condition (3), and (b) follows because S_i is independent of (W_1, W_2, X_1^n), and conditioning does not increase entropy, and follows from the definition of T_i.

References

[1] S. Gel'fand and M. Pinsker. Coding for channels with ramdom parameters. *Probl. Contr. Inform. Theory*, 9(1):19–31, January 1980.

[2] L. Zhang, T. Liu, and S. Cui. Symmetric Gaussian interference channel with state information. In *Proc. 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, September 2011.

[3] L. Zhang, J. Jiang, and S. Cui. Gaussian interference channel with state information. *IEEE Trans. Wireless Commun.*, 12(8):4058–4071, Aug 2013.

[4] S. Ghasemi-Goojani and H. Behroozi. On the achievable rate-regions for state-dependent Gaussian interference channel. *Available at http://arxiv.org/abs/1301.5535, submitted in January*, 2013.

[5] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdú. Cognitive interference channels with state information. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Toronto, Canada, July 2008.

[6] M. Kazemi and A. Vosoughi. On the capacity of the state-dependent cognitive interference channel. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Istanbul, Turkey, July 2013.

[7] W. Wu, S. Vishwanath, and A. Arapostathis. Capacity of a class of cognitive radio channels: Interference channels with degraded message sets. *IEEE Trans. Inform. Theory*, 53(11):4391–4399, November 2007.

[8] A. Jovicic and P. Viswanath. Cognitive radio: An information-theoretic perspective. *IEEE Trans. Inform. Theory*, 55(9):3945–3958, September 2009.

[9] I. Maric, A. Goldsmith, G. Kramer, and S. Shamai (Shitz). On the capacity of interference channels with one cooperating transmitter. *European Transactions on Telecommunications*, 19:405–420, April 2008.

[10] S. Rini, D. Tuninetti, and N. Devroye. New inner and outer bounds for the memoryless cognitive interference channel and some new capacity results. *IEEE Trans. Inform. Theory*, 57(7):4087–4109, July 2011.

[11] S. Rini, D. Tuninetti, and N. Devroye. Inner and outer bounds for the Gaussian cognitive interference channel and new capacity results. *IEEE Trans. Inform. Theory*, 58(2):820 – 848, July 2012.

[12] N. Devroye, P. Mitran, and V. Tarokh. Achievable rates in cognitive radio channels. *IEEE Trans. Inform. Theory*, 52(5):1813–1827, May 2006.

[13] J. Jiang and Y. Xin. On the achievable rate regions for interference channels with degraded message sets. *IEEE Trans. Inform. Theory*, 54(10):4707–4712, October 2008.

[14] Y. Liang, A. Somekh-Baruch, H. V. Poor, S. Shamai (Shitz), and S. Verdú. Capacity of cognitive interference channels with and without secrecy. *IEEE Trans. Inform. Theory*, 55(2):604–619, February 2009.

[15] A. Lapidoth and L. Wang. The state-dependent semideterministic broadcast channel. *IEEE Trans. Inform. Theory*, 59(4):2242–2251, April 2013.

[16] S. I. Bross and Y. Steinberg. The state-dependent interference channel with states available at a cribbing encoder and one receiver. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Saint Peterssurg, Russia, July 2011.

[17] Y. Steinberg. Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information. *IEEE Trans. Inform. Theory*, 51(8):2867–2877, August 2005.

[18] M. H. M. Costa. Writing on dirty paper. *IEEE Trans. Inform. Theory*, 29(3):439–441, May 1983.

[19] J. Körner and K. Marton. Comparison of two noisy channels. Topics in Inform. Theory (edited by I. Csiszar and P. Elias), Keszthely, Hungary, 411-423, August, 1975.

[20] A. Schrijver. *Theory of Linear and Integer Programming.* Wiley, New York, 1998.

[21] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdú. Cooperative multiple-access encoding with states available at one transmitter. *IEEE Trans. Inform. Theory*, 54(10):4448–4469, October 2008.

[22] I. Csiszár and J. Körner. *Information Theory.* Akademiai Kiado, Budapest, 1981.

[23] N. M. Blachman. The convolution inequality for entropy powers. *IEEE Trans. Inform. Theory*, 11(2):267–271, April 1965.