

Article

MIMO Gaussian State-Dependent Channels with a State-Cognitive Helper

Michael Dikshtein ^{1,*} , Ruchen Duan ^{2,*}, Yingbin Liang ^{3,*} and Shlomo Shamai (Shitz) ^{1,*}¹ Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel² Samsung Semiconductor Inc., San Jose, CA 95134, USA³ Department of ECE, The Ohio State University, Columbus, OH 43210, USA

* Correspondence: michaeldic@campus.technion.ac.il (M.D.); r.duan@samsung.com (R.D.); liang.889@osu.edu (Y.L.); sshlomo@ee.technion.ac.il (S.S.)

Received: 2 January 2019; Accepted: 5 March 2019; Published: 12 March 2019



Abstract: We consider the problem of channel coding over multiterminal state-dependent channels in which neither transmitters nor receivers but only a helper node has a non-causal knowledge of the state. Such channel models arise in many emerging communication schemes. We start by investigating the parallel state-dependent channel with the same but differently scaled state corrupting the receivers. A cognitive helper knows the state in a non-causal manner and wishes to mitigate the interference that impacts the transmission between two transmit–receive pairs. Outer and inner bounds are derived. In our analysis, the channel parameters are partitioned into various cases, and segments on the capacity region boundary are characterized for each case. Furthermore, we show that for a particular set of channel parameters, the capacity region is entirely characterized. In the second part of this work, we address a similar scenario, but now each channel is corrupted by an independent state. We derive an inner bound using a coding scheme that integrates single-bin Gel'fand–Pinsker coding and Marton's coding for the broadcast channel. We also derive an outer bound and further partition the channel parameters into several cases for which parts of the capacity region boundary are characterized.

Keywords: dirty paper coding; Gel'fand–Pinsker scheme; non-causal channel state information; network information theory

1. Introduction

Cellular communication systems are designed to allow multiple users to share the same communication medium. Traditionally, mobile networks have enabled this feature by dividing the physical resources (such as time, frequency, code, and space) in an orthogonal manner between users. An illustration of the typical methods, called Orthogonal Multiuser Access (OMA) is shown in Figure 1.

The future of cellular communications is facing exponential growth in bandwidth demand. Furthermore, increased popularity in Internet of Things (IoT) applications and the emergence of Vehicle-to-Vehicle (V2V) connectivity will further grow the number of network consumers. Hence, fifth-generation (5G) wireless networks are required to support extensive connectivity, low latency, and higher data rates. Such requirements cannot be satisfied using the traditional OMA methods and thus to sustain more users and higher transmission rates, non-orthogonal multiuser access (NOMA) has been intensively investigated, where interference mitigation is the key issue for non-orthogonal transmission. A comprehensive survey on NOMA from an information theoretic perspective is given in [1].

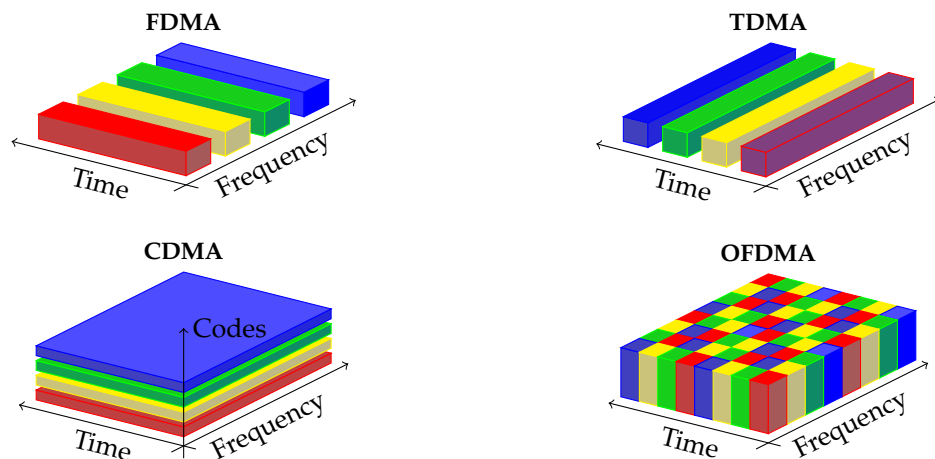


Figure 1. Orthogonal Multiple Access Techniques.

In this work, we study a particular communication model that can be used in future NOMA techniques. Specifically, we investigate a type of state-dependent channel with a helper, illustrated in Figure 2, in which two transmitters wish to send messages to their corresponding receivers over a parallel state-dependent channel. The state is not known to either transmitter or receiver but is non-causally (the side information in all times is given to the encoder before the block transmission) known to a state-cognitive helper, who tries to assist each receiver in mitigating the interference caused by the state. This model captures interference cancellation in various practical scenarios. For example, users in multi-cell systems may be interfered by a base station located in other cells. Such a base station, being as the source that causes the interference, clearly knows the information of the interference (modeled by state) and can serve as a helper to mitigate the interference. Alternatively, that base station can also convey the interference information to other base stations via the backhaul network so that other base stations can serve as helpers to reduce the interference. As another example, consider a situation where there are two Device to Device (D2D) links located in two distinct cells, and there is a downlink signal sent from the base station to some conventional mobile user in the cell. Also, there is some central unit that knows in a non-causal manner the signal to be sent by each base station, the helper in our model, and tries to assist the D2D communication links by mitigating the interference (see Figure 3). As a comparison, this type of state-dependent models differs from the original state-dependent channels studied in, e.g., [2,3], in that the state-cognitive helper is not informed of the transmitters' messages, and hence its state cancellation strategies are necessarily independent of message encoding at the transmitters.

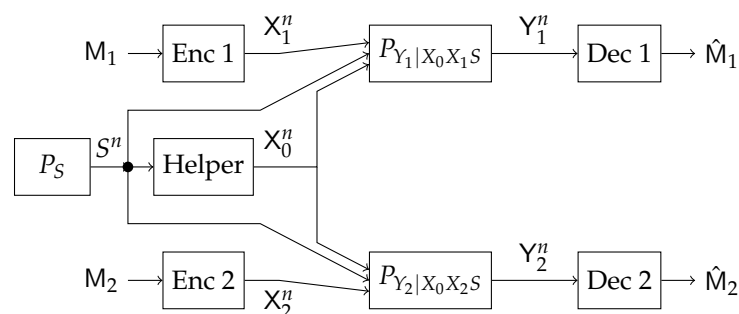


Figure 2. General State-Dependent Parallel Channel with a helper.

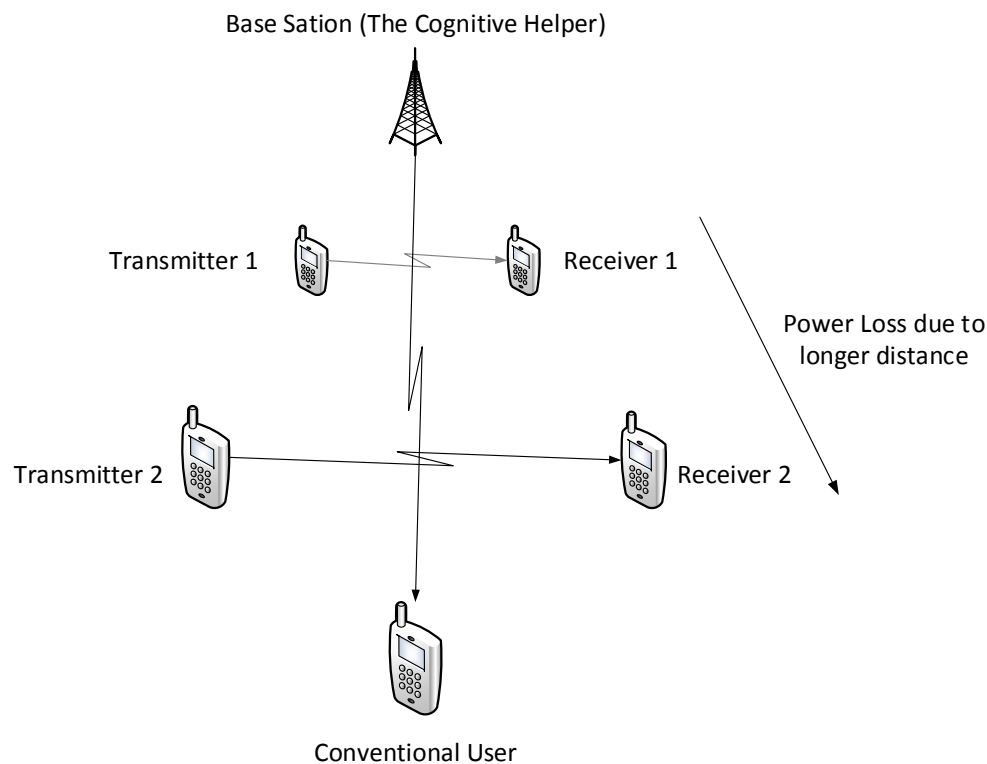


Figure 3. Particular NOMA configuration.

The study of channel coding in the presence of channel side information (CSI) was initiated by Shannon [4] who considered a discrete memoryless channel (DMC) channel with random parameters and side information provided causally to the transmitter. The single-letter expression for the capacity of the point-to-point DMC with non-causal CSI at the encoder (the G-P channel) was derived in the seminal work of Gel'fand and Pinsker [2]. One of the most interesting special cases of the G-P channel is the Gaussian additive noise and interference setting in which the additive interference plays the role of the state sequence, which is known non-causally to the transmitter. Costa showed in [3] that the capacity of this channel is equal to the capacity of the same channel without additive interference. The capacity achieving scheme of [3] (which is that of [2] applied to the Gaussian case) is termed “writing on dirty paper” (WDP), and consequently, the property of the channel where the known interference can be completely removed is dubbed “the WDP property”. Cohen and Lapidot [5] showed that any interference sequence can be removed entirely when the channel noise is ergodic and Gaussian.

The models we study in this work all have a broadcasting node. The *discrete memoryless broadcast channel* (DM-BC) was introduced by Cover [6]. The capacity region of the DM-BC is still an open problem. The largest known inner bound on the capacity region of the DM-BC with private messages was derived by Marton [7]. Liang [8] derived an inner bound on the capacity region of the DM-BC with an additional common message. The best outer bound for DM-BC with a common message is due to Nair and El Gamal [9]. There are, however, some special cases where the capacity region is fully characterized. For example, the capacity region of the degraded DM-BC was established by Gallager [10]. The capacity region of the Gaussian BC was derived by Bergmans [11]. An interesting result is the capacity region of the Gaussian MIMO BC which was established by Weingarten et al. [12]. The authors introduced a new notion of an *enhanced channel* and used it jointly with the Entropy Power Inequality (EPI) to show their result. The capacity achieving scheme relies on the dirty paper coding technique. Liu and Viswanath [13] developed an *extremal inequality* proof technique and showed that it can be used to establish a converse result in various Gaussian MIMO multiterminal networks, including the Gaussian MIMO BC with private messages. Recently, Geng and Nair [14] developed

a different technique to characterize the capacity region of Gaussian MIMO BC with common and private messages.

Degraded DM-BC with causal and non-causal side information was introduced by Steinberg [15]. Inner and outer bounds on the capacity region were derived. For the particular case in which the nondegraded user is informed about the channel parameters, it was shown that the bounds are tight, thus obtaining the capacity region for that case. The general DM-BC with non-causal CSI at the encoder was studied by Steinberg and Shamai [16]. An inner bound was derived, and it was shown to be tight for the Gaussian BC with private messages and independent additive interference at both channels. The latter setting was recently extended to the case of common and private messages in the Gaussian framework with K users in [17]. The special case where the transmitter sends only a common message to all receivers over an additive BC has been initially studied in [18] and has been recently extended to the compound setting in [19]. Outer bounds for DM-BC with CSI at the encoder were derived in [20].

The models addressed in this paper have a mismatched property, that is the state sequence is known only to some nodes, which differs from the classical study on state-dependent channels. The type of channels with mismatched property has been addressed in the past for various models, for example, in [21–25], the state-dependent *multiple access channel* (MAC) is studied with the state known at only one transmitter. The best outer bound for the Gaussian MAC setting was recently reported in [26]. The point-to-point helper channel studied in [27,28] can be considered as a special case of [25], where the cognitive transmitter does not send any message. Further in [28], the state-dependent MAC with an additional helper was studied, and the partial/full capacity region was characterized under various channel parameters. Moreover, some state-dependent relay channel models can also be viewed as an extension of the state-dependent channel with a helper, where the relay serves the role of the helper by knowing the state information. In [29], the state-dependent relay channel with state non-causally available at the relay is considered. An achievable rate was derived using a combination of decode-and-forward, Gel'fand–Pinsker (GP) binning and codeword splitting. Also, in [30], additional noiseless cooperation links with finite capacity were assumed between the transmitter and the relay, and various coding techniques were explored. The authors of [31] have recently considered a different scenario with a state-cognitive relay. The state-dependent Z-IC with a common state known in a non-causal manner only to the primary user was studied in [32]. A good tutorial on channel coding in the presence of CSI can be found in [33].

The basic state-dependent Gaussian channel with a helper is illustrated in Figure 4. It was first introduced in [27], where the capacity in the infinite power regime was characterized and was shown to be achievable by lattice coding. The capacity under arbitrary state power was established for some special cases in [28]. Based on a single-bin GP binning scheme the following lower bound was derived for the discrete memoryless case

$$R \leq \max_{P_{UX_0|S} P_X} \min\{I(X; Y|U), I(UX; Y) - I(U; S)\}.$$

This lower bound was further evaluated for Gaussian channel by appropriate choice of the maximizing input distribution. The surprising result of that study was that when the helper power is above some threshold, then the interference caused by the state is entirely canceled and the capacity of the channel without the state can be achieved. This threshold does not depend on the state power, and hence it was shown that this channel also has WDP property, that is the capacity of the channel is the same as the capacity of the similar channel without the interference (which is modeled as the state).

The most relevant work to this study is [34], in which the state-dependent parallel channel with a helper was studied, for the regime with infinite state power and with two receivers being corrupted by two independent states. A time-sharing scheme was proved to be capacity achieving under certain channel parameters. In contrast, in this study, we expand those results for the arbitrary state power regime. We also consider two extreme cases. At first, we address the problem where the two receivers of the parallel channel are corrupted by the same but differently scaled states, and in the second part,

those states are independent. For both cases, we show that the time-sharing scheme is no longer optimal. Our main contribution in this work is a derivation of inner bound, which is an extension of the Marton coding scheme for the discrete broadcast channel to the current model. We will apply this bound for the MIMO Gaussian setting and characterize the segments of the capacity region for various channel parameters. The material in this paper was presented in part at [35,36].

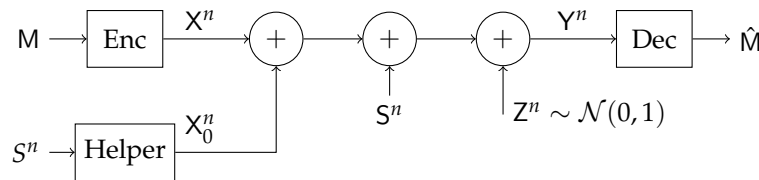


Figure 4. Point-to-Point Helper Channel.

2. Preliminaries

2.1. Notation Conventions

Throughout the paper, random variables are denoted using a sans-serif font, e.g., \mathbf{X} , their realizations are denoted by the respective lower-case letters, e.g., x , and their alphabets are denoted by the respective calligraphic letters, e.g., \mathcal{X} . Let \mathcal{X}^n stand for the set of all n -tuples of elements from \mathcal{X} . An element from \mathcal{X}^n is denoted by $x^n = (x_1, x_2, \dots, x_n)$ and substrings are denoted by $x_i^j = (x_i, x_{i+1}, \dots, x_j)$. The cardinality of a finite set, say \mathcal{X} , is denoted by $|\mathcal{X}|$. The probability distribution function of \mathbf{X} , the joint distribution function of \mathbf{X} and \mathbf{Y} , and the conditional distribution of \mathbf{X} given \mathbf{Y} are denoted by $P_{\mathbf{X}}$, $P_{\mathbf{X},\mathbf{Y}}$ and $P_{\mathbf{X}|\mathbf{Y}}$ respectively. The expectation of \mathbf{X} is denoted by $\mathbb{E}[\mathbf{X}]$. The probability of an event \mathcal{E} is denoted as $P\{\mathcal{E}\}$. The set of jointly ϵ -typical n -tuples (x^n, y^n) is defined as $\mathcal{T}_{\epsilon}^{(n)}(P_{\mathbf{X},\mathbf{Y}})$ [37]. A set of consecutive integers starting at 1 and ending in $\lceil 2^{nR} \rceil$ is denoted as $\mathcal{I}_R^{(n)} \triangleq \{1, 2, \dots, \lceil 2^{nR} \rceil\}$. We assume throughout this paper that 2^{nR} are integers, for any R and $n \rightarrow \infty$.

We denote the covariance of a zero mean vector \mathbf{X} by $\Sigma_{\mathbf{X}} \triangleq \mathbb{E}[\mathbf{X}\mathbf{X}^T]$, $\Sigma_{\mathbf{X}\mathbf{Y}} \triangleq \mathbb{E}[\mathbf{X}\mathbf{Y}^T]$ is the cross-correlation, and the conditional correlation matrix of \mathbf{X} given \mathbf{Y} as $M_{\mathbf{X}|\mathbf{Y}} \triangleq \Sigma_{\mathbf{X}} - \Sigma_{\mathbf{X}\mathbf{Y}}\Sigma_{\mathbf{Y}}^{-1}\Sigma_{\mathbf{Y}\mathbf{X}}$.

2.2. Definitions

Definition 1. Random variables $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ are said to form a Markov chain in that order (denoted by $\mathbf{X} \rightarrow \mathbf{Y} \rightarrow \mathbf{Z}$) if the conditional distribution of \mathbf{Z} depends only on \mathbf{Y} and is conditionally independent of \mathbf{X} . Specifically, \mathbf{X}, \mathbf{Y} and \mathbf{Z} form a Markov chain $\mathbf{X} \rightarrow \mathbf{Y} \rightarrow \mathbf{Z}$ if the joint probability mass function can be written as

$$P_{\mathbf{X}\mathbf{Y}\mathbf{Z}} = P_{\mathbf{X}}P_{\mathbf{Y}|\mathbf{X}}P_{\mathbf{Z}|\mathbf{Y}}. \quad (1)$$

2.3. Auxiliary Results

This section introduces some auxiliary results that are relevant to the analysis in this work [37].

Lemma 1 (Data-processing inequality). If $\mathbf{X} \rightarrow \mathbf{Y} \rightarrow \mathbf{Z}$, then

$$I(\mathbf{X}; \mathbf{Y}) \geq I(\mathbf{X}; \mathbf{Z}). \quad (2)$$

The following inequality will be frequently used in the proofs of outer bounds on the capacity regions.

Lemma 2 (Fano's Inequality). Let $(\mathbf{X}, \mathbf{Y}) \sim P_{\mathbf{X}\mathbf{Y}}$ and $P_e = \Pr(\mathbf{X} \neq \mathbf{Y})$. Then

$$H(\mathbf{X}|\mathbf{Y}) \leq H(P_e) + P_e \log |\mathcal{X}| \leq 1 + P_e \log |\mathcal{X}|. \quad (3)$$

The covering lemma and the packing lemma will be used in the achievability proofs throughout this paper.

Lemma 3 (Covering Lemma). Let $(U, X, \hat{X}) \sim P_{U\hat{X}}$ and $\epsilon' < \epsilon$. Let $(U^n, X^n) \sim P_{U^n X^n}$ be a pair of random sequences with

$$\lim_{n \rightarrow \infty} P\{(U^n, X^n) \in \mathcal{T}_{\epsilon'}^{(n)}(P_{U\hat{X}})\} = 1,$$

and let $\hat{X}^n(m)$, $m \in \mathcal{A}$, where $|\mathcal{A}| \geq 2^{nR}$, be random sequences, conditionally independent of each other and of X^n given U^n , each distributed according to $\prod_{i=1}^n P_{\hat{X}|U}(\hat{x}_i|u_i)$. Then, there exists $\delta(\epsilon)$ that approaches zero as $\epsilon \rightarrow 0$ such that

$$\lim_{n \rightarrow \infty} P\{(U^n, X^n, \hat{X}^n(m)) \notin \mathcal{T}_{\epsilon}^{(n)} \text{ for all } m \in \mathcal{A}\} = 0, \quad (4)$$

if $R > I(X; \hat{X}|U) + \delta(\epsilon)$.

Lemma 4 (Packing Lemma). Let $(U, X, Y) \sim P_{UXY}$. Let $(\tilde{U}^n, \tilde{Y}^n) \sim P_{\tilde{U}^n \tilde{Y}^n}$ be a pair of arbitrarily distributed random sequences, not necessarily distributed according to $\prod_{i=1}^n P_{UY}(\tilde{u}_i, \tilde{y}_i)$. Let $X^n(m)$, $m \in \mathcal{A}$, where $|\mathcal{A}| \leq 2^{nR}$, be random sequences, each distributed according to $\prod_{i=1}^n P_{X|U}(x_i|\tilde{u}_i)$. Further assume that $X^n(m)$, $m \in \mathcal{A}$, is pairwise conditionally independent of \tilde{Y}^n given \tilde{U}^n , but is arbitrarily dependent on other $X^n(m)$ sequences. Then, there exists $\delta(\epsilon)$ that approaches zero as $\epsilon \rightarrow 0$ such that

$$\lim_{n \rightarrow \infty} P\{(\tilde{U}^n, X^n(m), \tilde{Y}^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } m \in \mathcal{A}\} = 0,$$

if $R < I(X; Y|U) - \delta(\epsilon)$.

3. The MIMO Gaussian Channel with Same but Differently Scaled States

3.1. Channel Model

In this section, we study the state-dependent parallel network with a state-cognitive helper, in which two transmitters communicate with two corresponding receivers over a state-dependent parallel channel. The two receivers are corrupted by the same but differently scaled state, respectively. The state information is not known to either the transmitters or the receivers, but a helper non-causally. Hence, the helper assists these receivers to cancel the state interference (see Figure 5).

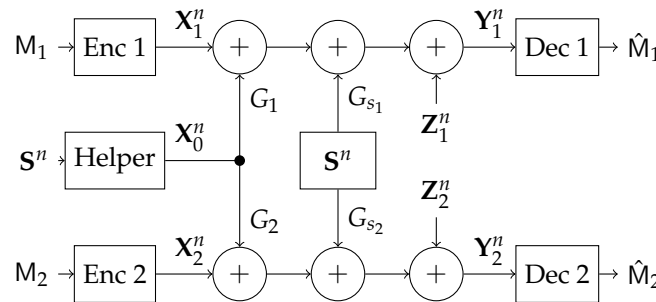


Figure 5. The state-dependent parallel channel with same but differently scaled states and a state-cognitive helper.

More specifically, the encoder at transmitter l , $f_l : \mathcal{I}_{R_l}^{(n)} \rightarrow \mathcal{X}_l^n$, maps a message $m_l \in \mathcal{I}_{R_l}^{(n)}$ to a codeword x_l^n , for $l = 1, 2$. The inputs x_1^n and x_2^n are sent respectively over the two subchannels of the parallel channel. The two receivers are corrupted by the same but differently scaled and identically distributed (i.i.d.) state sequence $s^n \in \mathcal{S}^n$, which is known to a common helper non-causally. Hence, the encoder at the helper, $f_0 : \mathcal{S}^n \rightarrow \mathcal{X}_0^n$, maps the state sequence $s^n \in \mathcal{S}^n$ into a codeword $x_0^n \in \mathcal{X}_0^n$. The channel transition probability is given by $P_{Y_1|X_0X_1S} \cdot P_{Y_2|X_0X_2S}$. The decoder at receiver l , $g_l : \mathcal{Y}_l^n \rightarrow \mathcal{I}_{R_l}^{(n)}$, maps a received sequence y_l^n into a message $\hat{m}_l \in \mathcal{I}_{R_l}^{(n)}$, for $l = 1, 2$. We assume that the

messages are uniformly distributed over the sets $\mathcal{I}_{R_1}^{(n)}$ and $\mathcal{I}_{R_2}^{(n)}$. We define the average probability of error for a length- n code as

$$P_e^{(n)} = \frac{1}{2^{n(R_1+R_2)}} \sum_{m_1=1}^{2^{nR_1}} \sum_{m_2=1}^{2^{nR_2}} P\{(\hat{m}_1, \hat{m}_2) \neq (m_1, m_2)\}. \quad (5)$$

Definition 2. A rate pair (R_1, R_2) is said to be achievable if there exist a sequence of message sets $\mathcal{I}_{R_1}^{(n)}$ and $\mathcal{I}_{R_2}^{(n)}$, and encoder-decoder tuples $(f_0^{(n)}, f_1^{(n)}, f_2^{(n)}, g_1^{(n)}, g_2^{(n)})$ such that the average probability of error $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$.

Definition 3. We define the capacity region of the channel as the closure of the set of all achievable rate pairs (R_1, R_2) .

In this section, we focus on the MIMO Gaussian channel, with the outputs at the two receivers for one channel use given by

$$\mathbf{Y}_l = G_l \mathbf{X}_0 + \mathbf{X}_l + G_{s_l} \mathbf{S} + \mathbf{Z}_l \quad l \in \{1, 2\}, \quad (6)$$

where $\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2, \mathbf{S}_2, \mathbf{Z}_1$ and \mathbf{Z}_2 are all real vectors of size $t \times 1$, and

- $\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2$ are the input vectors that are subject to the covariance matrix constraints $\frac{1}{n} \sum_{i=1}^n \mathbf{x}_{li} \mathbf{x}_{li}^T \preceq K_l, l \in \{0, 1, 2\}$,
- \mathbf{Y}_l is the output vector, $l \in \{1, 2\}$,
- \mathbf{S} is a real Gaussian random vector with zero mean and covariance matrix $K_S = \mathbb{E}[\mathbf{S}\mathbf{S}^T] \succeq 0$,
- \mathbf{Z}_l is a real Gaussian random vector with zero mean and an identity covariance matrix $K_{Z_l} = I$, for $l \in \{1, 2\}$.

Both the noise variables, and the state variable are i.i.d. over channel uses. G_{s_1} (G_{s_2}) is $t \times t$ real matrix that represents the channel matrix connecting the state source to the first (second) user. Similarly, G_1 (G_2) is a $t \times t$ real channel matrix connecting the helper to the first (second) user. Thus, our model captures a general scenario, where the helper's power and the state power can be arbitrary.

Our goal is to characterize the capacity region of the Gaussian channel under various channel parameters $(G_1, G_2, G_{s_1}, G_{s_2}, K_0, K_1, K_2, K_S)$.

3.2. Inner and Outer Bounds

In this section, we first derive inner and outer bounds on the capacity region for the state-dependent parallel channel with a helper. Then by comparing the inner and outer bounds, we characterize the segments on the capacity region boundary under various channel parameters.

We start by deriving an inner bound on the capacity region for the DMC based on the single-bin GP scheme.

Proposition 1. For the discrete memoryless state-dependent parallel channel with a helper under the same but differently scaled states at the two receivers, an inner bound on the capacity region consists of rate pairs (R_1, R_2) satisfying:

$$R_1 \leq \min \{I(\mathbf{W}, \mathbf{X}_1; \mathbf{Y}_1) - I(\mathbf{W}; \mathbf{S}), I(\mathbf{X}_1; \mathbf{Y}_1 | \mathbf{W})\}, \quad (7a)$$

$$R_2 \leq \min \{I(\mathbf{W}, \mathbf{X}_2; \mathbf{Y}_2) - I(\mathbf{W}; \mathbf{S}), I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{W})\}, \quad (7b)$$

for some distribution $P_{\mathbf{W}|\mathbf{S}} P_{\mathbf{X}_0|\mathbf{W}\mathbf{S}} P_{\mathbf{X}_1} P_{\mathbf{X}_2}$.

Proof. The proof is relegated to Appendix A. \square

We evaluate the inner bound for the Gaussian channel by choosing the joint Gaussian distribution for random variables as follows:

$$\begin{aligned}\mathbf{W} &= \mathbf{X}'_0 + \mathbf{A}\mathbf{S}, \\ \mathbf{X}_0 &= \mathbf{X}'_0 + \mathbf{B}\mathbf{S}, \\ \mathbf{X}'_0 &\sim \mathcal{N}(\mathbf{0}, K'_0), \\ \mathbf{X}_1 &\sim \mathcal{N}(\mathbf{0}, K_1) \quad \mathbf{X}_2 \sim \mathcal{N}(\mathbf{0}, K_2),\end{aligned}\tag{8}$$

where $\mathbf{X}'_0, \mathbf{X}_1, \mathbf{X}_2, \mathbf{S}$ are independent and $K'_0 \preceq K_0$.

Let $f_1(\cdot), g_1(\cdot), f_2(\cdot)$ and $g_2(\cdot)$ be defined as

$$\begin{aligned}f_1(A, B, K'_0) &= I(\mathbf{W}, \mathbf{X}_1; \mathbf{Y}_1) - I(\mathbf{W}; \mathbf{S}), \\ g_1(A, B, K'_0) &= I(\mathbf{X}_1; \mathbf{Y}_1 | \mathbf{W}), \\ f_2(A, B, K'_0) &= I(\mathbf{W}, \mathbf{X}_2; \mathbf{Y}_2) - I(\mathbf{W}; \mathbf{S}), \\ g_2(A, B, K'_0) &= I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{W}),\end{aligned}$$

where the mutual information terms are evaluated using the joint Gaussian distribution chosen in (8). Based on those definitions, we obtain an achievable region for the Gaussian channel.

Proposition 2. *An inner bound on the capacity region of the parallel state-dependent MIMO Gaussian channel with same but differently scaled states and a state-cognitive helper consists of rate pairs (R_1, R_2) satisfying:*

$$R_1 \leq \min\{f_1(A, B, K'_0), g_1(A, B, K'_0)\},\tag{9a}$$

$$R_2 \leq \min\{f_2(A, B, K'_0), g_2(A, B, K'_0)\},\tag{9b}$$

for some real matrices A, B and K'_0 satisfying $K'_0 \succeq 0, K'_0 + BK_5B^T \preceq K_0$.

We note that the above choice of the helper's signal incorporates two parts with \mathbf{X}'_0 designed using single-bin dirty paper coding, and $\mathbf{B}\mathbf{S}$ acting as direct state subtraction.

We next present an outer bound which applies the point-to-point channel capacity and the upper bound derived for the point-to-point channel with a helper in [27].

Denote

$$\begin{aligned}R_l^{\text{ub1}}(\Sigma_{\mathbf{X}_0\mathbf{S}}) &\triangleq \frac{1}{2} \log \left(\frac{|G_l K_0 G_l^T + K_l + G_l \Sigma_{\mathbf{X}_0\mathbf{S}} G_{s_l} + G_{s_l} \Sigma_{\mathbf{X}_0\mathbf{S}}^T G_l^T + G_{s_l} K_{\mathbf{S}} G_{s_l}^T + I|}{|G_l K_0 G_l^T + G_l \Sigma_{\mathbf{X}_0\mathbf{S}} G_{s_l}^T + G_{s_l} \Sigma_{\mathbf{X}_0\mathbf{S}}^T G_l^T + G_{s_l} K_{\mathbf{S}} G_{s_l}^T + I|} \right) \\ &\quad + \frac{1}{2} \log \left(|G_l (K_0 - \Sigma_{\mathbf{X}_0\mathbf{S}} K_{\mathbf{S}}^{-1} \Sigma_{\mathbf{X}_0\mathbf{S}}^T) G_l^T + I| \right).\end{aligned}\tag{10}$$

Proposition 3. *An outer bound on the capacity region of the state-dependent parallel MIMO Gaussian channel with a helper consists of rate pairs (R_1, R_2) satisfying:*

$$R_l \leq \min \left\{ R_l^{\text{ub1}}(\Sigma_{\mathbf{X}_0\mathbf{S}}), \frac{1}{2} \log (|K_l + I|) \right\},\tag{11}$$

for every $l \in \{1, 2\}$ and $\Sigma_{\mathbf{X}_0\mathbf{S}}$ that satisfies $\Sigma_{\mathbf{X}_0\mathbf{S}} K_{\mathbf{S}}^{-1} \Sigma_{\mathbf{X}_0\mathbf{S}}^T \preceq K_0$.

Proof. The second term in (11) is simply the capacity of a point-to-point channel without state. The first term is derived in Appendix B. \square

3.3. Capacity Region Characterization

In this section, we optimize A and B in Proposition 2, and compare the rate bounds with the outer bounds in Proposition 3 to characterize the points or segments on the capacity region boundary.

Since the inner bound in Proposition 2 is not convex, it is difficult to provide a closed form for the jointly optimized bounds. Therefore, we first optimize the bounds for R_1 and R_2 respectively, and then provide conditions on channel parameters such that these bounds match the outer bound. Based on the conditions, we partition the channel parameters into the sets, in which different segments of the capacity region boundary can be obtained.

We first consider the rate bound for R_1 in (9a). By setting

$$A^a \triangleq (G_1 K'_0 G_1^T + I)^{-1} K'_0 G_1^T (G_1 B + I), \quad B^a \triangleq \Sigma_{\mathbf{x}_0 \mathbf{s}}^* G_{s_1} K_{\mathbf{s}}^{-1},$$

$f_1(A, B, K'_0)$ takes the following form

$$f_1(A^a, B^a, K'_0) = \frac{1}{2} \log \left(\frac{|G_1 K_0 G_1^T + K_1 + G_1 \Sigma_{\mathbf{x}_0 \mathbf{s}}^* G_{s_1} + G_{s_1}^T \Sigma_{\mathbf{x}_0 \mathbf{s}}^* G_1^T + G_{s_1} K_{\mathbf{s}} G_{s_1}^T + I|}{|G_1 K_0 G_1^T + G_1 \Sigma_{\mathbf{x}_0 \mathbf{s}}^* G_{s_1} + G_{s_1}^T \Sigma_{\mathbf{x}_0 \mathbf{s}}^* G_1^T + G_{s_1} K_{\mathbf{s}} G_{s_1}^T + I|} \right) \\ + \frac{1}{2} \log (|G_1 K'_0 G_1^T + I|),$$

where $\Sigma_{\mathbf{x}_0 \mathbf{s}}^*$ maximizes $f_1(A^a, B(\Sigma_{\mathbf{x}_0 \mathbf{s}}), K'_0)$. In fact, A^a maximizes $f_1(A, B, K'_0)$ for fixed B , and B^a maximizes the function with $A = A^a$.

If $f_1(A^a, B^a, K'_0) \leq g_1(A^a, B^a, K'_0)$, $R_1 = f_1(A^a, B^a, K'_0)$ is achievable, and this matches the outer bound in (11). Thus, one segment of the capacity region is specified by

$$R_1 = f_1(A^a, B^a, K'_0), \quad (12a)$$

$$R_2 \leq \min\{f_2(A^a, B^a, K'_0), g_2(A^a, B^a, K'_0)\}. \quad (12b)$$

We further observe that the second term $g_1(A, B, K'_0)$ in (9a) is optimized by setting $A^b = B + G_1^{-1} G_{s_1}$, and hence

$$g_1(A^b, B, K'_0) = \frac{1}{2} \log(|K_1 + I|).$$

If $g_1(B - G_1^{-1} G_{s_1}, B, K'_0) \leq f_1(B - G_1^{-1} G_{s_1}, B, K'_0)$, i.e.,

$$K'_0 G_1 K'_0 G_1^T \succeq A K_{\mathbf{s}} A^T (K_1 + I) - K'_0 G_1 A K_{\mathbf{s}} A^T G_1^T, \quad (13)$$

then the inner bound for R_1 becomes $R_1 = \frac{1}{2} \log(|K_1 + I|)$, which is the capacity of the point-to-point channel without state and matches the outer bound in (11). Thus, another segment of the capacity is specified by

$$R_1 = \frac{1}{2} \log(|K_1 + I|), \quad (14a)$$

$$R_2 \leq \min\{f_2(A^b, B, K'_0), g_2(A^b, B, K'_0)\}. \quad (14b)$$

We then consider the rate bound for R_2 . Similarly, the following segments on the capacity boundary can be obtained. If $f_2(A^c, B^c, K'_0) \leq g_2(A^c, B^c, K'_0)$, one segment of the capacity region boundary is specified by

$$R_1 \leq \min\{f_1(A^c, B^c, K'_0), g_1(A^c, B^c, K'_0)\}, \quad (15a)$$

$$R_2 = \frac{1}{2} \log \left(\frac{|G_2 K_0 G_2^T + K_2 + G_2 \Sigma_{X_0 S}^* G_{s_2} + G_{s_2}^T \Sigma_{X_0 S}^* G_2^T + G_{s_2} K_S G_{s_2}^T + I|}{|G_2 K_0 G_2^T + G_2 \Sigma_{X_0 S}^* G_{s_2} + G_{s_2}^T \Sigma_{X_0 S}^* G_2^T + G_{s_2} K_S G_{s_2}^T + I|} \right) + \frac{1}{2} \log (|G_2 K'_0 G_2^T + I|), \quad (15b)$$

where

$$A^c \triangleq (G_2 K'_0 G_2^T + I)^{-1} K'_0 G_2^T (G_2 B + G_S), \quad B^c \triangleq \Sigma_{X_0 S}^{**} G_{s_1} K_S^{-1},$$

and $\Sigma_{X_0 S}^{**}$ maximizes $f_2(A^c, B^c, K'_0)$.

Furthermore, if $g_2(A, A - G_2^{-1} G_{s_2}, K'_0) \leq f_2(A, A - G_2^{-1} G_{s_2}, K'_0)$, one segment of the capacity region boundary is specified by

$$R_1 \leq \min \left\{ f_1(A, A - G_2^{-1} G_{s_2}, K'_0), g_1(A, A - G_2^{-1} G_{s_2}, K'_0) \right\} \quad (16a)$$

$$R_2 = \frac{1}{2} \log (|K_2 + I|). \quad (16b)$$

Appendix C describes how (A^a, B^a) , (A^b, B^b) , (A^c, B^c) and (A^d, B^d) were chosen.

Summarizing the above analysis, we obtain the following characterization of segments of the capacity region boundary.

Theorem 1. The channel parameters $(G_1, G_2, G_{s_1}, G_{s_2}, K_0, K_1, K_2, K_S)$ can be partitioned into the sets $\mathcal{A}_1, \mathcal{B}_1, \mathcal{C}_1$, where

$$\begin{aligned} \mathcal{A}_1 &= \{(G_1, G_2, G_{s_1}, G_{s_2}, K_0, K_1, K_2, K_S) : f_1(A^a, B^a, K'_0) \leq g_1(A^a, B^a, K'_0)\}, \\ \mathcal{C}_1 &= \{(G_1, G_2, G_{s_1}, G_{s_2}, K_0, K_1, K_2, K_S) : K'_0 G_1 K'_0 G_1^T \succeq A K_S A^T (K_1 + I) - K'_0 G_1 A K_S A^T G_1^T \\ &\quad \text{where } K'_0 = K_0 - (A - G_1^{-1} G_{s_1}) K_S (A - G_1^{-1} G_{s_1})^T, \text{ for some } A \in \Omega_A\}, \\ \mathcal{B}_1 &= (\mathcal{A}_1 \cup \mathcal{C}_1)^c. \end{aligned}$$

If $(G_1, G_2, G_{s_1}, G_{s_2}, K_0, K_1, K_2, K_S) \in \mathcal{A}_1$, then (12a)–(12b) captures one segment of the capacity region boundary, where the state cannot be fully canceled. If $(G_1, G_2, G_{s_1}, G_{s_2}, K_0, K_1, K_2, K_S) \in \mathcal{C}_1$, then (14a)–(14b) captures one segment of the capacity region boundary where the state is fully canceled. If $(G_1, G_2, G_{s_1}, G_{s_2}, K_0, K_1, K_2, K_S) \in \mathcal{B}_1$, then the R_1 segment of the capacity region boundary is not characterized.

The channel parameters $(G_1, G_2, G_{s_1}, G_{s_2}, K_0, K_1, K_2, K_S)$ can also be partitioned into the sets $\mathcal{A}_2, \mathcal{B}_2, \mathcal{C}_2$, where

$$\begin{aligned} \mathcal{A}_2 &= \{(G_1, G_2, G_{s_1}, G_{s_2}, K_0, K_1, K_2, K_S) : f_2(A^c, B^c, K'_0) \leq g_2(A^c, B^c, K'_0)\} \\ \mathcal{C}_2 &= \{(G_1, G_2, G_{s_1}, G_{s_2}, K_0, K_1, K_2, K_S) : K'_0 G_2 K'_0 G_2^T \succeq A K_S A^T (K_2 + I) - K'_0 G_2 A K_S A^T G_2^T \\ &\quad \text{where } K'_0 = K_0 - (A - G_2^{-1} G_{s_2}) K_S (A - G_2^{-1} G_{s_2})^T, \text{ for some } A \in \Omega_A\} \\ \mathcal{B}_2 &= (\mathcal{A}_2 \cup \mathcal{C}_2)^c. \end{aligned}$$

If $(G_1, G_2, G_{s_1}, G_{s_2}, K_0, K_1, K_2, K_S) \in \mathcal{A}_2$, then (15a)–(15b) captures one segment of the capacity region boundary, where the state cannot be fully canceled. If $(G_1, G_2, G_{s_1}, G_{s_2}, K_0, K_1, K_2, K_S) \in \mathcal{C}_2$, then (16a)–(16b) captures one segment of the capacity boundary where the state is fully canceled. If $(G_1, G_2, G_{s_1}, G_{s_2}, K_0, K_1, K_2, K_S) \in \mathcal{B}_2$, then the R_2 segment of the capacity region boundary is not characterized.

The above theorem describes two partitions of the channel parameters, respectively under which segments on the capacity region boundary corresponding to R_1 and R_2 can be characterized. Intersection of two sets, each from one partition, collectively characterizes the entire segments on the capacity region boundary.

Figure 6 lists all possible intersection of sets that the channel parameters can belong to. For each case in Figure 6, we use red solid line to represent the segments on the capacity region that are characterized in Theorem 1, and we also mark the value of the capacity that each segment corresponds to as characterized in Theorem 1. Please note that the case $\mathcal{B}_1 \cap \mathcal{B}_2$ is not illustrated in Figure 6 since no segments are characterized in this case.

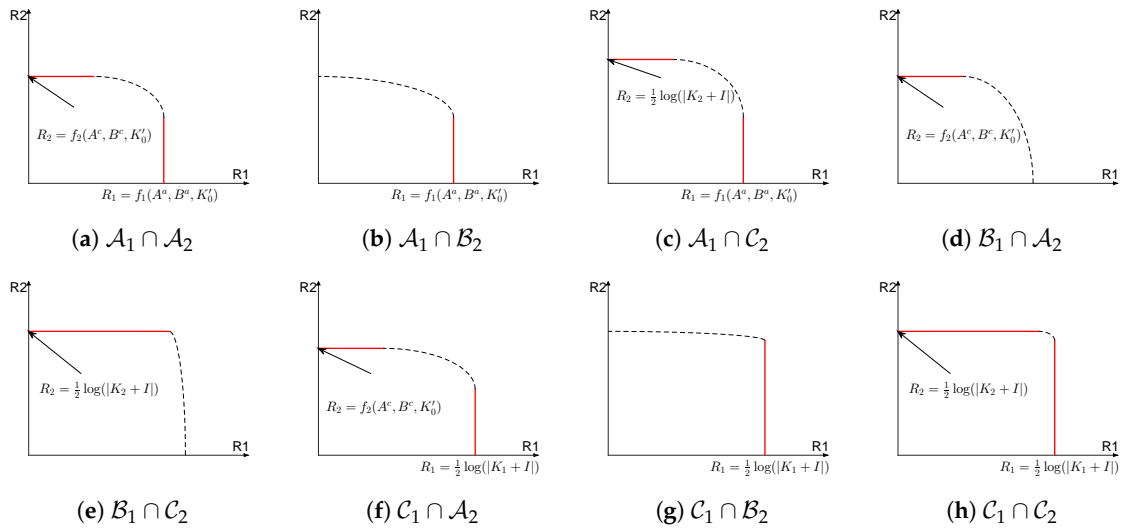


Figure 6. Segments of the capacity region for all cases of channel parameters.

One interesting example in Theorem 1 is the case with $G_1^{-1}G_{s1} = G_2^{-1}G_{s2}$, in which R_1 and R_2 are optimized with the same set of coefficients A and B when $(G_1, G_2, G_{s1}, G_{s2}, K_0, K_1, K_2, K_S) \in \mathcal{C}_1 \cap \mathcal{C}_2$. Thus, the point-to-point channel capacity is simultaneously obtained for both R_1 and R_2 , with state being fully canceled. We state this result in the following theorem.

Theorem 2. If $G_1^{-1}G_{s1} = G_2^{-1}G_{s2}$, $K_0'G_1K_0'G_1^T \succeq AK_S A^T(K_1 + I) - K_0'G_1AK_S A^T G_1^T$ and $K_0'G_2K_0'G_2^T \succeq AK_S A^T(K_2 + I) - K_0'G_2AK_S A^T G_2^T$ where $K_0' = K_0 - (A - G_1^{-1}G_{s1})K_S(A - G_1^{-1}G_{s1})^T$, for some $A \in \Omega_A$ then the capacity region of the state-dependent parallel Gaussian channel with a helper and under the same but differently scaled states contains (R_1, R_2) satisfying

$$\begin{aligned} R_1 &\leq 0.5 \log(|K_1 + I|), \\ R_2 &\leq 0.5 \log(|K_2 + I|). \end{aligned}$$

The channel conditions of Theorem 2 are not just of mathematical importance but also have a practical utility. Consider, for example, a scenario where the helper is also the interferer (see Figure 3), in such case it is reasonable to assume that $G_{s1} = G_1$ and $G_{s2} = G_2$, and thus the aforementioned conditions are satisfied.

3.4. Numerical Example

We now examine our results via simulations. In particular, we focus on the scalar channel case, i.e., $G_1 \leftarrow 1$, $G_2 \leftarrow b$, $G_{s1} \leftarrow 1$, $G_{s2} \leftarrow a$, $K_0 \leftarrow P_0$, $K_0' \leftarrow P_0'$, $K_1 \leftarrow P_1$, $K_2 \leftarrow P_2$ and $K_S \leftarrow Q$. Furthermore, we denote $A \leftarrow \alpha$, $B \leftarrow \beta$ and $\rho_{0S} \triangleq \beta\sqrt{P_0Q}$.

We set $P_0 = 6$, $P_1 = P_2 = 5$, $Q = 12$, and $b = 0.8$, and plot the inner and outer bounds for the capacity region (R_1, R_2) for two values of a . It can be observed from Figure 7 that the upper bound is defined by the rectangular region of channel without state. The inner bound, in the contrary, is susceptible to the value of a , such that in the case where $a = b$, our inner and outer bounds

coincide everywhere, while in the case $a \neq b$ they coincide only on some segments. Both observations corroborate the characterization of the capacity in Theorems 1 and 2.

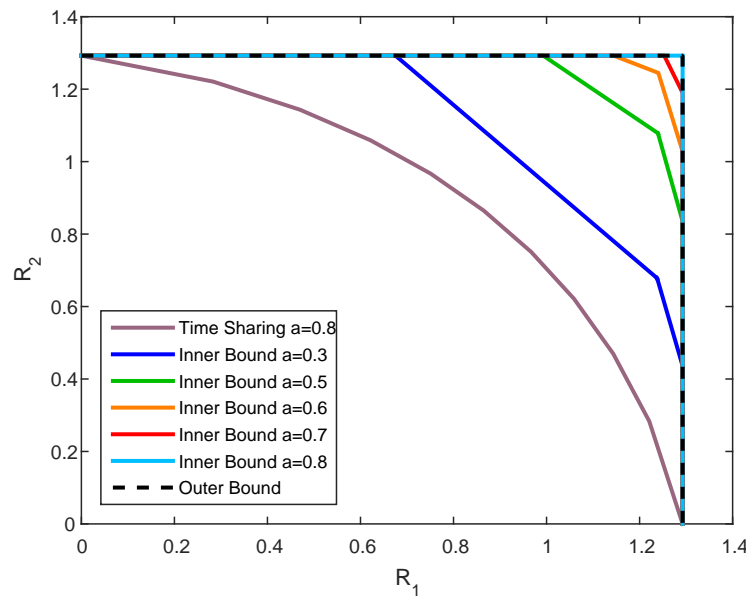


Figure 7. Capacity bounds for channel parameters $P_0 = 6, P_1 = P_2 = 5, Q = 12, b = 0.8$ and various state gain a .

It is also interesting to illustrate how the channel parameters (a, b) affect our ability to characterize the capacity region boundary. For this we propose the following setup:

- we choose α and β such that R_1 lies on the capacity region boundary;
- we further choose ρ_{0S} that maximizes the achievable R_2 , denoted as R_2^I ;
- we compare it to the outer bound of R_2 , R_2^O , and plot the gap $\Delta \triangleq R_2^O - R_2^I$.

Figure 8 shows the results of such simulation for two values of P_0 : $P_0 = 1$ for which the state is not fully canceled for user 1 and $P_0 = 6$, for which the state is canceled. We fix other parameters as before, that is $P_1 = P_2 = 5$ and $Q = 12$. The right figure shows that the capacity gap is small around the line $a = b$, this result is not surprising, and it appears in Theorem 2. The left Figure is also interesting. It shows that there is a curve $a \neq b$ for which the capacity gap is also near zero. The reason for this phenomenon is explained as follows.

- The chosen channel parameters satisfy $(a, b, P_0, P_1, P_2, Q) \in \mathcal{A}_1$, and hence

$$\alpha_1 = \frac{(1 + \beta_1)P'_0}{P'_0 + 1} \quad \beta_1 = \rho_{0S}^* \sqrt{\frac{P_0}{Q}} \quad (17)$$

- optimize R_1 .
- Thus, if $\frac{a}{b}$ satisfies

$$\frac{a}{b} = \alpha_1 - \beta_1, \quad (18)$$

and $b^2 P_0'^2 \geq \alpha_1^2 Q (P_2 + 1 - b^2 P_0')$, then $(a, b, P_0, P_1, P_2, Q) \in \mathcal{C}_2$, i.e., $R_2 = \frac{1}{2} \log(1 + P_2)$ is achievable.

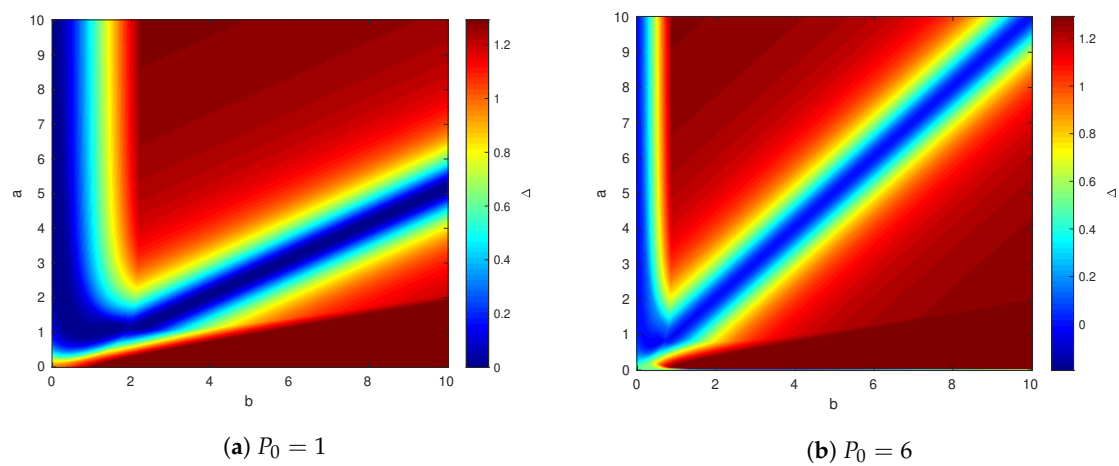


Figure 8. Capacity gap for fixed P_0 .

We illustrate this result in Figure 9, where we fixed the channel parameters $b = 1$, $P_1 = P_2 = 5$, $Q = 12$, and calculate the capacity gap for various values of a and P_0 . The shaded area is the region of P_0 where the capacity of the point-to-point helper channel is not characterized.

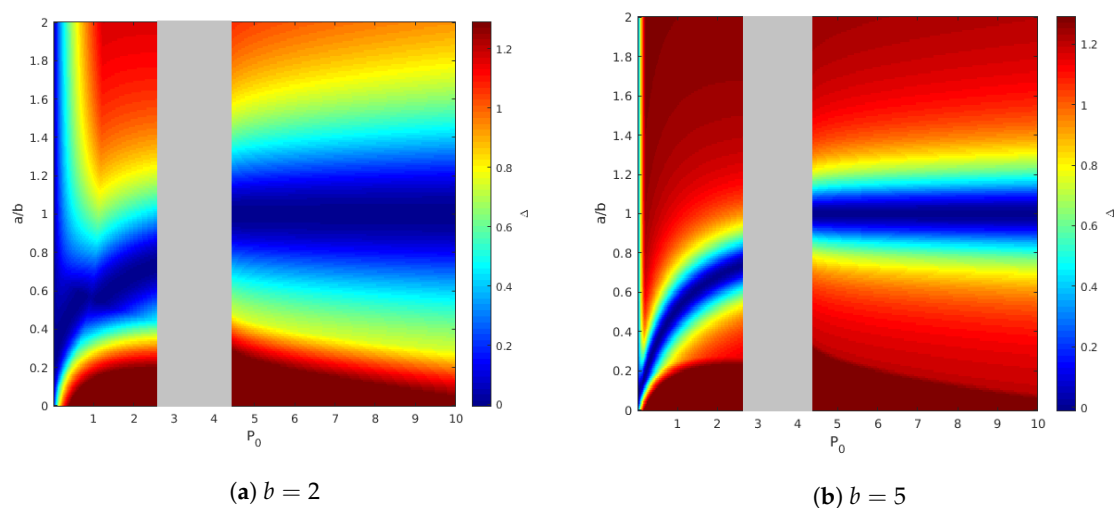


Figure 9. Capacity gap for fixed b .

In practical situations the channel parameters a and b are fixed but the helper can control P_0 . The results here imply that for a fixed (a, b) we can choose P_0 such that the capacity gap is close to zero. We emphasize this in Figure 10, where we plot the inner and outer bounds on achievable (R_1, R_2) with the following channel parameters

$$(a, b, P_0, P_1, P_2, Q) = (3.5, 5, 2.17, 5, 5, 12). \quad (19)$$

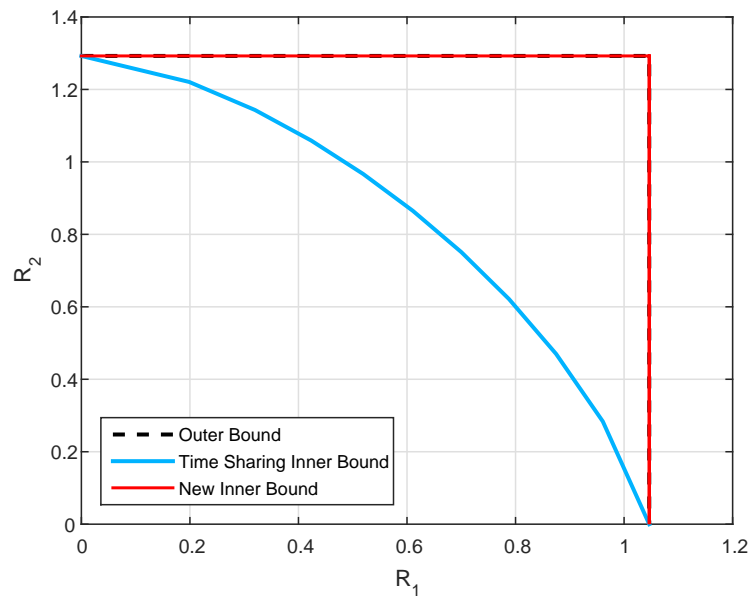


Figure 10. Inner and outer bounds for $(a, b, P_0, P_1, P_2, Q) = (3.5, 5, 2, 5, 5, 12)$.

4. MIMO Gaussian Channel with Independent States

In this section, we consider the problem of channel coding over MIMO Gaussian parallel state-dependent channel with a cognitive helper where the states are independent. We start with deriving an achievable region for a general discrete memoryless case. We then, evaluate this region for the Gaussian setting by choosing an appropriate jointly Gaussian input distribution.

4.1. Problem Formulation

Consider a 3-transmitter, 2-receiver *state-dependent parallel DMC* depicted in Figure 11, where Transmitter 1 wishes to communicate a message M_1 to Receiver 1, and similarly Transmitter 2 wishes to transmit a message M_2 to its corresponding Receiver 2. The messages M_1 and M_2 are independent. The communication takes over a parallel state-dependent channel characterized by a probability transition matrix $p(y_1, y_2 | x_0, x_1, x_2, s)$. The transmitter at the helper has non-causal knowledge of the state and tries to mitigate the interference caused in both channels. The state variable S is random taking values in \mathcal{S} and drawn from a discrete memoryless source (DMS)

$$P_{S^n}(s^n) = \prod_{i=1}^n P_S(s_i).$$

A $(2^{nR_1}, 2^{nR_2}, n)$ code for the parallel state-dependent channel with state known non-causally at the helper consists of

- two message sets $\mathcal{I}_{R_1}^{(n)}$ and $\mathcal{I}_{R_2}^{(n)}$,
- three encoders, where the encoder at the helper assigns a codeword $x_0^n(s^n)$ to each state sequence $s^n \in \mathcal{S}^n$, encoder 1 assigns a codeword $x_1^n(m_1)$ to each message $m_1 \in \mathcal{I}_{R_1}^{(n)}$ and encoder 2 assigns a codeword $x_2^n(m_2)$ to each message $m_2 \in \mathcal{I}_{R_2}^{(n)}$, and
- two decoders, where decoder 1 assigns an estimate $\hat{m}_1 \in \mathcal{I}_{R_1}^{(n)}$ or an error message e to each received sequence y_1^n , and decoder 2 assigns an estimate $\hat{m}_2 \in \mathcal{I}_{R_2}^{(n)}$ or an error message e to each received sequence y_2^n .

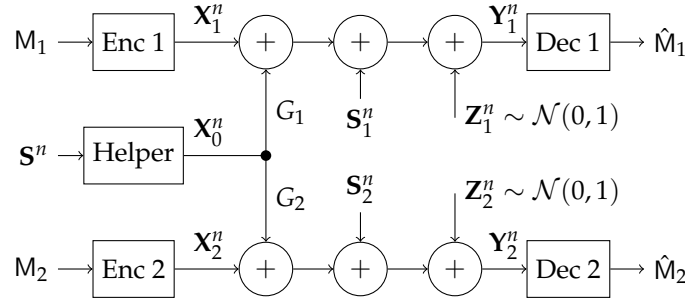


Figure 11. MIMO State-Dependent Parallel Channel with a Helper.

We assume that the message pair (M_1, M_2) is uniformly distributed over $\mathcal{I}_{R_1}^{(n)} \times \mathcal{I}_{R_2}^{(n)}$. The average probability of error for a length- n code is defined as

$$P_e^{(n)} = \mathbb{P}\{\hat{M}_1 \neq M_1 \text{ or } \hat{M}_2 \neq M_2\}. \quad (20)$$

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 $\lim_{n \rightarrow \infty} P_e^{(n)} = 0$. The capacity region \mathcal{C} is the closure of the set of all achievable rate pairs (R_1, R_2) .

We observe that due to the lack of cooperation between the receivers, the capacity region of this channel depends on the $p(y_1, y_2 | x_0, x_1, x_2, s)$ only through the conditional marginal PMFs $p(y_1 | x_0, x_1, s)$ and $p(y_2 | x_0, x_2, s)$. This observation is similar to the DM-BC ([37], Lemma 5.1).

Our goal is to characterize the capacity region \mathcal{C} for the state-dependent Gaussian parallel channel with additive state known at the helper. Here, the state $\mathbf{S} = (\mathbf{S}_1, \mathbf{S}_2)^T$. The channel is modeled by a Gaussian vector parallel state-dependent channel

$$\mathbf{Y}_l = G_l \mathbf{X}_0 + \mathbf{X}_l + \mathbf{S}_l + \mathbf{Z}_l, \quad l = 1, 2, \quad (21)$$

where G_1, G_2 are $t \times t$ channel gain matrices. $\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2$ are the helper and the noncognitive transmitters channel input signals, each subject to an average matrix power constraint

$$\frac{1}{n} \sum_{i=1}^n \mathbf{X}_{l,i} \mathbf{X}_{l,i}^T \preceq K_l, \quad l = 0, 1, 2. \quad (22)$$

The additive state variables \mathbf{S}_l and noise \mathbf{Z}_l are independent and identically distributed (i.i.d.) Gaussian with zero mean and strictly positive definite covariance matrix $K_{\mathbf{S}_l}$ and I respectively.

4.2. Outer and Inner Bounds

To characterize the capacity region of this channel, we first consider the following outer bound on the capacity region for the Gaussian setting.

Let,

$$K_{\mathbf{S}} \triangleq \begin{pmatrix} K_{\mathbf{S}_1} & 0 \\ 0 & K_{\mathbf{S}_2} \end{pmatrix}, \quad (23)$$

and

$$R_l^{\text{ub}_2}(\Sigma_{\mathbf{X}_0} \mathbf{S}) \triangleq \frac{1}{2} \log \left(\frac{|G_l K_0 G_l^T + K_l + G_l \Sigma_{\mathbf{X}_0} \mathbf{S}_l + \Sigma_{\mathbf{X}_0}^T \mathbf{S}_l G_l^T + K_{\mathbf{S}_l} + I|}{|G_l K_0 G_l^T + G_l \Sigma_{\mathbf{X}_0} \mathbf{S}_l + \Sigma_{\mathbf{X}_0}^T \mathbf{S}_l G_l^T + K_{\mathbf{S}_l} + I|} \right) + \frac{1}{2} \log \left(|G_l (K_0 - \Sigma_{\mathbf{X}_0} \mathbf{S} K_{\mathbf{S}}^{-1} \Sigma_{\mathbf{X}_0}^T) G_l^T + I| \right). \quad (24)$$

Proposition 4. Every achievable rate pair (R_1, R_2) of the state-dependent parallel Gaussian channel with a helper must satisfy the following inequalities

$$R_l \leq \min \left\{ R_l^{ub_2}(\Sigma_{\mathbf{x}_0 \mathbf{s}}), \frac{1}{2} \log(|K_l + I|) \right\}, \quad (25)$$

for $l = \{1, 2\}$ and some covariance matrices $(\Sigma_{\mathbf{x} \mathbf{s}_1}, \Sigma_{\mathbf{x} \mathbf{s}_2})$, such that $\Sigma_{\mathbf{x}_0 \mathbf{s}} K_{\mathbf{s}}^{-1} \Sigma_{\mathbf{x}_0 \mathbf{s}}^T \preceq K_0$, where

$$\Sigma_{\mathbf{x}_0 \mathbf{s}} \triangleq \begin{pmatrix} \Sigma_{\mathbf{x}_0 \mathbf{s}_1} & \Sigma_{\mathbf{x}_0 \mathbf{s}_2} \end{pmatrix}. \quad (26)$$

The proof of this outer bound is quite similar to the proof of the outer bound in Proposition 3 and is given in Appendix D.

The upper bound for each rate consists of two terms, the first one reflects the scenario when the interference cannot be completely canceled, and the second is simply the point-to-point capacity of the channel without the state. Furthermore, the individual rate bounds are connected through the choice of $\Sigma_{\mathbf{x}_0 \mathbf{s}_1}$ and $\Sigma_{\mathbf{x}_0 \mathbf{s}_2}$.

We next derive an achievable region for the channel based on an achievable scheme that integrates Marton's coding, single-bin dirty paper coding, and state cancelation. More specifically, we generate two auxiliary random variables, \mathbf{U} and \mathbf{V} to incorporate the state information so that Receiver 1 (and respectively 2) decodes \mathbf{U} (and respectively \mathbf{V}) and then decodes the respective transmitter information. Based on such an achievable scheme, we derive the following inner bound on the capacity region for the DM case.

Proposition 5. An inner bound on the capacity region of the discrete memoryless parallel state-dependent channel with a helper consists of rate pairs (R_1, R_2) satisfying:

$$R_1 \leq \min\{I(\mathbf{U}, \mathbf{X}_1; \mathbf{Y}_1) - I(\mathbf{U}; \mathbf{S}), I(\mathbf{X}_1; \mathbf{Y}_1 | \mathbf{U})\}, \quad (27a)$$

$$R_2 \leq \min\{I(\mathbf{V}, \mathbf{X}_2; \mathbf{Y}_2) - I(\mathbf{V}; \mathbf{S}), I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{V})\}, \quad (27b)$$

$$R_1 + R_2 \leq \min\{I(\mathbf{U}, \mathbf{X}_1; \mathbf{Y}_1) - I(\mathbf{U}; \mathbf{S}) + I(\mathbf{V}, \mathbf{X}_2; \mathbf{Y}_2) - I(\mathbf{V}; \mathbf{S}) - I(\mathbf{V}; \mathbf{U} | \mathbf{S}), \\ I(\mathbf{X}_1; \mathbf{Y}_1 | \mathbf{U}) + I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{V})\}, \quad (27c)$$

for some PMF $P_{\mathbf{U} \mathbf{V} \mathbf{X}_0 | \mathbf{S}} P_{\mathbf{X}_1} P_{\mathbf{X}_2}$.

Remark 1. The achievable region in Proposition 5 is equivalent to the following region

$$R_1 \leq \min\{I(\mathbf{U}, \mathbf{X}_1; \mathbf{Y}_1) - I(\mathbf{U}; \mathbf{S}), I(\mathbf{X}_1; \mathbf{Y}_1 | \mathbf{U})\}, \quad (28a)$$

$$R_2 \leq \min\{I(\mathbf{V}, \mathbf{X}_2; \mathbf{Y}_2) - I(\mathbf{V}; \mathbf{U}, \mathbf{S}), I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{V})\}, \quad (28b)$$

for some PMF $P_{\mathbf{U} \mathbf{V} \mathbf{X}_0 | \mathbf{S}} P_{\mathbf{X}_1} P_{\mathbf{X}_2}$.

Proof. The proof of the inner bound is relegated to Appendix E. \square

We evaluate the latter inner bound for the Gaussian channel by choosing the joint Gaussian distribution for random variables as follows:

$$\begin{aligned}
 \mathbf{U} &= \mathbf{X}'_{01} + A_{11}\mathbf{S}_1 + A_{12}\mathbf{S}_2, \\
 \mathbf{V} &= \mathbf{X}'_{02} + A_{20}\mathbf{X}'_{01} + A_{21}\mathbf{S}_1 + A_{22}\mathbf{S}_2, \\
 \mathbf{X}_0 &= \mathbf{X}'_{01} + B_1\mathbf{S}_1 + \mathbf{X}'_{02} + B_2\mathbf{S}_2, \\
 \mathbf{X}'_{01} &\sim \mathcal{N}(0, K'_{01}) \quad \mathbf{X}'_{02} \sim \mathcal{N}(0, K'_{02}), \\
 \mathbf{X}_1 &\sim \mathcal{N}(0, K_1) \quad \mathbf{X}_2 \sim \mathcal{N}(0, K_2),
 \end{aligned} \tag{29}$$

where $\mathbf{X}'_{01}, \mathbf{X}'_{02}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{S}_1, \mathbf{S}_2$ are independent. For simplicity of representation, denote $\bar{A}_1 = (A_{11}, A_{12})$, $\bar{A}_2 = (A_{20}, A_{21}, A_{22})$ and $\bar{B} = (B_1, B_2)$. Let $f_1(\cdot), g_1(\cdot), f_2(\cdot)$ and $g_2(\cdot)$ be defined as

$$\begin{aligned}
 f_1(\bar{A}_1, \bar{B}, K'_{01}, K'_{02}) &= I(\mathbf{U}, \mathbf{X}_1; \mathbf{Y}_1) - I(\mathbf{U}; \mathbf{S}), \\
 g_1(\bar{A}_1, \bar{B}, K'_{01}, K'_{02}) &= I(\mathbf{X}_1; \mathbf{Y}_1 | \mathbf{U}), \\
 f_2(\bar{A}_2, \bar{B}, K'_{01}, K'_{02}) &= I(\mathbf{V}, \mathbf{X}_2; \mathbf{Y}_2) - I(\mathbf{V}; \mathbf{U}, \mathbf{S}), \\
 g_2(\bar{A}_2, \bar{B}, K'_{01}, K'_{02}) &= I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{V}),
 \end{aligned}$$

where the mutual information terms are evaluated using the joint Gaussian distribution set at (29). Based on those definitions we obtain an achievable region for the Gaussian channel.

Proposition 6. *An inner bound on the capacity region of the parallel state-dependent Gaussian channel with a helper and with independent states, consists of rate pairs (R_1, R_2) satisfying;*

$$R_1 \leq \min\{f_1(\bar{A}_1, \bar{B}, K'_{01}, K'_{02}), g_1(\bar{A}_1, \bar{B}, K'_{01}, K'_{02})\}, \tag{30a}$$

$$R_2 \leq \min\{f_2(\bar{A}_2, \bar{B}, K'_{01}, K'_{02}), g_2(\bar{A}_2, \bar{B}, K'_{01}, K'_{02})\}, \tag{30b}$$

for some real matrices $A_{20}, A_{21}, A_{22}, B_1, B_2, K'_{01}$ and K'_{02} satisfying $K'_{01}, K'_{02} \succeq 0, K'_{01} + K'_{02} + B_1 K_{S_1} B_1^T + B_2 K_{S_2} B_2^T \preceq K_0$.

Now we provide our intuition behind such construction of the RVs in the proof of Proposition 6. \mathbf{X}_0 contains two parts, the one with $B_l, l = 1, 2$ controls the direct state cancelation of each state. The second part $\mathbf{X}'_{0l}, l = 1, 2$, is used for dirty paper coding via generation of the state-correlated auxiliary RVs \mathbf{U} and \mathbf{V} .

4.3. Capacity Region Characterization

In this section, we will characterize segments on the capacity boundary for various channel parameters using the inner and outer bounds that were derived in Section 4.2. Consider the inner bounds in (30a)–(30b). Each bound has two terms in the argument of min. We suggest optimizing each term independently and then comparing it to the outer bounds in (25). In the last step we will state the conditions under which those terms are valid. Our technique for optimal choice of $(A_{11}, A_{12}, A_{20}, A_{21}, A_{22})$ be such that cancels the respective interfering terms from the mutual information quantities. We explain how those matrices were chosen in Appendix F.

We begin by considering what choice of (A_{11}, A_{12}) can maximize $f_1(\bar{A}_1, \bar{B}, K'_{01}, K'_{02})$. Let

$$\begin{aligned}
 A_{11}^a &= (G_1(K'_{01} + K'_{02})G_1^T + I)^{-1}K'_{01}G_1^T(G_1B_1 + I), \\
 A_{12}^a &= (G_1(K'_{01} + K'_{02})G_1^T + I)^{-1}K'_{01}G_1^TG_1B_2.
 \end{aligned}$$

Then $f_1(\bar{A}_1, \bar{B}, K'_{01}, K'_{02})$ takes the following form

$$f_1(\bar{A}_1^a, \bar{B}, K'_{01}, K'_{02}) = \frac{1}{2} \log \left(\frac{|G_1 K_0 G_1^T + K_1 + G_1 B_1 K_{S_1} + K_{S_1} B_1^T G_1^T + K_{S_1} + I|}{|G_1 K_0 G_1^T + G_1 B_1 K_{S_1} + K_{S_1} B_1^T G_1^T + I|} \right) + \frac{1}{2} \log \left(\frac{|G_1 (K'_{01} + K'_{02}) G_1^T + I|}{|G_1 K'_{02} G_1^T + I|} \right). \quad (31)$$

If $f_1(\bar{A}_1^a, \bar{B}, K'_{01}, K'_{02}) \leq g_1(\bar{A}_1^a, \bar{B}, K'_{01}, K'_{02})$, then $R_1 = f_1(\bar{A}_1^a, \bar{B}, K'_{01}, K'_{02})$ is achievable. Moreover, if we choose $K'_{02} = 0$, then $R_1 = f_1(\bar{A}_1^a, \bar{B}, K'_{01}, 0)$ meets the outer bound (the first term in “min” in (25)) with $B_1 K_{S_1} = \Sigma_{x_0 s_1}$ and $B_2 K_{S_2} = \Sigma_{x_0 s_2}$. Furthermore, by setting

$$A_{11}^b = B_1 + G_1^{-1}, \quad A_{12}^b = B_2,$$

we obtain

$$g_1(\bar{A}_1^b, \bar{B}, K'_{01}, K'_{02}) = \frac{1}{2} \log \left(\frac{|G_1 K'_{02} G_1^T + K_1 + I|}{|G_1 K'_{02} G_1^T + I|} \right).$$

If $g_1(\bar{A}_1^b, \bar{B}, K'_{01}, K'_{02}) \leq f_1(\bar{A}_1^b, \bar{B}, K'_{01}, K'_{02})$, then

$$R_1 = \frac{1}{2} \log \left(\frac{|G_1 K'_{02} G_1^T + K_1 + I|}{|G_1 K'_{02} G_1^T + I|} \right)$$

is achievable. Similarly, by choosing $K'_{02} = 0$, then $R_1 = \frac{1}{2} \log |K_1 + I|$ is achievable and this meets the outer bound (the second term in “min” in (25)). Next we consider the bound on R_2 . Let

$$\begin{aligned} A_{20}^a &= (G_2 K'_{02} G_2^T + I)^{-1} K'_{02} G_2^T G_2, \\ A_{21}^a &= (G_2 K'_{02} G_2^T + I)^{-1} K'_{02} G_2^T G_2 B_1, \\ A_{22}^a &= (G_2 K'_{02} G_2^T + I)^{-1} K'_{02} G_2^T (G_2 B_2 + I). \end{aligned}$$

Then $f_2(\bar{A}_2, \bar{B}, K'_{01}, K'_{02})$ takes the following form

$$f_2(\bar{A}_2^a, \bar{B}, K'_{01}, K'_{02}) = \frac{1}{2} \log \left(\frac{|G_2 K_0 G_2^T + K_2 + G_2 B_2 K_{S_2} + K_{S_2} B_2^T G_2^T + K_{S_2} + I|}{|G_2 K_0 G_2^T + G_2 B_2 K_{S_2} + K_{S_2} B_2^T G_2^T + K_{S_2} + I|} \right) + \frac{1}{2} \log (|G_2 K'_{02} G_2^T + I|). \quad (32)$$

If $f_2(\bar{A}_2^a, \bar{B}, K'_{01}, K'_{02}) \leq g_2(\bar{A}_2^a, \bar{B}, K'_{01}, K'_{02})$, then $R_2 = f_2(\bar{A}_2^a, \bar{B}, K'_{01}, K'_{02})$ is achievable. Moreover, if we choose $K'_{01} = 0$, then $R_2 = f_2(\bar{A}_2^a, \bar{B}, 0, K'_{02})$ meets the outer bound (the first term in “min” in (25)).

Furthermore, we set

$$A_{20}^b = I, \quad A_{21}^b = B_1, \quad A_{22}^b = B_2 + G_2^{-1}, \quad (33)$$

and then obtain

$$g_2(\bar{A}_2^b, \bar{B}, K'_{01}, K'_{02}) = \frac{1}{2} \log (|K_2 + I|). \quad (34)$$

If $g_2(\bar{A}_2^b, \bar{B}, K'_{01}, K'_{02}) \leq f_2(\bar{A}_2^b, \bar{B}, K'_{01}, K'_{02})$, then $R_2 = \frac{1}{2} \log (|K_2 + I|)$ is achievable and this meets the outer bound. This also equals the maximum rate for R_2 when the channel is not corrupted by state.

Summarizing the above analysis, we obtain the following characterization of segments of the capacity region boundary.

Theorem 3. The channel parameters $(G_1, G_2, K_0, K_1, K_2, K_{S_1}, K_{S_2})$ can be partitioned into the sets $\mathcal{A}_1, \mathcal{B}_1, \mathcal{C}_1$, where

$$\begin{aligned}\mathcal{A}_1 &= \{(G_1, G_2, K_0, K_1, K_2, K_{S_1}, K_{S_2}) : f_1(\bar{A}_1^a, \bar{B}^a, K'_{01}, K'_{02}) \leq g_1(\bar{A}_1^a, \bar{B}^a, K'_{01}, K'_{02}), \\ \mathcal{C}_1 &= \{(G_1, G_2, K_0, K_1, K_2, K_{S_1}, K_{S_2}) : f_1(\bar{A}_1^b, \bar{B}, K'_{01}, K'_{02}) \geq g_1(\bar{A}_1^b, \bar{B}, K'_{01}, K'_{02})\}, \\ \mathcal{B}_1 &= (\mathcal{A}_1 \cup \mathcal{C}_1)^c.\end{aligned}$$

If $(G_1, G_2, K_0, K_1, K_2, K_{S_1}, K_{S_2}) \in \mathcal{A}_1$, then $R_1 = f_1(\bar{A}_1^a, \bar{B}, K'_0, 0)$ captures one segment of the capacity region boundary, where the state cannot be fully canceled. If $(G_1, G_2, K_0, K_1, K_2, K_{S_1}, K_{S_2}) \in \mathcal{C}_1$, then $R_1 = \frac{1}{2} \log |K_1 + I|$ captures one segment of the capacity region boundary where the state is fully canceled. If $(G_1, G_2, K_0, K_1, K_2, K_{S_1}, K_{S_2}) \in \mathcal{B}_1$, then the R_1 segment of the capacity region boundary is not characterized.

The channel parameters $(G_1, G_2, K_0, K_1, K_2, K_{S_1}, K_{S_2})$ can also be partitioned into the sets $\mathcal{A}_2, \mathcal{B}_2, \mathcal{C}_2$, where

$$\begin{aligned}\mathcal{A}_2 &= \{(G_1, G_2, K_0, K_1, K_2, K_{S_1}, K_{S_2}) : f_2(\bar{A}_2^a, \bar{B}, K'_{01}, K'_{02}) \leq g_2(\bar{A}_2^a, \bar{B}, K'_{01}, K'_{02}), \\ \mathcal{C}_2 &= \{(G_1, G_2, K_0, K_1, K_2, K_{S_1}, K_{S_2}) : f_2(\bar{A}_2^b, \bar{B}, K'_{01}, K'_{02}) \geq g_2(\bar{A}_2^b, \bar{B}, K'_{01}, K'_{02}), \\ \mathcal{B}_2 &= (\mathcal{A}_2 \cup \mathcal{C}_2)^c.\end{aligned}$$

If $(G_1, G_2, K_0, K_1, K_2, K_{S_1}, K_{S_2}) \in \mathcal{A}_2$, then $R_2 = f_2(\bar{A}_2^a, \bar{B}, 0, K'_0)$ captures one segment of the capacity region boundary, where the state cannot be fully canceled. If $(G_1, G_2, K_0, K_1, K_2, K_{S_1}, K_{S_2}) \in \mathcal{C}_2$, then $R_2 = \frac{1}{2} \log (|K_2 + I|)$ captures one segment of the capacity boundary where the state is fully canceled. If $(G_1, G_2, K_0, K_1, K_2, K_{S_1}, K_{S_2}) \in \mathcal{B}_2$, then the R_2 segment of the capacity region boundary is not characterized.

The above theorem describes two partitions of the channel parameters, respectively under which segments on the capacity region boundary corresponding to R_1 and R_2 can be characterized. Intersection of two sets, each from one partition, collectively characterizes the entire segments on the capacity region boundary.

We note that our inner bound can be tight for some set of channel parameters. As an example, assume that $(G_1, G_2, K_0, K_1, K_2, K_{S_1}, K_{S_2}) \in \mathcal{C}_1 \cap \mathcal{C}_2$. In such case, $R_1 = \frac{1}{2} \log \left(\frac{|G_1 K'_{02} G_1^T + K_1 + I|}{|G_1 K'_{02} G_1^T + I|} \right)$ and $R_2 = \frac{1}{2} \log (|K_2 + I|)$ are achievable. For the point-to-point helper channel [28], it was shown that if the helper power is above some threshold, the state is completely canceled, whereas in our model we have two parallel channels. If the helper power is high enough, it can split its signal, similarly as for the Gaussian BC, such that one part of it is intended for Receiver 2, where by using dirty paper coding it eliminates completely the interference caused by the state and the part of the signal intended for Receiver 1. In the same time the part of the helper signal intended for Receiver 1, can only cancel the interference caused by the state while the part intended to Receiver 2 is treated as noise.

4.4. Numerical Results

In this section, we provide specific numerical examples to illustrate the bounds obtained in the previous sections. In particular, we focus on scalar Gaussian channel setting, such that: $G_1 \leftarrow \eta_1$; $G_2 \leftarrow \eta_2$; $K_0 \leftarrow P_0$; $K'_{01} \leftarrow P'_{01}$; $K'_{02} \leftarrow P'_{02}$; $K_1 \leftarrow P_1$; $K_2 \leftarrow P_2$; $K_{S_1} \leftarrow Q_1$; $K_{S_2} \leftarrow Q_2$. We also denote $(A_{11}, A_{12}, A_{20}, A_{21}, A_{22}, B_1, B_2) \leftarrow (\alpha_{11}, \alpha_{12}, \alpha_{20}, \alpha_{21}, \alpha_{22}, \beta_1, \beta_2)$. We plot the inner and outer bounds for various values of helper power P_0 , channel gains, η_1 and η_2 and different state power. The results are shown in Figure 12. The outer bound is based on Proposition 4. The inner bound is the convex hull of all the achievable regions, with interchange between the roles of the decoders. The time-sharing inner bound is according to point-to-point helper channel achievable region [28]. The scenario where the helper power is less than the users power is depicted in Figure 12a,b, while the channel gains in Figure 12a are equal, they are mismatched in Figure 12b. Please note that in both cases our inner

bound outperforms the time-sharing bound, especially in the mismatched case, and some segments of the capacity region are characterized.

The scenario with helper power being higher than the user power and matched and mismatched channel gain is depicted in Figure 12c,d respectively. Similar to for low helper power regime, our proposed achievability scheme performs better than time-sharing.

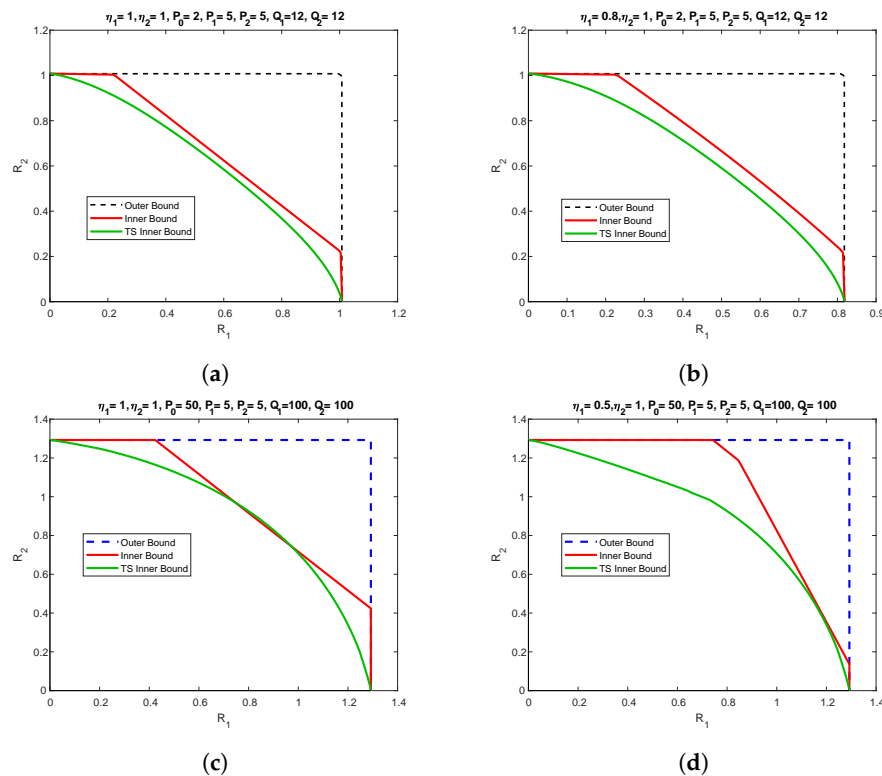


Figure 12. Numerical Results.

5. Conclusions

In the first part of this paper, we have studied the parallel state-dependent Gaussian channel with a state-cognitive helper and with same but differently scaled states. An inner bound was derived and was compared to an upper bound, and the segments of the capacity region boundary were characterized for various channel parameters. We have shown that if the channel gain matrices satisfy a certain symmetry property, the full rectangular capacity region of the two point-to-point channels without the state can be achieved. Furthermore, for the scalar channel case, we have shown that for a given ratio of state gain over the helper signal gain, a/b , one can find a value of the helper power— P_0 , such that the capacity region is fully characterized.

A different model of the parallel state-dependent Gaussian channel with a state-cognitive helper and independent states was considered in the second part of this study. Inner and outer bounds were derived, and segments of the capacity region boundary were characterized for various channel parameters. We have also demonstrated our results using numerical simulation and have shown that our achievability scheme outperforms time-sharing that was shown to be optimal for the infinite state power regime in [34].

These two models represent a special case of a more general scenario with correlated states, our results in both studies imply that as the states get more correlated, it is easier to mitigate the interference. Furthermore, the gap between the inner bound and the outer bound in this work suggests that a new techniques for outer bound derivation is needed as we believe that the inner bounds consisting of pairs $(R_1, R_2) = (f_1(\bar{A}_1', \bar{B}, K_{01}', K_{02}'), f_2(\bar{A}_2', \bar{B}, K_{01}', K_{02}'))$ is indeed tight for some set of channel parameters.

Author Contributions: Conceptualization, Y.L.; Formal analysis, M.D.; Supervision, S.S.; Validation, R.D.; Writing—original draft, M.D.; Writing—review & editing, M.D.

Funding: The work of M.D. and S.S. has been supported by the European Union’s Horizon 2020 Research And Innovation Programme, grant agreement no. 694630, and by the Heron consortium via the Israel minister of economy and science. The work of Y.L. was supported in part by the U.S. National Science Foundation under Grant CCF-1801846.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Proposition 1

Fix the following joint PMF

$$P_{SWX_0X_1X_2Y_1Y_2} = P_S P_{W|S} P_{X_0|SW} P_{X_1} P_{X_2} P_{Y_1|SX_0X_1} P_{Y_2|SX_0X_2}. \quad (A1)$$

Appendix A.1. Codebook Generation

Randomly and independently generate $2^{n\tilde{R}}$ sequences $w^n(\tilde{m})$, $\tilde{m} \in \mathcal{I}_{\tilde{R}}^{(n)}$, each according to $\prod_{i=1}^n P_W(w_i)$. Similarly, for $l = \{1, 2\}$, generate 2^{nR_l} sequences $x_l^n(m_l)$, $m_l \in \mathcal{I}_{R_l}^{(n)}$, each according to $\prod_{i=1}^n P_{X_l}(x_{li})$. These sequences constitute the codebook, which is revealed to the encoders and the decoders.

Appendix A.2. Encoding

Appendix A.2.1. Encoder at the Helper

Fix $\epsilon' > 0$. Given s^n , find \tilde{m}' , such that $(w^n(\tilde{m}'), s^n) \in \mathcal{T}_{\epsilon'}^{(n)}(P_{SW})$. If no such sequence exists, it declares an error. Then, given $(w^n(\tilde{m}'), s^n)$, generate x_0^n according to $\prod_{i=1}^n P_{X_0|SW}(x_{0i}|s_i, w_i(\tilde{m}'))$. The encoder at the helper then transmits x_{0i} at time $i \in [1 : n]$.

Appendix A.2.2. Encoder at Transmitter 1

To send message m_l , encoder l transmits $x_l^n(m_l)$, for $l = \{1, 2\}$.

Appendix A.3. Decoding

Let $\epsilon > \epsilon'$ and $l \in \{1, 2\}$. Upon receiving y_l^n , the decoder at Receiver l declares that $\hat{m}_l \in \mathcal{I}_{R_l}^{(n)}$ is sent if it is the unique message such that $(w^n(\hat{m}), x_l^n(\hat{m}_l), y_l^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{WX_lY_l})$ for some $\hat{m} \in \mathcal{I}_{\tilde{R}}^{(n)}$; otherwise it declares an error.

Appendix A.4. Analysis of the Probability of Error

The encoder at the helper declares an error if the following event occurs

$$\mathcal{E}_0 = \{(S^n, W^n(\tilde{m})) \notin \mathcal{T}_{\epsilon'}^{(n)}(P_{SW}) \text{ for all } \tilde{m} \in \mathcal{I}_{\tilde{R}}^{(n)}\}. \quad (A2)$$

By the covering lemma (Section 2.3), with setting the original random variables (U, X, \hat{X}) as (\emptyset, S, W) respectively, and $\mathcal{A} = \mathcal{I}_{\tilde{R}}^{(n)}$, $P\{\mathcal{E}_0\}$ tends to zero as $n \rightarrow \infty$ if

$$\tilde{R} > I(W; S). \quad (A3)$$

Assume without loss of generality that $(M_1, M_2) = (1, 1)$, condition (A3) holds, and let \tilde{M} denote the index of the chosen w^n sequence for s^n . The decoder at Receiver l makes an error only if one or more of the following events occur:

$$\begin{aligned}\mathcal{E}_{l1} &= \{(W^n(\tilde{M}), X_l^n(1), Y_l^n) \notin \mathcal{T}_\epsilon^{(n)}(P_{W X_l Y_l})\}, \\ \mathcal{E}_{l2} &= \{(W^n(\tilde{M}), X_l^n(m_l), Y_l^n) \in \mathcal{T}_\epsilon^{(n)}(P_{W X_l Y_l}) \text{ for some } m_l \neq l\}, \\ \mathcal{E}_{l3} &= \{(W^n(\tilde{m}), X_l^n(m_l), Y_l^n) \in \mathcal{T}_\epsilon^{(n)}(P_{W X_l Y_l}) \text{ for some } m_l \neq 1 \text{ and } \tilde{m} \neq \tilde{M}\}.\end{aligned}$$

Thus, by the union of events bound,

$$P\{\mathcal{E}_l\} = P\{\mathcal{E}_{l1} \cup \mathcal{E}_{l2} \cup \mathcal{E}_{l3}\} \leq P\{\mathcal{E}_{l1}\} + P\{\mathcal{E}_{l2}\} + P\{\mathcal{E}_{l3}\}. \quad (\text{A4})$$

By the LLN, the first term $P\{\mathcal{E}_{l1}\}$ tend to zero as $n \rightarrow \infty$. For the second term, note that for $m_l \neq l$,

$$\begin{aligned}& p(w^n(\tilde{M}), x_l^n(m_l), y_l^n) \\ &= \sum_{x_l^n(1), x_0^n(\tilde{M}), s^n} p(w^n(\tilde{M}), x_0^n(\tilde{M}), x_l^n(1), x_l^n(m_l), s^n, y_l^n) \\ &= p(x_l^n(m_l)) \sum_{x_l^n(1), x_0^n(\tilde{M}), s^n} p(w^n(\tilde{M}), x_0^n(\tilde{M}), s^n) p(x_l^n(1)) p(y_l^n | x_0^n(\tilde{M}), x_l^n(1), s^n) \\ &= \prod_{i=1}^n p(x_{li}(m_l)) \sum_{x_l^n(1), x_0^n(\tilde{M}), s^n} \prod_{i=1}^n p(w_i(\tilde{M}), x_{0i}(\tilde{M}), s_i) p(x_{li}(1)) p(y_{li} | x_{0i}(\tilde{M}), x_{li}(1), s_i) \\ &= \prod_{i=1}^n p(x_{li}(m_l)) p(w_i(\tilde{M}), y_{li}).\end{aligned}$$

Hence, by the packing lemma, choosing the original random variable (U, X, Y) as $(\emptyset, X_l, (W, Y_l))$ respectively, $\mathcal{A} = \mathcal{I}_{R_l}^{(n)}$, $P\{\mathcal{E}_{l2}\}$ tends to zero as $n \rightarrow \infty$ if $R_l < I(X_l; Y_l, W)$. Since X_l and W are independent, $R_l < I(X_l; Y_l | W)$. Finally, for the third term, note that for $m_l \neq 1$ and $\tilde{m} \neq \tilde{M}$

$$\begin{aligned}& p(w^n(\tilde{m}), x_l^n(m_l), y_l^n) \\ &= \sum_{w^n(\tilde{M}), x_l^n(1), x_0^n(\tilde{M}), s^n} p(w^n(\tilde{m}), w^n(\tilde{M}), x_0^n(\tilde{M}), x_l^n(1), x_l^n(m_l), s^n, y_l^n) \\ &= p(x_l^n(m_l)) p(w^n(\tilde{m})) \sum_{w^n(\tilde{M}), x_l^n(1), x_0^n(\tilde{M}), s^n} p(w^n(\tilde{M}), x_0^n(\tilde{M}), s^n) p(x_l^n(1)) p(y_l^n | x_0^n(\tilde{M}), x_l^n(1), s^n) \\ &= \prod_{i=1}^n p(x_{li}(m_l)) p(w_i(\tilde{m})) p(y_{li}).\end{aligned}$$

Again, by the packing lemma, choosing the original random variable (U, X, Y) as $(\emptyset, (W, X_l), Y_l)$ respectively, $\mathcal{A} = \mathcal{I}_{R_l}^{(n)} \times \mathcal{I}_{\tilde{R}}^{(n)}$, $P\{\mathcal{E}_{l3}\}$ tends to zero as $n \rightarrow \infty$ if $\tilde{R} + R_l < I(W, X_l; Y_l)$.

Appendix B. Proof of Proposition 3

We prove for a general $l \in \{1, 2\}$. By Fano's inequality (Lemma 2),

$$H(M_l | Y_l^n) \leq n R_l P_e^{(n)} + 1 \leq n \epsilon_n$$

where ϵ_n tends to zero as $n \rightarrow \infty$ by the assumption that $\lim_{n \rightarrow \infty} P_e^{(n)} = 0$.

Now consider

$$\begin{aligned}
 & nR_l + I(\mathbf{S}^n; \mathbf{Y}_l^n | \mathbf{M}_l) \\
 &= H(\mathbf{M}_l) - H(\mathbf{M}_l | \mathbf{Y}_l^n) + H(\mathbf{M}_l | \mathbf{Y}_l^n) + I(\mathbf{S}^n; \mathbf{Y}_l^n | \mathbf{M}_l) \\
 &\leq I(\mathbf{M}_l, \mathbf{S}^n; \mathbf{Y}_l^n) + n\epsilon_n \\
 &= h(\mathbf{Y}_l^n) - h(\mathbf{Y}_l^n | \mathbf{M}_l, \mathbf{S}^n) + n\epsilon_n \\
 &\leq h(\mathbf{Y}_l^n) - h(\mathbf{Y}_l^n | \mathbf{M}_l, \mathbf{X}_l^n, \mathbf{X}_0^n, \mathbf{S}^n) + n\epsilon_n \\
 &= h(\mathbf{Y}_l^n) - h(\mathbf{Z}_l^n) + n\epsilon_n \\
 &\leq \frac{n}{2} \log \left(|G_l K_0 G_l^T + K_l + G_l \Sigma_{\mathbf{X}_0 \mathbf{S}} G_{s_l}^T + G_{s_l} \Sigma_{\mathbf{X}_0 \mathbf{S}}^T G_l^T + G_{s_l} K_{\mathbf{S}} G_{s_l}^T + I| \right) + n\epsilon_n.
 \end{aligned}$$

$I(\mathbf{S}^n; \mathbf{Y}_l^n | \mathbf{M}_l)$ can be lower bounded as follows:

$$I(\mathbf{S}^n; \mathbf{Y}_l^n | \mathbf{M}_l) = h(\mathbf{S}^n) - h(\mathbf{S}^n | \tilde{\mathbf{Y}}_l^n), \quad (\text{A5})$$

where $\tilde{\mathbf{Y}}_l^n \triangleq G_l \mathbf{X}_0^n + G_{s_l} \mathbf{S}^n + \mathbf{Z}_l^n$. The conditional differential entropy can be upper bounded as follows

$$\begin{aligned}
 h(\mathbf{S}^n | \tilde{\mathbf{Y}}_l^n) &\leq \sum_{i=1}^n h(\mathbf{S}_i | \tilde{\mathbf{Y}}_{li}) \\
 &\leq \frac{n}{2} \log(2\pi e)^t \left| K_{\mathbf{S}} - \Sigma_{\mathbf{S} \tilde{\mathbf{Y}}_l} \Sigma_{\tilde{\mathbf{Y}}_l}^{-1} \Sigma_{\mathbf{S} \tilde{\mathbf{Y}}_l}^T \right|
 \end{aligned} \quad (\text{A6})$$

where $\Sigma_{\mathbf{S} \tilde{\mathbf{Y}}_l} = \mathbb{E} [\mathbf{S} \tilde{\mathbf{Y}}_l^T] = K_{\mathbf{S}} G_{s_l}^T + \Sigma_{\mathbf{X}_0 \mathbf{S}}^T G_l^T$, and

$$\Sigma_{\tilde{\mathbf{Y}}_l} = G_l K_0 G_l^T + G_l \Sigma_{\mathbf{X}_0 \mathbf{S}} G_{s_l}^T + G_{s_l} \Sigma_{\mathbf{X}_0 \mathbf{S}}^T G_l^T + G_{s_l} K_{\mathbf{S}} G_{s_l}^T + I.$$

Now we apply Sylvester's Determinant Theorem [38] to have

$$\begin{aligned}
 \left| K_{\mathbf{S}} - \Sigma_{\mathbf{S} \tilde{\mathbf{Y}}_l} \Sigma_{\tilde{\mathbf{Y}}_l}^{-1} \Sigma_{\mathbf{S} \tilde{\mathbf{Y}}_l}^T \right| &= |K_{\mathbf{S}}| \left| I - K_{\mathbf{S}}^{-1} \Sigma_{\mathbf{S} \tilde{\mathbf{Y}}_l} \Sigma_{\tilde{\mathbf{Y}}_l}^{-1} \Sigma_{\mathbf{S} \tilde{\mathbf{Y}}_l}^T \right| \\
 &= |K_{\mathbf{S}}| \left| I - \Sigma_{\mathbf{S} \tilde{\mathbf{Y}}_l}^T K_{\mathbf{S}}^{-1} \Sigma_{\mathbf{S} \tilde{\mathbf{Y}}_l} \Sigma_{\tilde{\mathbf{Y}}_l}^{-1} \right| \\
 &= |K_{\mathbf{S}}| \left| \Sigma_{\tilde{\mathbf{Y}}_l} - \Sigma_{\mathbf{S} \tilde{\mathbf{Y}}_l}^T K_{\mathbf{S}}^{-1} \Sigma_{\mathbf{S} \tilde{\mathbf{Y}}_l} \right| \left| \Sigma_{\tilde{\mathbf{Y}}_l}^{-1} \right|.
 \end{aligned}$$

Consider the argument of the middle determinant. Since $K_{\mathbf{S}}^{-1} \Sigma_{\mathbf{S} \tilde{\mathbf{Y}}_l} = G_{s_l}^T + K_{\mathbf{S}}^{-1} \Sigma_{\mathbf{X}_0 \mathbf{S}}^T G_l^T$, it follows that

$$\begin{aligned}
 \Sigma_{\mathbf{S} \tilde{\mathbf{Y}}_l}^T K_{\mathbf{S}}^{-1} \Sigma_{\mathbf{S} \tilde{\mathbf{Y}}_l} &= [G_{s_l} K_{\mathbf{S}} + G_l \Sigma_{\mathbf{X}_0 \mathbf{S}}] \left[G_{s_l}^T + K_{\mathbf{S}}^{-1} \Sigma_{\mathbf{X}_0 \mathbf{S}}^T G_l^T \right] \\
 &= G_{s_l} K_{\mathbf{S}} G_{s_l}^T + G_{s_l} \Sigma_{\mathbf{X}_0 \mathbf{S}}^T G_l^T + G_l \Sigma_{\mathbf{X}_0 \mathbf{S}} G_{s_l}^T + G_l \Sigma_{\mathbf{X}_0 \mathbf{S}} K_{\mathbf{S}}^{-1} \Sigma_{\mathbf{X}_0 \mathbf{S}}^T G_l^T,
 \end{aligned}$$

and

$$\Sigma_{\tilde{\mathbf{Y}}_l} - \Sigma_{\mathbf{S} \tilde{\mathbf{Y}}_l}^T K_{\mathbf{S}}^{-1} \Sigma_{\mathbf{S} \tilde{\mathbf{Y}}_l} = G_l (K_0 - \Sigma_{\mathbf{X}_0 \mathbf{S}} K_{\mathbf{S}}^{-1} \Sigma_{\mathbf{X}_0 \mathbf{S}}^T) G_l^T + I.$$

Finally, by collecting terms,

$$I(\mathbf{S}^n; \mathbf{Y}_l^n | \mathbf{M}_l) \geq \frac{n}{2} \log \left(\frac{|G_l K_0 G_l^T + G_l \Sigma_{\mathbf{X}_0 \mathbf{S}} G_{s_l}^T + G_{s_l} \Sigma_{\mathbf{X}_0 \mathbf{S}}^T G_l^T + G_{s_l} K_{\mathbf{S}} G_{s_l}^T + I|}{|G_l (K_0 - \Sigma_{\mathbf{X}_0 \mathbf{S}} K_{\mathbf{S}}^{-1} \Sigma_{\mathbf{X}_0 \mathbf{S}}^T) G_l^T + I|} \right).$$

Thus, the bound in (11) is satisfied.

It remains to show that $\Sigma_{\mathbf{X}_0\mathbf{S}}K_{\mathbf{S}}^{-1}\Sigma_{\mathbf{X}_0\mathbf{S}}^T \preceq K_0$. We use the non-negativity property of the covariance matrix of the vector $(\mathbf{X}_0, \mathbf{S})^T$

$$\begin{aligned}\det \mathbb{E} \left[(\mathbf{X}_0, \mathbf{S})^T (\mathbf{X}_0, \mathbf{S}) \right] &= \det \begin{bmatrix} K_0 & \Sigma_{\mathbf{X}_0\mathbf{S}} \\ \Sigma_{\mathbf{X}_0\mathbf{S}}^T & K_{\mathbf{S}} \end{bmatrix} \\ &= |K_{\mathbf{S}}| \cdot \left| K_0 - \Sigma_{\mathbf{X}_0\mathbf{S}}K_{\mathbf{S}}^{-1}\Sigma_{\mathbf{X}_0\mathbf{S}}^T \right| \\ &\geq 0,\end{aligned}$$

where the last inequality follows since any covariance matrix ($\Sigma_{\mathbf{X}_0\mathbf{S}}$) is by definition positive definite. Now we arrange parts to have:

$$\Sigma_{\mathbf{X}_0\mathbf{S}}K_{\mathbf{S}}^{-1}\Sigma_{\mathbf{X}_0\mathbf{S}}^T \preceq K_0.$$

This completes the proof of Proposition 3.

Appendix C. Optimal Coefficients for the MIMO Gaussian with Differently Scaled States Channel

We first consider the bound on R_1 . Consider the first argument in min of (7a)

$$\begin{aligned}I(\mathbf{W}, \mathbf{X}_1; \mathbf{Y}_1) - I(\mathbf{W}; \mathbf{S}) &= I(\mathbf{X}_1; \mathbf{Y}_1) + I(\mathbf{W}; \mathbf{Y}_1 | \mathbf{X}_1) - I(\mathbf{W}; \mathbf{S} | \mathbf{X}_1) \\ &= I(\mathbf{X}_1; \mathbf{Y}_1) + h(\mathbf{W} | \mathbf{S}, \mathbf{X}_1) - h(\mathbf{W} | \mathbf{X}_1, \mathbf{Y}_1).\end{aligned}$$

It is straightforward to show that

$$I(\mathbf{X}_1; \mathbf{Y}_1) = \frac{1}{2} \log \left(\frac{|G_1 K_0 G_1^T + K_1 + G_1 B K_{\mathbf{S}} G_{s_1}^T + G_{s_1} K_{\mathbf{S}} B^T G_1^T + G_{s_1} K_{\mathbf{S}} G_{s_1}^T + I|}{|G_1 K_0 G_1^T + G_1 B K_{\mathbf{S}} G_{s_1}^T + G_{s_1} K_{\mathbf{S}} B^T G_1^T + G_{s_1} K_{\mathbf{S}} G_{s_1}^T + I|} \right),$$

and

$$h(\mathbf{W} | \mathbf{S}, \mathbf{X}_1) = h(\mathbf{X}'_0).$$

As for the third term, denote $\tilde{\mathbf{Y}}_1 = \mathbf{Y}_1 - \mathbf{X}_1$, thus

$$\begin{aligned}h(\mathbf{W} | \mathbf{X}_1, \mathbf{Y}_1) &= h(\mathbf{W} | \tilde{\mathbf{Y}}_1) \\ &= h(\mathbf{W} - M_{\mathbf{W} | \tilde{\mathbf{Y}}_1} \tilde{\mathbf{Y}}_1) \\ &= h(\mathbf{X}'_0 + A\mathbf{S} - M_{\mathbf{W} | \tilde{\mathbf{Y}}_1} (G_1 (\mathbf{X}'_0 + B\mathbf{S}) + G_{s_1} \mathbf{S} + \mathbf{Z}_1)).\end{aligned}$$

We require that term \mathbf{S} in the argument of the differential entropy be completely canceled, therefore we choose

$$A^a = M_{\mathbf{W} | \tilde{\mathbf{Y}}_1} (G_1 B + G_{s_1}).$$

With the above choice of A , we have

$$h(\mathbf{W} | \mathbf{X}_1, \mathbf{Y}_1) = h(\mathbf{X}'_0 - M_{\mathbf{W} | \tilde{\mathbf{Y}}_1} (G_1 \mathbf{X}'_0 + \mathbf{Z}_1)).$$

Finally, we demand that $M_{\mathbf{W} | \tilde{\mathbf{Y}}_1}$ be the MMSE of \mathbf{X}'_0 given $G_1 \mathbf{X}'_0 + \mathbf{Z}_1$, i.e.

$$M_{\mathbf{W} | \tilde{\mathbf{Y}}_1} = (G_1 K'_0 G_1^T + I)^{-1} K'_0 G_1^T.$$

In such case

$$A^a = (G_1 K'_0 G_1^T + I)^{-1} K'_0 G_1^T (G_1 B + G_{s_1}).$$

Hence

$$h(\mathbf{W}|\mathbf{X}_1, \mathbf{Y}_1) = h(\mathbf{X}'_0|G_1 \mathbf{X}'_0 + \mathbf{Z}_1),$$

and thus

$$\begin{aligned} h(\mathbf{W}|\mathbf{S}, \mathbf{X}_1) - h(\mathbf{W}|\mathbf{X}_1, \mathbf{Y}_1) &= h(\mathbf{X}'_0) - h(\mathbf{X}'_0|G_1 \mathbf{X}'_0 + \mathbf{Z}_1) \\ &= I(\mathbf{X}'_0; G_1 \mathbf{X}'_0 + \mathbf{Z}_1) \\ &= \frac{1}{2} \log \left(|G_1 K'_0 G_1^T + I| \right). \end{aligned}$$

Furthermore, if we choose $A^b = B + G_1^{-1} G_{s_1}$, then

$$\begin{aligned} I(\mathbf{X}_1; \mathbf{Y}_1|\mathbf{W}) &= I(\mathbf{X}_1; G_1(\mathbf{X}'_0 + B\mathbf{S}) + G_{s_1}\mathbf{S} + \mathbf{X}_1 + \mathbf{Z}_1|\mathbf{X}'_0 + (B + G_1^{-1} G_{s_1})\mathbf{S}) \\ &= I(\mathbf{X}_1; \mathbf{X}_1 + \mathbf{Z}_1) \\ &= \frac{1}{2} \log (|K_1 + I|). \end{aligned}$$

With this choice of A , $h(\mathbf{W}|\tilde{\mathbf{Y}}_1)$ is equal to

$$\begin{aligned} h(\mathbf{W}|\tilde{\mathbf{Y}}_1) &= h(\mathbf{X}'_0 + A\mathbf{S}|G_1(\mathbf{X}'_0 + (A - G_1^{-1} G_{s_1})\mathbf{S}) + G_{s_1}\mathbf{S} + \mathbf{Z}_1) \\ &= h(\mathbf{X}'_0 + A\mathbf{S}|G_1(\mathbf{X}'_0 + A\mathbf{S}) + \mathbf{Z}_1) \\ &= \frac{1}{2} \log(2\pi e)^t \left| K'_0 + AK_S A^T - (K'_0 G_1^T + AK_S A^T G_1^T)(G_1(K'_0 + AK_S A^T)G_1^T + I)^{-1}(K'_0 G_1^T + AK_S A^T G_1^T)^T \right| \\ &= \frac{1}{2} \log(2\pi e)^t \left| K'_0 + AK_S A^T - (K'_0 + AK_S A^T)G_1^T (G_1(K'_0 + AK_S A^T)G_1^T + I)^{-1}G_1(K'_0 + AK_S A^T)^T \right| \\ &= \frac{1}{2} \log(2\pi e)^t \left| K'_0 + AK_S A^T \right| \left| I - G_1^T (G_1(K'_0 + AK_S A^T)G_1^T + I)^{-1}G_1(K'_0 + AK_S A^T)^T \right| \\ &= \frac{1}{2} \log(2\pi e)^t \left| K'_0 + AK_S A^T \right| \left| I - (G_1(K'_0 + AK_S A^T)G_1^T + I)^{-1}G_1(K'_0 + AK_S A^T)^T G_1^T \right| \\ &= \frac{1}{2} \log(2\pi e)^t \frac{|K'_0 + AK_S A^T|}{|G_1(K'_0 + AK_S A^T)G_1^T + I|} \left| (G_1(K'_0 + AK_S A^T)G_1^T + I) - G_1(K'_0 + AK_S A^T)^T G_1^T \right| \\ &= \frac{1}{2} \log(2\pi e)^t \frac{|K'_0 + AK_S A^T|}{|G_1(K'_0 + AK_S A^T)G_1^T + I|}. \end{aligned}$$

Thus

$$I(\mathbf{W}, \mathbf{X}_1; \mathbf{Y}_1) - I(\mathbf{W}; \mathbf{S}) = \frac{1}{2} \log \left(\frac{|K'_0| |G_1(K'_0 + AK_S A^T)G_1^T + K_1 + I|}{|K'_0 + AK_S A^T|} \right).$$

We would like to obtain a condition under which $g_1(A^b, B, K'_0) \leq f_1(A^b, B, K'_0)$, i.e.

$$|K'_0| |G_1(K'_0 + AK_S A^T)G_1^T + K_1 + I| \geq |K'_0 + AK_S A^T| |K_1 + I|$$

that is equivalent to

$$K'_0 G_1 (K'_0 + AK_S A^T) G_1^T + K'_0 K_1 + K'_0 \succeq K'_0 K_1 + K'_0 + AK_S A^T K_1 + AK_S A^T.$$

Furthermore, after rearranging terms, we have

$$K'_0 G_1 K'_0 G_1^T \succeq A K_S A^T (K_1 + I) - K'_0 G_1 A K_S A^T G_1^T.$$

The choices of A^c and A^d for the achievability proof of R_2 follows using similar steps by interchanging the indices $1 \leftarrow 2$.

Appendix D. Proof of Proposition 4

This proof relies greatly on the proof of the outer bound in the differently scaled states scenario in Appendix B. The main differences are:

- $\mathbf{S} = (\mathbf{S}_1, \mathbf{S}_2)^T$. Since \mathbf{S}_1 and \mathbf{S}_2 are independent, the covariance matrix of \mathbf{S} is block-diagonal

$$K_S = \begin{pmatrix} K_{S_1} & 0 \\ 0 & K_{S_2} \end{pmatrix},$$

- the helper signal \mathbf{X}_0 correlates with \mathbf{S}_1 and \mathbf{S}_2 and it is characterized by the cross-covariance matrices $\Sigma_{\mathbf{X}_0 \mathbf{S}_1}$ and $\Sigma_{\mathbf{X}_0 \mathbf{S}_2}$ respectively,
- the state gain matrices, G_{S_1} and G_{S_2} are unity matrices.

Hence, in the independent-states case, we have the following upper bound on $nR_1 + I(\mathbf{S}^n; \mathbf{Y}_1^n | \mathbf{M}_1)$

$$nR_1 + I(\mathbf{S}^n; \mathbf{Y}_1^n | \mathbf{M}_1) \leq \frac{n}{2} \log \left(|G_1 K_0 G_1^T + K_1 + G_1 \Sigma_{\mathbf{X}_0 \mathbf{S}_1} + \Sigma_{\mathbf{X}_0 \mathbf{S}_1}^T G_1^T + K_{S_1} + I| \right) + n\epsilon_n. \quad (\text{A7})$$

Let $\tilde{\mathbf{Y}}_1^n \triangleq G_1 \mathbf{X}_0^n + \mathbf{S}_1^n + \mathbf{Z}_1^n$. We proceed to lower bound $I(\mathbf{S}^n; \mathbf{Y}_1^n | \mathbf{M}_1) = h(\mathbf{S}^n) - h(\mathbf{S}^n | \tilde{\mathbf{Y}}_1^n)$ in similar fashion to (A6), the conditional differential entropy can be upper bounded as follows

$$h(\mathbf{S}^n | \tilde{\mathbf{Y}}_1^n) \leq \frac{n}{2} \log(2\pi e)^{2t} |K_S| \left| \Sigma_{\tilde{\mathbf{Y}}_1} - K_{S\tilde{\mathbf{Y}}_1}^T \Sigma_S^{-1} \Sigma_{S\tilde{\mathbf{Y}}_1} \right| \left| \Sigma_{\tilde{\mathbf{Y}}_1}^{-1} \right|, \quad (\text{A8})$$

where

$$\Sigma_{S\tilde{\mathbf{Y}}_1} = \mathbb{E} [\mathbf{S} \tilde{\mathbf{Y}}_1^T] = \begin{bmatrix} K_{S_1} + \Sigma_{\mathbf{X}_0 \mathbf{S}_1}^T G_1^T \\ \Sigma_{\mathbf{X}_0 \mathbf{S}_2}^T G_1^T \end{bmatrix},$$

and $\Sigma_{\tilde{\mathbf{Y}}_1} = G_1 K_0 G_1^T + G_1 \Sigma_{\mathbf{X}_0 \mathbf{S}_1} + \Sigma_{\mathbf{X}_0 \mathbf{S}_1}^T G_1^T + K_{S_1} + I$. The power $2t$ in (A8) is due to the size of the vector $(\mathbf{S}_1, \mathbf{S}_2)^T$. The argument of the inner determinant in (A8) can be further evaluated as follows,

$$\begin{aligned} \Sigma_S^{-1} \Sigma_{S\tilde{\mathbf{Y}}_1} &= \begin{bmatrix} K_{S_1}^{-1} & 0 \\ 0 & K_{S_2}^{-1} \end{bmatrix} \begin{bmatrix} K_{S_1} + \Sigma_{\mathbf{X}_0 \mathbf{S}_1}^T G_1^T \\ \Sigma_{\mathbf{X}_0 \mathbf{S}_2}^T G_1^T \end{bmatrix} \\ &= \begin{bmatrix} I + K_{S_1}^{-1} \Sigma_{\mathbf{X}_0 \mathbf{S}_1}^T G_1^T \\ K_{S_2}^{-1} \Sigma_{\mathbf{X}_0 \mathbf{S}_2}^T G_1^T \end{bmatrix}, \end{aligned}$$

and

$$\begin{aligned} \Sigma_{S\tilde{\mathbf{Y}}_1}^T \Sigma_S^{-1} \Sigma_{S\tilde{\mathbf{Y}}_1} &= \begin{bmatrix} K_{S_1} + G_1 \Sigma_{\mathbf{X}_0 \mathbf{S}_1} & G_1 \Sigma_{\mathbf{X}_0 \mathbf{S}_2} \end{bmatrix} \begin{bmatrix} I + K_{S_1}^{-1} \Sigma_{\mathbf{X}_0 \mathbf{S}_1}^T G_1^T \\ K_{S_2}^{-1} \Sigma_{\mathbf{X}_0 \mathbf{S}_2}^T G_1^T \end{bmatrix} \\ &= K_{S_1} + G_1 \Sigma_{\mathbf{X}_0 \mathbf{S}_1} + \Sigma_{\mathbf{X}_0 \mathbf{S}_1}^T G_1^T + G_1 \Sigma_{\mathbf{X}_0 \mathbf{S}_1} K_{S_1}^{-1} \Sigma_{\mathbf{X}_0 \mathbf{S}_1}^T G_1^T + G_1 \Sigma_{\mathbf{X}_0 \mathbf{S}_2} K_{S_2}^{-1} \Sigma_{\mathbf{X}_0 \mathbf{S}_2}^T G_1^T. \end{aligned}$$

Thus,

$$\begin{aligned} \Sigma_{\tilde{\mathbf{Y}}_1} - \Sigma_{S\tilde{\mathbf{Y}}_1}^T \Sigma_S^{-1} \Sigma_{S\tilde{\mathbf{Y}}_1} &= G_1 (K_0 - \Sigma_{\mathbf{X}_0 \mathbf{S}_1} K_{S_1}^{-1} \Sigma_{\mathbf{X}_0 \mathbf{S}_1}^T - \Sigma_{\mathbf{X}_0 \mathbf{S}_2} K_{S_2}^{-1} \Sigma_{\mathbf{X}_0 \mathbf{S}_2}^T) G_1^T + I \\ &= G_1 (K_0 - \Sigma_{\mathbf{X}_0 \mathbf{S}} K_S^{-1} \Sigma_{\mathbf{X}_0 \mathbf{S}}^T) G_1^T + I, \end{aligned}$$

where in the last equality we used the definition of $\Sigma_{\mathbf{X}_0\mathbf{S}}$ from (26). Consequently, we established an upper bound on $I(\mathbf{S}^n; \mathbf{Y}_1^n | \mathbf{M}_1)$:

$$I(\mathbf{S}^n; \mathbf{Y}_1^n | \mathbf{M}_1) \geq \frac{n}{2} \log \left(\frac{|G_1 K_0 G_1^T + G_1 \Sigma_{\mathbf{X}_0\mathbf{S}_1} + \Sigma_{\mathbf{X}_0\mathbf{S}_1}^T G_1^T + K_{\mathbf{S}_1} + I|}{|G_1 (K_0 - \Sigma_{\mathbf{X}_0\mathbf{S}} K_{\mathbf{S}}^{-1} \Sigma_{\mathbf{X}_0\mathbf{S}}^T) G_1^T + I|} \right). \quad (\text{A9})$$

Finally, collecting (A7) and (A9), the bound in (25), with $l = 1$, is satisfied. The bound (25) for $l = 2$ follows from similar considerations. It remains to show that $\Sigma_{\mathbf{X}_0\mathbf{S}} K_{\mathbf{S}}^{-1} \Sigma_{\mathbf{X}_0\mathbf{S}}^T \preceq K_0$. We use the non-negativity property of the covariance matrix of the vector $(\mathbf{X}_0, \mathbf{S}_1, \mathbf{S}_2)^T$

$$\begin{aligned} \det \mathbb{E} \left[(\mathbf{X}_0, \mathbf{S}_1, \mathbf{S}_2)^T (\mathbf{X}_0, \mathbf{S}_1, \mathbf{S}_2) \right] &= \det \begin{bmatrix} K_0 & \Sigma_{\mathbf{X}_0\mathbf{S}_1} & \Sigma_{\mathbf{X}_0\mathbf{S}_2} \\ \Sigma_{\mathbf{X}_0\mathbf{S}_1}^T & K_{\mathbf{S}_1} & 0 \\ \Sigma_{\mathbf{X}_0\mathbf{S}_2}^T & 0 & K_{\mathbf{S}_2} \end{bmatrix} \\ &= \left| \begin{bmatrix} K_{\mathbf{S}_1} & 0 \\ 0 & K_{\mathbf{S}_2} \end{bmatrix} \right| \cdot \left| K_0 - \begin{bmatrix} \Sigma_{\mathbf{X}_0\mathbf{S}_1} & \Sigma_{\mathbf{X}_0\mathbf{S}_2} \end{bmatrix} \begin{bmatrix} K_{\mathbf{S}_1} & 0 \\ 0 & K_{\mathbf{S}_2} \end{bmatrix}^{-1} \begin{bmatrix} \Sigma_{\mathbf{X}_0\mathbf{S}_1}^T \\ \Sigma_{\mathbf{X}_0\mathbf{S}_2}^T \end{bmatrix} \right| \\ &= \left| \begin{bmatrix} K_{\mathbf{S}_1} & 0 \\ 0 & K_{\mathbf{S}_2} \end{bmatrix} \right| \cdot \left| K_0 - \begin{bmatrix} \Sigma_{\mathbf{X}_0\mathbf{S}_1} & \Sigma_{\mathbf{X}_0\mathbf{S}_2} \end{bmatrix} \begin{bmatrix} K_{\mathbf{S}_1}^{-1} & 0 \\ 0 & K_{\mathbf{S}_2}^{-1} \end{bmatrix} \begin{bmatrix} \Sigma_{\mathbf{X}_0\mathbf{S}_1}^T \\ \Sigma_{\mathbf{X}_0\mathbf{S}_2}^T \end{bmatrix} \right| \\ &= \left| \begin{bmatrix} K_{\mathbf{S}_1} & 0 \\ 0 & K_{\mathbf{S}_2} \end{bmatrix} \right| \cdot \left| K_0 - \Sigma_{\mathbf{X}_0\mathbf{S}_1} K_{\mathbf{S}_1}^{-1} \Sigma_{\mathbf{X}_0\mathbf{S}_1}^T - \Sigma_{\mathbf{X}_0\mathbf{S}_2} K_{\mathbf{S}_2}^{-1} \Sigma_{\mathbf{X}_0\mathbf{S}_2}^T \right| \\ &\stackrel{(b)}{\geq} 0, \end{aligned}$$

where the last inequality follows from non-negativity of covariance matrix. Now we arrange parts to have:

$$\Sigma_{\mathbf{X}_0\mathbf{S}_1} K_{\mathbf{S}_1}^{-1} \Sigma_{\mathbf{X}_0\mathbf{S}_1}^T + \Sigma_{\mathbf{X}_0\mathbf{S}_2} K_{\mathbf{S}_2}^{-1} \Sigma_{\mathbf{X}_0\mathbf{S}_2}^T = \Sigma_{\mathbf{X}_0\mathbf{S}} K_{\mathbf{S}}^{-1} \Sigma_{\mathbf{X}_0\mathbf{S}}^T \preceq K_0.$$

This completes the proof of Proposition 4.

Appendix E. Proof of Proposition 5

We use random codes and fix the following joint distribution:

$$P_{SUVX_0X_1X_2Y_1Y_2} = P_{SUV} P_{X_0|SUV} P_{X_1} P_{X_2} P_{Y_1|SX_0X_1} P_{Y_2|SX_0X_2}.$$

Appendix E.1. Codebook Generation

Generate $2^{n\tilde{R}_U}$ randomly and independently generated sequences $u^n(r)$, $r \in \mathcal{I}_{\tilde{R}_U}^{(n)}$, each according to $\prod_{i=1}^n P_U(u_i)$. Similarly, generate $2^{n\tilde{R}_V}$ randomly and independently generated sequences $v^n(t)$, $t \in \mathcal{I}_{\tilde{R}_V}^{(n)}$ according to $\prod_{i=1}^n P_V(v_i)$.

Let $l \in \{1, 2\}$. Randomly and independently generate 2^{nR_l} sequences $x_1^n(m_l)$, $m_l \in \mathcal{I}_{R_l}^{(n)}$, each according to $\prod_{i=1}^n P_{X_l}(x_{li})$.

These sequences constitute the codebook, which is revealed to the encoders and the decoders.

Appendix E.2. Encoding

Fix $\epsilon'' > \epsilon' > 0$. The encoder at the helper, given s^n , finds \tilde{r} such that

$$(s^n, u^n(\tilde{r})) \in \mathcal{T}_{\epsilon'}^{(n)}(P_{SU}),$$

if there is more than one such \tilde{r} , choose the smallest one. If no such \tilde{r} can be found declare an error. Next, given $s^n, u^n(\tilde{r})$, find \tilde{t} such that

$$(s^n, u^n(\tilde{r}), v^n(\tilde{t})) \in \mathcal{T}_{\epsilon''}^{(n)}(P_{SUV}),$$

if there is more than one such \tilde{t} , choose the smallest one. If no such \tilde{t} can be found declare an error. Then, given $s^n, u^n(\tilde{r})$ and $v^n(\tilde{t})$, generate x_0^n with i.i.d. components according to $\prod_{i=1}^n P_{X_0|SUV}(x_{0i}|s_i, u_i, v_i)$. Let (m_1, m_2) be the messages to be sent. The encoder at transmitter l transmits $x_l^n(m_l)$.

Appendix E.3. Decoding

Fix $\epsilon > \epsilon''$. Given y_1^n , decoder 1 declares that \hat{m}_1 was sent if it is the unique message such that

$$(u^n(\hat{r}), x_1^n(\hat{m}_1), y_1^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{UX_1Y_1}).$$

If no or more than one such \hat{m}_1 can be found, it declares an error.

Similarly, given y_2^n , decoder 2 finds the unique message \hat{m}_2 such that

$$(v^n(\hat{t}), x_2^n(\hat{m}_2), y_2^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{VX_2Y_2}).$$

If no or more than one such \hat{m}_2 can be found, it declares an error.

Appendix E.4. Analysis of the Probability of Error

Assume without loss of generality that the message pair $(M_1, M_2) = (1, 1)$ was sent and let r_0 be the chosen index for u^n and t_0 be the chosen index for v^n . The encoder at the helper makes an error only if one or both of the following errors occur:

$$\begin{aligned} \mathcal{E}_{01} &= \{(S^n, U^n(r)) \notin \mathcal{T}_{\epsilon'}^{(n)}(P_{SU}) \text{ for all } r \in \mathcal{I}_{\tilde{R}_U}^{(n)}\}, \\ \mathcal{E}_{02} &= \{(S^n, U^n(r_0), V^n(t)) \notin \mathcal{T}_{\epsilon''}^{(n)}(P_{SUV}) \text{ for all } t \in \mathcal{I}_{\tilde{R}_V}^{(n)}\}. \end{aligned}$$

Thus, by the union of events bound, the probability that the encoder at the helper makes an error, can be upper bounded as

$$\Pr(\mathcal{E}_0) = \Pr(\mathcal{E}_{01} \cup \mathcal{E}_{02}) \leq \Pr(\mathcal{E}_{01}) + \Pr(\mathcal{E}_{01}^c \cap \mathcal{E}_{02}).$$

By the covering lemma with $U = \emptyset, X \leftarrow S, \hat{X} \leftarrow U$, and $\mathcal{A} = \mathcal{I}_{\tilde{R}_U}^{(n)}$, $\Pr(\mathcal{E}_{01})$ tends to zero as $n \rightarrow \infty$ if $\tilde{R}_U > I(U; S) + \delta(\epsilon')$.

Similarly, using the covering lemma with $U = \emptyset, X \leftarrow (S, U), \hat{X} \leftarrow V$, and $\mathcal{A} = \mathcal{I}_{\tilde{R}_V}^{(n)}$, $\Pr(\mathcal{E}_{01}^c \cap \mathcal{E}_{02})$ tends to zero as $n \rightarrow \infty$ if $\tilde{R}_V > I(V; S, U) + \delta(\epsilon'')$.

The decoder at receiver 1 makes an error only if one or more of the following events occur

$$\begin{aligned} \mathcal{E}_{11} &= \{(U^n(r_0), X_1^n(1), Y_1^n) \notin \mathcal{T}_{\epsilon}^{(n)}(P_{UX_1Y_1})\}, \\ \mathcal{E}_{12} &= \{(U^n(r_0), X_1^n(m_1), Y_1^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{UX_1Y_1}) \text{ for some } m_1 \neq 1\}, \\ \mathcal{E}_{13} &= \{(U^n(r), X_1^n(m_1), Y_1^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{UX_1Y_1}) \text{ for some } r \neq r_0 \text{ and } m_1 \neq 1\}. \end{aligned}$$

Again, by the union of events bound, the probability that the decoder at receiver 1 makes an error, can be upper bounded as

$$\begin{aligned}\Pr(\mathcal{E}_1) &= \Pr(\mathcal{E}_{11} \cup \mathcal{E}_{12} \cup \mathcal{E}_{13}) \\ &\leq \Pr(\mathcal{E}_{01} \cup \mathcal{E}_{11} \cup \mathcal{E}_{12} \cup \mathcal{E}_{13}) \\ &\leq \Pr(\mathcal{E}_{01}) + \Pr(\mathcal{E}_{01}^c \cap \mathcal{E}_{11}) + \Pr(\mathcal{E}_{01}^c \cap \mathcal{E}_{12}) + \Pr(\mathcal{E}_{13}).\end{aligned}\quad (\text{A10})$$

We have already shown that $\Pr(\mathcal{E}_{01})$ tends to zero as $n \rightarrow \infty$ if $\tilde{R}_U > I(U; S) + \delta(\epsilon')$. Next, note that

$$\mathcal{E}_{01}^c = \{(S^n, U^n(r_0)) \in \mathcal{T}_{\epsilon'}^{(n)}(P_{SU})\} = \{(S^n, U^n(r_0), X_0^n) \in \mathcal{T}_{\epsilon'}^{(n)}(P_{SUX_0})\},$$

and

$$\begin{aligned}P_{Y_1^n|S^n U^n(r_0) X_0^n X_1^n(1)}(y_1^n|s^n, u^n, x_0^n, x_1^n) &= \prod_{i=1}^n P_{Y_1|S U X_0 X_1}(y_{1i}|s_i, u_i, x_{0i}, x_{1i}) \\ &= \prod_{i=1}^n P_{Y_1|S X_0 X_1}(y_{1i}|s_i, x_{0i}, x_{1i}).\end{aligned}$$

Hence, by the conditionally typicality lemma, $\Pr(\mathcal{E}_{01}^c \cap \mathcal{E}_{11})$ tends to zero as $n \rightarrow \infty$.

As for the probability of the event $(\mathcal{E}_{01}^c \cap \mathcal{E}_{12})$, $X_1^n(m_1)$ is independent of $(U^n(r_0), Y_1^n) \sim \prod_{i=1}^n P_{UY_1}(u_i, y_{1i})$. Hence, by the packing lemma, with $U = \emptyset$, $X \leftarrow X_1$, $Y \leftarrow (U, Y_1)$ and $\mathcal{A} = [2 : 2^{nR_1}]$, $\Pr(\mathcal{E}_{01}^c \cap \mathcal{E}_{12})$ tends to zero as $n \rightarrow \infty$ if $R_1 < I(X_1; U, Y_1) - \delta(\epsilon)$. X_1 and U are mutually independent, hence the latter condition is equivalent to $R_1 < I(X_1; Y_1|U) - \delta(\epsilon)$.

Finally, since for $m_1 \neq 1$, $r \neq r_0$, $(X_1^n(m_1), U^n(r))$ is independent of $(X_1^n(1), U^n(r_0), Y_1^n)$, again by the packing lemma with $U = \emptyset$, $X \leftarrow (U, X_1)$, $Y \leftarrow Y_1$ and $\mathcal{A} = [2 : 2^{nR_1}] \times [1 : 2^{n\tilde{R}_U}]/r_0$, $\Pr(\mathcal{E}_{13})$ tends to zero as $n \rightarrow \infty$ if $\tilde{R}_U + R_1 < I(U, X_1; Y_1) - \delta(\epsilon)$.

Next consider the average probability of error for decoder 2. The decoder at receiver 2 makes an error only if one or more of the following events occur

$$\begin{aligned}\mathcal{E}_{21} &= \{(V^n(t_0), X_2^n(1), Y_2^n) \notin \mathcal{T}_{\epsilon}^{(n)}(P_{VX_2Y_2})\}, \\ \mathcal{E}_{22} &= \{(V^n(t_0), X_2^n(m_2), Y_2^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{VX_2Y_2}) \text{ for some } m_2 \neq 1\}, \\ \mathcal{E}_{23} &= \{(V^n(t), X_2^n(m_2), Y_2^n) \in \mathcal{T}_{\epsilon}^{(n)}(P_{VX_2Y_2}) \text{ for some } t \neq t_0 \text{ and } m_2 \neq 1\}.\end{aligned}$$

Similar to (A10), the probability that the decoder at receiver 2 makes an error, can be upper bounded as

$$\Pr(\mathcal{E}_2) \leq \Pr(\mathcal{E}_0) + \Pr(\mathcal{E}_0^c \cap \mathcal{E}_{21}) + \Pr(\mathcal{E}_0^c \cap \mathcal{E}_{22}) + \Pr(\mathcal{E}_{23}).$$

In a very similar fashion as was shown for decoder 1, it can be shown that $\Pr(\mathcal{E}_2)$ tends to zero as $n \rightarrow \infty$ if

$$\begin{aligned}\tilde{R}_V &\geq I(V; S, U) + \delta(\epsilon''), \\ R_2 &\leq I(X_2; Y_2|V) - \delta(\epsilon), \\ R_2 + \tilde{R}_V &\leq I(V, X_2; Y_2) - \delta(\epsilon).\end{aligned}$$

Finally, combining the aforementioned bounds yields the following achievable region:

$$\begin{aligned}R_1 &\leq \min \{I(U, X_1; Y_1) - I(U; S), I(X_1; Y_1|U)\}, \\ R_2 &\leq \min \{I(V, X_2; Y_2) - I(V; U, S), I(X_2; Y_2|V)\}.\end{aligned}$$

This completes the proof of achievability.

Appendix F. Optimal Coefficients for the MIMO Gaussian with Independent States Channel

We first consider the bound on R_1 . Consider the first argument in min of (28a)

$$\begin{aligned} I(\mathbf{U}; \mathbf{X}_1; \mathbf{Y}_1) - I(\mathbf{U}; \mathbf{S}) &= I(\mathbf{X}_1; \mathbf{Y}_1) + I(\mathbf{U}; \mathbf{Y}_1 | \mathbf{X}_1) - I(\mathbf{U}; \mathbf{S} | \mathbf{X}_1) \\ &= I(\mathbf{X}_1; \mathbf{Y}_1) + h(\mathbf{U} | \mathbf{S}, \mathbf{X}_1) - h(\mathbf{U} | \mathbf{X}_1, \mathbf{Y}_1). \end{aligned} \quad (\text{A11})$$

It is straightforward to show that

$$I(\mathbf{X}_1; \mathbf{Y}_1) = \frac{1}{2} \log \left(\frac{|G_1 K_0 G_1^T + K_1 + G_1 B_1 K_{S_1} + K_{S_1} B_1^T G_1^T + K_{S_1} + I|}{|G_1 K_0 G_1^T + G_1 B_1 K_{S_1} + K_{S_1} B_1^T G_1^T + I|} \right),$$

and

$$h(\mathbf{U} | \mathbf{S}, \mathbf{X}_1) = h(\mathbf{X}'_{01}) = \frac{1}{2} \log(2\pi e)^t |K'_{01}|.$$

As for the third term in (A11), denote $\tilde{\mathbf{Y}}_1 = \mathbf{Y}_1 - \mathbf{X}_1$,

$$\begin{aligned} h(\mathbf{U} | \mathbf{X}_1, \mathbf{Y}_1) &= h(\mathbf{U} | \tilde{\mathbf{Y}}_1) \\ &= h(\mathbf{U} - M_{\mathbf{U}|\tilde{\mathbf{Y}}_1} \tilde{\mathbf{Y}}_1) \\ &= h(\mathbf{X}'_{01} + A_{11} \mathbf{S}_1 + A_{12} \mathbf{S}_2 - M_{\mathbf{U}|\tilde{\mathbf{Y}}_1} (G_1 (\mathbf{X}'_{01} + \mathbf{X}'_{02} + B_1 \mathbf{S}_1 + B_2 \mathbf{S}_2) + \mathbf{S}_1 + \mathbf{Z}_1)). \end{aligned}$$

We require that the terms \mathbf{S}_1 and \mathbf{S}_2 in the argument of the differential entropy be completely canceled, hence we choose

$$A_{11}^a = M_{\mathbf{U}|\tilde{\mathbf{Y}}_1} (G_1 B_1 + I), \quad A_{12}^a = M_{\mathbf{U}|\tilde{\mathbf{Y}}_1} G_1 B_2.$$

With the above choice of (A_{11}, A_{12}) , we have

$$h(\mathbf{U} | \mathbf{X}_1, \mathbf{Y}_1) = h(\mathbf{X}'_{01} - M_{\mathbf{U}|\tilde{\mathbf{Y}}_1} (G_1 (\mathbf{X}'_{01} + \mathbf{X}'_{02}) + \mathbf{Z}_1)).$$

Finally, we demand that $M_{\mathbf{U}|\tilde{\mathbf{Y}}_1}$ be the MMSE of \mathbf{X}'_{01} given $G_1 (\mathbf{X}'_{01} + \mathbf{X}'_{02}) + \mathbf{Z}_1$, i.e.,

$$M_{\mathbf{U}|\tilde{\mathbf{Y}}_1} = \left(G_1 (K'_{01} + K'_{02}) G_1^T + I \right)^{-1} K'_{01} G_1^T.$$

In such case

$$\begin{aligned} A_{11}^a &= \left(G_1 (K'_{01} + K'_{02}) G_1^T + I \right)^{-1} K'_{01} G_1^T (G_1 B_1 + I), \\ A_{12}^a &= \left(G_1 (K'_{01} + K'_{02}) G_1^T + I \right)^{-1} K'_{01} G_1^T G_1 B_2. \end{aligned}$$

Hence

$$h(\mathbf{U} | \mathbf{X}_1, \mathbf{Y}_1) = h(\mathbf{X}'_{01} | G_1 (\mathbf{X}'_{01} + \mathbf{X}'_{02}) + \mathbf{Z}_1),$$

and thus

$$\begin{aligned} h(\mathbf{U} | \mathbf{S}, \mathbf{X}_1) - h(\mathbf{U} | \mathbf{X}_1, \mathbf{Y}_1) &= h(\mathbf{X}'_{01}) - h(\mathbf{X}'_{01} | G_1 (\mathbf{X}'_{01} + \mathbf{X}'_{02}) + \mathbf{Z}_1) \\ &= I(\mathbf{X}'_{01}; G_1 (\mathbf{X}'_{01} + \mathbf{X}'_{02}) + \mathbf{Z}_1) \\ &= \frac{1}{2} \log \left(\frac{|G_1 (K'_{01} + K'_{02}) G_1^T + I|}{|G_1 K'_{02} G_1^T + I|} \right). \end{aligned}$$

Furthermore, if we choose $A_{11}^b = B_1 + G_1^{-1}$ and $A_{12}^b = B_2$, then

$$\begin{aligned} I(\mathbf{X}_1; \mathbf{Y}_1 | \mathbf{U}) &= I(\mathbf{X}_1; G_1 (\mathbf{X}'_{01} + \mathbf{X}'_{02} + B_1 \mathbf{S}_1 + B_2 \mathbf{S}_2) + \mathbf{X}_1 + \mathbf{S}_1 + \mathbf{Z}_1 | \mathbf{X}'_{01} + A_{11}^b \mathbf{S}_1 + A_{12}^b \mathbf{S}_2) \\ &= I(\mathbf{X}_1; G_1 \mathbf{X}'_{02} + \mathbf{X}_1 + \mathbf{Z}_1) \\ &= \frac{1}{2} \log \left(\frac{|G_1 K_{02} G_1^T + K_1 + I|}{|G_1 K_{02} G_1^T + I|} \right). \end{aligned}$$

We next consider the bound on R_2 . Consider the first argument in min of (28b)

$$\begin{aligned} I(\mathbf{V}; \mathbf{X}_2; \mathbf{Y}_2) - I(\mathbf{V}; \mathbf{U}, \mathbf{S}) &= I(\mathbf{X}_2; \mathbf{Y}_2) + I(\mathbf{V}; \mathbf{Y}_2 | \mathbf{X}_2) - I(\mathbf{V}; \mathbf{U}, \mathbf{S} | \mathbf{X}_2) \\ &= I(\mathbf{X}_2; \mathbf{Y}_2) + h(\mathbf{V} | \mathbf{U}, \mathbf{S}, \mathbf{X}_2) - h(\mathbf{V} | \mathbf{Y}_2, \mathbf{X}_2). \end{aligned} \quad (\text{A12})$$

It is straightforward to show that

$$I(\mathbf{X}_2; \mathbf{Y}_2) = \frac{1}{2} \log \left(\frac{|G_2 K_0 G_2^T + K_2 + G_2 B_2 K_{S_2} + K_{S_2} B_2^T G_2^T + K_{S_2} + I|}{|G_2 K_0 G_2^T + G_2 B_2 K_{S_2} + K_{S_2} B_2^T G_2^T + K_{S_2} + I|} \right),$$

and $h(\mathbf{V} | \mathbf{U}, \mathbf{S}, \mathbf{X}_2) = h(\mathbf{X}'_{02})$. As for the third term in (A12), denote $\tilde{\mathbf{Y}}_2 = \mathbf{Y}_2 - \mathbf{X}_2$

$$\begin{aligned} h(\mathbf{V} | \mathbf{Y}_2, \mathbf{X}_2) &= h(\mathbf{V} | \tilde{\mathbf{Y}}_2) = h(\mathbf{V} - M_{\mathbf{V} | \tilde{\mathbf{Y}}_2} \tilde{\mathbf{Y}}_2) \\ &= h(\mathbf{X}'_{02} + A_{20} \mathbf{X}'_{01} + A_{21} \mathbf{S}_1 + A_{22} \mathbf{S}_2 - M_{\mathbf{V} | \tilde{\mathbf{Y}}_2} (G_2 (\mathbf{X}'_{01} + \mathbf{X}'_{02} + B_1 \mathbf{S}_1 + B_2 \mathbf{S}_2) + \mathbf{S}_2 + \mathbf{Z}_2)). \end{aligned}$$

We require that terms \mathbf{X}'_{01} , \mathbf{S}_1 and \mathbf{S}_2 in the argument of the differential entropy be completely canceled, hence we choose

$$A_{20}^a = M_{\mathbf{V} | \tilde{\mathbf{Y}}_2} G_2, \quad A_{21}^a = M_{\mathbf{V} | \tilde{\mathbf{Y}}_2} G_2 B_1, \quad A_{22}^a = M_{\mathbf{V} | \tilde{\mathbf{Y}}_2} (G_2 B_2 + I).$$

With the above choice of (A_{20}, A_{21}, A_{22}) , we have

$$h(\mathbf{V} | \mathbf{Y}_2, \mathbf{X}_2) = h(\mathbf{X}'_{02} - M_{\mathbf{V} | \tilde{\mathbf{Y}}_2} (G_2 \mathbf{X}'_{02} + \mathbf{Z}_2)).$$

Following this we demand that $M_{\mathbf{V} | \tilde{\mathbf{Y}}_2}$ be the MMSE of \mathbf{X}'_{02} given $G_2 \mathbf{X}'_{02} + \mathbf{Z}_2$, i.e.,

$$M_{\mathbf{V} | \tilde{\mathbf{Y}}_2} = (G_2 K_{02} G_2^T + I)^{-1} K_{02} G_2^T.$$

In such case

$$\begin{aligned} A_{20}^a &= (G_2 K_{02} G_2^T + I)^{-1} K_{02} G_2^T G_2, \\ A_{21}^a &= (G_2 K_{02} G_2^T + I)^{-1} K_{02} G_2^T G_2 B_1, \\ A_{22}^a &= (G_2 K_{02} G_2^T + I)^{-1} K_{02} G_2^T (G_2 B_2 + I). \end{aligned}$$

Hence

$$h(\mathbf{V} | \mathbf{Y}_2, \mathbf{X}_2) = h(\mathbf{X}'_{02} | G_2 \mathbf{X}'_{02} + \mathbf{Z}_2),$$

and thus

$$\begin{aligned} h(\mathbf{V} | \mathbf{U}, \mathbf{S}, \mathbf{X}_2) - h(\mathbf{V} | \mathbf{Y}_2, \mathbf{X}_2) &= h(\mathbf{X}'_{02}) - h(\mathbf{X}'_{02} | G_2 \mathbf{X}'_{02} + \mathbf{Z}_2) \\ &= I(\mathbf{X}'_{02}; G_2 \mathbf{X}'_{02} + \mathbf{Z}_2) \\ &= \frac{1}{2} \log (|G_2 K_{02} G_2^T + I|). \end{aligned}$$

Furthermore, if we choose $A_{20}^b = I$, $A_{21}^b = B_1$ and $A_{22}^b = B_2 + G_2^{-1}$, then

$$I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{V}) = I(\mathbf{X}_2; \mathbf{X}_2 + \mathbf{Z}_2) = \frac{1}{2} \log(|K_2 + I|).$$

References

1. Vaezi, M.; Vincent Poor, H., NOMA: An Information-Theoretic Perspective. In *Multiple Access Techniques for 5G Wireless Networks and Beyond*; Springer International Publishing: Cham, Switzerland, 2019; pp. 167–193. [\[CrossRef\]](#)
2. Gel'fand, S.; Pinsker, M. Coding for channels with random parameters. *Probl. Contr. Inf. Theory* **1980**, *9*, 19–31.
3. Costa, M. Writing on dirty paper (Corresp.). *IEEE Trans. Inf. Theory* **1983**, *29*, 439–441. [\[CrossRef\]](#)
4. Shannon, C.E. Channels with Side Information at the Transmitter. *IBM J. Res. Dev.* **1958**, *2*, 289–293. [\[CrossRef\]](#)
5. Cohen, A.S.; Lapidith, A. Generalized writing on dirty paper. In Proceedings of the IEEE International Symposium on Information Theory, Lausanne, Switzerland, 30 June–5 July 2002; p. 227. [\[CrossRef\]](#)
6. Cover, T. Broadcast channels. *IEEE Trans. Inf. Theory* **1972**, *18*, 2–14. [\[CrossRef\]](#)
7. Marton, K. A coding theorem for the discrete memoryless broadcast channel. *IEEE Trans. Inf. Theory* **1979**, *25*, 306–311. [\[CrossRef\]](#)
8. Liang, Y. Multiuser Communications with Relaying and User Cooperation. Ph.D. Thesis, University of Illinois, Urbana-Champaign, IL, USA, 2005.
9. Nair, C.; Gamal, A.E. An Outer Bound to the Capacity Region of the Broadcast Channel. *IEEE Trans. Inf. Theory* **2007**, *53*, 350–355. [\[CrossRef\]](#)
10. Gallager, R.G. Capacity and coding for degraded broadcast channels. *Probl. Pereda. Inf.* **1974**, *10*, 3–14.
11. Bergmans, P. A simple converse for broadcast channels with additive white Gaussian noise (Corresp.). *IEEE Trans. Inf. Theory* **1974**, *20*, 279–280. [\[CrossRef\]](#)
12. Weingarten, H.; Steinberg, Y.; Shamai, S. The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel. *IEEE Trans. Inf. Theory* **2006**, *52*, 3936–3964. [\[CrossRef\]](#)
13. Liu, T.; Viswanath, P. An Extremal Inequality Motivated by Multiterminal Information-Theoretic Problems. *IEEE Trans. Inf. Theory* **2007**, *53*, 1839–1851. [\[CrossRef\]](#)
14. Geng, Y.; Nair, C. The Capacity Region of the Two-Receiver Gaussian Vector Broadcast Channel with Private and Common Messages. *IEEE Trans. Inf. Theory* **2014**, *60*, 2087–2104. [\[CrossRef\]](#)
15. Steinberg, Y. Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information. *IEEE Trans. Inf. Theory* **2005**, *51*, 2867–2877. [\[CrossRef\]](#)
16. Steinberg, Y.; Shamai, S. Achievable rates for the broadcast channel with states known at the transmitter. In Proceedings of the International Symposium on Information Theory, Adelaide, Australia, 4–9 September 2005; pp. 2184–2188. [\[CrossRef\]](#)
17. Ghabeli, L. On the capacity of a class of K-user Gaussian broadcast channel with states known at the transmitter. *IET Commun.* **2018**, *12*, 787–795. [\[CrossRef\]](#)
18. Khisti, A.; Erez, U.; Lapidith, A.; Wornell, G.W. Carbon Copying Onto Dirty Paper. *IEEE Trans. Inf. Theory* **2007**, *53*, 1814–1827. [\[CrossRef\]](#)
19. Rini, S.; Shamai, S. On the Capacity of the Carbon Copy onto Dirty Paper Channel. *IEEE Trans. Inf. Theory* **2017**, *63*, 5907–5922. [\[CrossRef\]](#)
20. Khosravi-Farsani, R.; Marvasti, F. Capacity bounds for multiuser channels with non-causal channel state information at the transmitters. In Proceedings of the 2011 IEEE Information Theory Workshop, Paraty, Brazil, 16–20 October 2011; pp. 195–199. [\[CrossRef\]](#)
21. Kotagiri, S.; Laneman, J.N. Achievable rates for multiple access channels with state information known at one encoder. In Proceedings of the Allerton Conference on Communications, Control, and Computing, Monticello, IL, USA, 29 September–1 October 2004.
22. Kotagiri, S.; Laneman, J.N. Multiaccess Channels with State Known to One Encoder: A Case of Degraded Message Sets. In Proceedings of the 2007 IEEE International Symposium on Information Theory, Nice, France, 24–29 June 2007; pp. 1566–1570. [\[CrossRef\]](#)

23. Kotagiri, S.P.; Laneman, J.N. Multiaccess Channels with State Known to Some Encoders and Independent Messages. *EURASIP J. Wirel. Commun. Netw.* **2008**, *2008*, 26:1–26:14. [\[CrossRef\]](#)
24. Somekh-Baruch, A.; Shamai, S.; Verdu, S. Cooperative Multiple-Access Encoding With States Available at One Transmitter. *IEEE Trans. Inf. Theory* **2008**, *54*, 4448–4469. [\[CrossRef\]](#)
25. Zaidi, A.; Kotagiri, S.P.; Laneman, J.N.; Vandendorpe, L. Multiaccess channels with state known to one encoder: Another case of degraded message sets. In Proceedings of the 2009 IEEE International Symposium on Information Theory, Seoul, Korea, 28 June–3 July 2009; pp. 2376–2380. [\[CrossRef\]](#)
26. Yang, W.; Liang, Y.; Shitz, S.S.; Poor, H.V. State-Dependent Gaussian Multiple Access Channels: New Outer Bounds and Capacity Results. *IEEE Trans. Inf. Theory* **2018**, *64*, 7866–7882. [\[CrossRef\]](#)
27. Mallik, S.; Koetter, R. Helpers for Cleaning Dirty Papers. In Proceedings of the 7th International ITG Conference on Source and Channel Coding, Ulm, Germany, 14–16 January 2008; pp. 1–5.
28. Sun, Y.; Duan, R.; Liang, Y.; Khisti, A.; Shamai, S. Capacity Characterization for State-Dependent Gaussian Channel With a Helper. *IEEE Trans. Inf. Theory* **2016**, *62*, 7123–7134. [\[CrossRef\]](#)
29. Zaidi, A.; Kotagiri, S.P.; Laneman, J.N.; Vandendorpe, L. Cooperative Relaying with State Available Noncausally at the Relay. *IEEE Trans. Inf. Theory* **2010**, *56*, 2272–2298. [\[CrossRef\]](#)
30. Li, M.; Simeone, O.; Yener, A. Message and State Cooperation in a Relay Channel When Only the Relay Knows the State. *ArXiv* **2011**, arXiv:1102.0768.
31. Boostanpour, J.; Hodtani, G.A. Impact of relay side information on the coverage region for the wireless relay channel with correlated noises. *IET Commun.* **2018**, *12*, 776–786. [\[CrossRef\]](#)
32. Ghasemi-Goojani, S.; Papadimitratos, P. On the capacity of Gaussian “dirty” Z-interference channel with common state. In Proceedings of the 2018 52nd Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA, 21–23 March 2018; pp. 1–6. [\[CrossRef\]](#)
33. Keshet, G.; Steinberg, Y.; Merhav, N. Channel coding in the presence of side information. *Found. Trends Commun. Inf. Theory* **2008**, *4*, 445–586. [\[CrossRef\]](#)
34. Duan, R.; Liang, Y.; Khisti, A.; Shamai, S. State-Dependent Parallel Gaussian Networks With a Common State-Cognitive Helper. *IEEE Trans. Inf. Theory* **2015**, *61*, 6680–6699. [\[CrossRef\]](#)
35. Dikshtein, M.; Duan, R.; Liang, Y.; Shamai, S. Parallel Gaussian Channels Corrupted by Independent States With a State-Cognitive Helper. In Proceedings of the IEEE International Conference on the Science of Electrical Engineering (ICSEE), Eilat, Israel, 12–14 December 2018.
36. Dikshtein, M.; Duan, R.; Liang, Y.; Shamai, S. State-Dependent Parallel Gaussian Channels with a State-Cognitive Helper. In Proceedings of the International Zurich Seminar on Information and Communication (IZS 2018), ETH Zurich, Zurich, Switzerland, 21–23 February 2018.
37. El Gamal, A.; Kim, Y. *Network Information Theory*; Cambridge University Press: Cambridge, UK, 2011.
38. Pozrikidis, C. *An Introduction to Grids, Graphs, and Networks*; Oxford University Press: Oxford, UK, 2014.

