

State-Dependent Gaussian Multiple Access Channels: New Outer Bounds and Capacity Results

Wei Yang, *Member, IEEE*, Yingbin Liang, *Senior Member, IEEE*,
Shlomo Shamai (Shitz) *Fellow, IEEE*, and H. Vincent Poor, *Fellow, IEEE*

Abstract

This paper studies a two-user state-dependent Gaussian multiple-access channel (MAC) with state noncausally known at one encoder. Two scenarios are considered: i) each user wishes to communicate an independent message to the common receiver, and ii) the two encoders send a common message to the receiver and the non-cognitive encoder (i.e., the encoder that does not know the state) sends an independent individual message (this model is also known as the MAC with degraded message sets). For both scenarios, new outer bounds on the capacity region are derived, which improve uniformly over the best known outer bounds. In the first scenario, the two corner points of the capacity region as well as the sum rate capacity are established, and it is shown that a single-letter solution is adequate to achieve both the corner points and the sum rate capacity. Furthermore, the full capacity region is characterized in situations in which the sum rate capacity is equal to the capacity of the helper problem. The proof exploits the optimal-transportation idea of Polyanskiy and Wu (which was used previously to establish an outer bound on the capacity region of the interference channel) and the worst-case Gaussian noise result for the case in which the input and the noise are dependent.

The work of W. Yang and H. V. Poor was supported by the U. S. National Science Foundation under Grants ECCS-1343210 and ECCS-1647198. The work of Y. Liang was supported by the U. S. National Science Foundation under Grant CCF-1618127. The work of S. Shamai (Shitz) was supported by the European Union's Horizon 2020 Research And Innovation Programme, under grant agreement no. 694630. The material of this paper will be presented in part at the IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, June 2017.

W. Yang and H. V. Poor are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (email: {weiy, poor}@princeton.edu).

Y. Liang is with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 USA (email: yliang06@syr.edu).

S. Shamai (Shitz) is with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Technion City, Haifa 32000, Israel (email: sshlomo@ee.technion.ac.li).

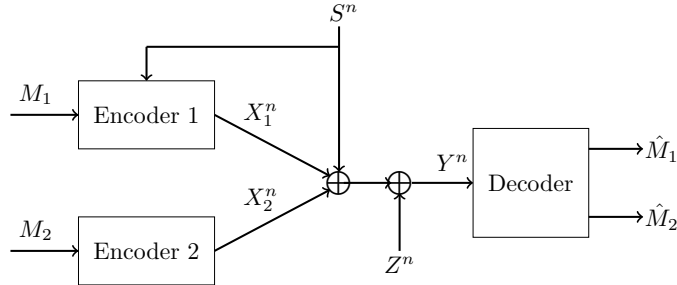


Figure 1. State-dependent Gaussian MAC with state available noncausally at one encoder without degraded message sets.

I. INTRODUCTION

We study a two-user state-dependent Gaussian multiple-access channel (MAC) with the state noncausally known at one encoder. The channel input-output relationship for a single channel use is given by

$$Y = X_1 + X_2 + S + Z \quad (1)$$

where $Z \sim \mathcal{N}(0, 1)$ denotes the additive white Gaussian noise, and X_1 and X_2 are the channel inputs from two users, which are subject to the (average) power constraints P_1 and P_2 , respectively. The state $S \sim \mathcal{N}(0, Q)$ is known noncausally at encoder 1 (state-cognitive user), but is not known at encoder 2 (non-cognitive user) nor at the decoder. This channel model generalizes Costa’s dirty-paper channel [1] to the multiple-access setting, and is also known as “dirty MAC” or “MAC with a single dirty user” [2]. In this paper, we consider the following two scenarios:

- i) Each user wishes to communicate an *independent* message to the common receiver, where the state-cognitive user sends the message M_1 and the non-cognitive user sends M_2 (see Fig. 1);
- ii) The state-cognitive encoder sends the message M_1 and the non-cognitive encoder sends both M_1 and M_2 (see Fig. 2). In this case, the message M_1 can be also viewed as a common message.

We shall refer to the first setting as the “dirty MAC without degraded message sets”, and the second setting as the “dirty MAC with degraded message sets”.

Although the dirty MAC (with and without degraded message sets) described in (1) has been studied extensively in the literature [2]–[5], no single-letter expression for the capacity

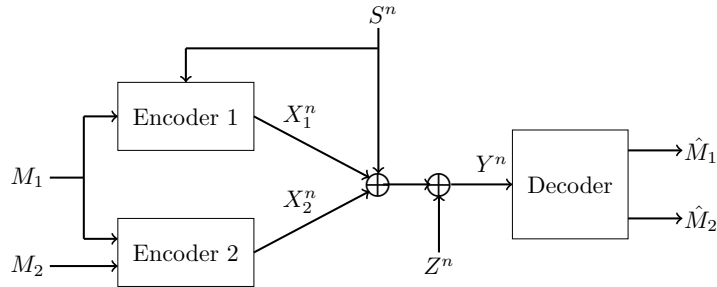


Figure 2. State-dependent Gaussian MAC with state available noncausally at one encoder with degraded message sets.

region is characterized to date. For the dirty MAC without degraded message sets, Kotagiri and Laneman [3] derived an inner bound on the capacity region using a generalized dirty paper coding scheme at the cognitive encoder, which allows arbitrary correlation between the input X_1 and the state S . Philosof *et al.* [2] showed that the same rate region can be achieved by using lattice-based transmission. In general, it is not clear whether a single-letter solution (i.e., random coding/random binning using independent and identically distributed (i.i.d.) copies of a certain scalar distribution) is optimal for the dirty MAC (1). However, as [2] and [4] demonstrated, a single-letter solution is suboptimal for the *doubly-dirty* MAC, in which the output is corrupted by two states, each known at one encoder noncausally (see also [6]). In this case, (linear) structured lattice coding outperforms the best known single-letter solution. An inner bound for the dirty MAC with degraded message sets was derived in [5], which uses superposition coding at the non-cognitive encoder to send the two messages M_1 and M_2 .

On the converse side, all existing outer bounds for the dirty MAC without degraded message sets are obtained by assuming that a genie provides auxiliary information to the encoders/decoder. For example, by revealing the state to the decoder, one obtains an outer bound given by the capacity region of the Gaussian MAC without state dependence. In [5], Zaidi *et al.* derived an outer bound on the capacity region of the dirty MAC with degraded message sets, which also serves as an outer bound for the dirty MAC without degraded message sets. Somekh-Baruch *et al.* [7] considered the setting in which the cognitive encoder knows the message of the non-cognitive encoder (i.e., the roles of the two encoders are reversed), and derived the exact capacity region (see also [8]). Interestingly, this capacity region remains valid if the non-cognitive encoder

processes strictly causal state information [9].

Different variants of the dirty MAC model in (1) have also been investigated in the literature. A special case of the dirty MAC model is the “helper problem” [10], in which the cognitive user does not send any information, and its goal is to help the non-cognitive user. For the helper problem, the capacity (of the non-cognitive user) is known for a wide range of channel parameters [11]. The authors in [12] and [13] considered the case in which the state is known only strictly causally or causally at the cognitive encoder, and derived inner and outer bounds on the capacity region. The capacity region of the MAC with action-dependent states was established in Dikstein *et al.* [14]. Finally, Wang [15] characterized the capacity region of the K -user dirty MAC to within a bounded gap. For a general account of state-dependent multiuser models, we refer the reader to [16] and [17].

The main contributions of this paper are the establishment of new outer bounds on the capacity region of the dirty MAC given in (1) with and without degraded message sets. In both scenarios, our bounds improve uniformly over the best known outer bounds (see Fig. 3–Fig. 6 for numerical examples). For the dirty MAC without degraded message sets, the new outer bounds allow us to characterize the two corner points of the capacity region as well as the sum rate capacity (note that, unlike [2], we do not assume $Q \rightarrow \infty$). In this case, a single-letter solution is shown to be adequate to achieve both the corner points and the sum rate capacity. Furthermore, the full capacity region of the dirty MAC without degraded message sets is established in situations in which the sum rate capacity coincides with the capacity of the helper problem.

The proof of our outer bounds builds on a recent technique proposed by Polyanskiy and Wu [18] that bounds the difference of the differential entropies of two probability distributions via their quadratic Wasserstein distance and via Talagrand’s transportation inequality [19]. It also relies on a generalized version of the worst-case Gaussian noise result, in which the Gaussian input and the noise are dependent (but are uncorrelated) [20]–[22]. We anticipate that these techniques can be useful more broadly for other state-dependent multiuser models, such as state-dependent interference channels and relay channels.

II. PROBLEM SETUP AND PREVIOUS RESULTS

A. Problem Setup

Consider the Gaussian MAC (1) with additive Gaussian state noncausally known at encoder 1 depicted in Fig. 1 and Fig. 2. The state $S \sim \mathcal{N}(0, Q)$ is independent of the additive white Gaussian noise $Z \sim \mathcal{N}(0, 1)$ and of the input X_2 of the non-cognitive encoder. The state and the noise are i.i.d. over channel uses. For the dirty MAC without degraded message sets (Fig. 1), we assume that encoder 1 and encoder 2 must satisfy the (average) power constraints¹

$$\sum_{i=1}^n \mathbb{E}[X_{1,i}^2(M_1, S^n)] \leq nP_1 \quad (2)$$

$$\sum_{i=1}^n \mathbb{E}[X_{2,i}^2(M_2)] \leq nP_2 \quad (3)$$

where the index i denotes the channel use, and M_1 and M_2 denote the transmitted messages, which are independently and uniformly distributed. The decoder reconstructs the transmitted messages M_1 and M_2 from the channel output, and outputs \hat{M}_1 and \hat{M}_2 . The (average) probability of error is defined as

$$P_e \triangleq \mathbb{P}[(M_1, M_2) \neq (\hat{M}_1, \hat{M}_2)]. \quad (4)$$

If the message sets are degraded (Fig. 2), then the power constraint (3) becomes

$$\sum_{i=1}^n \mathbb{E}[X_{2,i}^2(M_1, M_2)] \leq nP_2. \quad (5)$$

The capacity regions for the dirty MAC with and without degraded message sets are denoted by $\mathcal{C}_{\text{deg}}(P_1, P_2, Q)$ and $\mathcal{C}(P_1, P_2, Q)$, respectively. Note that, by definition,

$$\mathcal{C}(P_1, P_2, Q) \subseteq \mathcal{C}_{\text{deg}}(P_1, P_2, Q). \quad (6)$$

In both scenarios, a single-letter characterization for the capacity region is not known in the literature. In Section II-B below, we review the existing inner and outer bounds on $\mathcal{C}_{\text{deg}}(P_1, P_2, Q)$ and $\mathcal{C}(P_1, P_2, Q)$.

¹Note that, the authors of [2] and [7] assumed *per-codeword* power constraints, i.e., for all messages m_1 and m_2 , the codewords x_1^n and x_2^n satisfy $\sum_{i=1}^n x_{1,i}^2(m_1, S^n) \leq nP_1$ and $\sum_{i=1}^n X_{2,i}^2(m_2) \leq nP_2$ almost surely. Clearly, every outer bound for the average power constraint is also a valid outer bound for the *per-codeword* power constraint.

B. Previous Results

For the dirty MAC without degraded message sets, the best known achievable rate region was derived by Kotagiri and Laneman [3], and is given by the convex hull of the rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(U; Y|X_2) - I(U; S) \quad (7)$$

$$R_2 \leq I(X_2; Y|U) \quad (8)$$

$$R_1 + R_2 \leq I(U, X_2; Y) - I(U, S) \quad (9)$$

for some joint probability distribution $P_{U X_1 | S} P_{X_2}$. A computable inner bound was obtained in [3] from (7)–(9) by setting

$$P_{X_1 | S=s} = \mathcal{N}\left(\rho\sqrt{P_1/Q}s, P_1(1 - \rho^2)\right) \quad (10)$$

$$P_{X_2} = \mathcal{N}(0, P_2) \quad (11)$$

$$U = X_1 - \rho\sqrt{\frac{P_1}{Q}}S + \alpha \left(1 + \rho\sqrt{\frac{P_1}{Q}}\right) S \quad (12)$$

for some $\rho \in [-1, 0]$ and $\alpha \in \mathbb{R}$. This choice of input distribution is also known as *generalized dirty paper coding*. Unlike in the point-to-point setting [1], allowing a (negative) correlation between X_1 and S may be beneficial since it partially cancels the state for the non-cognitive encoder. However, it is not clear whether the Gaussian distribution optimizes the bounds in (7)–(9).

The best known outer bound is given by the region of rate pairs (R_1, R_2) satisfying²

$$R_1 \leq \frac{1}{2} \log(1 + P_1(1 - \rho_1^2 - \rho_s^2)) \quad (13)$$

$$R_2 \leq \frac{1}{2} \log\left(1 + \frac{P_2(1 - \rho_1^2 - \rho_s^2)}{1 - \rho_s^2}\right) \quad (14)$$

$$R_1 + R_2 \leq \frac{1}{2} \log(1 + P_1(1 - \rho_1^2 - \rho_s^2)) + \frac{1}{2} \log\left(1 + \frac{(\sqrt{P_2} + \rho_1\sqrt{P_1})^2}{1 + P_1(1 - \rho_1^2 - \rho_s^2) + (\sqrt{Q} + \rho_s\sqrt{P_1})^2}\right) \quad (15)$$

$$R_1 + R_2 \leq \frac{1}{2} \log(1 + P_1 + P_2) \quad (16)$$

²In this paper, the logarithm (log) and exponential (exp) functions are taken with respect to an arbitrary basis.

for some $\rho_1 \in [0, 1]$ and $\rho_s \in [-1, 0]$ that satisfy $\rho_1^2 + \rho_s^2 \leq 1$. This outer bound is a combination of several (genie-aided) outer bounds established in the literature:

- The bounds (14) and (15) form the outer bound in [5] on $\mathcal{C}_{\text{deg}}(P_1, P_2, Q)$, and hence on $\mathcal{C}(P_1, P_2, Q)$.
- The bounds (13) and (15) characterize the capacity region of the dirty MAC under the assumption that the cognitive user knows the message of the non-cognitive user [7].
- The bound (16) upper-bounds the sum rate of the Gaussian MAC without state dependence.

For the dirty MAC with degraded message sets, inner and outer bounds on the capacity region were derived in [5]. As reviewed above, the capacity region $\mathcal{C}_{\text{deg}}(P_1, P_2, Q)$ is outer-bounded by the region with rate pairs (R_1, R_2) satisfying (14) and (15). This outer bound follows from the following single-letter outer region [5, Th. 2]:

$$R_2 \leq I(X_2; Y|S, X_1) \quad (17)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y|S) - I(S; X_2|Y) \quad (18)$$

where the joint probability distributions of X_1 , X_2 , and S must be of the form $P_S P_{X_2} P_{X_1|X_2, S}$. The inner bound in [5] consists of rate pairs (R_1, R_2) satisfying

$$R_2 \leq I(X_2; Y|U_1, U_2) \quad (19)$$

$$R_2 \leq I(X_2, U_2; Y|U_1) - I(U_2; S|U_1) \quad (20)$$

$$R_1 + R_2 \leq I(X_2, U_1, U_2; Y) - I(U_2; S|U_1) \quad (21)$$

for some joint probability distributions $P_S P_{U_1} P_{X_2|U_1} P_{U_2|U_1, S} P_{X_1|U_1, U_2, S}$ that satisfy

$$I(U_2; Y|U_1, X_1) - I(U_2; S|U_1) \geq 0. \quad (22)$$

This inner bound is evaluated in [5] for the case in which (X_1, X_2, U_1, U_2, S) are jointly Gaussian distributed. Again, it is not known whether the Gaussian input optimizes the bound.

C. The Helper Problem

As reviewed in the introduction, the dirty MAC model includes the helper problem as a special case. More specifically, in the helper problem, the cognitive user (also known as the *helper*) does

not send any information, and its goal is to assist the non-cognitive user by canceling the state. The capacity of the helper problem is defined as

$$C_{\text{helper}} \triangleq \max\{R_2 : (0, R_2) \in \mathcal{C}(P_1, P_2, Q)\} \quad (23)$$

$$= \max\{R_2 : (0, R_2) \in \mathcal{C}_{\text{deg}}(P_1, P_2, Q)\}. \quad (24)$$

The equivalence between (23) and (24) follows since $I(M_1; X_2^n) = 0$ regardless of whether the message sets are degraded or not.

The capacity of the helper problem was studied in [10] and [11], and is known for a wide range of channel parameters. More specifically, it was shown that [11, Th. 2]

$$C_{\text{helper}} = \frac{1}{2} \log(1 + P_2) \quad (25)$$

provided that P_1 , P_2 , and Q satisfy the following condition.

Condition 1: There exists an $\alpha \in [1 - \sqrt{P_1/Q}, 1 + \sqrt{P_1/Q}]$ such that

$$(P_1 - (\alpha - 1)^2 Q)^2 \geq \alpha^2 Q (P_2 + 1 - P_1 + (\alpha - 1)^2 Q). \quad (26)$$

In other words, if Condition 1 is satisfied, then the state can be perfectly canceled, and the non-cognitive user achieves the channel capacity without state dependence. Note that, to satisfy Condition 1 it is not necessary that $P_1 \geq Q$ (e.g., (26) holds as long as $P_1 \geq P_2 + 1$, regardless of the value of Q).

III. MAIN RESULTS

The main results of this paper are the establishment of several new outer bounds on the capacity region of the dirty MAC (1) with and without degraded message sets. For notational convenience, we denote

$$C_1 \triangleq \frac{1}{2} \log(1 + P_1), \quad C_2 \triangleq \frac{1}{2} \log(1 + P_2). \quad (27)$$

A. Dirty MAC Without Degraded Message Sets

1) *New outer bounds:* In this section, we present two outer bounds on $\mathcal{C}(P_1, P_2, Q)$.

Theorem 1: The capacity region $\mathcal{C}(P_1, P_2, Q)$ of the dirty MAC without degraded message sets is outer-bounded by the region with rate pairs (R_1, R_2) satisfying

$$R_2 \leq C_{\text{helper}} \quad (28)$$

and

$$R_1 \leq \min_{0 \leq \delta \leq 1} \left\{ \frac{1}{2} \log \left(1 + \frac{1 + P_2 - \delta}{P_2 \delta} g(R_2) \right) + f(\delta) \right\} \quad (29)$$

where

$$g(R_2) \triangleq \exp \left(2c_1 \sqrt{C_2 - R_2} + 2(C_2 - R_2) \right) - 1 \quad (30)$$

with

$$c_1 \triangleq \frac{3\sqrt{1 + (\sqrt{P_1} + \sqrt{Q})^2 + P_2} + 4(\sqrt{P_1} + \sqrt{Q})}{\sqrt{(1 + P_2)/(2 \log e)}} \quad (31)$$

and

$$f(\delta) \triangleq \max_{\rho \in [-1, 0]} \frac{1}{2} \left\{ \log \frac{1 + P_2 + P_1 + Q + 2\rho\sqrt{P_1 Q}}{\delta + P_1 + Q + 2\rho\sqrt{P_1 Q}} + \log \frac{\delta + (1 - \rho^2)P_1}{1 + P_2} \right\}. \quad (32)$$

Proof: See Section IV-A. ■

Remark 1: The objective function on the right-hand side (RHS) of (32) is concave in ρ for every $\delta \in [0, 1]$.

Remark 2: The upper bound (29) can be slightly improved by replacing Q on the RHS of (29) with $\tilde{Q} \leq Q$ and by minimizing over \tilde{Q} . This follows because, for a fixed rate R_2 , the maximum achievable R_1 is monotonically non-increasing in Q , whereas the RHS of (29) is not.

We next illustrate the main intuition behind Theorem 1. To concentrate ideas, we assume that the channel parameters P_1 , P_2 , and Q satisfy Condition 1, which implies that $C_{\text{helper}} = C_2$ [11, Th. 2]. Consider two auxiliary channels

$$Y_G^n \triangleq X_1^n + S^n + G^n + Z^n \quad (33)$$

$$Y_\delta^n \triangleq X_1^n + S^n + \sqrt{\delta} Z^n \quad (34)$$

where $G^n \sim \mathcal{N}(0, P_2 \mathbf{I}_n)$ is a Gaussian vector having the same power as X_2^n , and $\delta \in (0, 1)$ is a constant. In words, Y_G^n is obtained from Y^n by replacing the codeword X_2^n with Gaussian interference of the same power, and Y_δ^n is obtained from Y^n by removing the interference X_2^n and by increasing the signal-to-noise ratio (SNR). Therefore, the channel $M_1 \rightarrow Y_G^n$ is worse than the original channel whereas the channel $M_1 \rightarrow Y_\delta^n$ is better than the original one. In fact, we argue next that, when the non-cognitive user is communicating at a rate close to its maximum rate C_2 , the three channels have approximately the same rate for the cognitive user.

Indeed, suppose that $R_2 \approx C_2$. Then, on the one hand, the distribution of X_2^n is close to that of G^n , and hence

$$I(X_1^n + S^n; Y_G^n) \approx I(X_1^n + S^n; Y^n). \quad (35)$$

On the other hand, since the receiver is able to decode the message of the non-cognitive user, it follows that

$$I(X_1^n + S^n; Y^n) \approx I(X_1^n + S^n; Y^n | X_2^n) \quad (36)$$

$$= I(X_1^n + S^n; X_1^n + S^n + Z^n). \quad (37)$$

Combining (35) and (37), we conclude that

$$\begin{aligned} & I(X_1^n + S^n; X_1^n + S^n + G^n + Z^n) \\ & \approx I(X_1^n + S^n; X_1^n + S^n + Z^n). \end{aligned} \quad (38)$$

In other words, reducing the power of the Gaussian noise (from $1 + P_2$ to 1) does not (significantly) increase the mutual information between $X_1^n + S^n$ and the output. By further reducing the noise power, we obtain

$$I(X_1^n + S^n; Y^n) \approx I(X_1^n + S^n; Y_G^n) \approx I(X_1^n + S^n; Y_\delta^n). \quad (39)$$

The errors in the estimation (39) can be bounded via Costa's entropy power inequality [23] or the I-MMSE relation [24].

To see how the relation (39) can be used to upper-bound R_1 , we note that by standard manipulations of mutual information,

$$nR_1 \leq I(X_1^n + S^n; Y^n) - I(S^n; Y^n). \quad (40)$$

By (39), we may replace the two Y^n 's on the RHS of (40) with Y_G^n and Y_δ^n , respectively, and obtain

$$nR_1 \lesssim I(X_1^n + S^n; Y_G^n) - I(S^n; Y_\delta^n) \quad (41)$$

$$\lesssim n \max_{P_{X_1|S}} \left\{ I(X_1 + S; Y_G) - I(S; Y_\delta) \right\} \quad (42)$$

where

$$Y_G \triangleq X_1 + S + G + Z \quad (43)$$

$$Y_\delta = X_1 + S + \sqrt{\delta}Z \quad (44)$$

are the single-letter versions of Y_G^n and Y_δ^n , respectively. By the Gaussian saddle point property (namely, the Gaussian distribution is the best input distribution for Gaussian noise, and is the worst noise distribution for a Gaussian input), we expect that the RHS of (42) is maximized when (X_1, S) are jointly Gaussian. The maximum of the objective function on the RHS of (42) is precisely the $f(\delta)$ defined in (32), whereas the logarithm term on the RHS of (29) quantifies the error in the approximation (39), which vanishes as $R_2 \rightarrow C_2$. The rigorous proof of Theorem 1 which builds upon the above intuition can be found in Section IV-A.

The outer bound provided in Theorem 1 improves the best known outer bound in the regime where R_2 is close to C_2 (provided that C_{helper} is also close to C_2). The next theorem provides a tighter upper bound on the sum rate than (15) and (16).

Theorem 2: The capacity region $\mathcal{C}(P_1, P_2, Q)$ of the dirty MAC without degraded message sets is outer-bounded by the region with rate pairs (R_1, R_2) satisfying

$$R_1 \leq \frac{1}{2} \log(1 + P_1(1 - \rho^2)) \quad (45)$$

$$R_2 \leq C_2 \quad (46)$$

$$R_1 + R_2 \leq \frac{1}{2} \log\left(1 + \frac{P_2}{1 + P_1 + Q + 2\rho\sqrt{P_1Q}}\right) + \frac{1}{2} \log(1 + P_1(1 - \rho^2)) \quad (47)$$

for some $\rho \in [-1, 0]$.

Proof: The proof of Theorem 2 follows from the following single-letter outer bound on the capacity region.

Proposition 3: The capacity region $\mathcal{C}(P_1, P_2, Q)$ of the dirty MAC without degraded message sets is outer-bounded by the region with rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(X_1; Y|X_2, S) \quad (48)$$

$$R_2 \leq I(X_2; Y|X_1, S) \quad (49)$$

$$R_1 + R_2 \leq I(X_1; Y|X_2, S) + I(X_2; Y) \quad (50)$$

for some joint distributions $P_S P_{X_1|S} P_{X_2}$ that satisfy the power constraint

$$\mathbb{E}[X_1^2] \leq P_1 \text{ and } \mathbb{E}[X_2^2] \leq P_2. \quad (51)$$

Proof: See Section IV-B. ■

It is not difficult to show that the outer bound in Proposition 3 is maximized when S , X_1 , and X_2 are jointly Gaussian distributed (proof omitted). Evaluating (48)–(50) for Gaussian distributions $P_S P_{X_1|S} P_{X_2}$, we obtain the outer bound in Theorem 2. ■

2) *Sum rate capacity*: Let C_{sum} be the sum rate capacity of the dirty MAC (1) without degraded message sets, i.e.,

$$C_{\text{sum}} \triangleq \max\{R_1 + R_2 : (R_1, R_2) \in \mathcal{C}(P_1, P_2, Q)\}. \quad (52)$$

By comparing the inner bound (9) (evaluated using Gaussian inputs) and the outer bound (47), we establish the sum rate capacity C_{sum} .

Theorem 4: The sum rate capacity of the dirty MAC without degraded message sets is given by

$$C_{\text{sum}} = \max_{\rho \in [-1, 0]} \frac{1}{2} \left\{ \log \left(1 + \frac{P_2}{1 + P_1 + Q + 2\rho\sqrt{P_1 Q}} \right) + \frac{1}{2} \log(1 + P_1(1 - \rho^2)) \right\} \quad (53)$$

or equivalently,

$$C_{\text{sum}} = C_2 + f(1). \quad (54)$$

Proof: The converse part of (53) follows directly from (47). Since the objective function on the RHS of (53) is continuous and concave in $\rho \in [-1, 0]$ (see Remark 1), it has a unique maximizer on $[-1, 0]$, which we denote by ρ^* . It follows that the rate pair

$$\bar{R}_1 \triangleq \frac{1}{2} \log(1 + P_1(1 - (\rho^*)^2)) \quad (55)$$

$$\bar{R}_2 \triangleq \frac{1}{2} \log \left(1 + \frac{P_2}{1 + P_1 + Q + 2\rho^*\sqrt{P_1 Q}} \right) \quad (56)$$

is achievable by treating the interference $X_1 + S$ as noise for the non-cognitive user, and by using generalized dirty paper coding for the cognitive user with $\rho = \rho^*$ and

$$\alpha = \frac{P_1(1 - (\rho^*)^2)}{P_1(1 - (\rho^*)^2) + 1} \quad (57)$$

in (10)–(12). The choice of α in (57) is the usual dirty paper coding coefficient for the equivalent channel (obtained by canceling the interference X_2 from the non-cognitive user)

$$\tilde{Y} = X_0 + \left(1 - \rho^* \sqrt{\frac{P_1}{Q}} \right) S + Z \quad (58)$$

where $X_0 \triangleq X_1 - \rho^* \sqrt{P_1/Q} S \sim \mathcal{N}(0, P_1(1 - (\rho^*)^2))$ is independent of S . The rate pair in (55) and (56) achieves the sum rate capacity (53). The equivalence between (53) and (54) is straightforward to establish. ■

The next result shows that, if $C_{\text{helper}} = C_{\text{sum}}$, then the outer bound in Theorem 2 matches the inner bound in (7)–(9) evaluated for Gaussian inputs. In this case, we obtain a complete characterization of the capacity region $\mathcal{C}(P_1, P_2, Q)$.

Corollary 5: For the dirty MAC without degraded messages, if $C_{\text{helper}} = C_{\text{sum}}$, then the capacity region is given by the convex hull of the set of rate pairs (R_1, R_2) satisfying

$$R_1 \leq \frac{1}{2} \log(1 + P_1(1 - \rho^2)) \quad (59)$$

$$R_1 + R_2 \leq \frac{1}{2} \log\left(1 + \frac{P_2}{1 + P_1 + Q + 2\rho\sqrt{P_1Q}}\right) + \frac{1}{2} \log(1 + P_1(1 - \rho^2)) \quad (60)$$

for some $\rho \in [-1, 0]$.

Proof: By Theorem 2, the rate region characterized by (59) and (60), which we denote by $\mathcal{R}^*(P_1, P_2, Q)$, is an outer bound on the capacity region $\mathcal{C}(P_1, P_2, Q)$.

To prove Corollary 5, it suffices to show that the rate region $\mathcal{R}^*(P_1, P_2, Q)$ is achievable. Observe that, by the hypothesis $C_{\text{helper}} = C_{\text{sum}}$, the sum rate capacity is achieved with the rate pairs $(0, C_{\text{helper}})$ and (\bar{R}_1, \bar{R}_2) , where \bar{R}_1 and \bar{R}_2 are defined in (55) and (56), respectively. Let now (R_1, R_2) be an arbitrary point that lies on the boundary of $\mathcal{R}^*(P_1, P_2, Q)$. If $R_1 \leq \bar{R}_1$, then the rate pair $(R_1, C_{\text{sum}} - R_1)$ is achievable using time sharing. Since, by (60), $R_2 \leq C_{\text{sum}} - R_1$, we conclude that the rate pair $(R_1, C_{\text{sum}} - R_1)$ coincides with (R_1, R_2) . If $\bar{R}_1 \leq R_1 \leq C_1$, it follows that there exists an $\rho_0 \in [\rho^*, 0]$ which satisfies $R_1 = \frac{1}{2} \log(1 + P_1(1 - \rho_0^2))$. In this case, we have

$$R_2 = \frac{1}{2} \log\left(1 + \frac{P_2}{1 + P_1 + Q + 2\rho_0\sqrt{P_1Q}}\right). \quad (61)$$

This rate pair is again achievable by treating interference as noise for the non-cognitive user, and by using generalized dirty paper coding for the cognitive user. ■

For the case when $C_{\text{helper}} < C_{\text{sum}}$, the outer bound in Theorem 2 matches the inner bound only for R_1 values greater than a threshold $R_{1,\text{th}}$. This threshold is given by

$$R_{1,\text{th}} = I(U^*; Y) - I(U^*; S) \quad (62)$$

where X_1^* , X_2^* , and U^* are given in (10)–(12) with ρ and α chosen as in the proof of Theorem 4. It is also not difficult to check that $R_{1,\text{th}} > 0$ if and only if $C_{\text{helper}} < C_{\text{sum}}$.

3) *Corner points*: The bounds in Theorems 1 and 2 allow us to characterize the corner points of the capacity region, which are defined as

$$\tilde{C}_1(P_1, P_2, Q) \triangleq \max\{R_1 : (R_1, C_2) \in \mathcal{C}(P_1, P_2, Q)\} \quad (63)$$

$$\tilde{C}_2(P_1, P_2, Q) \triangleq \max\{R_2 : (C_1, R_2) \in \mathcal{C}(P_1, P_2, Q)\}. \quad (64)$$

Corollary 6: For every P_1 , every P_2 , and every Q , we have

$$\tilde{C}_2(P_1, P_2, Q) = \frac{1}{2} \log \left(1 + \frac{P_2}{1 + P_1 + Q} \right). \quad (65)$$

Furthermore, if P_1 , P_2 , and Q satisfy Condition 1, then

$$\tilde{C}_1(P_1, P_2, Q) = f(0) \quad (66)$$

where $f(\cdot)$ is defined in (32).

Proof: The corner point (65) follows from (45) and (47) (with $\rho = 0$), and (66) follows from (29) by setting $R_2 = C_2$, and by taking $\delta \rightarrow 0$. ■

A few remarks are in order.

- The bottom corner point (C_1, \tilde{C}_2) also follows from the (genie-aided) outer bound (13) and (15) developed in [7].
- In the asymptotic limit of strong state power (i.e., $Q \rightarrow \infty$), the two corner points become

$$\lim_{Q \rightarrow \infty} \tilde{C}_1(P_1, P_2, Q) = \frac{1}{2} \log \frac{P_1}{1 + P_2} \quad (67)$$

$$\lim_{Q \rightarrow \infty} \tilde{C}_2(P_1, P_2, Q) = 0. \quad (68)$$

For comparison, existing outer bounds in [2] and [5] only yield the upper bound

$$\lim_{Q \rightarrow \infty} \tilde{C}_1(P_1, P_2, Q) \leq \frac{1}{2} \log \frac{1 + P_1}{1 + P_2}. \quad (69)$$

- The top corner point (\tilde{C}_1, C_2) is achieved by using generalized dirty paper coding with $U = X_1 + S$ and by treating the interference X_2 as noise for the cognitive user. The proof of Theorem 1 suggests that there is essentially no other alternative. Indeed, if $R_2 = C_2 + o(1)$ as $n \rightarrow \infty$, then by (39) and the I-MMSE relation [24], the minimum mean-square error (MMSE) in estimating $X_1^n + S^n$ given Y_G^n satisfies

$$\text{MMSE}(X_1^n + S^n | Y_G^n) = o(n). \quad (70)$$

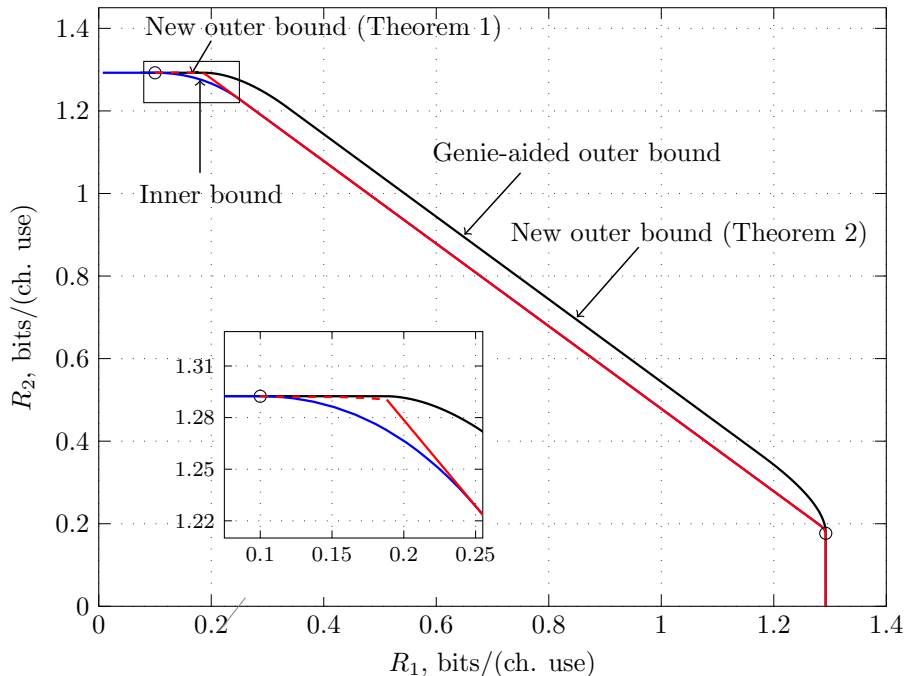


Figure 3. Inner and outer bounds on the capacity region $\mathcal{C}(P_1, P_2, Q)$ with $P_1 = 5$, $P_2 = 5$, and $Q = 12$.

This implies that, in order to achieve $R_2 = C_2 + o(1)$, it is necessary for the decoder to “decode” $X_1^n + S^n$ without knowing the codebook of the non-cognitive user (recall that Y_G^n is obtained from Y^n by replacing the codeword X_2^n with Gaussian interference of the same power).

4) *Numerical results:* In Fig. 3, we compare our new bounds in Theorems 1 and 2 with the inner and outer bounds reviewed in Section II for $P_1 = 5$, $P_2 = 5$, and $Q = 12$. It is not difficult to verify that this set of parameters satisfy Condition 1. We make the following observations from Fig. 3.

- The top corner point of the capacity region is given by the rate pair $(1.29, 0.1)$.
- The outer bound in Theorem 2 matches the inner bound when $R_1 \geq R_{1,\text{th}} = 0.25$ bits/(ch. use).
- In the regime $R_1 \in (0.1, 0.25)$, there is a gap between our outer bounds and the inner bound. This regime can be further divided into two regimes: if $R_1 \in (0.1, 0.19)$, then Theorem 1 yields a tighter upper bound on R_2 ; if $R_1 \in (0.19, 0.25)$, then the bound in Theorem 2 is tighter.

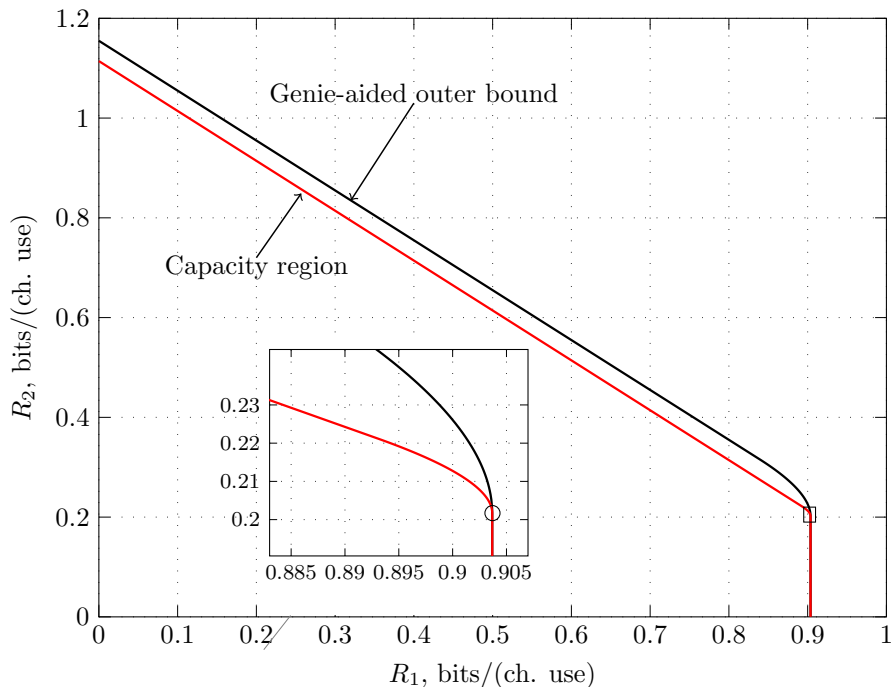


Figure 4. A comparison between the capacity region $\mathcal{C}(P_1, P_2, Q)$ and the genie-aided outer bound with $P_1 = 2.5$, $P_2 = 5$, and $Q = 12$.

Overall, our outer bounds provide a substantial improvement over the genie-aided outer bound in (13)–(16).

In Fig. 4, we consider another set of parameters with $P_1 = 2.5$, $P_2 = 5$, and $Q = 12$. In this case, we have $C_{\text{helper}} = C_{\text{sum}} = 1.11$ bits/(ch. use), and the capacity region $\mathcal{C}(P_1, P_2, Q)$ is completely characterized by Corollary 5. As explained in the proof of Corollary 5, the capacity region consists of three pieces: a straight line connecting the two points $(0, C_{\text{helper}})$ and (\bar{R}_1, \bar{R}_2) , where $\bar{R}_1 = 0.89$ bits/(ch. use) and $\bar{R}_2 = 0.22$ bits/(ch. use), a curved line connecting (\bar{R}_1, \bar{R}_2) and the bottom corner point $(0.9, 0.2)$, and a vertical line connecting the bottom corner point $(0.9, 0.2)$ and $(0.9, 0)$.

5) *Generalization to MAC with non-Gaussian state:* In the proofs of Theorems 1–4, the only place where we have used the Gaussianity of S^n is to optimize appropriate mutual information terms over $P_{X_1|S}$ (see, e.g., (42)). If the state sequence S^n is non-Gaussian but is i.i.d., then the upper bound (29) remains valid if $f(\delta)$ is replaced by

$$\tilde{f}(\delta) \triangleq \max_{P_{X_1|S}} \left\{ I(X_1 + S; Y_G) - I(S; Y_\delta) \right\}. \quad (71)$$

In this case, the top corner point becomes

$$\tilde{C}_1 = \max_{P_{X_1|S}} \{I(X_1 + S; Y_G) - I(X_1 + S; S)\} \quad (72)$$

and the sum rate capacity becomes

$$C_{\text{sum}} = \max_{P_{X_1|S} P_{X_2}} (I(X_1; Y|X_2, S) + I(X_2; Y)). \quad (73)$$

Furthermore, both (53) and (73) can be achieved by treating interference as noise for the non-cognitive user, and by using generalized dirty paper coding for the cognitive user (recall that, in the dirty paper coding problem, the state S does not need to be Gaussian; see, e.g., [25, Sec. 7.7]).

B. Dirty MAC with Degraded Message Sets

Theorem 7 below extends the outer bound in Theorem 1 to the dirty MAC with degraded message sets.

Theorem 7: The capacity region $\mathcal{C}_{\text{deg}}(P_1, P_2, Q)$ of the dirty MAC with degraded message sets is outer-bounded by the region with rate pairs (R_1, R_2) satisfying

$$R_2 \leq C_{\text{helper}} \quad (74)$$

and

$$R_1 \leq \min_{0 \leq \delta \leq 1} \left\{ \frac{1}{2} \log \left(1 + \frac{1 + P_2 - \delta}{P_2 \delta} \tilde{g}(R_2) \right) + f(\delta) \right\} + (c_2 + c_3)(C_2 - R_2) \quad (75)$$

where $f(\cdot)$ is defined in (32),

$$\tilde{g}(R_2) \triangleq \exp \left(2c_2 \sqrt{C_2 - R_2} + 2(C_2 - R_2) \right) - 1 \quad (76)$$

with

$$c_2 \triangleq \frac{3\sqrt{1 + (\sqrt{P_1} + \sqrt{P_2} + \sqrt{Q})^2} + 4(\sqrt{P_1} + \sqrt{Q})}{\sqrt{(1 + P_2)/(2 \log e)}} \quad (77)$$

and

$$c_3 \triangleq \sqrt{2(1 + P_2) \log e} \cdot \left(3\sqrt{1 + (\sqrt{P_1} + \sqrt{P_2} + \sqrt{Q})^2} + 4(\sqrt{P_1} + \sqrt{P_2} + \sqrt{Q}) \right). \quad (78)$$

Proof: See Section IV-C. ■

As a corollary of Theorem 7, we establish that under Condition 1, the top corner point established in (66) is unchanged even if the non-cognitive user knows the message of the cognitive user. Formally, the top corner point is defined as

$$\tilde{C}_{\text{deg},1}(P_1, P_2, Q) \triangleq \max\{R_1 : (R_1, C_2) \in \mathcal{C}_{\text{deg}}(P_1, P_2, Q)\}. \quad (79)$$

Corollary 8: For the dirty MAC with degraded message sets, if P_1 , P_2 , and Q satisfy Condition 1, then

$$\tilde{C}_{\text{deg},1}(P_1, P_2, Q) = f(0) \quad (80)$$

with $f(\cdot)$ defined in (32).

Note that, for the dirty MAC with degraded message sets, both the bottom corner point and the sum rate capacity can be established from the inner and outer bounds in [5].

The next theorem provides an outer bound, which is uniformly tighter than the one in (14) and (15) derived in [5, Th. 4].

Theorem 9: The capacity region of the dirty MAC with degraded message set is outer-bounded by the region with rate pairs (R_1, R_2) satisfying

$$R_2 \leq \frac{1}{2} \log(1 + P_2(1 - \rho_2^2)) \quad (81)$$

$$R_2 \leq \frac{1}{2} \log(1 + P_1(1 - \rho_1^2 - \rho_s^2)) + \frac{1}{2} \log\left(1 + \frac{P_2(1 - \rho_2^2)}{1 + (\sqrt{Q} + \rho_s\sqrt{P_1})^2 + P_1(1 - \rho_1^2 - \rho_s^2)}\right) \quad (82)$$

$$R_1 + R_2 \leq \frac{1}{2} \log(1 + P_1(1 - \rho_1^2 - \rho_s^2)) + \frac{1}{2} \log\left(1 + \frac{P_2(1 - \rho_2^2) + (\rho_2\sqrt{P_2} + \rho_1\sqrt{P_1})^2}{1 + (\sqrt{Q} + \rho_s\sqrt{P_1})^2 + P_1(1 - \rho_1^2 - \rho_s^2)}\right) \quad (83)$$

for some $\rho_1 \in [0, 1]$, $\rho_2 \in [0, 1]$, $\rho_s \in [-1, 0]$ that satisfy

$$\rho_1^2 + \rho_s^2 \leq 1. \quad (84)$$

Proof: The proof of Theorem 9 follows from the following single-letter outer bound on the capacity region, whose proof is given in Section IV-D.

Proposition 10: The capacity region of the dirty MAC with degraded message set is outer-bounded by the region with rate pairs (R_1, R_2) satisfying

$$R_2 \leq I(X_2; Y|X_1, S, U) \quad (85)$$

$$R_2 \leq I(X_1; Y|X_2, S, U) + I(X_2; Y|U) \quad (86)$$

$$R_1 + R_2 \leq I(X_1; Y|X_2, S, U) + I(X_2, U; Y) \quad (87)$$

for some joint distributions $P_{X_1, X_2, S, U}$ that satisfy

- X_1 and X_2 are conditionally independent given U ;
- U and X_2 are independent of S ;
- $\mathbb{E}[X_1^2] \leq P_1$ and $\mathbb{E}[X_2^2] \leq P_2$.

To prove Theorem 9, it remains to show that the bounds in (85)–(87) are maximized when U, S, X_1 , and X_2 are jointly Gaussian. The proof of this result is provided in the appendix. ■

Next, we explain how the outer bound in Proposition 10 improves upon (17) and (18). Observe that (18) can be rewritten as

$$R_1 + R_2 \leq I(X_1; Y|S, X_2) + I(X_2; Y) \quad (88)$$

where the joint probability distribution of S, X_1 , and X_2 has the form $P_S P_{X_2} P_{X_1|X_2, S}$. The key difference between Proposition 10 and the outer bound in (17) and (18) is the introduction of the auxiliary random variable U in Proposition 10. The intuition for this auxiliary random variable is as follows. Since the non-cognitive user knows both messages M_1 and M_2 , its input X_2 must contain two parts, where each part depends only on one message. The auxiliary random variable U in Proposition 10 captures precisely the part of X_2 that depends on M_1 . Since the input X_1 of the cognitive user depends on X_2 only through the message M_1 , and hence through U , we see that X_1 and X_2 are conditionally independent given U , as stated in the proposition. For comparison, the bounds (17) and (18), which allow arbitrary dependence between X_1 and X_2 , is looser than the bound in Proposition 10 (unless $R_2 = 0$, in which case $U = X_2$).

In Figs. 5 and 6, we compare our new outer bound in Theorem 9 with the inner and outer bounds in [5] for different values of P_1, P_2 , and Q . In both figures, the red solid curve denotes our new outer bound in Theorem 9, and the blue dashed curve and the black curve denote the inner and outer bounds obtained in [5]. As expected, our new outer bound is tighter than the outer bound in [5, Th. 4], and is almost on top of the inner bound for the parameters considered

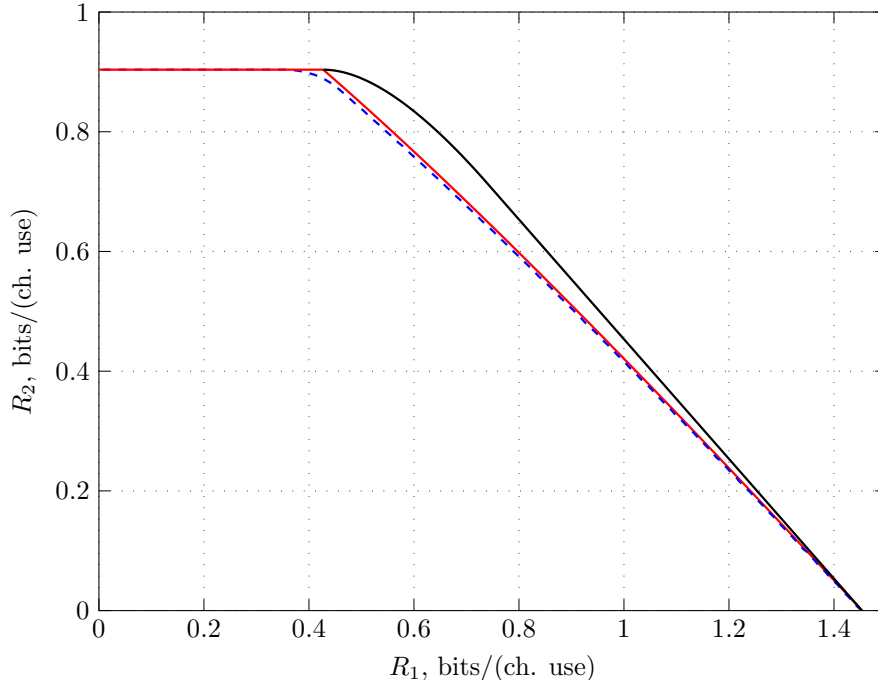


Figure 5. Inner and outer bounds for the capacity region of the dirty MAC with degraded message sets for $P_1 = 4$, $P_2 = 2.5$, and $Q = 5$. The red solid curve denotes our new outer bound in Theorem 9, the blue dashed curve and the black curve denote the inner and outer bounds obtained in [5].

in Figs. 5 and 6. For the scenario considered in Fig. 5, our outer bound does not match the inner bound (unless $R_2 = 0$). Numerically, we observe that the gap between the inner bound and our outer bound is less than 0.013 bits/(ch. use). For the scenario considered in Fig. 6, our outer bound matches the inner bound if either $R_1 \leq 0.1$ or $R_2 = 0$. The gap between the inner and outer bounds in this scenario is less than 3.4×10^{-3} bits/(ch. use).

C. The helper problem

The outer bound in Theorem 1 also yields an upper bound on the capacity of the helper problem as shown in the next result.

Theorem 11: For the helper problem, we have

$$C_{\text{helper}} \leq \max \left\{ R_2 : R_2 \leq C_2, \text{ and } \min_{0 \leq \delta \leq 1} \left\{ \frac{1}{2} \log \left(1 + \frac{1 + P_2 - \delta}{P_2 \delta} g(R_2) \right) + f(\delta) \right\} \geq 0 \right\} \quad (89)$$

where $g(\cdot)$ and $f(\cdot)$ are defined in (30) and (32), respectively.

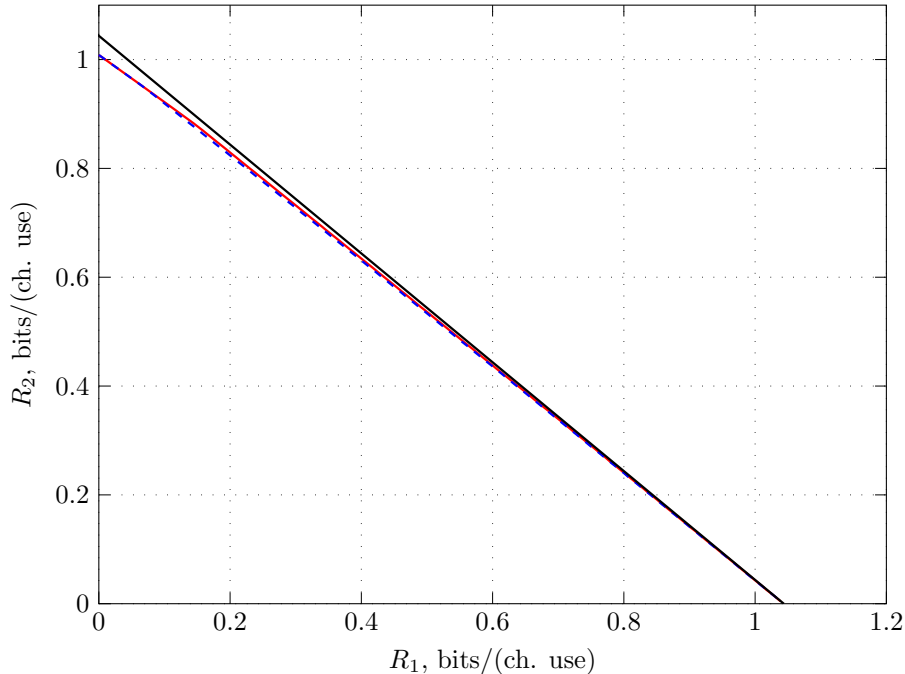


Figure 6. Inner and outer bounds for the capacity region of the dirty MAC with degraded message sets for $P_1 = 2$, $P_2 = 5$, and $Q = 12$. The red solid curve denotes our new outer bound in Theorem 9, the blue dashed curve and the black curve denote the inner and outer bounds obtained in [5].

Proof: Setting $R_1 = 0$ in the outer bound (29) in Theorem 1, we conclude that the rate R_2 of the non-cognitive user must satisfy

$$\min_{0 \leq \delta \leq 1} \left\{ \frac{1}{2} \log \left(1 + \frac{1 + P_2 - \delta}{P_2 \delta} g(R_2) \right) + f(\delta) \right\} \geq 0. \quad (90)$$

This implies (89). ■

A simple consequence of Theorem 11 is the following result, which shows that Condition 1 is both necessary and sufficient for the non-cognitive user to achieve the channel capacity without state dependence.

Corollary 12: For the helper problem, the following two statements are equivalent:

- 1) $C_{\text{helper}} = \frac{1}{2} \log(1 + P_2)$;
- 2) The channel parameters P_1 , P_2 , and Q satisfy Condition 1;
- 3) $f(0) \geq 0$, where $f(\cdot)$ is defined in (32).

In Fig. 7, we compare the new upper bound in Theorem 11 with the upper and lower bounds

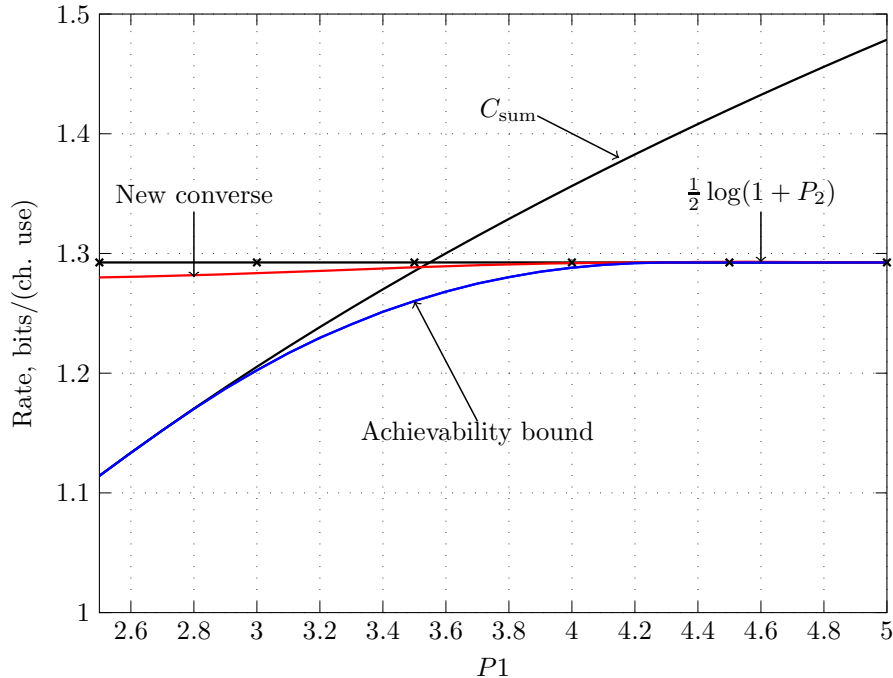


Figure 7. Upper and lower bounds on C_{helper} as a function of P_1 for $P_2 = 5$ and $Q = 12$.

in [11]. The two upper bounds reported in [11, Lemmas 2 and 3] correspond to

$$C_{\text{helper}} \leq C_{\text{sum}} \quad (91)$$

and

$$C_{\text{helper}} \leq \frac{1}{2} \log(1 + P_2) \quad (92)$$

respectively. The lower bound (achievability bound) is [11, Th. 1]. As observed in [11], the upper bound (91) is tight (i.e., $C_{\text{helper}} = C_{\text{sum}}$) if $P_1 \leq 2.5$, and the bound (92) is tight (i.e., $C_{\text{helper}} = \frac{1}{2} \log(1 + P_2)$) if $P_1 \geq 4.5$. Our new upper bound is tighter than (91) and (92) for $P_1 \in [3.5, 4.5]$.

IV. TECHNICAL PROOFS

A. Proof of Theorem 1

The upper bound (28) is straightforward. The proof of (29), which builds upon the intuition described in Section III-A, consists of four steps.

1) We derive an upper bound on

$$I_\delta \triangleq I(X_1^n + S^n; Y_\delta^n) - I(X_1^n + S^n; Y_G^n) \quad (93)$$

that holds for all $X_1^n(M_1, S^n)$ such that the uninformed user is able to communicate at rate R_2 with vanishing error probability. Here, Y_G^n and Y_δ^n are defined in (34) and (33), respectively. The derivation relies on an elegant argument of Polyanskiy and Wu [18], used in the derivation of the outer bound on the capacity region of Gaussian interference channels.

- 2) We obtain a lower bound on I_δ that involves R_1 . Combining this lower bound with the upper bound obtained in the first step, we obtain a multi-letter upper bound on R_1 that depends on the joint distribution of X_1^n and S^n but not on X_2^n .
- 3) We single-letterize the upper bound obtained in Step 2.
- 4) We show that the upper bound obtained in Step 3 is maximized when X_1 and S are jointly Gaussian.

1) *Step 1: Upper-bounding I_δ :* The derivation follows closely the proof of [18, Th. 7]. Let

$$R_1 \triangleq \frac{1}{n} I(M_1; Y^n) \quad (94)$$

$$R_2 \triangleq \frac{1}{n} I(X_2^n; Y^n). \quad (95)$$

As explained in [18], this definition of rate agrees with the operational definition (i.e., the ratio between the logarithm of the number of messages and the blocklength) asymptotically. Without loss of generality, we assume that X_1^n and X_2^n have zero mean. Let

$$N_S(\gamma) \triangleq \exp\left\{\frac{2}{n} h(X_1^n + S^n + \sqrt{\gamma} Z^n)\right\} \quad (96)$$

where $Z^n \sim \mathcal{N}(0, I_n)$ is independent of X_1^n and S^n . By Costa's entropy power inequality [23], the function $N_S(\cdot)$ is concave. The term I_δ in (93) can be expressed in terms of $N_S(\cdot)$ as

$$I_\delta = \frac{n}{2} \log \frac{N_S(\delta)}{N_S(1 + P_2)} + \frac{n}{2} \log \frac{1 + P_2}{\delta}. \quad (97)$$

Repeating the steps in [18, Eqs. (41)–(43)], we obtain (recall that $G^n \sim \mathcal{N}(0, P_2 I_n)$)

$$D(P_{X_2^n + Z^n} \| P_{G^n + Z^n}) \leq n(C_2 - R_2) \quad (98)$$

where $D(\cdot\|\cdot)$ denotes the relative entropy between two distributions, and

$$nR_2 = I(X_2^n; Y^n) \quad (99)$$

$$= h(Y^n) - h(Y_G^n) + h(Y_G^n) - h(X_1^n + S^n + Z^n) \quad (100)$$

$$= h(Y^n) - h(Y_G^n) + \frac{n}{2} \log \frac{N_S(1 + P_2)}{N_S(1)}. \quad (101)$$

Note that $\mathbb{E}[X_1^n + S^n] = \mathbf{0}$, $\mathbb{E}[X_2^n] = \mathbf{0}$, $\mathbb{E}[\|X_2^n\|^2] \leq nP_2$, and

$$\begin{aligned} & \mathbb{E}[\|X_1^n + S^n\|^2] \\ &= \mathbb{E}[\|X_1^n\|^2] + \mathbb{E}[\|S^n\|^2] + 2\mathbb{E}[\langle X_1^n, S^n \rangle] \end{aligned} \quad (102)$$

$$\leq nP_1 + nQ + 2\mathbb{E}[\|X_1^n\| \|S^n\|] \quad (103)$$

$$\leq nP_1 + nQ + 2\sqrt{\mathbb{E}[\|X_1^n\|^2] \mathbb{E}[\|S^n\|^2]} \quad (104)$$

$$\leq n(\sqrt{P_1} + \sqrt{Q})^2. \quad (105)$$

By [18, Prop. 2], the random variable Y_G^n is $(\frac{3\log e}{1+P_2}, \frac{4(\sqrt{P_1}+\sqrt{Q})\log e}{1+P_2})$ -regular, i.e., the probability density function $p_{Y_G^n}(y^n)$ of Y_G^n satisfies

$$\|\nabla \log p_{Y_G^n}(y^n)\| \leq \frac{3\log e}{1+P_2} \|y^n\| + \frac{4(\sqrt{P_1} + \sqrt{Q})\log e}{1+P_2}, \quad \forall y^n \in \mathbb{R}^n. \quad (106)$$

Therefore, by [18, Prop. 1], the entropy difference between Y^n and Y_G^n can be bounded via the Wasserstein distance $W_2(P_{Y^n}, P_{Y_G^n})$ (see [26, p. 12] for the definition of W_2) as

$$\begin{aligned} & h(Y^n) - h(Y_G^n) \\ & \leq \left(3\sqrt{1 + (\sqrt{P_1} + \sqrt{Q})^2 + P_2} + 4(\sqrt{P_1} + \sqrt{Q}) \right) \\ & \quad \cdot \frac{\sqrt{n} \log e}{1 + P_2} \cdot W_2(P_{Y^n} \| P_{Y_G^n}). \end{aligned} \quad (107)$$

Furthermore, we have

$$W_2(P_{Y^n} \| P_{Y_G^n}) \leq W_2(P_{X_2^n + Z^n} \| P_{G^n + Z^n}) \quad (108)$$

$$\leq \sqrt{\frac{2(1 + P_2)}{\log e} D(P_{X_2^n + Z^n} \| P_{G^n + Z^n})} \quad (109)$$

$$\leq \sqrt{\frac{2n(1 + P_2)}{\log e} (C_2 - R_2)}. \quad (110)$$

Here, (108) follows because the $W_2(\cdot, \cdot)$ distance is non-decreasing under convolutions, (109) follows by using Talagrand's inequality [19], and (110) follows from (98). Substituting (110) into (107), and then (107) into (101), we conclude that

$$\log \frac{N_S(1)}{N_S(1+P_2)} \leq 2c_1 \sqrt{C_2 - R_2} + 2(C_2 - R_2) - \log(1+P_2) \quad (111)$$

where c_1 is defined in (175), or equivalently,

$$\frac{N_S(1)}{N_S(1+P_2)} \leq \frac{\exp(2c_1 \sqrt{C_2 - R_2} + 2(C_2 - R_2))}{1+P_2}. \quad (112)$$

Let $\alpha \triangleq P_2/(1+P_2-\delta)$ be such that

$$\alpha\delta + (1-\alpha)(1+P_2) = 1. \quad (113)$$

By the concavity of $N_S(\cdot)$, we have

$$\alpha N_S(\delta) + (1-\alpha)N_S(1+P_2) \leq N_S(1) \quad (114)$$

which implies that

$$\begin{aligned} & \frac{N_S(\delta)}{N_S(1+P_2)} \\ & \leq \frac{1}{\alpha} \frac{N_S(1) - (1-\alpha)N_S(1+P_2)}{N_S(1+P_2)} \end{aligned} \quad (115)$$

$$\leq \frac{1}{\alpha} \left(\frac{\exp(2c_1 \sqrt{C_2 - R_2} + 2(C_2 - R_2))}{1+P_2} - 1 + \alpha \right). \quad (116)$$

Substituting (116) into (97), we conclude that

$$I_\delta \leq \frac{n}{2} \log \left(1 + \frac{1+P_2-\delta}{P_2\delta} g(R_2) \right) \quad (117)$$

where $g(R_2)$ is defined in (30).

2) *Step 2: Lower-bounding I_δ* : We next derive a lower bound on I_δ . Consider the following chain of (in)equalities:

$$I_\delta = I(X_1^n + S^n; Y_\delta^n) - I(X_1^n + S^n; Y_G^n) \quad (118)$$

$$= I(X_1^n, S^n; Y_\delta^n) - I(X_1^n + S^n; Y_G^n) \quad (119)$$

$$= I(X_1^n, S^n; Y_\delta^n, M_1) - I(X_1^n, S^n; M_1 | Y_\delta^n) \\ - I(X_1^n + S^n; Y_G^n) \quad (120)$$

$$= I(X_1^n, S^n; M_1) + I(X_1^n, S^n; Y_\delta^n | M_1) \\ - H(M_1 | Y_\delta^n) - I(X_1^n + S^n; Y_G^n) \quad (121)$$

$$= nR_1 + I(S^n; Y_\delta^n | M_1) + I(X_1^n; Y_\delta^n | S^n, M_1) \\ - H(M_1 | Y_\delta^n) - I(X_1^n + S^n; Y_G^n) \quad (122)$$

$$= nR_1 + I(S^n; Y_\delta^n, M_1) - H(M_1 | Y_\delta^n) \\ - I(X_1^n + S^n; Y_G^n) \quad (123)$$

$$\geq nR_1 + I(S^n; Y_\delta^n) - H(M_1 | Y_\delta^n) \\ - I(X_1^n + S^n; Y_G^n). \quad (124)$$

Here, (119) follows because $(X_1^n, S^n) \rightarrow X_1^n + S^n \rightarrow Y_\delta^n$ forms a Markov chain; (121) follows because $H(M_1 | X_1^n, S^n, Y_\delta^n) = 0$; and finally, (123) follows because S^n is independent of M_1 .

Observe now that the channel $M_1 \rightarrow Y^n$ is stochastically degraded with respect to the channel $M_1 \rightarrow Y_\delta^n$, since Y^n has the same distribution as $Y_\delta^n + X_2^n + \sqrt{1 - \delta^2} \tilde{Z}^n$, where $\tilde{Z}^n \sim \mathcal{N}(0, I_n)$. This implies that a receiver that observes Y_δ^n is able to decode M_1 with vanishing error probability. By Fano's inequality,

$$H(M_1 | Y_\delta^n) = o(n). \quad (125)$$

Here, the $o(n)$ term depends on R_1 and the error probability of the cognitive encoder, but not on the joint probability distribution of X_1^n and S^n . Using (125) in (124) we obtain that

$$I_\delta \geq nR_1 + I(S^n; Y_\delta^n) - I(X_1^n + S^n; Y_G^n) + o(n). \quad (126)$$

Combining the lower bound (126) with the upper bound (117), we conclude that

$$\begin{aligned} nR_1 &\leq I(X_1^n + S^n; Y_G^n) - I(S^n; Y_\delta^n) \\ &\quad + \frac{n}{2} \log \left(1 + \frac{1 + P_2 - \delta}{P_2 \delta} g(R_2) \right) + o(n). \end{aligned} \quad (127)$$

It remains to upper-bound the first two terms on the RHS of (127). This is done in the next two sections.

3) *Step 3: Single-letterization:* Observe that

$$\begin{aligned} I(X_1^n + S^n; Y_G^n) &= \sum_{i=1}^n (h(Y_{G,i} | Y_G^{i-1}) - h(Y_{G,i} | X_{1,i}, S_i)) \end{aligned} \quad (128)$$

$$\leq \sum_{i=1}^n (h(Y_{G,i}) - h(Y_{G,i} | X_{1,i}, S_i)) \quad (129)$$

$$= \sum_{i=1}^n I(X_{1,i} + S_i; Y_{G,i}) \quad (130)$$

and

$$I(S^n; Y_\delta^n) = h(S^n) - h(S^n | Y_\delta^n) \quad (131)$$

$$= \sum_{i=1}^n (h(S_i) - h(S_i | Y_\delta^n, S^{i-1})) \quad (132)$$

$$\geq \sum_{i=1}^n (h(S_i) - h(S_i | Y_{\delta,i})) \quad (133)$$

$$= \sum_{i=1}^n I(S_i; Y_{\delta,i}) \quad (134)$$

where both (129) and (133) follow because conditioning reduces entropy. Combining (130) and (134), we obtain

$$\begin{aligned} I(X_1^n + S^n; Y_G^n) - I(S^n; Y_\delta^n) &\leq \sum_{i=1}^n (I(X_{1,i} + S_i; Y_{G,i}) - I(S_i; Y_{\delta,i})) \end{aligned} \quad (135)$$

where the RHS of (135) depends on $P_{X_1^n | S^n}$ only through the (marginal) conditional distributions $\{P_{X_{1,i} | S_i}\}$.

Now, a critical observation is that the functional $P_{X_1 | S} \mapsto I(X_1 + S; Y_G) - I(S; Y_\delta)$ is concave (recall that Y_G and Y_δ are defined in (43) and (44), respectively). This follows because, for

a fixed channel, mutual information is concave in the input distribution, and for a fixed input distribution, mutual information is convex in the channel (see, e.g., [27, Th. 2.7.3]). Furthermore, both the state sequence S^n and noise sequence Z^n are i.i.d.. This allows us to conclude that

$$\begin{aligned} & I(X_1^n + S^n; Y_G^n) - I(S^n; Y_\delta^n) \\ & \leq n \max_{P_{X_1|S}: \mathbb{E}[X_1^2] \leq P_1} \left\{ I(X_1 + S; Y_G) - I(S; Y_\delta) \right\}. \end{aligned} \quad (136)$$

4) *Optimality of Gaussian inputs:* As explained in the intuitive argument after Theorem 1, we will invoke the Gaussian saddle-point property to solve the maximization problem in (136). Lemma 13 below generalizes the well-known worst-case Gaussian noise result [20], [21] to the case in which the noise and the Gaussian input are dependent.

Lemma 13 ([22, Th. 1]): Let $\mathbf{X}_G \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_x)$ and $\mathbf{Z}_G \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_z)$ be Gaussian random vectors in \mathbb{R}^d . Let \mathbf{Z} be a random vector in \mathbb{R}^d with the same covariance matrix as \mathbf{Z}_G . Assume that \mathbf{X}_G is independent of \mathbf{Z}_G , and that

$$\mathbb{E}[\mathbf{X}_G \mathbf{Z}^T] = \mathbf{0}_{d \times d} \quad (137)$$

where the superscript $(\cdot)^T$ denotes transposition. Then

$$I(\mathbf{X}_G; \mathbf{X}_G + \mathbf{Z}_G) \leq I(\mathbf{X}_G; \mathbf{X}_G + \mathbf{Z}). \quad (138)$$

We proceed as follows. For a given $P_{X_1|S}$, let $\rho \triangleq \mathbb{E}[X_1 S] / \sqrt{P_1 Q}$ be the correlation coefficient between X_1 and S . Denote

$$\tilde{X}_1 \triangleq X_1 - \rho \sqrt{P_1/Q} S \quad (139)$$

$$\tilde{S} \triangleq (1 + \rho \sqrt{P_1/Q}) S. \quad (140)$$

It is not difficult to verify that $\mathbb{E}[\tilde{X}_1 \tilde{S}] = 0$ and $\tilde{X}_1 + \tilde{S} = X_1 + S$. Therefore, we have

$$I(X_1 + S; Y_G) = I(\tilde{X}_1 + \tilde{S}; \tilde{X}_1 + \tilde{S} + \sqrt{1 + P_2/Q} Z) \quad (141)$$

and

$$I(S; Y_\delta) \geq I(\tilde{S}; Y_\delta) = I(\tilde{S}; \tilde{S} + \tilde{X}_1 + \sqrt{\delta} Z) \quad (142)$$

where the inequality holds with equality if $\rho \sqrt{P_1/Q} \neq -1$.

Observe now that, for a fixed ρ and $b \triangleq \mathbb{E}[\tilde{X}_1^2]$, the mutual information term in (141) is maximized when \tilde{X}_1 is Gaussian and is independent of S . Furthermore, by Lemma 13, the

mutual information term on the RHS of (142) is minimized also when \tilde{X}_1 is Gaussian and is independent of S . Therefore, we conclude that

$$\begin{aligned} & \max_{P_{X_1|S}: \mathbb{E}[X_1^2] \leq P_1} \left\{ I(X_1 + S; Y_G) - I(S; Y_\delta) \right\} \\ & \leq \max_{b, \rho} \frac{1}{2} \log \frac{(1 + P_2 + b + (1 + \rho\sqrt{P_1/Q})^2 Q)(\delta^2 + b)}{(\delta^2 + b + (1 + \rho\sqrt{P_1/Q})^2 Q)(1 + P_2)} \end{aligned} \quad (143)$$

where the maximization on the RHS is over all pair (b, ρ) satisfying

$$b \geq 0, \quad \text{and} \quad b + \rho^2 P_1 \leq P_1. \quad (144)$$

By examining the Karush-Kuhn-Tucker (KKT) necessary conditions [28, Sec. 5.5.3], it can be shown that the constraint $b + P_1 \rho^2 \leq P_1$ is always binding (namely, the optimal (b^*, ρ^*) pair must satisfy this inequality with equality), and that the optimal ρ^* must be non-positive. As a result, the maximization problem on the RHS of (143) can be simplified to the one dimensional one in (32). The desired bound (29) follows by substituting (32) and (143) into (136), then (136) into (127), and by optimizing over δ .

B. Proof of Proposition 3

It is straightforward to show the bounds

$$nR_1 \leq \sum_{i=1}^n I(X_{1,i}; Y_i | X_{2,i}, S_i) \quad (145)$$

and

$$nR_2 \leq \sum_{i=1}^n I(X_{2,i}; Y_i | X_{1,i}, S_i). \quad (146)$$

The counterpart of (50) can be proved as follows. As in the proof of Theorem 1, we define the rates R_1 and R_2 as in (94) and (95) without loss of generality. We have

$$n(R_1 + R_2) = I(M_1; Y^n) + I(X_2^n; Y^n) \quad (147)$$

$$= I(M_1, X_2^n; Y^n) - I(X_2^n; M_1 | Y^n) \quad (148)$$

$$\leq h(Y^n) - h(Y^n | M_1, X_2^n) \quad (149)$$

$$\leq \sum_{i=1}^n h(Y_i) - h(Y^n | M_1, X_2^n). \quad (150)$$

Here, (148) follows because X_2^n and M_1 are independent. The conditional differential entropy term $h(Y^n|M_1, X_2^n)$ can be further lower-bounded as follows:

$$h(Y^n|M_1, X_2^n) \tag{151}$$

$$= h(Y^n, S^n|M_1, X_2^n) - h(S^n|Y^n, M_1, X_2^n) \tag{152}$$

$$= h(S^n|M_1, X_2^n) + h(Y^n|M_1, X_2^n, S^n) - h(S^n|Y^n, M_1, X_2^n) \tag{153}$$

$$= h(S^n) + h(Y^n|X_1^n, S^n, X_2^n) - h(S^n|Y^n, M_1, X_2^n) \tag{154}$$

$$\geq h(S^n) + h(Y^n|X_1^n, S^n, X_2^n) - h(S^n|Y^n, X_2^n) \tag{155}$$

$$\geq \sum_{i=1}^n (h(S_i) + h(Y_i|X_{1,i}, S_i, X_{2,i}) - h(S_i|Y_i, X_{2,i})). \tag{156}$$

Here, both (155) and (156) hold because conditioning does not increase differential entropy. Substituting (156) into (150), we conclude that

$$n(R_1 + R_2) \leq \sum_{i=1}^n \left(h(Y_i) - h(Y_i|X_{1,i}, S_i, X_{2,i}) - h(S_i) + h(S_i|Y_i, X_{2,i}) \right) \tag{157}$$

$$= \sum_{i=1}^n \left(h(Y_i) - h(Y_i|X_{1,i}, S_i, X_{2,i}) - h(Y_i|X_{2,i}) + h(Y_i|S_i, X_{2,i}) \right) \tag{158}$$

$$= \sum_{i=1}^n \left(I(X_{1,i}; Y_i|X_{2,i}, S_i) + I(X_{2,i}; Y_i) \right). \tag{159}$$

Here, (158) follows because S_i and $X_{2,i}$ are independent.

Introducing the time-sharing random variable Q , which is uniformly distributed over the integers $\{1, \dots, n\}$, we obtain the following outer bound

$$R_1 \leq I(X_1; Y|X_2, S, Q) \tag{160}$$

$$R_2 \leq I(X_2; Y|X_1, S, Q) \tag{161}$$

$$R_1 + R_2 \leq I(X_1; Y|X_2, S, Q) + I(X_2; Y|Q). \tag{162}$$

Using the concavity of mutual information and the fact that Q is independent of S , it can be shown that the above region is equivalent to the one stated in the proposition (without the time sharing random variable Q). This concludes the proof.

C. Proof of Theorem 7

The proof uses techniques similar to the ones used in the proof of Theorem 1. The main twist in this case compared with Theorem 1 is that X_2^n and X_1^n are not independent. To circumvent this, we need to modify the steps in (98)–(116) by conditioning on M_1 , and by using the fact that X_1^n and X_2^n are conditionally independent given M_1 . In particular, the counterpart of I_δ in (93) is defined as

$$\tilde{I}_\delta \triangleq I(X_1^n + S^n; Y_\delta^n | M_1) - I(X_1^n + S^n; Y_G^n | M_1) \quad (163)$$

$$= \frac{n}{2} \mathbb{E}_{M_1} \left[\log \frac{\tilde{N}_S(\delta | M_1)}{\tilde{N}_S(1 + P_2 | M_1)} \right] + \frac{n}{2} \log \frac{1 + P_2}{\delta} \quad (164)$$

where

$$\tilde{N}_S(\gamma | m) \triangleq \exp \left\{ \frac{2}{n} h(X_1^n + S^n + \sqrt{\gamma} Z^n | M_1 = m) \right\}. \quad (165)$$

The function $\tilde{N}_S(\gamma | m)$ inherits all the properties of $N_S(\gamma)$ that are used in Section IV-A, such as monotonicity and concavity. In the remaining part of the proof, we omit the mechanical details and only highlight the steps that differ from the ones in Section IV-A.

As in Section IV-A, we first upper-bound \tilde{I}_δ . Let

$$R_1 \triangleq I(M_1; Y^n) \quad (166)$$

$$R_2 \triangleq I(X_2^n; Y^n | M_1). \quad (167)$$

Again, by Fano's inequality, the definitions of the rates in (166) and (167) agree with the operational ones. With the conditioning on M_1 , the bounds (98) and (101) become

$$D(P_{X_2^n + Z^n | M_1} \| P_{G^n + Z^n} | P_{M_1}) \leq n(C_2 - R_2) \quad (168)$$

and

$$\begin{aligned} nR_2 &= h(Y^n | M_1) - h(Y_G^n | M_1) \\ &\quad + \mathbb{E}_{M_1} \left[\frac{n}{2} \log \frac{\tilde{N}_S(1 + P_2 | M_1)}{\tilde{N}_S(1 | M_1)} \right]. \end{aligned} \quad (169)$$

Here, $D(P_{X_2^n + Z^n | M_1} \| P_{G^n + Z^n} | P_{M_1})$ denotes the conditional relative entropy

$$D(P_{X_2^n + Z^n | M_1} \| P_{G^n + Z^n} | P_{M_1}) \triangleq \mathbb{E}_{M_1} [D(P_{X_2^n + Z^n | M_1} \| P_{G^n + Z^n})]. \quad (170)$$

Using [18, Props. 1 and 2] and (168), we bound the difference $h(Y^n|M_1) - h(Y_G^n|M_1)$ as follows:

$$\begin{aligned} & h(Y^n|M_1) - h(Y_G^n|M_1) \\ & \leq \frac{\log e}{1+P_2} \mathbb{E}_{M_1} \left[W_2(P_{Y_G^n|M_1}, P_{Y^n|M_1}) \left(4\mathbb{E}[\|X_1^n + S^n\||M_1] \right. \right. \\ & \quad \left. \left. + \frac{3}{2}\sqrt{\mathbb{E}[\|Y_G^n\|^2|M_1]} + \frac{3}{2}\sqrt{\mathbb{E}[\|Y^n\|^2|M_1]} \right) \right] \end{aligned} \quad (171)$$

$$\begin{aligned} & \leq \frac{\log e}{1+P_2} \mathbb{E}_{M_1} \left[\sqrt{\frac{2(1+P_2)}{\log e} D(P_{X_2^n+Z^n|M_1} \| P_{G^n+Z^n})} \left(4\sqrt{\mathbb{E}[\|X_1^n + S^n\|^2|M_1]} \right. \right. \\ & \quad \left. \left. + \frac{3}{2}\sqrt{\mathbb{E}[\|Y_G^n\|^2|M_1]} + \frac{3}{2}\sqrt{\mathbb{E}[\|Y^n\|^2|M_1]} \right) \right] \end{aligned} \quad (172)$$

$$\begin{aligned} & \leq \frac{\log e}{1+P_2} \sqrt{\frac{2(1+P_2)}{\log e} D(P_{X_2^n+Z^n|M_1} \| P_{G^n+Z^n} | P_{M_1})} \\ & \quad \cdot \left(4\sqrt{\mathbb{E}[\|X_1^n + S^n\|^2]} + \frac{3}{2}\sqrt{\mathbb{E}[\|Y_G^n\|^2]} + \frac{3}{2}\sqrt{\mathbb{E}[\|Y^n\|^2]} \right) \end{aligned} \quad (173)$$

$$\leq c_2 n \sqrt{C_2 - R_2} \quad (174)$$

where

$$c_2 \triangleq \frac{3\sqrt{1 + (\sqrt{P_1} + \sqrt{P_2} + \sqrt{Q})^2} + 4(\sqrt{P_1} + \sqrt{Q})}{\sqrt{(1+P_2)/(2\log e)}}. \quad (175)$$

Here, (171) follows from [18, Props. 1 and 2]; (172) follows because for every message m ,

$$\mathbb{E}[\|X_1^n + S^n\||M_1 = m] \leq \sqrt{\mathbb{E}[\|X_1^n + S^n\|^2|M_1 = m]} \quad (176)$$

and

$$W_2(P_{Y_G^n|M_1=m}, P_{Y^n|M_1=m}) \leq W_2(P_{X_2^n+Z^n|M_1=m}, P_{G^n+Z^n}) \quad (177)$$

$$\leq \sqrt{\frac{2(1+P_2)}{\log e} D(P_{X_2^n+Z^n|M_1=m} \| P_{G^n+Z^n})} \quad (178)$$

where (177) follows because the $W_2(\cdot, \cdot)$ distance is non-decreasing under convolutions and because $X_1^n + S^n$ and X_2^n are conditionally independent given M_1 , and the bound (178) follows from Talagrand's inequality [19]; (173) follows from the Cauchy-Schwarz inequality; and finally (174) follows from (168), (105), and because

$$\frac{1}{n} \mathbb{E}[\|Y^n\|^2] \leq 1 + (\sqrt{P_1} + \sqrt{P_2} + \sqrt{Q})^2 \quad (179)$$

$$\frac{1}{n} \mathbb{E}[\|Y_G^n\|^2] \leq 1 + (\sqrt{P_1} + \sqrt{P_2} + \sqrt{Q})^2. \quad (180)$$

Substituting (174) into (169), we conclude that

$$\mathbb{E}_{M_1} \left[\log \frac{\tilde{N}_S(1 + P_2|M_1)}{\tilde{N}_S(1|M_1)} \right] \leq 2c_2\sqrt{C_2 - R_2} + 2(C_2 - R_2) - \log(1 + P_2). \quad (181)$$

Letting $\alpha \triangleq P_2/(1 + P_2 - \delta)$ as in Section IV-A, we obtain

$$\begin{aligned} & \mathbb{E}_{M_1} \left[\log \frac{\tilde{N}_S(\delta|M_1)}{\tilde{N}_S(1 + P_2|M_1)} \right] \\ & \leq \mathbb{E}_{M_1} \left[\log \left(\frac{\tilde{N}_S(1|M_1)}{\tilde{N}_S(1 + P_2|M_1)} - 1 + \alpha \right) \right] - \log \alpha \end{aligned} \quad (182)$$

$$\leq \log \left(\frac{\exp(2c_2\sqrt{C_2 - R_2} + 2(C_2 - R_2))}{1 + P_2} - 1 + \alpha \right) - \log \alpha. \quad (183)$$

Here, (182) follows from the concavity of $\gamma \mapsto \tilde{N}_S(\gamma|M_1)$, and (183) follows from Jensen's inequality and because the function $x \mapsto \log(\exp(x) - (1 - \alpha))$ is concave. Finally, substituting (183) into (164), we conclude that

$$\tilde{I}_\delta \leq \frac{n}{2} \log \left(1 + \frac{1 + P_2 - \delta}{P_2\delta} \tilde{g}(R_2) \right) \quad (184)$$

where $\tilde{g}(R_2)$ is defined in (76).

We next relate \tilde{I}_δ to R_1 . This part is quite different from the steps in Section IV-A2, since for the dirty MAC with degraded message sets, the information about the message M_1 is contained in both X_1^n and X_2^n . Consider the following chain:

$$\tilde{I}_\delta = I(X_1^n, S^n; Y_\delta^n | M_1) - I(X_1^n + S^n, M_1; Y_G^n) + I(M_1; Y_G^n) \quad (185)$$

$$= I(S^n; Y_\delta^n, M_1) + I(X_1^n; Y_\delta^n | S^n, M_1) - I(X_1^n + S^n, M_1; Y_G^n) + I(M_1; Y_G^n) \quad (186)$$

$$\geq I(S^n; Y_\delta^n) - I(X_1^n + S^n; Y_G^n) + I(M_1; Y_G^n) \quad (187)$$

$$= I(S^n; Y_\delta^n) - I(X_1^n + S^n; Y_G^n) + I(M_1; Y_G^n) - I(M_1; Y^n) + nR_1. \quad (188)$$

Here, the penultimate step follows because $M_1 \rightarrow X_1^n + S^n \rightarrow Y_G^n$ forms a Markov chain. The first two terms on the RHS of (188) can be single-letterized and bounded in the same way as in Section IV-A3 and Section IV-A4, i.e.,

$$I(S^n; Y_\delta^n) - I(X_1^n + S^n; Y_G^n) \geq -nf(\delta) \quad (189)$$

where $f(\cdot)$ was defined in (32).

To conclude the proof, it remains to lower-bound $I(M_1; Y_G^n) - I(M_1; Y^n)$. To this end, we rewrite it as

$$I(M_1; Y_G^n) - I(M_1; Y^n) = h(Y_G^n) - h(Y^n) + h(Y^n|M_1) - h(Y_G^n|M_1). \quad (190)$$

The differences $h(Y_G^n) - h(Y^n)$ and $h(Y^n|M_1) - h(Y_G^n|M_1)$ can be bounded via steps similar to those in (171)–(174). More specifically, we have

$$h(Y_G^n|M_1) - h(Y^n|M_1) \leq c_3 n \sqrt{C_2 - R_2} \quad (191)$$

and

$$h(Y^n) - h(Y_G^n) \leq c_2 n \sqrt{C_2 - R_2} \quad (192)$$

where c_3 was defined in (78). Here, to prove (192), we have used

$$D(P_{Y^n} \| P_{Y_G^n}) \leq D(P_{Y^n|M_1} \| P_{Y_G^n|M_1} | P_{M_1}) \quad (193)$$

$$\leq D(P_{X_2^n+Z^n|M_1} \| P_{G^n+Z^n} | P_{M_1}) \quad (194)$$

$$\leq n(C_2 - R_2) \quad (195)$$

where (193) follows from the data processing inequality, (194) follows from the data processing inequality and because $X_1^n + S^n$ and X_2^n are conditionally independent given M_1 , and (195) follows from (168). Substituting (191) and (192) into (190), then (190) and (189) into (188), and combining (188) with (184), we conclude the proof of (75).

D. Proof of Proposition 10

The key idea of the proof is to identify the auxiliary random variables $U \triangleq (M_1, Q)$, where Q denotes the time-sharing random variable that is uniformly distributed over the integers $\{1, \dots, n\}$. We have

$$nR_2 = I(X_2^n; Y^n | M_1) \quad (196)$$

$$\leq I(X_2^n; Y^n, X_1^n, S^n | M_1) \quad (197)$$

$$= I(X_2^n; Y^n | X_1^n, S^n, M_1) \quad (198)$$

$$= h(Y^n | X_1^n, S^n, M_1) - h(Y^n | X_1^n, X_2^n, S^n, M_1) \quad (199)$$

$$\leq \sum_{i=1}^n h(Y_i | X_{1,i}, S_i, M_1) - h(Y_i | X_{1,i}, X_{2,i}, S_i, M_1) \quad (200)$$

$$= \sum_{i=1}^n I(X_{2,i}; Y_i | X_{1,i}, S_i, M_1) \quad (201)$$

$$= I(X_2; Y | X_1, S, U). \quad (202)$$

This yields the upper bound in (85).

To prove (86), we observe that

$$R_2 = I(X_2^n; Y^n | M_1) \quad (203)$$

$$= h(Y^n | M_1) - h(Y^n | M_1, X_2^n) \quad (204)$$

$$\leq \sum_{i=1}^n h(Y_i | M_1) - h(Y^n | M_1, X_2^n). \quad (205)$$

Proceeding as in (150)–(162) while keeping the conditioning on M_1 , we conclude that

$$R_2 \leq \sum_{i=1}^n \left(I(X_{1,i}; Y_i | X_{2,i}, S_i, M_1) + I(X_{2,i}; Y_i | M_1) \right) \quad (206)$$

$$= I(X_1; Y | X_2, S, M_1, Q) + I(X_2; Y | Q, M_1) \quad (207)$$

$$\leq I(X_1; Y | X_2, S, U) + I(X_2; Y | U). \quad (208)$$

Finally, we prove (87). We proceed again as in (147)–(156) and keep the conditioning on M_1 whenever appropriate. This yields

$$\begin{aligned} & n(R_1 + R_2) \\ & \leq \sum_{i=1}^n \left(h(Y_i) - h(Y_i | X_{1,i}, S_i, X_{2,i}) \right) \\ & \quad - h(S^n | M_1) + h(S^n | Y^n, M_1, X_2^n) \end{aligned} \quad (209)$$

$$\begin{aligned} & \leq \sum_{i=1}^n \left(h(Y_i) - h(Y_i | X_{1,i}, S_i, X_{2,i}, M_1) \right) \\ & \quad - h(S_i | M_1, X_{2,i}) + h(S_i | M_1, Y_i, X_{2,i}) \end{aligned} \quad (210)$$

$$= \sum_{i=1}^n \left(I(X_{1,i}; Y_i | X_{2,i}, S_i, M_1) + I(X_{2,i}, M_1; Y_i) \right) \quad (211)$$

$$= I(X_1; Y | X_2, S, M_1, Q) + I(X_2, M_1; Y | Q) \quad (212)$$

$$\leq I(X_1; Y | X_2, S, U) + I(X_2, U; Y). \quad (213)$$

Here, (210) follows because S_i is independent of M_1 and $X_{2,i}$, and because conditioning does not increase entropy. The proof is concluded by observing that the auxiliary random variable U and the random variables X_1, X_2, S satisfy the conditions listed in the theorem.

V. CONCLUSION

In this paper, we have studied a two-user state-dependent Gaussian MAC with state noncausally known at one encoder and with and without degraded message sets. We have derived several new outer bounds on the capacity region, which provide substantial improvements over the best previously known outer bounds. For the dirty MAC without degraded message sets, our outer bounds yield the following:

- The characterization of the sum rate capacity;
- The establishment of the two corner points of the capacity region;
- The characterization of the full capacity region in the special case in which the sum rate capacity is equal to the capacity C_{helper} of the helper problem;
- A new upper bound on C_{helper} , and a necessary and sufficient condition to achieve $C_{\text{helper}} = \frac{1}{2} \log(1 + P_2)$.

We have shown that a single-letter solution is adequate to achieve both the corner points and the sum rate capacity. In addition, we have generalized our outer bounds to the case of additive non-Gaussian states.

There are several possible generalizations of the results in this paper.

- The outer bounds derived in this paper can be readily generalized to the discrete and to the multiple-input multiple-output (MIMO) setting. This is unlike the *doubly dirty Gaussian MAC setting*, in which additional difficulties arise when extending from the single-input single-output to the MIMO setting [29].
- In this paper, we assume that the state is not known at the non-cognitive user. It would be interesting to investigate whether revealing the state information strictly causally to the non-cognitive user can increase the capacity region. As shown in [30], strictly causal state information enables cooperations between the two encoders (e.g., by letting the encoders convey the past state information jointly to the decoder).
- In the proofs of Theorem 1 and Theorem 7, we have essentially transformed the dirty MAC into a state-dependent Z -interference channel with input-output relationship

$$Y_1 = X_1 + S + \sqrt{\delta}Z_1 \quad (214)$$

$$Y_2 = X_1 + X_2 + S + Z_2 \quad (215)$$

where the Gaussian noises $Z_1, Z_2 \sim \mathcal{N}(0, 1)$ are independent. This suggests that our techniques may yield tighter outer bounds on the capacity region of the state-dependent Gaussian Z -interference channel than the ones derived in [31].

- Another related setting is the state-dependent relay channel with state available noncausally at the relay considered in [32]. It would be interesting to see whether our techniques can lead to any improvement over the bounds there.

APPENDIX

GAUSSIAN INPUTS MAXIMIZE (85)–(87)

We shall prove that the outer region provided in Proposition 10 is maximized when U, S, X_1 , and X_2 are jointly Gaussian distributed. Differently from [5, Th. 4], the presence of the auxiliary random variable U complicates the proof substantially.

Consider an arbitrary distribution $P_{USX_1X_2}$ that satisfies the conditions stated in the proposition. Without loss of generality, we assume that $P_{USX_1X_2}$ satisfies the following conditions, in addition to the ones stated in Proposition 10:

- U has zero mean and unit variance;
- $\mathbb{E}[X_1^2] = P_1$ and $\mathbb{E}[X_2] = P_2$.

The first assumption comes without loss of generality since U does not appear in the channel input-output relation $Y = X_1 + X_2 + S + Z$, and the second assumption comes without loss of generality because we do not assume X_1 and X_2 to have zero mean. We next introduce the following notation:

$$\mu_k(u) \triangleq \mathbb{E}[X_k|U = u] \tag{216}$$

$$\sigma_k(u) \triangleq \sqrt{\text{Var}[X_k|U = u]} \tag{217}$$

$$\rho_k \triangleq \sqrt{\mathbb{E}[\mu_k^2(U)] / P_k} \tag{218}$$

$$\mu_s(u) \triangleq \mathbb{E}[X_1 S | U = u] / \sqrt{Q} \tag{219}$$

$$\rho_s \triangleq \mathbb{E}[\mu_s(U)] / \sqrt{P_1} \tag{220}$$

where $k \in \{1, 2\}$. It follows that

$$R_1 \leq I(X_2; Y | X_1, S, U) \tag{221}$$

$$\leq \frac{1}{2} \mathbb{E}[\log(1 + \sigma_2(U)^2)] \quad (222)$$

$$\leq \frac{1}{2} \log(1 + \mathbb{E}[\sigma_2(U)^2]) \quad (223)$$

$$= \frac{1}{2} \log(1 + P_2(1 - \rho_2^2)). \quad (224)$$

Here, (223) follows from Jensen's inequality, and (224) follows because

$$\mathbb{E}[\sigma_2^2(U)] = \mathbb{E}[\mathbb{E}[X_2^2|U] - \mu_2(U)^2] = P_2 - \rho_2^2 P_2. \quad (225)$$

This proves (81).

To prove (82), we proceed as follows:

$$R_2 \leq I(X_1; Y|X_2, S, U) + I(X_2; Y|U) \quad (226)$$

$$= I(X_1, X_2, S; Y|U) - I(S; Y|U, X_2). \quad (227)$$

To upper-bound $I(X_1, X_2, S; Y|U)$, we observe that

$$\begin{aligned} \text{Var}[X_1 + X_2 + S|U = u] \\ = \sigma_1^2(u) + \sigma_2^2(u) + Q + 2\sqrt{Q}\mu_s(u) \end{aligned} \quad (228)$$

where we have used (217) and (219), and that X_1 and X_2 are conditionally independent given U .

It thus follows that

$$\begin{aligned} I(X_1 + X_2 + S; Y|U) \\ \leq \frac{1}{2} \mathbb{E}[\log(1 + \sigma_1^2(U) + \sigma_2^2(U) + Q + 2\sqrt{Q}\mu_s(U))] \end{aligned} \quad (229)$$

$$\leq \frac{1}{2} \log\left(1 + \mathbb{E}[\sigma_1^2(U) + \sigma_2^2(U) + Q + 2\sqrt{Q}\mu_s(U)]\right) \quad (230)$$

$$= \frac{1}{2} \log\left(1 + P_1(1 - \rho_1^2) + P_2(1 - \rho_2^2) + Q + 2\rho_s\sqrt{QP_1}\right). \quad (231)$$

Here, in (231) we have used the following identity:

$$\mathbb{E}[\sigma_k^2(U)] = \mathbb{E}[\text{Var}[X_k|U]] \quad (232)$$

$$= \text{Var}[X_k] - \text{Var}[\mu_k(U)] \quad (233)$$

$$= \text{Var}[X_k] - \mathbb{E}[\mu_k(U)^2] + \mathbb{E}[X_k]^2 \quad (234)$$

$$= P_k - P_k\sigma_k^2, \quad k \in \{1, 2\} \quad (235)$$

where (233) follows from the law of total variance.

We next bound the second term on the RHS of (227). Let

$$\tilde{X}_1 \triangleq X_1 - \mu_1(U) - \frac{\mu_s(U)S}{\sqrt{Q}}. \quad (236)$$

It follows that

$$\mathbb{E}[\tilde{X}_1 S | U = u] = \mathbb{E}[X_1 S | U = u] - \mu_s(u)\sqrt{Q} = 0. \quad (237)$$

Since S is Gaussian distributed, by Lemma 13,

$$I(S; Y | X_2, U) \quad (238)$$

$$= \mathbb{E}_U \left[I(S; (1 + \mu_s(U)/\sqrt{Q})S + \tilde{X}_1 + Z | U) \right] \quad (239)$$

$$\geq \frac{1}{2} \mathbb{E} \left[\log \left(1 + \frac{(\sqrt{Q} + \mu_s(U))^2}{1 + \sigma_1^2(U) - \mu_s(U)^2} \right) \right]. \quad (240)$$

By (217), (236), and (237),

$$\sigma_1^2(u) = \mathbb{E}[X_1^2 | U = u] - \mu_1^2(u) \quad (241)$$

$$= \mathbb{E}[\tilde{X}_1^2 | U = u] + \mu_s(u)^2 \geq \mu_s(u)^2. \quad (242)$$

Now, observe that the function

$$\xi(a, b) \triangleq \frac{1}{2} \log \left(1 + \frac{(\sqrt{Q} - a)^2}{1 + b - a^2} \right) \quad (243)$$

is jointly convex in (a, b) as long as $a^2 \leq b$. Indeed, let \mathbf{H} be the Hessian matrix of $\xi(a, b)$. It follows that

$$H_{11} = \frac{\partial^2 \xi}{\partial a^2} \quad (244)$$

$$= \frac{(\sqrt{Q} - a)^2 ((\sqrt{Q} - a)^2 + 2 + 2b - 2a^2)}{(1 + b - a^2)^2 (\sqrt{Q} - a)^2 + 1 + b - a^2} \quad (245)$$

$$\geq 0 \quad (246)$$

and that

$$\text{Det}[\mathbf{H}] = \frac{(\sqrt{Q} - a)^4}{(1 + b - a^2)^3 (\sqrt{Q} - a)^2 + 1 + b - a^2} \quad (247)$$

$$\geq 0. \quad (248)$$

Therefore, H is positive semi-definite for all (a, b) satisfying $a^2 \leq b$, which implies that the function $\xi(a, b)$ is convex. Therefore, by Jensen's inequality,

$$I(S; X_1 + S + Z|U) \quad (249)$$

$$\geq \frac{1}{2} \log \left(1 + \frac{(\sqrt{Q} + \mathbb{E}[\mu_s(U)])^2}{1 + \mathbb{E}[\sigma_1^2(U)] - \mathbb{E}[\mu_s(U)]^2} \right) \quad (250)$$

$$= \frac{1}{2} \log \left(1 + \frac{(\sqrt{Q} + \rho_s \sqrt{P_1})^2}{1 + P_1 - \rho_1^2 P_1 - \rho_s^2 P_1} \right). \quad (251)$$

Here, in (251) we have used (235). Substituting (231) and (251) into (227) and rearranging the terms, we obtain (82).

The proof of (83) follows steps analogous to those in the proof of (82). More specifically, we obtain from (87) that

$$\begin{aligned} R_1 + R_2 &\leq h(Y|X_2, S, U) - h(Y|X_1, X_2, S, U) \\ &\quad + h(Y) - h(Y|X_2, U) \end{aligned} \quad (252)$$

$$= I(X_1 + X_2 + S; Y) - I(S; X_1 + S + Z|U). \quad (253)$$

The term $I(S; X_1 + S + Z|U)$ on the RHS of (253) has been lower-bounded in (251). To upper-bound $I(X_1 + X_2 + S; Y)$, we bound $\mathbb{E}[(X_1 + X_2 + S)^2]$ as

$$\begin{aligned} &\mathbb{E}[(X_1 + X_2 + S)^2] \\ &= P_1 + P_2 + Q + 2\mathbb{E}[X_1 S] + 2\mathbb{E}[X_1 X_2] \end{aligned} \quad (254)$$

$$= P_1 + P_2 + Q + 2\rho_s \sqrt{P_1 Q} + 2\mathbb{E}[\mathbb{E}[X_1|U] \mathbb{E}[X_2|U]] \quad (255)$$

$$\leq P_1 + P_2 + Q + 2\rho_s \sqrt{P_1 Q} + 2\rho_1 \rho_2 \sqrt{P_1 P_2}. \quad (256)$$

Here, (255) follows because X_1 and X_2 are conditionally independent given U , and (256) follows because

$$\mathbb{E}[\mathbb{E}[X_1|U] \mathbb{E}[X_2|U]] = \mathbb{E}[\mu_1(U) \mu_2(U)] \quad (257)$$

$$\leq \sqrt{\mathbb{E}[\mu_1(U)^2] \mathbb{E}[\mu_2(U)^2]} \quad (258)$$

$$= \rho_1 \rho_2 \sqrt{P_1 P_2}. \quad (259)$$

It thus follows that

$$\begin{aligned} I(X_1 + X_2 + S; Y) &\leq \frac{1}{2} \log \left(1 + P_1 + P_2 + Q \right. \\ &\quad \left. + 2\rho_s \sqrt{P_1 Q} + 2\rho_1 \rho_2 \sqrt{P_1 P_2} \right). \end{aligned} \quad (260)$$

Substituting (260) and (251) into (253), we obtain (83).

Finally, observe from (235) and (242) that

$$P_1 - P_1\sigma_1^2 = \mathbb{E}[\sigma_1^2(U)] \geq \mathbb{E}[\mu_s(U)^2] \geq \mathbb{E}[\mu_s(U)]^2 \geq P_1\rho_s^2 \quad (261)$$

which implies the condition (84). This concludes the proof.

REFERENCES

- [1] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
- [2] T. Philosof, R. Zamir, U. Erez, and A. J. Khisti, "Lattice strategies for the dirty multiple access channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5006–5035, Aug. 2011.
- [3] S. Kotagiri and J. N. Laneman, "Multiaccess channels with state known to some encoders and independent messages," *EURASIP J. Wireless Commun. Netw.*, vol. 2008, Mar. 2008.
- [4] T. Philosof and R. Zamir, "On the loss of single-letter characterization: The dirty multiple access channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2442–2454, Jun. 2009.
- [5] A. Zaidi, S. P. Kotagiri, J. N. Laneman, and L. Vandendorpe, "Multiaccess channels with state known to one encoder: Another case of degraded message sets," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, Korea, Jun. 2009.
- [6] M. F. Pourbabae, M. J. Emadi, A. G. Davoodi, and M. R. Aref, "Lattice coding for multiple access channels with common message and additive interference," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Lausanne, Switzerland, Sep. 2012, pp. 412–416.
- [7] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdú, "Cooperative multiple access encoding with states available at one transmitter," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4448–4469, Oct. 2008.
- [8] S. Kotagiri and J. N. Laneman, "Multiaccess channels with state known to one encoder: A case of degraded message sets," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Nice, France, Jun. 2007.
- [9] A. Zaidi, P. Piantanida, and S. Shamai (Shitz), "Capacity region of cooperative multiple-access channel with states," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6153–6174, Oct. 2013.
- [10] S. Mallik and R. Kotter, "Helpers for cleaning dirty papers," in *Proc. Int. ITG Conf. Sour. Channel Coding (SCC)*, Ulm, Germany, Jan. 2008.
- [11] Y. Sun, R. Duan, Y. Liang, and S. Shamai (Shitz), "Capacity characterization for state-dependent Gaussian channel with a helper," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7123–7134, Dec. 2016.
- [12] A. Lapidoth and Y. Steinberg, "The multiple access channel with two independent states each known causally to one encoder," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Austin, TX, USA, Jun. 2010.
- [13] M. Li, O. Simeone, and A. Yener, "Multiple access channels with states causally known at transmitters," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1394–1404, Mar. 2013.
- [14] L. Dikstein, H. Permuter, and S. Shamai (Shitz), "MAC with action-dependent state information at one encoder," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 173–188, Jan. 2015.
- [15] I.-H. Wang, "Approximate capacity of the dirty multiple-access channel with partial state information at the encoders," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2781–2787, May 2012.
- [16] S. Jafar, "Capacity with causal and noncausal side information: A unified view," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5468–5474, Dec. 2006.

- [17] G. Keshet, Y. Steinberg, and N. Merhav, “Channel coding in the presence of side information,” *Foundations and Trends Commun. Inf. Theory*, vol. 4, no. 6, pp. 445–586, 2008.
- [18] Y. Polyanskiy and Y. Wu, “Wasserstein continuity of entropy and outer bounds for interference channels,” *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3992–4002, Jul. 2016.
- [19] M. Talagrand, “Transportation cost for Gaussian and other product measures,” *Geometric and Functional Analysis*, vol. 6, no. 3, pp. 587–600, May 1996.
- [20] S. Ihara, “On the capacity of channels with additive non-Gaussian noise,” *Inform. Contr.*, vol. 37, no. 1, pp. 34–39, Apr. 1978.
- [21] S. N. Diggavi and T. M. Cover, “The worst additive noise under a covariance constraint,” *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 3072–3081, Nov. 2001.
- [22] B. Hassibi and B. M. Hochwald, “How much training is needed in multiple-antenna wireless links?” *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [23] M. H. M. Costa, “A new entropy power inequality,” *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 751–760, Nov. 1985.
- [24] D. Guo, S. Shamai (Shitz), and S. Verdú, “Mutual information and minimum mean-square error in Gaussian channels,” *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.
- [25] A. El Gamal and Y. Kim, *Network Information Theory*. Cambridge, UK: Cambridge University Press, 2011.
- [26] C. Villani, *Topics in Optimal Transportation*. Providence, RI: American Mathematical Society, 2003.
- [27] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New Jersey: Wiley, 2006.
- [28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [29] A. Khina, Y. Kochman, and U. Erez, “The dirty MIMO multiple-access channel,” *IEEE Trans. Inf. Theory*, 2017. [Online]. Available: <https://arxiv.org/pdf/1510.08018.pdf>
- [30] A. Lapidoth and Y. Steinberg, “The multiple-access channel with causal side information: Common state,” *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 32–50, Jan. 2013.
- [31] R. Duan, Y. Liang, A. Khisti, and S. Shamai (Shitz), “State-dependent parallel Gaussian networks with a common state-cognitive helper,” *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6680–6699, Dec. 2015.
- [32] A. Zaidi, S. P. Kotagiri, J. N. Laneman, and L. Vandendorpe, “Cooperative relaying with state available noncausally at the relay,” *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2272–2298, May 2010.