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: 1) each user wishes to communicate an independent message to the common receiver; and 2) 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.**

***Index Terms*— Channel with states, capacity region, dirty paper coding, multiple access channel, outer bound.**

I. Introduction

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

$$Y = X_1 + X_2 + S + Z \tag{1}$$

Manuscript received May 3, 2017; revised March 28, 2018; accepted May 7, 2018. Date of publication May 22, 2018; date of current version November 20, 2018. W. Yang and H. V. Poor were supported by the U.S. National Science Foundation under Grant ECCS-1343210 and Grant ECCS-1647198. Y. Liang was supported by the U.S. National Science Foundation under Grant CCF-1801846. S. Shamai (Shitz) was supported by the European Union's Horizon 2020 Research and Innovation Programme under Grant 694630. This paper was presented in part at the 2017 IEEE International Symposium on Information Theory.

W. Yang is with Qualcomm Technologies, Inc., San Diego, CA 92121 USA (e-mail: weiyang@qti.qualcomm.com).

Y. Liang is with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: liang.889@osu.edu).

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

H. V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Communicated by J.-F. Chamberland, Associate Editor for Communications.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2018.2839643

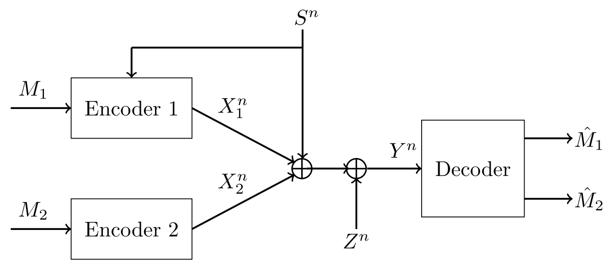

Fig. 1. State-dependent Gaussian MAC with state available noncausally at one encoder without degraded message sets.

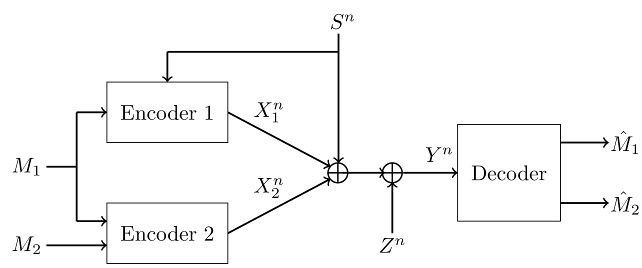

Fig. 2. State-dependent Gaussian MAC with state available noncausally at one encoder with degraded message sets.

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 the "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 region has been 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. Zaidi *et al.* [5] 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 possesses 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]. Lapidoth and Steinberg [12] and Li *et al.* [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]. Inner and outer bounds on the capacity region of the state-dependent MAC with rate-limited decoder side information were derived in [15]. Finally, Wang [16] 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 [17] and [18].

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 $Q \to \infty$). 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 new outer bounds derived in this paper also lead to a new upper bound on the capacity of the helper problem.

The proof of our outer bounds builds on two sets of techniques that are quite different from each other. Interestingly, each set of techniques yields one corner point of the capacity region. The first set of techniques is *algebraic* in nature, exploiting certain algebraic properties of mutual information and differential entropy. Among others, we make use of a recent technique proposed by Polyanskiy and Wu [19] that bounds the difference of the differential entropies of two probability distributions via their quadratic Wasserstein distance and via Talagrand's transportation inequality [20]. The second set of techniques are standard outer-bounding techniques in network information theory, including a generalized version of the worst-case Gaussian noise result, in which the Gaussian input and the noise are dependent (but are uncorrelated) [21]–[23]. The improvement of our bounds over existing ones in, e.g., [3] and [5] mainly lies in the way we apply these techniques, and in a novel identification of a certain auxiliary random variable in the degraded-message-set case. 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.

In addition to the outer-bounding techniques reviewed above, the idea of *generalization* (or *extension*) has been applied several times in this manuscript. In a nutshell, this idea allows us to solve a problem by either generalizing it or extending its domain. For example, to establish one of the outer bounds on the capacity region of the state-dependent MAC, we generalize the MAC to an interference channel. It is this generalization that allows us to use the recent advances in the study of interference channels (including Polyanskiy and Wu's technique [19]) on the state-dependent MAC. As another instance of this powerful idea, we show how an outer bound on the capacity region of the state-dependent MAC leads to an upper bound on the capacity of the helper problem. We believe that this idea of generalization/extension will be very beneficial in information theory. It also parallels the "reductionist" view of information theory that was propounded recently in [24].

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 constraints[1]

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

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

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

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

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

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

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,

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

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 $\mathcal{C}(P_1, P_2, Q)$.

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

$$R_1 \le I(U; Y|X_2) - I(U; S) \tag{7}$$

$$R_2 \le I(X_2; Y|U) \tag{8}$$

$$R_1 + R_2 \le I(U, X_2; Y) - I(U, S) \tag{9}$$

for some joint probability distribution $P_{UX_1|S}P_{X_2}$. A computable inner bound was obtained in [3] from (7)–(9) by setting

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

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

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

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) satisfying[2]

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

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

$$\begin{aligned} R_1 + R_2 &\le \frac{1}{2}\log(1 + P_1(1-\rho_1^2-\rho_s^2)) \\ &\quad + \frac{1}{2}\log\left(1 + \frac{(\sqrt{P_2}+\rho_1\sqrt{P_1})^2}{1 + P_1(1-\rho_1^2-\rho_s^2) + (\sqrt{Q}+\rho_s\sqrt{P_1})^2}\right) \end{aligned} \tag{15}$$

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

for some $\rho_1 \in [0, 1]$ and $\rho_s \in [-1, 0]$ that satisfy $\rho_1^2 + \rho_s^2 \le 1$. 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]:

$$R_2 \le I(X_2; Y|S, X_1) \tag{17}$$

$$R_1 + R_2 \le I(X_1, X_2; Y|S) - I(S; X_2|Y) \tag{18}$$

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

$$R_2 \le I(X_2; Y|U_1, U_2) \tag{19}$$

$$R_2 \le I(X_2, U_2; Y|U_1) - I(U_2; S|U_1) \tag{20}$$

$$R_1 + R_2 \le I(X_2, U_1, U_2; Y) - I(U_2; S|U_1) \tag{21}$$

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

$$I(U_2; Y|U_1, X_1) - I(U_2; S|U_1) \ge 0. \tag{22}$$

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.

[1]Note that, Philosof *et al.* [2] and Somekh-Baruch *et al.* [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 $\sum_{i=1}^n x_{1,i}^2(m_1, S^n) \le nP_1$ and $\sum_{i=1}^n X_{2,i}^2(m_2) \le nP_2$ almost surely. Clearly, every outer bound for the average power constraint is also a valid outer bound for the *per-codeword* power constraint.

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

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

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

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

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]

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

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

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

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).

The following upper bound on C_{helper}, which holds for all parameters, was derived in [10]:

$$C_{\text{helper}} \leq \max_{-1\leq\rho\leq 0}\left\{\frac{1}{2}\log\left(1 + \frac{P_2}{1 + P_1 + Q + 2\rho\sqrt{P_1 Q}}\right) + \frac{1}{2}\log(1 + P_1(1-\rho^2))\right\}. \tag{27}$$

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

$$C_1 \triangleq \frac{1}{2}\log(1 + P_1), \quad C_2 \triangleq \frac{1}{2}\log(1 + P_2). \tag{28}$$

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

$$R_2 \leq C_{\text{helper}} \tag{29}$$

and

$$R_1 \leq \min_{0\leq\delta\leq 1}\left\{\frac{1}{2}\log\left(1 + \frac{1 + P_2 - \delta}{P_2\delta}g(R_2)\right) + f(\delta)\right\} \tag{30}$$

where

$$g(R_2) \triangleq \exp\left(2c_1\sqrt{C_2 - R_2} + 2(C_2 - R_2)\right) - 1 \tag{31}$$

with

$$c_1 \triangleq \frac{3\sqrt{1 + (\sqrt{P_1} + \sqrt{Q})^2 + P_2} + 4(\sqrt{P_1} + \sqrt{Q})}{\sqrt{(1 + P_2)/(2\log e)}} \tag{32}$$

and

$$f(\delta) \triangleq \max_{\rho\in[-1,0]} \frac{1}{2}\left\{\log\frac{1 + P_2 + P_1 + Q + 2\rho\sqrt{P_1 Q}}{\delta + P_1 + Q + 2\rho\sqrt{P_1 Q}} + \log\frac{\delta + (1-\rho^2)P_1}{1 + P_2}\right\}. \tag{33}$$

Proof: See Section IV-A. ■

Remark 1: The objective function on the right-hand side (RHS) of (33) is concave in ρ for every $\delta \in [0, 1]$.

Remark 2: The upper bound (30) can be slightly improved by replacing Q on the RHS of (30) with $\widetilde{Q} \leq Q$ and by minimizing over $\widetilde{Q}$. This follows because, for a fixed rate R_2, the maximum achievable R_1 is monotonically non-increasing in Q, whereas the RHS of (30) 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

$$Y_G^n \triangleq X_1^n + S^n + G^n + Z^n \tag{34}$$

$$Y_\delta^n \triangleq X_1^n + S^n + \sqrt{\delta}Z^n \tag{35}$$

where $G^n \sim \mathcal{N}(0, P_2\mathsf{I}_n)$ 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 $M_1 \to Y_G^n$ is worse than the original channel whereas the channel $M_1 \to Y_\delta^n$ 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

$$I(X_1^n + S^n; Y_G^n) \approx I(X_1^n + S^n; Y^n). \tag{36}$$

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

$$I(X_1^n + S^n; Y^n) \approx I(X_1^n + S^n; Y^n | X_2^n) \tag{37}$$
$$= I(X_1^n + S^n; X_1^n + S^n + Z^n). \tag{38}$$

Combining (36) and (38), we conclude that

$$I(X_1^n + S^n; X_1^n + S^n + G^n + Z^n) \approx I(X_1^n + S^n; X_1^n + S^n + Z^n). \tag{39}$$

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

$$I(X_1^n + S^n; Y^n) \approx I(X_1^n + S^n; Y_G^n) \approx I(X_1^n + S^n; Y_\delta^n). \tag{40}$$

The errors in the estimation (40) can be bounded via Costa's entropy power inequality [25] or the I-MMSE relation [26].

To see how the relation (40) can be used to upper-bound R_1, we note that by standard manipulations of mutual information,

$$nR_1 \leq I(X_1^n + S^n; Y^n) - I(S^n; Y^n). \tag{41}$$

By (40), we may replace the two Y^n's on the RHS of (41) with Y_G^n and Y_δ^n, respectively, and obtain

$$nR_1 \lessapprox I(X_1^n + S^n; Y_G^n) - I(S^n; Y_\delta^n) \tag{42}$$
$$\lessapprox n \max_{P_{X_1|S}} \left\{ I(X_1 + S; Y_G) - I(S; Y_\delta) \right\} \tag{43}$$

where

$$Y_G \triangleq X_1 + S + G + Z \tag{44}$$
$$Y_\delta = X_1 + S + \sqrt{\delta} Z \tag{45}$$

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 (43) is maximized when (X_1, S) are jointly Gaussian. The maximum of the objective function on the RHS of (43) is precisely the $f(\delta)$ defined in (33), whereas the logarithm term on the RHS of (30) quantifies the error in the approximation (40), which vanishes as $R_2 \to C_2$. 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

$$R_1 \leq \frac{1}{2}\log(1 + P_1(1-\rho^2)) \tag{46}$$
$$R_2 \leq C_2 \tag{47}$$
$$R_1 + R_2 \leq \frac{1}{2}\log\left(1 + \frac{P_2}{1 + P_1 + Q + 2\rho\sqrt{P_1 Q}}\right) + \frac{1}{2}\log(1 + P_1(1-\rho^2)) \tag{48}$$

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

Remark 3: The upper bound (48) on the sum rate coincides with the upper bound (27) on the capacity of the helper problem. In fact, the proof of (48) generalizes the proof techniques used in the derivation of (27).

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

$$R_1 \leq I(X_1; Y | X_2, S) \tag{49}$$
$$R_2 \leq I(X_2; Y | X_1, S) \tag{50}$$
$$R_1 + R_2 \leq I(X_1; Y | X_2, S) + I(X_2; Y) \tag{51}$$

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

$$\mathbb{E}\left[X_1^2\right] \leq P_1 \text{ and } \mathbb{E}\left[X_2^2\right] \leq P_2. \tag{52}$$

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 (proof omited). Evaluating (49)–(51) 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.,

$$C_{\text{sum}} \triangleq \max\{R_1 + R_2 : (R_1, R_2) \in \mathcal{C}(P_1, P_2, Q)\}. \tag{53}$$

By comparing the inner bound (9) (evaluated using Gaussian inputs) and the outer bound (48), 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

$$C_{\text{sum}} = \max_{\rho \in [-1,0]} \frac{1}{2}\left\{ \log\left(1 + \frac{P_2}{1 + P_1 + Q + 2\rho\sqrt{P_1 Q}}\right) + \frac{1}{2}\log(1 + P_1(1-\rho^2)) \right\} \tag{54}$$

or equivalently,

$$C_{\text{sum}} = C_2 + f(1). \tag{55}$$

Proof: The converse part of (54) follows directly from (48). Since the objective function on the RHS of (54) 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

$$\bar{R}_1 \triangleq \frac{1}{2}\log(1 + P_1(1 - (\rho^*)^2)) \tag{56}$$
$$\bar{R}_2 \triangleq \frac{1}{2}\log\left(1 + \frac{P_2}{1 + P_1 + Q + 2\rho^*\sqrt{P_1 Q}}\right) \tag{57}$$

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

$$\alpha = \frac{P_1(1 - (\rho^*)^2)}{P_1(1 - (\rho^*)^2) + 1} \tag{58}$$

in (10)–(12). The choice of α in (58) is the usual dirty paper coding coefficient for the equivalent channel (obtained by canceling the interference X_2 from the non-cognitive user)

$$\widetilde{Y} = X_0 + \left(1 - \rho^* \sqrt{\frac{P_1}{Q}}\right) S + Z \tag{59}$$

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 (56) and (57) achieves the sum rate capacity (54). The equivalence between (54) and (55) 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

$$R_1 \leq \frac{1}{2} \log\big(1 + P_1(1 - \rho^2)\big) \tag{60}$$

$$R_1 + R_2 \leq \frac{1}{2} \log\left(1 + \frac{P_2}{1 + P_1 + Q + 2\rho\sqrt{P_1 Q}}\right) + \frac{1}{2} \log(1 + P_1(1 - \rho^2)) \tag{61}$$

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

Proof: By Theorem 2, the rate region characterized by (60) and (61), 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 (56) and (57), 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 (61), $R_2 \leq C_{\text{sum}} - R_1$, 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

$$R_2 = \frac{1}{2} \log\left(1 + \frac{P_2}{1 + P_1 + Q + 2\rho_0\sqrt{P_1 Q}}\right). \tag{62}$$

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

$$R_{1,\text{th}} = I(U^*; Y) - I(U^*; S) \tag{63}$$

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

$$\widetilde{C}_1(P_1, P_2, Q) \triangleq \max\{R_1 : (R_1, C_2) \in \mathcal{C}(P_1, P_2, Q)\} \tag{64}$$

$$\widetilde{C}_2(P_1, P_2, Q) \triangleq \max\{R_2 : (C_1, R_2) \in \mathcal{C}(P_1, P_2, Q)\}. \tag{65}$$

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

$$\widetilde{C}_2(P_1, P_2, Q) = \frac{1}{2} \log\left(1 + \frac{P_2}{1 + P_1 + Q}\right). \tag{66}$$

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

$$\widetilde{C}_1(P_1, P_2, Q) = f(0) \tag{67}$$

where $f(\cdot)$ is defined in (33).

Proof: The corner point (66) follows from (46) and (48) (with $\rho = 0$), and (67) follows from (30) by setting $R_2 = C_2$, and by taking $\delta = 0$. ■

A few remarks are in order.

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

$$\lim_{Q\to\infty} \widetilde{C}_1(P_1, P_2, Q) = \frac{1}{2} \log \frac{P_1}{1 + P_2} \tag{68}$$

$$\lim_{Q\to\infty} \widetilde{C}_2(P_1, P_2, Q) = 0. \tag{69}$$

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

$$\lim_{Q\to\infty} \widetilde{C}_1(P_1, P_2, Q) \leq \frac{1}{2} \log \frac{1 + P_1}{1 + P_2}. \tag{70}$$

- The top corner point $(\widetilde{C}_1, C_2)$ 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 $n \to \infty$, then by (40) and the I-MMSE relation [26], the minimum mean-square error (MMSE) in estimating $X_1^n + S^n$ given Y_G^n satisfies

$$\mathsf{MMSE}(X_1^n + S^n | Y_G^n) = o(n). \tag{71}$$

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 outer bounds in Theorem 1 (dashed red curve) and Theorem 2 (solid red curve) with the inner (solid blue curve) and outer bounds (solid black curve) 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 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.
- The improvement of Theorem 1 (dashed red curve) over the genie aided outer bound (solid black curve) is not clearly visible in the figure. However, numerically the outer bound in Theorem 1 indicates that R_2 is strictly

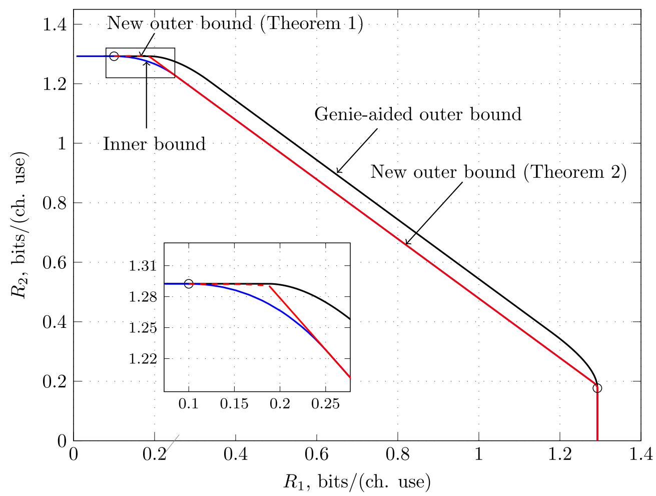

Fig. 3. Inner and outer bounds on the capacity region region $\mathcal{C}(P_1, P_2, Q)$ with $P_1 = 5$, $P_2 = 5$, and $Q = 12$.

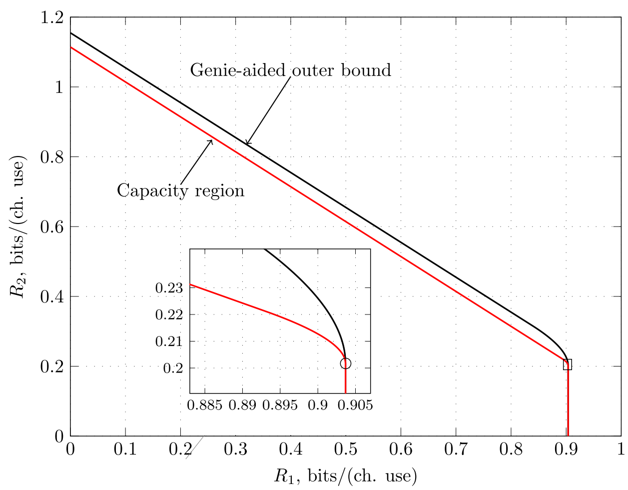

Fig. 4. A comparison between the capacity region $\mathcal{C}(P_1, P_2, Q)$ and the genie-aided outer bound with $P_1 = 2.5$, $P_2 = 5$, and $Q = 12$.

below $C_2 = 1.29$ when $R_1 > 0.1$. This implies that the top corner point of the capacity region is given by the rate pair $(0.1, 1.29)$, which confirms the corner point result in Section III-A.3.

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 $Q = 12$, which do not satisfy Condition 1. 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., (43)). If the state sequence S^n is non-Gaussian but is i.i.d., then the upper bound (30) remains valid if $f(\delta)$ is replaced by

$$\tilde{f}(\delta) \triangleq \max_{P_{X_1|S}} \Big\{ I(X_1 + S; Y_G) - I(S; Y_\delta) \Big\}. \quad (72)$$

In this case, the top corner point becomes

$$\widetilde{C}_1 = \max_{P_{X_1|S}} \{ I(X_1 + S; Y_G) - I(X_1 + S; S) \} \quad (73)$$

and the sum rate capacity becomes

$$C_{\text{sum}} = \max_{P_{X_1|S} P_{X_2}} \Big(I(X_1; Y|X_2, S) + I(X_2; Y) \Big). \quad (74)$$

Furthermore, both (54) and (74) 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 [27, 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

$$R_2 \le C_{\text{helper}} \quad (75)$$

and

$$R_1 \le \min_{0 \le \delta \le 1} \left\{ \frac{1}{2} \log\left(1 + \frac{1 + P_2 - \delta}{P_2 \delta} \tilde{g}(R_2) \right) + f(\delta) \right\} + (c_2 + c_3)(C_2 - R_2) \quad (76)$$

where $f(\cdot)$ is defined in (33),

$$\tilde{g}(R_2) \triangleq \exp\Big(2c_2 \sqrt{C_2 - R_2} + 2(C_2 - R_2) \Big) - 1 \quad (77)$$

with

$$c_2 \triangleq \frac{3\sqrt{1 + (\sqrt{P_1} + \sqrt{P_2} + \sqrt{Q})^2} + 4(\sqrt{P_1} + \sqrt{Q})}{\sqrt{(1 + P_2)/(2 \log e)}} \quad (78)$$

and

$$c_3 \triangleq \sqrt{2(1 + P_2) \log e} \cdot \Big(3\sqrt{1 + (\sqrt{P_1} + \sqrt{P_2} + \sqrt{Q})^2} + 4(\sqrt{P_1} + \sqrt{P_2} + \sqrt{Q}) \Big). \quad (79)$$

Proof: See Section IV-C. ∎

As a corollary of Theorem 7, we establish that under Condition 1, the top corner point established in (67) is unchanged even if the non-cognitive user knows the message of the cognitive user. Formally, the top corner point is defined as

$$\widetilde{C}_{\text{deg},1}(P_1, P_2, Q) \triangleq \max\{ R_1 : (R_1, C_2) \in \mathcal{C}_{\text{deg}}(P_1, P_2, Q) \} \quad (80)$$

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

$$\widetilde{C}_{\text{deg},1}(P_1, P_2, Q) = f(0) \quad (81)$$

with $f(\cdot)$ defined in (33).

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

$$R_2 \leq \frac{1}{2}\log(1 + P_2(1-\rho_2^2)) \tag{82}$$

$$R_2 \leq \frac{1}{2}\log(1 + P_1(1-\rho_1^2-\rho_s^2)) + \frac{1}{2}\log\Big(1+\frac{P_2(1-\rho_2^2)}{1+(\sqrt{Q}+\rho_s\sqrt{P_1})^2+P_1(1-\rho_1^2-\rho_s^2)}\Big) \tag{83}$$

$$R_1 + R_2 \leq \frac{1}{2}\log(1 + P_1(1-\rho_1^2-\rho_s^2)) + \frac{1}{2}\log\Big(1+\frac{P_2(1-\rho_2^2)+(\rho_2\sqrt{P_2}+\rho_1\sqrt{P_1})^2}{1+(\sqrt{Q}+\rho_s\sqrt{P_1})^2+P_1(1-\rho_1^2-\rho_s^2)}\Big) \tag{84}$$

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

$$\rho_1^2 + \rho_s^2 \leq 1. \tag{85}$$

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

$$R_2 \leq I(X_2; Y|X_1, S, U) \tag{86}$$

$$R_2 \leq I(X_1; Y|X_2, S, U) + I(X_2; Y|U) \tag{87}$$

$$R_1 + R_2 \leq I(X_1; Y|X_2, S, U) + I(X_2, U; Y) \tag{88}$$

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;
- $\mathbb{E}\big[X_1^2\big] \leq P_1$ and $\mathbb{E}\big[X_2^2\big] \leq P_2$.

To prove Theorem 9, it remains to show that the bounds in (86)–(88) 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

$$R_1 + R_2 \leq I(X_1; Y|S, X_2) + I(X_2; Y) \tag{89}$$

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 bound determined by (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 $U = X_2$).

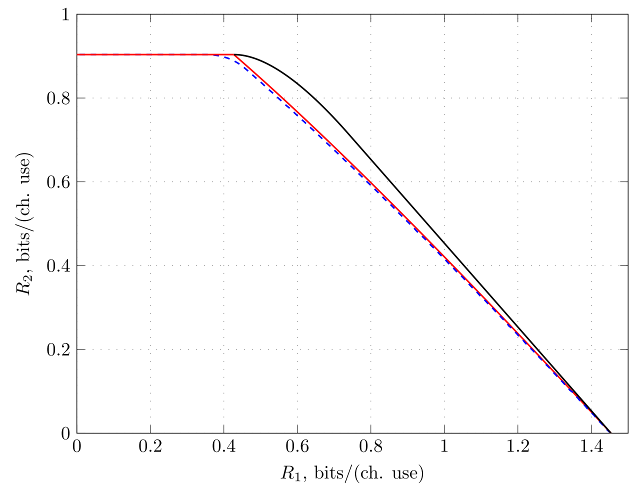

Fig. 5. Inner and outer bounds for the capacity region of the dirty MAC with degraded message sets for $P_1 = 4$, $P_2 = 2.5$, and $Q = 5$. The red solid curve denotes our new outer bound in Theorem 9, the blue dashed curve and the black curve denote the inner and outer bounds obtained in [5].

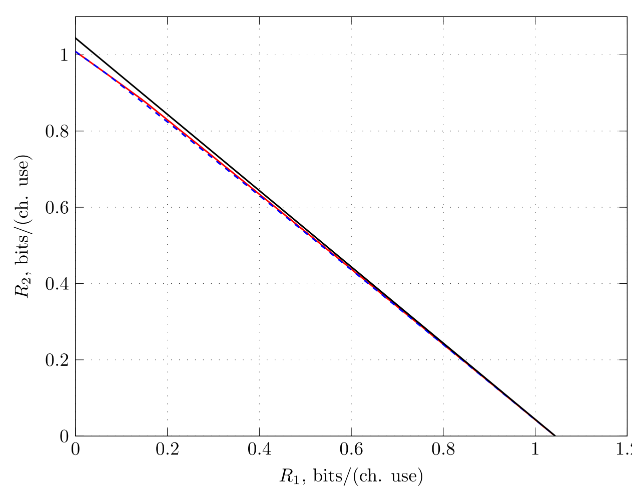

Fig. 6. Inner and outer bounds for the capacity region of the dirty MAC with degraded message sets for $P_1 = 2$, $P_2 = 5$, and $Q = 12$. The red solid curve denotes our new outer bound in Theorem 9, the blue dashed curve and the black curve denote the inner and outer bounds obtained in [5].

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 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 $R_1 \le 0.1$ 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

$$C_{\text{helper}} \le \max\Bigg\{ R_2 : R_2 \le C_2, \text{ and } \min_{0\le\delta\le 1} \left\{ \frac{1}{2}\log\left(1 + \frac{1+P_2-\delta}{P_2\delta} g(R_2)\right) + f(\delta) \right\} \ge 0 \Bigg\} \tag{90}$$

where $g(\cdot)$ and $f(\cdot)$ are defined in (31) and (33), respectively.

Proof: Setting $R_1 = 0$ in the outer bound (30) in Theorem 1, we conclude that the rate R_2 of the non-cognitive user must satisfy

$$\min_{0\le\delta\le 1} \left\{ \frac{1}{2}\log\left(1 + \frac{1+P_2-\delta}{P_2\delta} g(R_2)\right) + f(\delta) \right\} \ge 0. \tag{91}$$

This implies (90). ∎

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 three 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) $f(0) \ge 0$, where $f(\cdot)$ is defined in (33).

In Fig. 7, we compare the new upper bound in Theorem 11 with the upper and lower bounds in [11]. The two upper bounds reported in [11, Lemmas 2 and 3] correspond to

$$C_{\text{helper}} \le C_{\text{sum}} \tag{92}$$

and

$$C_{\text{helper}} \le \frac{1}{2}\log(1+P_2) \tag{93}$$

respectively. The lower bound (achievability bound) is [11, Th. 1]. As observed in [11], the upper bound (92) is tight (i.e., $C_{\text{helper}} = C_{\text{sum}}$) if $P_1 \le 2.5$, and the bound (93) is tight (i.e., $C_{\text{helper}} = \frac{1}{2}\log(1+P_2)$) if $P_1 \ge 4.5$. Our new upper bound is tighter than (92) and (93) for $P_1 \in [3.5, 4.5]$.

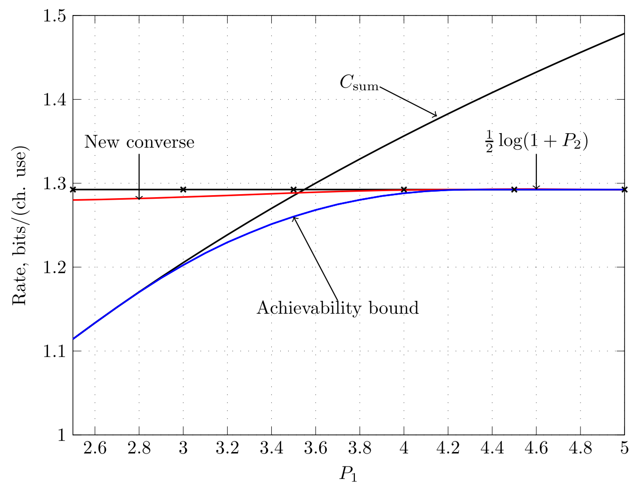

Fig. 7. Upper and lower bounds on C_{helper} as a function of P_1 for $P_2 = 5$ and $Q = 12$.

IV. Technical Proofs

A. Proof of Theorem 1

Define

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

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

By Fano's inequality, there is no difference asymptotically between this definition of the rate and the operational one (i.e., the ratio between the logarithm of the number of messages and the blocklength). Without loss of generality, we further assume that X_1^n and X_2^n have zero mean.

The upper bound (29) is straightforward. The proof of (30), which builds upon the intuition described in Section III-A, consists of three steps.

1) We derive the following upper bound on R_1 by standard manipulations of the mutual information terms:

$$nR_1 \le \underbrace{I(X_1^n + S^n; Y_\delta^n) - I(X_1^n + S^n; Y_G^n)}_{\triangleq I_\delta} + \underbrace{I(X_1^n + S^n; Y_G^n) - I(S^n; Y_\delta^n)}_{\triangleq J_\delta}. \tag{96}$$

Here, Y_G^n and Y_δ^n are defined in (34) and (35), respectively. Note that, the upper bound on R_1 in (96) depends on the joint distribution of X_1^n and S^n but not on X_2^n.

2) We upper-bound the term I_δ in (96) as follows:

$$I_\delta \le \frac{n}{2}\log\left(1 + \frac{1+P_2-\delta}{P_2\delta} g(R_2)\right) \tag{97}$$

where $g(R_2)$ is defined in (31). The derivation relies on an elegant argument of Polyanskiy and Wu [19], used in the derivation of the outer bound on the capacity region of Gaussian interference channels.

3) We show that the term J_δ in (96) can be single-letterized as

$$J_\delta \le n \max_{P_{X_1|S}:\mathbb{E}[X_1^2]\le P_1} \Big\{ I(X_1+S;Y_G) - I(S;Y_\delta) \Big\}. \tag{98}$$

Then, we show that the expression on the RHS of (98) is maximized when X_1 and S are jointly Gaussian. Substituting the upper bounds on I_δ and J_δ derived in Step 2 and Step 3 into (96), we obtain the desired upper bound (30).

1) Step 1: Proof of (96)*:* We start by observing that the channel $M_1 \to Y^n$ is stochastically degraded with respect to the channel $M_1 \to Y_\delta^n$, since Y^n has the same distribution as $Y_\delta^n + X_2^n + \sqrt{1-\delta^2}\tilde{Z}^n$, where $\tilde{Z}^n \sim \mathcal{N}(0, \mathsf{I}_n)$. This implies that

$$nR_1 = I(M_1;Y^n) \le I(M_1;Y_\delta^n). \tag{99}$$

The mutual information $I(M_1;Y_\delta^n)$ can be upper-bounded as follows:

$$\begin{aligned} I(M_1;Y_\delta^n) &= I(M_1,S^n;Y_\delta^n) - I(S^n;Y_\delta^n|M_1) && (100)\\ &\le I(M_1,X_1^n,S^n;Y_\delta^n) - I(S^n;Y_\delta^n|M_1) && (101)\\ &= I(X_1^n,S^n;Y_\delta^n) + I(M_1;Y_\delta^n|X_1^n,S^n) \\ &\quad - I(S^n;Y_\delta^n,M_1) + I(S^n;M_1) && (102)\\ &= I(X_1^n,S^n;Y_\delta^n) - I(S^n;Y_\delta^n,M_1) && (103)\\ &= I(X_1^n+S^n;Y_\delta^n) - I(S^n;Y_\delta^n) - I(S^n;M_1|Y_\delta^n) && (104)\\ &\le I(X_1^n+S^n;Y_\delta^n) - I(S^n;Y_\delta^n). && (105)\end{aligned}$$

Here, in (103) we used that $I(M_1;Y_\delta^n|X_1^n,S^n) = 0$, which follows because $M_1 \to (X_1,S^n) \to Y_\delta^n$ forms a Markov chain; (104) follows because $I(X_1^n,S^n;Y_\delta^n) = I(X_1^n+S^n;Y_\delta^n)$; and (105) follows because $I(S^n;M_1|Y_\delta^n) \ge 0$. It is not difficult to verify that (105) is equivalent to (96).

2) Step 2: Upper-Bounding I_δ: We next upper-bound the term I_δ defined in (96). The derivation follows closely the proof of [19, Th. 7]. Let

$$N_S(\gamma) \triangleq \exp\left\{\frac{2}{n}h(X_1^n+S^n+\sqrt{\gamma}Z^n)\right\} \tag{106}$$

where $Z^n \sim \mathcal{N}(0,\mathsf{I}_n)$ is independent of X_1^n and S^n. By Costa's entropy power inequality [25], the function $N_S(\cdot)$ is concave. The term I_δ can be expressed in terms of $N_S(\cdot)$ as

$$I_\delta = \frac{n}{2}\log\frac{N_S(\delta)}{N_S(1+P_2)} + \frac{n}{2}\log\frac{1+P_2}{\delta}. \tag{107}$$

Repeating the steps in [19, eqs. (41)–(43)], we obtain (recall that $G^n \sim \mathcal{N}(0, P_2\mathsf{I}_n)$)

$$D(P_{X_2^n+Z^n}\|P_{G^n+Z^n}) \le n(C_2 - R_2) \tag{108}$$

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

$$\begin{aligned} nR_2 &= I(X_2^n;Y^n) && (109)\\ &= h(Y^n) - h(Y_G^n) + h(Y_G^n) - h(X_1^n+S^n+Z^n) && (110)\\ &= h(Y^n) - h(Y_G^n) + \frac{n}{2}\log\frac{N_S(1+P_2)}{N_S(1)}. && (111)\end{aligned}$$

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

$$\begin{aligned} &\mathbb{E}\Big[\|X_1^n+S^n\|^2\Big] \\ &= \mathbb{E}\Big[\|X_1^n\|^2\Big] + \mathbb{E}\Big[\|S^n\|^2\Big] + 2\mathbb{E}\big[\langle X_1^n,S^n\rangle\big] && (112)\\ &\le nP_1 + nQ + 2\mathbb{E}\big[\|X_1^n\|\|S^n\|\big] && (113)\\ &\le nP_1 + nQ + 2\sqrt{\mathbb{E}\big[\|X_1^n\|^2\big]\,\mathbb{E}\big[\|S^n\|^2\big]} && (114)\\ &\le n(\sqrt{P_1}+\sqrt{Q})^2. && (115)\end{aligned}$$

By [19, Propostion. 2], the random variable Y_G^n is $\big(\frac{3\log e}{1+P_2}, \frac{4(\sqrt{P_1}+\sqrt{Q})\log e}{1+P_2}\big)$-regular, i.e., the probability density function $p_{Y_G^n}(y^n)$ of Y_G^n satisfies

$$\|\nabla \log p_{Y_G^n}(y^n)\| \le \frac{3\log e}{1+P_2}\|y^n\| + \frac{4(\sqrt{P_1}+\sqrt{Q})\log e}{1+P_2}, \quad \forall y^n \in \mathbb{R}^n. \tag{116}$$

Therefore, by [19, Propostion 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 [28, p. 12] for the definition of W_2) as

$$\begin{aligned} &h(Y^n) - h(Y_G^n) \\ &\le \left(3\sqrt{1+(\sqrt{P_1}+\sqrt{Q})^2+P_2} + 4(\sqrt{P_1}+\sqrt{Q})\right) \\ &\quad \cdot \frac{\sqrt{n}\log e}{1+P_2}\cdot W_2(P_{Y^n}\|P_{Y_G^n}). && (117)\end{aligned}$$

Furthermore, we have

$$\begin{aligned} W_2(P_{Y^n}\|P_{Y_G^n}) &\le W_2(P_{X_2^n+Z^n}\|P_{G^n+Z^n}) && (118)\\ &\le \sqrt{\frac{2(1+P_2)}{\log e}D(P_{X_2^n+Z^n}\|P_{G^n+Z^n})} && (119)\\ &\le \sqrt{\frac{2n(1+P_2)}{\log e}(C_2-R_2)}. && (120)\end{aligned}$$

Here, (118) follows because the $W_2(\cdot,\cdot)$ distance is non-increasing under convolutions, (119) follows by using Talagrand's inequality [20], and (120) follows from (108). Substituting (120) into (117), and then (117) into (111), we conclude that

$$\log\frac{N_S(1)}{N_S(1+P_2)} \le 2c_1\sqrt{C_2-R_2} + 2(C_2-R_2) - \log(1+P_2) \tag{121}$$

where c_1 is defined in (32), or equivalently,

$$\frac{N_S(1)}{N_S(1+P_2)} \le \frac{\exp\big(2c_1\sqrt{C_2-R_2}+2(C_2-R_2)\big)}{1+P_2}. \tag{122}$$

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

$$\alpha\delta + (1-\alpha)(1+P_2) = 1. \tag{123}$$

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

$$\alpha N_S(\delta) + (1-\alpha)N_S(1+P_2) \le N_S(1) \tag{124}$$

which implies that

$$\begin{aligned}\frac{N_S(\delta)}{N_S(1+P_2)} &\le \frac{1}{\alpha}\frac{N_S(1)-(1-\alpha)N_S(1+P_2)}{N_S(1+P_2)} \quad (125)\\ &\le \frac{1}{\alpha}\left(\frac{\exp\big(2c_1\sqrt{C_2-R_2}+2(C_2-R_2)\big)}{1+P_2}-1+\alpha\right). \quad (126)\end{aligned}$$

Substituting (126) into (107), we conclude the desired upper bound (97) on I_δ.

3) Step 3: Upper-Bounding J_δ: We proceed to upper-bound the term J_δ defined in (96). Observe that

$$\begin{aligned}I(X_1^n+S^n;Y_G^n) &= \sum_{i=1}^n \Big(h(Y_{G,i}|Y_G^{i-1})-h(Y_{G,i}|X_{1,i},S_i)\Big) \quad (127)\\ &\le \sum_{i=1}^n \big(h(Y_{G,i})-h(Y_{G,i}|X_{1,i},S_i)\big) \quad (128)\\ &= \sum_{i=1}^n I(X_{1,i}+S_i;Y_{G,i}) \quad (129)\end{aligned}$$

and

$$\begin{aligned}I(S^n;Y_\delta^n) &= h(S^n)-h(S^n|Y_\delta^n) \quad (130)\\ &= \sum_{i=1}^n \Big(h(S_i)-h(S_i|Y_\delta^n,S^{i-1})\Big) \quad (131)\\ &\ge \sum_{i=1}^n \big(h(S_i)-h(S_i|Y_{\delta,i})\big) \quad (132)\\ &= \sum_{i=1}^n I(S_i;Y_{\delta,i}) \quad (133)\end{aligned}$$

where both (128) and (132) follow because conditioning reduces entropy. Combining (129) and (133), we obtain

$$I(X_1^n+S^n;Y_G^n)-I(S^n;Y_\delta^n) \le \sum_{i=1}^n \big(I(X_{1,i}+S_i;Y_{G,i})-I(S_i;Y_{\delta,i})\big) \quad (134)$$

where the RHS of (134) 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 (44) and (45), 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 [29, 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 the single-letter upper bound (98) on J_δ.

To solve the maximization problem in (98), we next invoke the Gaussian saddle-point property as explained in the intuitive argument after Theorem 1. Lemma 13 below generalizes the well-known worst-case Gaussian noise result [21], [22] to the case in which the noise and the Gaussian input are dependent.

Lemma 13 ([23, Th. 1]): Let $\boldsymbol{X}_G \sim \mathcal{N}(\mathbf{0},\mathsf{K}_x)$ and $\boldsymbol{Z}_G \sim \mathcal{N}(\mathbf{0},\mathsf{K}_z)$ be Gaussian random vectors in $\mathbb{R}^d$. Let $\boldsymbol{Z}$ be a random vector in $\mathbb{R}^d$ with the same covariance matrix as $\boldsymbol{Z}_G$. Assume that $\boldsymbol{X}_G$ is independent of $\boldsymbol{Z}_G$, and that

$$\mathbb{E}\Big[\boldsymbol{X}_G\boldsymbol{Z}^{\mathrm{T}}\Big]=\mathbf{0}_{d\times d} \quad (135)$$

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

$$I(\boldsymbol{X}_G;\boldsymbol{X}_G+\boldsymbol{Z}_G) \le I(\boldsymbol{X}_G;\boldsymbol{X}_G+\boldsymbol{Z}). \quad (136)$$

We proceed as follows. For a given $P_{X_1|S}$, let $\rho \triangleq \mathbb{E}[X_1\, S]/\sqrt{P_1Q}$ be the correlation coefficient between X_1 and S. Denote

$$\begin{aligned}\widetilde{X}_1 &\triangleq X_1-\rho\sqrt{P_1/Q}S \quad (137)\\ \widetilde{S} &\triangleq (1+\rho\sqrt{P_1/Q})S. \quad (138)\end{aligned}$$

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

$$I(X_1+S;Y_G)=I(\widetilde{X}_1+\widetilde{S};\widetilde{X}_1+\widetilde{S}+\sqrt{1+P_2}Z) \quad (139)$$

and

$$I(S;Y_\delta) \ge I(\widetilde{S};Y_\delta)=I(\widetilde{S};\widetilde{S}+\widetilde{X}_1+\sqrt{\delta}Z) \quad (140)$$

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

Observe now that, for a fixed ρ and $b \triangleq \mathbb{E}\big[\widetilde{X}_1^2\big]$, the mutual information term in (139) is maximized when $\widetilde{X}_1$ is Gaussian and is independent of S. Furthermore, by Lemma 13, the mutual information term on the RHS of (140) is minimized also when $\widetilde{X}_1$ is Gaussian and is independent of S. Therefore, we conclude that

$$\begin{aligned}&\max_{P_{X_1|S}:\mathbb{E}[X_1^2]\le P_1}\Big\{I(X_1+S;Y_G)-I(S;Y_\delta)\Big\}\\ &\quad\le \max_{b,\rho}\frac{1}{2}\log\frac{(1+P_2+b+(1+\rho\sqrt{P_1/Q})^2Q)(\delta^2+b)}{(\delta^2+b+(1+\rho\sqrt{P_1/Q})^2Q)(1+P_2)} \quad (141)\end{aligned}$$

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

$$b\ge 0, \quad \text{and} \quad b+\rho^2P_1 \le P_1. \quad (142)$$

By examining the Karush-Kuhn-Tucker (KKT) necessary conditions [30, Sec. 5.5.3], it can be shown that the constraint $b+P_1\rho^2 \le P_1$ 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 (141) can be simplified to the one dimensional one in (33). In other words, we have proved that

$$J_\delta \le nf(\delta). \quad (143)$$

Finally, substituting (97) and (143) into (96), and optimizing the resulting bound over δ, we conclude the desired result (30).

B. Proof of Proposition 3

It is straightforward to show the bounds

$$nR_1 \le \sum_{i=1}^{n} I(X_{1,i}; Y_i | X_{2,i}, S_i) \tag{144}$$

and

$$nR_2 \le \sum_{i=1}^{n} I(X_{2,i}; Y_i | X_{1,i}, S_i). \tag{145}$$

The counterpart of (51) 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

$$\begin{aligned} n(R_1+R_2) &= I(M_1; Y^n) + I(X_2^n; Y^n) && (146)\\ &= I(M_1, X_2^n; Y^n) - I(X_2^n; M_1 | Y^n) && (147)\\ &\le h(Y^n) - h(Y^n | M_1, X_2^n) && (148)\\ &\le \sum_{i=1}^{n} h(Y_i) - h(Y^n | M_1, X_2^n). && (149)\end{aligned}$$

Here, (147) 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:

$$\begin{aligned} & h(Y^n | M_1, X_2^n) && (150)\\ &= h(Y^n, S^n | M_1, X_2^n) - h(S^n | Y^n, M_1, X_2^n) && (151)\\ &= h(S^n | M_1, X_2^n) + h(Y^n | M_1, X_2^n, S^n) \\ &\quad - h(S^n | Y^n, M_1, X_2^n) && (152)\\ &= h(S^n) + h(Y^n | X_1^n, S^n, X_2^n) - h(S^n | Y^n, M_1, X_2^n) && (153)\\ &\ge h(S^n) + h(Y^n | X_1^n, S^n, X_2^n) - h(S^n | Y^n, X_2^n) && (154)\\ &\ge \sum_{i=1}^{n} \left(h(S_i) + h(Y_i | X_{1,i}, S_i, X_{2,i}) - h(S_i | Y_i, X_{2,i}) \right). && (155)\end{aligned}$$

Here, both (154) and (155) hold because conditioning does not increase differential entropy. Substituting (155) into (149), we conclude that

$$\begin{aligned} n(R_1+R_2) &\le \sum_{i=1}^{n} \Big(h(Y_i) - h(Y_i | X_{1,i}, S_i, X_{2,i}) \\ &\qquad - h(S_i) + h(S_i | Y_i, X_{2,i}) \Big) && (156)\\ &= \sum_{i=1}^{n} \Big(h(Y_i) - h(Y_i | X_{1,i}, S_i, X_{2,i}) \\ &\qquad - h(Y_i | X_{2,i}) + h(Y_i | S_i, X_{2,i}) \Big) && (157)\\ &= \sum_{i=1}^{n} \Big(I(X_{1,i}; Y_i | X_{2,i}, S_i) + I(X_{2,i}; Y_i) \Big). && (158)\end{aligned}$$

Here, (157) follows because S_i and $X_{2,i}$ are independent.

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

$$\begin{aligned} R_1 &\le I(X_1; Y | X_2, S, Q) && (159)\\ R_2 &\le I(X_2; Y | X_1, S, Q) && (160)\\ R_1 + R_2 &\le I(X_1; Y | X_2, S, Q) + I(X_2; Y | Q). && (161)\end{aligned}$$

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 (108)–(126) 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 (96) is defined as

$$\begin{aligned} \tilde{I}_\delta &\triangleq I(X_1^n + S^n; Y_\delta^n | M_1) - I(X_1^n + S^n; Y_G^n | M_1) && (162)\\ &= \frac{n}{2} \mathbb{E}_{M_1}\left[\log \frac{\widetilde{N}_S(\delta | M_1)}{\widetilde{N}_S(1+P_2 | M_1)} \right] + \frac{n}{2} \log \frac{1+P_2}{\delta} && (163)\end{aligned}$$

where

$$\widetilde{N}_S(\gamma | m) \triangleq \exp\left\{ \frac{2}{n} h(X_1^n + S^n + \sqrt{\gamma} Z^n | M_1 = m) \right\}. \tag{164}$$

The function $\widetilde{N}_S(\gamma|m)$ 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 $\tilde{I}_\delta$. Let

$$\begin{aligned} R_1 &\triangleq I(M_1; Y^n) && (165)\\ R_2 &\triangleq I(X_2^n; Y^n | M_1). && (166)\end{aligned}$$

Again, by Fano's inequality, the definitions of the rates in (165) and (166) agree with the operational ones. With the conditioning on M_1, the bounds (108) and (111) become

$$D(P_{X_2^n + Z^n | M_1} \| P_{G^n + Z^n} | P_{M_1}) \le n(C_2 - R_2) \tag{167}$$

and

$$\begin{aligned} nR_2 &= h(Y^n | M_1) - h(Y_G^n | M_1) \\ &\quad + \mathbb{E}_{M_1}\left[\frac{n}{2} \log \frac{\tilde{N}_S(1+P_2 | M_1)}{\tilde{N}_S(1 | M_1)} \right]. && (168)\end{aligned}$$

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

$$\begin{aligned} & D(P_{X_2^n + Z^n | M_1} \| P_{G^n + Z^n} | P_{M_1}) \\ &\qquad \triangleq \mathbb{E}_{M_1}\Big[D(P_{X_2^n + Z^n | M_1} \| P_{G^n + Z^n}) \Big]. && (169)\end{aligned}$$

Using [19, Propositions 1 and 2] and (167), we bound the difference $h(Y^n|M_1) - h(Y_G^n|M_1)$ as follows:

$$\begin{aligned}
&h(Y^n|M_1) - h(Y_G^n|M_1) \\
&\le \frac{\log e}{1+P_2}\mathbb{E}_{M_1}\Big[W_2(P_{Y_G^n|M_1}, P_{Y^n|M_1})\Big(4\mathbb{E}\big[\|X_1^n + S^n\||M_1\big] \\
&\qquad + \frac{3}{2}\sqrt{\mathbb{E}\big[\|Y_G^n\|^2|M_1\big]} + \frac{3}{2}\sqrt{\mathbb{E}\big[\|Y^n\|^2|M_1\big]}\Big)\Big] \quad (170)\\
&\le \frac{\log e}{1+P_2}\mathbb{E}_{M_1}\Big[\sqrt{\frac{2(1+P_2)}{\log e}D(P_{X_2^n+Z^n|M_1}\|P_{G^n+Z^n})}\cdot \\
&\qquad \Big(4\sqrt{\mathbb{E}\big[\|X_1^n + S^n\|^2|M_1\big]} + \frac{3}{2}\sqrt{\mathbb{E}\big[\|Y_G^n\|^2|M_1\big]} \\
&\qquad + \frac{3}{2}\sqrt{\mathbb{E}\big[\|Y^n\|^2|M_1\big]}\Big)\Big] \quad (171)\\
&\le \frac{\log e}{1+P_2}\sqrt{\frac{2(1+P_2)}{\log e}D(P_{X_2^n+Z^n|M_1}\|P_{G^n+Z^n}|P_{M_1})}\cdot \\
&\qquad \Big(4\sqrt{\mathbb{E}\big[\|X_1^n + S^n\|^2\big]} + \frac{3}{2}\sqrt{\mathbb{E}\big[\|Y_G^n\|^2\big]} + \frac{3}{2}\sqrt{\mathbb{E}\big[\|Y^n\|^2\big]}\Big) \quad (172)\\
&\le c_2 n\sqrt{C_2 - R_2} \quad (173)
\end{aligned}$$

where

$$c_2 \triangleq \frac{3\sqrt{1+(\sqrt{P_1}+\sqrt{P_2}+\sqrt{Q})^2} + 4(\sqrt{P_1}+\sqrt{Q})}{\sqrt{(1+P_2)/(2\log e)}}. \quad (174)$$

Here, (170) follows from [19, Propositions 1 and 2]; (171) follows because for every message m,

$$\mathbb{E}\big[\|X_1^n + S^n\||M_1 = m\big] \le \sqrt{\mathbb{E}\big[\|X_1^n + S^n\|^2|M_1 = m\big]} \quad (175)$$

and

$$\begin{aligned}
&W_2(P_{Y_G^n|M_1=m}, P_{Y^n|M_1=m}) \\
&\quad \le W_2(P_{X_2^n+Z^n|M_1=m}, P_{G^n+Z^n}) \quad (176)\\
&\quad \le \sqrt{\frac{2(1+P_2)}{\log e}D(P_{X_2^n+Z^n|M_1=m}\|P_{G^n+Z^n})} \quad (177)
\end{aligned}$$

where (176) 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 (177) follows from Talagrand's inequality [20]; (172) follows from the Cauchy-Schwarz inequality; and finally (173) follows from (167), (115), and because

$$\frac{1}{n}\mathbb{E}\Big[\|Y^n\|^2\Big] \le 1 + (\sqrt{P_1}+\sqrt{P_2}+\sqrt{Q})^2 \quad (178)$$

$$\frac{1}{n}\mathbb{E}\Big[\|Y_G^n\|^2\Big] \le 1 + (\sqrt{P_1}+\sqrt{P_2}+\sqrt{Q})^2. \quad (179)$$

Substituting (173) into (168), we conclude that

$$\begin{aligned}
&\mathbb{E}_{M_1}\left[\log\frac{\tilde{N}_S(1+P_2|M_1)}{\tilde{N}_S(1|M_1)}\right] \\
&\quad \le 2c_2\sqrt{C_2-R_2} + 2(C_2-R_2) - \log(1+P_2). \quad (180)
\end{aligned}$$

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

$$\begin{aligned}
&\mathbb{E}_{M_1}\left[\log\frac{\widetilde{N}_S(\delta|M_1)}{\widetilde{N}_S(1+P_2|M_1)}\right] \\
&\le \mathbb{E}_{M_1}\left[\log\left(\frac{\widetilde{N}_S(1|M_1)}{\widetilde{N}_S(1+P_2|M_1)} - 1 + \alpha\right)\right] - \log\alpha \quad (181)\\
&\le \log\left(\frac{\exp\big(2c_2\sqrt{C_2-R_2}+2(C_2-R_2)\big)}{1+P_2} - 1 + \alpha\right) - \log\alpha. \quad (182)
\end{aligned}$$

Here, (181) follows from the concavity of $\gamma \mapsto \widetilde{N}_S(\gamma|M_1)$, and (182) follows from Jensen's inequality and because the function $x \mapsto \log(\exp(x) - (1-\alpha))$ is concave. Finally, substituting (182) into (163), we conclude that

$$\tilde{I}_\delta \le \frac{n}{2}\log\left(1 + \frac{1+P_2-\delta}{P_2\delta}\tilde{g}(R_2)\right) \quad (183)$$

where $\tilde{g}(R_2)$ is defined in (77).

We next relate $\tilde{I}_\delta$ to R_1. This part is quite different from the steps in Section IV-A.1, 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:

$$\begin{aligned}
\tilde{I}_\delta &= I(X_1^n, S^n; Y_\delta^n|M_1) - I(X_1^n + S^n, M_1; Y_G^n) \\
&\quad + I(M_1; Y_G^n) \quad (184)\\
&= I(S^n; Y_\delta^n, M_1) + I(X_1^n; Y_\delta^n|S^n, M_1) \\
&\quad - I(X_1^n + S^n, M_1; Y_G^n) + I(M_1; Y_G^n) \quad (185)\\
&\ge I(S^n; Y_\delta^n) - I(X_1^n + S^n; Y_G^n) + I(M_1; Y_G^n) \quad (186)\\
&= I(S^n; Y_\delta^n) - I(X_1^n + S^n; Y_G^n) + I(M_1; Y_G^n) \\
&\quad - I(M_1; Y^n) + nR_1. \quad (187)
\end{aligned}$$

Here, the penultimate step follows because $M_1 \to X_1^n + S^n \to Y_G^n$ forms a Markov chain. The first two terms on the RHS of (187) can be single-letterized and bounded in the same way as in Section IV-A.3, i.e.,

$$I(S^n; Y_\delta^n) - I(X_1^n + S^n; Y_G^n) \ge -nf(\delta) \quad (188)$$

where $f(\cdot)$ was defined in (33).

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

$$\begin{aligned}
&I(M_1; Y_G^n) - I(M_1; Y^n) \\
&\quad = h(Y_G^n) - h(Y^n) + h(Y^n|M_1) - h(Y_G^n|M_1). \quad (189)
\end{aligned}$$

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 (170)–(173). More specifically, we have

$$h(Y_G^n|M_1) - h(Y^n|M_1) \le c_3 n\sqrt{C_2 - R_2} \quad (190)$$

and

$$h(Y^n) - h(Y_G^n) \le c_2 n\sqrt{C_2 - R_2} \quad (191)$$

where c_3 was defined in (79). Here, to prove (191), we have used

$$\begin{aligned}
D(P_{Y^n}\|P_{Y_G^n}) &\le D(P_{Y^n|M_1}\|P_{Y_G^n|M_1}|P_{M_1}) \quad (192)\\
&\le D(P_{X_2^n+Z^n|M_1}\|P_{G^n+Z^n}|P_{M_1}) \quad (193)\\
&\le n(C_2 - R_2) \quad (194)
\end{aligned}$$

where (192) follows from the data processing inequality, (193) follows from the data processing inequality and because $X_1^n + S^n$ and X_2^n are conditionally independent given M_1, and (194) follows from (167). Substituting (190) and (191) into (189), then (189) and (188) into (187), and combining (187) with (183), we conclude the proof of (76).

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 $\{1, \ldots, n\}$. We have

$$\begin{aligned}
nR_2 &= I(X_2^n; Y^n|M_1) && (195)\\
&\leq I(X_2^n; Y^n, X_1^n, S^n|M_1) && (196)\\
&= I(X_2^n; Y^n|X_1^n, S^n, M_1) && (197)\\
&= h(Y^n|X_1^n, S^n, M_1) - h(Y^n|X_1^n, X_2^n, S^n, M_1) && (198)\\
&\leq \sum_{i=1}^{n} h(Y_i|X_{1,i}, S_i, M_1) - h(Y_i|X_{1,i}, X_{2,i}, S_i, M_1) && (199)\\
&= \sum_{i=1}^{n} I(X_{2,i}; Y_i|X_{1,i}, S_i, M_1) && (200)\\
&= I(X_2; Y|X_1, S, U). && (201)
\end{aligned}$$

This yields the upper bound in (86).

To prove (87), we observe that

$$\begin{aligned}
R_2 &= I(X_2^n; Y^n|M_1) && (202)\\
&= h(Y^n|M_1) - h(Y^n|M_1, X_2^n) && (203)\\
&\leq \sum_{i=1}^{n} h(Y_i|M_1) - h(Y^n|M_1, X_2^n). && (204)
\end{aligned}$$

Proceeding as in (149)–(161) while keeping the conditioning on M_1, we conclude that

$$\begin{aligned}
R_2 &\leq \sum_{i=1}^{n} \Big(I(X_{1,i}; Y_i|X_{2,i}, S_i, M_1) + I(X_{2,i}; Y_i|M_1) \Big) && (205)\\
&= I(X_1; Y|X_2, S, M_1, Q) + I(X_2; Y|Q, M_1) && (206)\\
&\leq I(X_1; Y|X_2, S, U) + I(X_2; Y|U). && (207)
\end{aligned}$$

Finally, we prove (88). We proceed again as in (146)–(155) and keep the conditioning on M_1 whenever appropriate. This yields

$$\begin{aligned}
&n(R_1 + R_2)\\
&\leq \sum_{i=1}^{n} \Big(h(Y_i) - h(Y_i|X_{1,i}, S_i, X_{2,i}) \Big)\\
&\quad - h(S^n|M_1) + h(S^n|Y^n, M_1, X_2^n) && (208)\\
&\leq \sum_{i=1}^{n} \Big(h(Y_i) - h(Y_i|X_{1,i}, S_i, X_{2,i}, M_1)\\
&\qquad - h(S_i|M_1, X_{2,i}) + h(S_i|M_1, Y_i, X_{2,i}) \Big) && (209)\\
&= \sum_{i=1}^{n} \Big(I(X_{1,i}; Y_i|X_{2,i}, S_i, M_1) + I(X_{2,i}, M_1; Y_i) \Big) && (210)\\
&= I(X_1; Y|X_2, S, M_1, Q) + I(X_2, M_1; Y|Q) && (211)\\
&\leq I(X_1; Y|X_2, S, U) + I(X_2, U; Y). && (212)
\end{aligned}$$

Here, (209) 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 and 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; and
- 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 case and to the multiple-input multiple-output (MIMO) setting. This is unlike the *doublely dirty Gaussian MAC setting*, in which additional difficulties arise when extending from the single-input single-output channel to the MIMO setting [31].
- 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 [32], strictly causal state information enables cooperation between the two encoders (e.g., by letting the encoders convey the past state information jointly to the decoder).
- In this paper, we have considered two state-dependent MAC settings: the dirty MAC without degraded message sets and the dirty MAC with degraded message sets. In the latter case, we assumed that the state non-cognitive encoder has access to the message of the state-cognitive encoder. As mentioned in the introduction, the setting in which the state-cognitive encoder knows the message of the non-cognitive encoder was studied in Somekh-Baruch *et al.* [7], where the exact capacity region was established. A general model that unifies all the three settings listed above is the state-dependent MAC with common and private messages. In this model, the state-cognitive and state non-cognitive encoders

each has a private message to transmit, in addition to a common message that is known at both encoders. This model generalizes the Slepian-Wolf model of the MAC with correlated sources [33] to the state-dependent setting. Establishing the capacity region of this model is an interesting open problem.

- 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

$$Y_1 = X_1 + S + \sqrt{\delta} Z_1 \tag{213}$$
$$Y_2 = X_1 + X_2 + S + Z_2 \tag{214}$$

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 [34].

- Another related setting is the state-dependent relay channel with state available noncausally at the relay considered in [35]. It would be interesting to see whether our techniques can lead to any improvement over the bounds there.

ACKNOWLEDGEMENT

The authors would like to thank the associate editor and two anonymous reviewers for their suggestions and comments of the paper, which helped greatly to improve the quality and the presentation of the paper. In particular, the authors would also like to thank one anonymous reviewer for simplifying the proof of (105) and the other anonymous reviewer for suggesting the generalized model of the state-dependent MAC with private and common messages discussed in Section V.

APPENDIX
GAUSSIAN INPUTS MAXIMIZE (86)–(88)

We shall prove that the outer region provided in Proposition 10 is maxmized 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;
- $\mathbb{E}\left[X_1^2\right] = P_1$ 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:

$$\mu_k(u) \triangleq \mathbb{E}[X_k | U = u] \tag{215}$$
$$\sigma_k(u) \triangleq \sqrt{\mathrm{Var}[X_k | U = u]} \tag{216}$$
$$\rho_k \triangleq \sqrt{\mathbb{E}\left[\mu_k^2(U)\right] / P_k} \tag{217}$$
$$\mu_s(u) \triangleq \mathbb{E}[X_1 S | U = u] / \sqrt{Q} \tag{218}$$
$$\rho_s \triangleq \mathbb{E}[\mu_s(U)] / \sqrt{P_1} \tag{219}$$

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

$$\begin{aligned} R_1 &\le I(X_2; Y | X_1, S, U) && (220)\\ &\le \frac{1}{2}\mathbb{E}\left[\log(1 + \sigma_2(U)^2\right] && (221)\\ &\le \frac{1}{2}\log\left(1 + \mathbb{E}\left[\sigma_2(U)^2\right]\right) && (222)\\ &= \frac{1}{2}\log(1 + P_2(1 - \rho_2^2)). && (223)\end{aligned}$$

Here, (222) follows from Jensen's inequality, and (223) follows because

$$\mathbb{E}\left[\sigma_2^2(U)\right] = \mathbb{E}\left[\mathbb{E}\left[X_2^2 | U\right] - \mu_2(U)^2\right] = P_2 - \rho_2^2 P_2. \tag{224}$$

This proves (82).

To prove (83), we proceed as follows:

$$\begin{aligned} R_2 &\le I(X_1; Y | X_2, S, U) + I(X_2; Y | U) && (225)\\ &= I(X_1, X_2, S; Y | U) - I(S; Y | U, X_2). && (226)\end{aligned}$$

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

$$\begin{aligned}&\mathrm{Var}[X_1 + X_2 + S | U = u]\\ &\qquad = \sigma_1^2(u) + \sigma_2^2(u) + Q + 2\sqrt{Q}\mu_s(u) && (227)\end{aligned}$$

where we have used (216) and (218), and that X_1 and X_2 are conditionally independent given U. It thus follows that

$$\begin{aligned}&I(X_1 + X_2 + S; Y | U)\\ &\le \frac{1}{2}\mathbb{E}\left[\log(1 + \sigma_1^2(U) + \sigma_2^2(U) + Q + 2\sqrt{Q}\mu_s(U))\right] && (228)\\ &\le \frac{1}{2}\log\left(1 + \mathbb{E}\left[\sigma_1^2(U) + \sigma_2^2(U) + Q + 2\sqrt{Q}\mu_s(U)\right]\right) && (229)\\ &= \frac{1}{2}\log\left(1 + P_1(1 - \rho_1^2) + P_2(1 - \rho_2^2) + Q + 2\rho_s\sqrt{QP_1}\right). && (230)\end{aligned}$$

Here, in (230) we have used the following identity:

$$\begin{aligned}\mathbb{E}\left[\sigma_k^2(U)\right] &= \mathbb{E}[\mathrm{Var}[X_k | U]] && (231)\\ &= \mathrm{Var}[X_k] - \mathrm{Var}[\mu_k(U)] && (232)\\ &= \mathrm{Var}[X_k] - \mathbb{E}\left[\mu_k(U)^2\right] + \mathbb{E}[X_k]^2 && (233)\\ &= P_k - P_k\sigma_k^2, \quad k \in \{1, 2\} && (234)\end{aligned}$$

where (232) follows from the law of total variance.

We next bound the second term on the RHS of (226). Let

$$\widetilde{X}_1 \triangleq X_1 - \mu_1(U) - \frac{\mu_s(U)S}{\sqrt{Q}}. \tag{235}$$

It follows that

$$\mathbb{E}\left[\widetilde{X}_1 S | U = u\right] = \mathbb{E}[X_1 S | U = u] - \mu_s(u)\sqrt{Q} = 0. \tag{236}$$

Since S is Gaussian distributed, by Lemma 13,

$$\begin{aligned}&I(S; Y | X_2, U) && (237)\\ &= \mathbb{E}_U\left[I(S; (1 + \mu_s(U)/\sqrt{Q})S + \widetilde{X}_1 + Z | U)\right] && (238)\\ &\ge \frac{1}{2}\mathbb{E}\left[\log\left(1 + \frac{(\sqrt{Q} + \mu_s(U))^2}{1 + \sigma_1^2(U) - \mu_s(U)^2}\right)\right]. && (239)\end{aligned}$$

By (216), (235), and (236),

$$\sigma_1^2(u) = \mathbb{E}\Big[X_1^2|U=u\Big] - \mu_1^2(u) \tag{240}$$
$$= \mathbb{E}\Big[\widetilde{X}_1^2|U=u\Big] + \mu_s(u)^2 \geq \mu_s(u)^2. \tag{241}$$

Now, observe that the function

$$\xi(a,b) \triangleq \frac{1}{2}\log\left(1+\frac{(\sqrt{Q}-a)^2}{1+b-a^2}\right) \tag{242}$$

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

$$H_{11} = \frac{\partial^2 \xi}{\partial a^2} \tag{243}$$
$$= \frac{(\sqrt{Q}-a)^2((\sqrt{Q}-a)^2+2+2b-2a^2)}{(1+b-a^2)^2(\sqrt{Q}-a)^2+1+b-a^2)^2} \tag{244}$$
$$\geq 0 \tag{245}$$

and that

$$\mathrm{Det}[\mathsf{H}] = \frac{(\sqrt{Q}-a)^4}{(1+b-a^2)^3(\sqrt{Q}-a)^2+1+b-a^2)^2} \tag{246}$$
$$\geq 0. \tag{247}$$

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,

$$I(S; X_1+S+Z|U) \tag{248}$$
$$\geq \frac{1}{2}\log\left(1+\frac{(\sqrt{Q}+\mathbb{E}[\mu_s(U)])^2}{1+\mathbb{E}\big[\sigma_1^2(U)\big]-\mathbb{E}[\mu_s(U)]^2}\right) \tag{249}$$
$$= \frac{1}{2}\log\left(1+\frac{(\sqrt{Q}+\rho_s\sqrt{P_1})^2}{1+P_1-\rho_1^2P_1-\rho_s^2P_1}\right). \tag{250}$$

Here, in (250) we have used (234). Substituting (230) and (250) into (226) and rearranging the terms, we obtain (83).

The proof of (84) follows steps analogous to those in the proof of (83). More specifically, we obtain from (88) that

$$R_1+R_2 \leq h(Y|X_2,S,U) - h(Y|X_1,X_2,S,U) + h(Y) - h(Y|X_2,U) \tag{251}$$
$$= I(X_1+X_2+S;Y) - I(S;X_1+S+Z|U). \tag{252}$$

The term $I(S;X_1+S+Z|U)$ on the RHS of (252) has been lower-bounded in (250). To upper-bound $I(X_1+X_2+S;Y)$, we bound $\mathbb{E}\big[(X_1+X_2+S)^2\big]$ as

$$\mathbb{E}\Big[(X_1+X_2+S)^2\Big]$$
$$= P_1+P_2+Q+2\mathbb{E}[X_1S]+2\mathbb{E}[X_1X_2] \tag{253}$$
$$= P_1+P_2+Q+2\rho_s\sqrt{P_1Q}+2\mathbb{E}[\mathbb{E}[X_1|U]\,\mathbb{E}[X_2|U]] \tag{254}$$
$$\leq P_1+P_2+Q+2\rho_s\sqrt{P_1Q}+2\rho_1\rho_2\sqrt{P_1P_2}. \tag{255}$$

Here, (254) follows because X_1 and X_2 are conditionally independent given U, and (255) follows because

$$\mathbb{E}[\mathbb{E}[X_1|U]\,\mathbb{E}[X_2|U]] = \mathbb{E}[\mu_1(U)\mu_2(U)] \tag{256}$$
$$\leq \sqrt{\mathbb{E}\big[\mu_1(U)^2\big]\,\mathbb{E}\big[\mu_2(U)^2\big]} \tag{257}$$
$$= \rho_1\rho_2\sqrt{P_1P_2}. \tag{258}$$

It thus follows that

$$I(X_1+X_2+S;Y) \leq \frac{1}{2}\log\Big(1+P_1+P_2+Q+2\rho_s\sqrt{P_1Q}+2\rho_1\rho_2\sqrt{P_1P_2}\Big). \tag{259}$$

Substituting (259) and (250) into (252), we obtain (84).

Finally, observe from (234) and (241) that

$$P_1 - P_1\sigma_1^2 = \mathbb{E}\Big[\sigma_1^2(U)\Big] \geq \mathbb{E}\Big[\mu_s(U)^2\Big] \geq \mathbb{E}[\mu_s(U)]^2 \geq P_1\rho_s^2 \tag{260}$$

which implies the condition (85). This concludes the proof.

REFERENCES

[1] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-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. P. Kotagiri and J. N. Laneman, "Multiaccess channels with state known to some encoders and independent messages," *EURASIP J. Wireless Commun. Netw.*, vol. 2008, Dec. 2008, p. 450680.

[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, South Korea, Jun./Jul. 2009, pp. 2376–2380.

[6] M. F. Pourbabaee, 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, pp. 1566–1570.

[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. Koetter, "Helpers for cleaning dirty papers," in *Proc. 7th Int. ITG Conf. Source Channel Coding (SCC)*, Ulm, Germany, Jan. 2008, pp. 1–5.

[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, pp. 480–484.

[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. 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] S. I. Bross and A. Lapidoth, "The Gaussian state-dependent channel with rate-limited decoder state-information," in *Proc. IEEE Int. Conf. Sci. Elect. Eng. (ICSEE)*, Eilat, Israel, Nov. 2016, pp. 1–5.

[16] 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.

[17] 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.

[18] G. Keshet, Y. Steinberg, and N. Merhav, "Channel coding in the presence of side information," *Found. Trends Commun. Inf. Theory*, vol. 4, no. 6, pp. 445–586, Jun. 2008.

[19] 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.

[20] M. Talagrand, “Transportation cost for Gaussian and other product measures,” *Geometric Funct. Anal.*, vol. 6, no. 3, pp. 587–600, May 1996.
[21] S. Ihara, “On the capacity of channels with additive non-Gaussian noise,” *Inf. Control*, vol. 37, no. 1, pp. 34–39, Apr. 1978.
[22] 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.
[23] 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.
[24] M. Effros and M. Langberg, “On a reductionist view of information theory,” in *Proc. 7th Joint Workshop Coding Commun. (JWCC)*, Barcelona, Spain, Nov. 2014, p. 20.
[25] M. H. M. Costa, “A new entropy power inequality,” *IEEE Trans. Inf. Theory*, vol. IT-31, no. 6, pp. 751–760, Nov. 1985.
[26] 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.
[27] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[28] C. Villani, *Topics in Optimal Transportation*. Providence, RI, USA: Amer. Math. Soc., 2003.
[29] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.
[30] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[31] A. Khina, Y. Kochman, and U. Erez, “The dirty MIMO multiple-access channel,” *IEEE Trans. Inf. Theory*, vol. 63, no. 9, pp. 6031–6040, Sep. 2017.
[32] 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.
[33] D. Slepian and J. K. Wolf, “A coding theorem for multiple access channels with correlated sources,” *Bell Syst. Tech. J.*, vol. 52, no. 7, pp. 1037–1076, 1973.
[34] 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.
[35] 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.

Wei Yang (S’09–M’15) received the B.E. degree in communication engineering and M.E. degree in communication and information systems from the Beijing University of Posts and Telecommunications, Beijing, China, in 2008 and 2011, and the Ph.D. degree in Electrical Engineering from Chalmers University of Technology, Gothenburg, Sweden, in 2015. In the summers of 2012 and 2014, he was a visiting student at the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA. From 2015 to 2017, he was a postdoctoral research associate at Princeton University, Princeton, NJ. In September 2017, he joined Qualcomm Research, San Diego, CA, where he is now a senior engineer.

Dr. Yang is the recipient of a Student Paper Award at the 2012 IEEE International Symposium on Information Theory (ISIT), Cambridge, MA, and the 2013 IEEE Sweden VT-COM-IT joint chapter best student conference paper award. His research interests are in the areas of information theory, communication theory, and wireless communication systems.

Yingbin Liang (S’01–M’05–SM’16) received the Ph.D. degree in Electrical Engineering from the University of Illinois at Urbana-Champaign in 2005. In 2005–2007, she was working as a postdoctoral research associate at Princeton University. In 2008–2009, she was an assistant professor at University of Hawaii. In 2010–2017, she was an assistant and then an associate professor at Syracuse University. Since August 2017, she has been an associate professor at the Department of Electrical and Computer Engineering at the Ohio State University. Dr. Liang’s research interests include machine learning, statistical signal processing, optimization, information theory, and wireless communication and networks.

Dr. Liang was a Vodafone Fellow at the University of Illinois at Urbana-Champaign during 2003–2005, and received the Vodafone-U.S. Foundation Fellows Initiative Research Merit Award in 2005. She also received the M. E. Van Valkenburg Graduate Research Award from the ECE department, University of Illinois at Urbana-Champaign, in 2005. In 2009, she received the National Science Foundation CAREER Award, and the State of Hawaii Governor Innovation Award. In 2014, she received EURASIP Best Paper Award for the *EURASIP Journal on Wireless Communications and Networking*. She served as an Associate Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY during 2013–2015.

Shlomo Shamai (Shitz) (F’94) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, in 1975, 1981 and 1986 respectively. During 1975–1985 he was with the Communications Research Labs, in the capacity of a Senior Research Engineer. Since 1986 he is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, where he is now a Technion Distinguished Professor, and holds the William Fondiller Chair of Telecommunications. His research interests encompasses a wide spectrum of topics in information theory and statistical communications.

Dr. Shamai (Shitz) is an IEEE Fellow, an URSI Fellow, a member of the Israeli Academy of Sciences and Humanities and a foreign member of the US National Academy of Engineering. He is the recipient of the 2011 Claude E. Shannon Award, the 2014 Rothschild Prize in Mathematics/Computer Sciences and Engineering and the 2017 IEEE Richard W. Hamming Medal. He has been awarded the 1999 van der Pol Gold Medal of the Union Radio Scientifique Internationale (URSI), and is a co-recipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003, and the 2004 joint IT/COM societies paper award, the 2007 IEEE Information Theory Society Paper Award, the 2009 and 2015 European Commission FP7, Network of Excellence in Wireless COMmunications (NEWCOM++, NEWCOM#) Best Paper Awards, the 2010 Thomson Reuters Award for International Excellence in Scientific Research, the 2014 EURASIP Best Paper Award (for the *EURASIP Journal on Wireless Communications and Networking*), and the 2015 IEEE Communications Society Best Tutorial Paper Award. He is also the recipient of 1985 Alon Grant for distinguished young scientists and the 2000 Technion Henry Taub Prize for Excellence in Research. He has served as Associate Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY, and has also served twice on the Board of Governors of the Information Theory Society. He has also served on the Executive Editorial Board of the IEEE TRANSACTIONS ON INFORMATION THEORY and on the IEEE Information Theory Society Nominations and Appointments Committee.

H. Vincent Poor (S’72–M’77–SM’82–F’87) received the Ph.D. degree in electrical engineering and computer science from Princeton University in 1977. From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990 he has been on the faculty at Princeton, where he is the Michael Henry Strater University Professor of Electrical Engineering. During 2006 to 2016, he served as Dean of Princeton’s School of Engineering and Applied Science. He has also held visiting appointments at several other institutions, most recently at Berkeley and Cambridge. His research interests are in the areas of information theory and signal processing, and their applications in wireless networks, energy systems and related fields. Among his publications in these areas is the recent book *Information Theoretic Security and Privacy of Information Systems* (Cambridge University Press, 2017).

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, and is a foreign member of the Chinese Academy of Sciences, the Royal Society, and other national and international academies. In 1990, he served as President of the IEEE Information Theory Society, in 2004–07 as the Editor-in-Chief of these TRANSACTIONS, and in 2009 as General Co-chair of the IEEE International Symposium on Information Theory, held in Seoul, South Korea. Recent recognition of his work includes the 2017 IEEE Alexander Graham Bell Medal, Honorary Professorships from Peking University and Tsinghua University, both conferred in 2017, and a D.Sc. *honoris causa* from Syracuse University awarded in 2017.