Interference Channels With Common Information

Jinhua Jiang, *Student Member, IEEE*, Yan Xin, *Member, IEEE*, and Hari Krishna Garg, *Senior Member, IEEE*

***Abstract*—In this paper, the interference channel with common information (ICC), in which two senders need deliver not only private messages but also certain common messages to their *corresponding* receivers, is investigated. An achievable rate region for such a channel is obtained by applying a superposition coding scheme that consists of successive encoding and simultaneous decoding. It is shown that the derived achievable rate region includes or extends several existing results for the interference channels with or without common information. The rate region is then specialized to a class of ICCs in which one sender has no private information to transmit, and a class of deterministic interference channels with common information (DICCs). In particular, the derived rate region is found to be the capacity region for this class of DICCs. Last, the achievable rate region derived for the discrete memoryless ICC is extended to the Gaussian case, in which a numerical example is provided to illustrate the improvement of our rate region over an existing result.**

***Index Terms*—Capacity region, common information, interference channel, multiple-access channel (MAC), simultaneous decoding, superposition coding.**

I. Introduction

The interference channel (IC) is one of the fundamental building blocks in communication networks, in which the transmissions between each sender and its corresponding receiver (each sender–receiver pair) take place simultaneously and interfere with each other. The information-theoretic study of such a channel was initiated by Shannon [1], and has been continued by many others [2]–[16]. So far, the capacity region of the general IC remains unknown except for some special cases, such as the IC with strong interference (SIC) [3], [6], [9], [10], [12], a class of discrete additive degraded ICs [8], and a class of deterministic ICs [11]. However, various achievable rate regions serving as inner bounds on the capacity region have been derived for the general IC [5], [7], [9], [15].

Notably, Carleial [7] obtained an achievable rate region for the discrete memoryless IC by employing a limited form of the superposition coding scheme [17], *successive* encoding and decoding. Subsequently, Han and Kobayashi [9] established the best achievable rate region known to date by applying the superposition coding scheme comprising of *simultaneous* encoding and decoding. Indeed, the improvement of the Han–Kobayashi (HK) region [9] over the Carleial region [7] is primarily due to the use of the simultaneous decoding. This has been validated in [15] and [16], in which Chong *et al.* obtained a so called Chong–Motani–Garg (CMG) rate region identical with the HK region but with a much simplified description, by using a hybrid of the successive encoding and simultaneous decoding. Moreover, Carleial [7] introduced the notion of the partial cross-observability of each sender's private information, which means that each receiver is able to decode part of the private information sent from its nonpairing sender. The derivation of the HK region and the CMG region followed this notion but Chong *et al.* have made the important observation that the decoding errors of the crossly observed information can be excluded in computing the probability of error [15]. With an introduction of the partial cross-observability, the IC can be viewed as a compound channel consisting of two associated multiple access channels (MACs) (strictly speaking, MAC-alike channels), and thus its achievable rate region can be obtained by exploiting existing techniques used for MACs. However, the converse for either the HK region or the CMG region has not been established. Very recently, a notable variant of the IC, namely the IC with degraded message sets (IC-DMS) [18]–[22], has attracted considerable research attention due to its applicability to model certain realistic communication scenarios in cognitive radio networks or wireless sensor networks. From an information-theoretic viewpoint, the IC-DMS is fundamentally different from the IC since the capacity regions of IC and IC-DMS, if any, do not necessarily imply each other.

Most of the prior work on the ICs assumes the statistical independence of the source messages [2]–[16]. However, the assumption becomes invalid in an IC where the senders need transmit not only the private information but also certain common information to their corresponding receivers. Such a scenario is generally modeled as the IC with common information (ICC) [23]–[25]. The ICC was first studied by Tan in his original work [23], where inner and outer bounds on the capacity region have been derived. In particular, when no common information is present, the inner bound (the achievable rate region) in [23] reduces to the Carleial region in [3]. More recently, Maric *et al.* [24] derived the capacity region for a special case of the ICC, the strong interference channel with common information (SICC), and showed that the derived capacity result includes the capacity region of the strong interference channel (with no common information) [12] as a special case. Parallel to the case of the IC, the study of the ICC is closely related to the previous work on the MAC with common information (MACC) that has been thoroughly studied by Slepian and Wolf [26] and Willems [27]. As an example, an achievable rate region for the SICC is an intersection of the rate regions for its two corresponding MACCs, and the

Manuscript received October 3, 2006; revised May 1, 2007. The material in this paper is presented in part at the 40th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, October/November 2006 and the IEEE International Conference on Acoustics, Speech and Signal Processing, Honolulu, HI, April 2007.

The authors are with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore, 117576, Republic of Singapore (e-mail: Jinhua.Jiang@nus.edu.sg; elexy@nus.edu.sg; and eleghk@nus.edu.sg).

Communicated by A. J. Goldsmith, Associate Editor for Communications.

Color version of Figure 6 in this paper is available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2007.911257

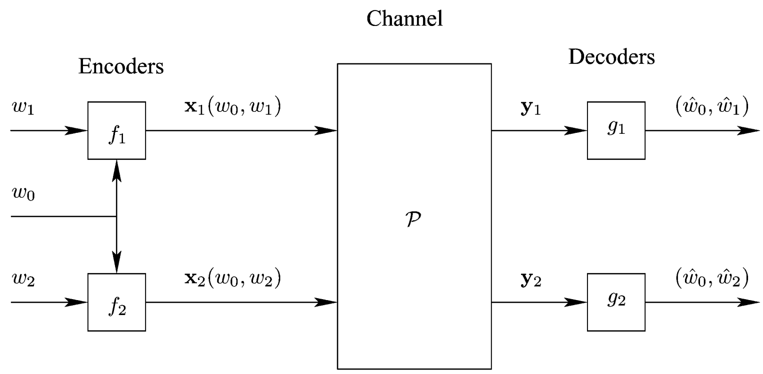

Fig. 1. The interference channel with common information.

capacity region of the SICC is the union of all such achievable rate regions.

In this paper, we begin with studying the *general* two-user ICC problem. We propose an encoding scheme that extends the idea of the Carleial's successive encoding for the ICC. With this encoding scheme, we allow the senders' common information to be conveyed through the channel in a cooperative manner. Exploiting the proposed encoding scheme along with the simultaneous decoding scheme [9], [15], we derive a new achievable rate region for the discrete memoryless ICC. We show that the derived achievable rate region contains the one in [23] as a *proper* subregion under some specific setting, and reduces to the CMG region [15] as well as the capacity region of the SICC [24] in their respective channel settings. We further investigate a class of deterministic interference channels with common information (DICCs), which can be viewed as a generalization of the class of deterministic ICs (DICs) in [11]. We show that under certain assumptions, our achievable rate region is the capacity region for this class of the DICCs.

In the review process of the present paper, we learned of the independent work by Cao *et al.* [28] from one referee. The achievable rate region in [28] is essentially the same as ours, even though, compared with the one presented in [28], the description of our achievable rate region is more compact in view of the number of constraints involved.

The rest of the paper is organized as follows. In Section II, we introduce the channel models. In Section III, we present the achievable rate region for the general discrete memoryless ICC in both implicit and explicit forms. In Section IV, we discuss the relations between our achievable rate region and several existing results in [15], [23], [24], [29]. In Section V, we investigate two special cases of the ICC. In Section VI, we extend our achievable rate region for the discrete memoryless ICC to the Gaussian case. The paper is concluded in Section VII.

The notations used in this paper are as follows. Random variables and their realizations are denoted by upper case letters and lower case letters respectively, e.g., X and x. Bold fonts are used to indicate vectors, e.g., $\mathbf{X}$ and $\mathbf{x}$.

II. Channel Models and Preliminaries

In this section, we present the channel models of the ICC, including the general ICC and a modified ICC. The modified ICC serves to reveal the information flow through its associated ICC, and facilitates the derivation of the achievable rate region for the associated ICC.

A. *Discrete Memoryless Interference Channel With Common Information*

A discrete memoryless IC is usually defined by a quintuple $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{P}, \mathcal{Y}_1, \mathcal{Y}_2)$, where $\mathcal{X}_t$ and $\mathcal{Y}_t, t = 1, 2$, denote the finite channel input and output alphabets respectively, and $\mathcal{P}$ denotes the collection of the conditional probabilities $p(y_1, y_2|x_1, x_2)$ on $(y_1, y_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2$ given $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$. The channel is memoryless in the sense that for n channel uses, we have

$$p(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x}_1, \mathbf{x}_2) = \prod_{i=1}^{n} p(y_{1i}, y_{2i}|x_{1i}, x_{2i})$$

where $\mathbf{x}_t = (x_{t1}, \ldots, x_{tn}) \in \mathcal{X}_t^n$ and $\mathbf{y}_t = (y_{t1}, \ldots, y_{tn}) \in \mathcal{Y}_t^n$ for $t = 1, 2$. The marginal distributions of y_1 and y_2 are given by

$$p_1(y_1|x_1, x_2) = \sum_{y_2 \in \mathcal{Y}_2} p(y_1, y_2|x_1, x_2)$$
$$p_2(y_2|x_1, x_2) = \sum_{y_1 \in \mathcal{Y}_1} p(y_1, y_2|x_1, x_2).$$

Building upon an IC, we depict an ICC in Fig. 1. Sender $t, t = 1, 2$, is to send a private message $w_t \in \mathcal{M}_t = \{1, 2, \ldots, M_t\}$ together with a common message $w_0 \in \mathcal{M}_0 = \{1, 2, \ldots, M_0\}$ to its pairing receiver. All the three messages are assumed to be independently and uniformly generated over their respective ranges.

Let $\boldsymbol{C}$ denote the discrete memoryless ICC defined above. An (M_0, M_1, M_2, n, P_e) code exists for the channel $\boldsymbol{C}$, if and only if there exist two encoding functions

$$f_1 : \mathcal{M}_0 \times \mathcal{M}_1 \to \mathcal{X}_1^n, \quad f_2 : \mathcal{M}_0 \times \mathcal{M}_2 \to \mathcal{X}_2^n$$

and two decoding functions

$$g_1 : \mathcal{Y}_1^n \to \mathcal{M}_0 \times \mathcal{M}_1, \quad g_2 : \mathcal{Y}_2^n \to \mathcal{M}_0 \times \mathcal{M}_2$$

such that $\max\left\{P_{e,1}^{(n)}, P_{e,2}^{(n)}\right\} \leq P_e$, where $P_{e,t}^{(n)}, t = 1, 2$, denotes the average decoding error probability of decoder t, and is computed by one of the following expressions:

$$P_{e,1}^{(n)} = \frac{1}{M_{\text{Prod}}} \sum_{w_0 w_1 w_2} p((\hat{w}_0, \hat{w}_1) \neq (w_0, w_1)|(w_0, w_1, w_2))$$
$$P_{e,2}^{(n)} = \frac{1}{M_{\text{Prod}}} \sum_{w_0 w_1 w_2} p((\hat{w}_0, \hat{w}_2) \neq (w_0, w_2)|(w_0, w_1, w_2))$$

where $M_{\text{Prod}} := M_0 M_1 M_2$.

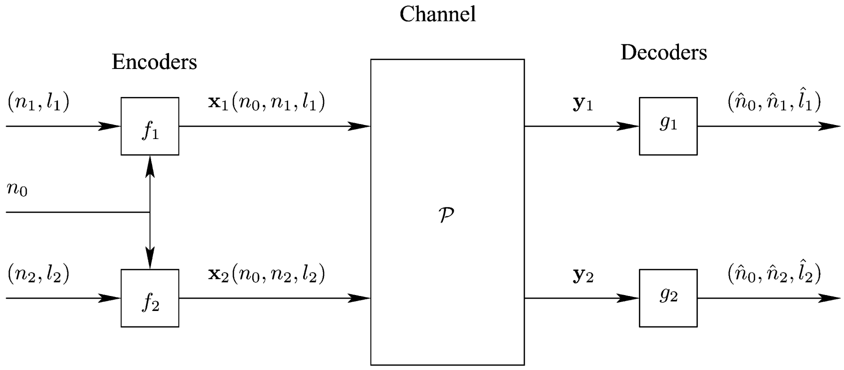

Fig. 2. The modified interference channel with common information.

A nonnegative rate triple (R_0, R_1, R_2) is achievable for the channel $\boldsymbol{C}$ if for any given $0 < P_e < 1$, and for any sufficiently large n, there exists a $\left(2^{nR_0}, 2^{nR_1}, 2^{nR_2}, n, P_e\right)$ code.

The capacity region for the channel $\boldsymbol{C}$ is defined as the closure of the set of all the achievable rate triples, while an achievable rate region for the channel $\boldsymbol{C}$ is a subset of the capacity region.

B. Modified Discrete Memoryless Interference Channel With Common Information

The modified ICC, as depicted in Fig. 2, inherits the same channel characteristics from its associated ICC, but it has five streams of messages instead of three in the associated ICC. The five streams of messages n_0, n_1, l_1, n_2, and l_2 are assumed to be independently and uniformly generated over the finite sets $\mathcal{N}_0 = \{1, \ldots, N_0\}$, $\mathcal{N}_1 = \{1, \ldots, N_1\}$, $\mathcal{L}_1 = \{1, \ldots, L_1\}$, $\mathcal{N}_2 = \{1, \ldots, N_2\}$, and $\mathcal{L}_2 = \{1, \ldots, L_2\}$, respectively. Denote the modified ICC by $\boldsymbol{C}_m$.

An $(N_0, N_1, L_1, N_2, L_2, n, P_e)$ code exists for the channel $\boldsymbol{C}_m$ if and only if there exist two encoding functions

$$f_1 : \mathcal{N}_0 \times \mathcal{N}_1 \times \mathcal{L}_1 \to \mathcal{X}_1^n, \quad f_2 : \mathcal{N}_0 \times \mathcal{N}_2 \times \mathcal{L}_2 \to \mathcal{X}_2^n$$

and two decoding functions

$$g_1 : \mathcal{Y}_1^n \to \mathcal{N}_0 \times \mathcal{N}_1 \times \mathcal{L}_1, \quad g_2 : \mathcal{Y}_2^n \to \mathcal{N}_0 \times \mathcal{N}_2 \times \mathcal{L}_2$$

such that $\max\left\{P_{e,1}^{(n)}, P_{e,2}^{(n)}\right\} \le P_e$, where the average probabilities of decoding error denoted by $P_{e,1}^{(n)}$ and $P_{e,2}^{(n)}$ are computed as

$$P_{e,1}^{(n)} = \frac{1}{N_{\text{Prod}}} \sum_{n_0 n_1 l_1 n_2 l_2} p((\hat{n}_0, \hat{n}_1, \hat{l}_1) \neq (n_0, n_1, l_1) | (n_0, n_1, l_1, n_2, l_2))$$

$$P_{e,2}^{(n)} = \frac{1}{N_{\text{Prod}}} \sum_{n_0 n_1 l_1 n_2 l_2} p((\hat{n}_0, \hat{n}_2, \hat{l}_2) \neq (n_0, n_2, l_2) | (n_0, n_1, l_1, n_2, l_2))$$

where $N_{\text{Prod}} := N_0 N_1 L_1 N_2 L_2$.

A nonnegative rate quintuple $(R_0, R_{12}, R_{11}, R_{21}, R_{22})$ is achievable for the channel $\boldsymbol{C}_m$ if for any given $0 < P_e < 1$ and any sufficiently large n, there exists a $\left(2^{nR_0}, 2^{nR_{12}}, 2^{nR_{11}}, 2^{nR_{21}}, 2^{nR_{22}}, n, P_e\right)$ code for the channel $\boldsymbol{C}_m$.

Remark 1: It should be noted that compared with Fig. 2 in [9], our modified channel depicted in Fig. 2 does not include the index $\hat{n}_2$ (or $\hat{n}_1$) in the decoded message vector at decoder 1 (or decoder 2). This is due to the observation made in [15] that, although receiver 1 (or receiver 2) attempts to decode the crossly observable private message n_2 (or n_1), it is not necessary to include decoding errors of such information in calculating probability of error at the respective receiver. This is also the reason why we term the two associated channels of an ICC as MACC-alike channels instead of MACCs.

Lemma 1: If $(R_0, R_{12}, R_{11}, R_{21}, R_{22})$ is achievable for the channel $\boldsymbol{C}_m$, then $(R_0, R_{12} + R_{11}, R_{21} + R_{22})$ is achievable for the associated ICC.

Remark 2: With the aid of Lemma 1, an achievable rate region for the modified ICC can be easily extended to one for the associated ICC.

III. Discrete Memoryless Interference Channel With Common Information

In this section, we derive a new achievable rate region for the discrete memoryless ICC introduced in Section II. The derived rate region is presented in both implicit and explicit forms.

A. An Achievable Rate Region for the Discrete Memoryless ICC

We first introduce three auxiliary random variables U_0, U_1, and U_2 that are defined over arbitrary finite sets $\mathcal{U}_0$, $\mathcal{U}_1$, and $\mathcal{U}_2$, respectively. Denote by $\mathcal{P}^*$ the set of all joint probability distributions $p(\cdot)$ that factor as

$$p(u_0, u_1, u_2, x_1, x_2, y_1, y_2) = p(u_0)p(u_1|u_0)p(u_2|u_0) \cdot p(x_1|u_1, u_0)p(x_2|u_2, u_0)p(y_1, y_2|x_1, x_2). \quad (1)$$

Let $\mathcal{R}_m(p)$ denote the set of all nonnegative rate quintuples $(R_0, R_{12}, R_{11}, R_{21}, R_{22})$ such that

$$R_{11} \le I(X_1; Y_1|U_0 U_1 U_2) \quad (2)$$

$$R_{12} + R_{11} \le I(X_1; Y_1|U_0 U_2) \quad (3)$$

$$R_{11} + R_{21} \le I(X_1 U_2; Y_1|U_0 U_1) \quad (4)$$

$$R_{12} + R_{11} + R_{21} \le I(X_1 U_2; Y_1|U_0) \quad (5)$$

$$R_0 + R_{12} + R_{11} + R_{21} \le I(U_0 X_1 U_2; Y_1) \quad (6)$$

$$R_{22} \le I(X_2; Y_2|U_0 U_2 U_1) \quad (7)$$

$$R_{21} + R_{22} \le I(X_2; Y_2|U_0 U_1) \quad (8)$$

$$R_{22} + R_{12} \le I(X_2 U_1; Y_2|U_0 U_2) \quad (9)$$

$$R_{21} + R_{22} + R_{12} \le I(X_2U_1; Y_2|U_0) \tag{10}$$
$$R_0 + R_{21} + R_{22} + R_{12} \le I(U_0X_2U_1; Y_2) \tag{11}$$

for some fixed joint probability distribution $p(\cdot) \in \mathcal{P}^*$. Note that each of the mutual information terms is computed with respect to the given fixed joint distribution.

Lemma 2: Any element $(R_0, R_{12}, R_{11}, R_{21}, R_{22}) \in \mathcal{R}_m(p)$ is achievable for the modified ICC $\boldsymbol{C}_m$ for a fixed joint probability distribution $p(\cdot) \in \mathcal{P}^*$.

Remark 3: The lengthy proof is relegated to Appendix A. Lemma 2 lays a foundation for us to establish an achievable rate region for the general ICC. One can interpret this achievable rate region as an intersection between the achievable rate regions of the two associated MACC-alike channels. Specifically, inequalities (2)–(6) depict an achievable rate region for one MACC-alike channel, and inequalities (7)–(11) depict one for the other.

Theorem 1: The rate region $\mathcal{R}_m$ is achievable for the channel $\boldsymbol{C}_m$ with $\mathcal{R}_m := \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}_m(p)$.

Remark 4: Theorem 1 is a direct extension of Lemma 2. The proof is straightforward and thus omitted. Note that the rate region $\mathcal{R}_m$ is convex, and therefore no convex hull operation or time sharing is necessary. The proof of the convexity is given in Appendix B.

Let us fix a joint distribution $p(\cdot) \in \mathcal{P}^*$, and denote by $\mathcal{R}_{\text{impl}}(p)$ the set of all the nonnegative rate triples (R_0, R_1, R_2) such that $R_1 = R_{12} + R_{11}$ and $R_2 = R_{21} + R_{22}$ for some $(R_0, R_{12}, R_{11}, R_{21}, R_{22}) \in \mathcal{R}_m(p)$.

Theorem 2: $\mathcal{R}_{\text{impl}}$ is an achievable rate region for the channel $\boldsymbol{C}$ with $\mathcal{R}_{\text{impl}} := \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}_{\text{impl}}(p)$.

Proof: It suffices to prove that $\mathcal{R}_{\text{impl}}(p)$ is an achievable rate region for $\boldsymbol{C}$ for any fixed joint probability distribution $p(\cdot) \in \mathcal{P}^*$, while the achievability of any rate triple $(R_0, R_1, R_2) \in \mathcal{R}_{\text{impl}}(p)$ follows immediately from Lemma 1 and Lemma 2. □

Remark 5: The main idea, as mentioned before, is that we allow the common information (of rate R_0) to be cooperatively transmitted by the two senders, on top of which we treat the private information at each sender as two parts. One part (of rate R_{12} or R_{21}) of the private information at each sender is crossly observable to the nonpairing receiver, but not the other part (of rate R_{11} or R_{22}). However, for each receiver, the crossly observed information is not required to be decoded correctly [15]. Details can be found in the proof of Lemma 2 in Appendix A.

Remark 6: One can observe that the rate of the common information, R_0, is bounded by only one inequality at each decoder. This is similar to the case of the MACC [26], [27], where the rate of the common information is bounded by only one inequality. This is due to the perfect cooperation of the two senders in transmitting the common information, and the simultaneous decoding. Details are illustrated in the proof of Lemma 2.

Remark 7: The region $\mathcal{R}_{\text{impl}}$ is also convex, which can be proven by following the same procedure as used in the proof of the convexity of $\mathcal{R}_m$ in Appendix B.

B. Explicit Description of the Achievable Rate Region

In order to reveal the geometric shape of the region $\mathcal{R}_{\text{impl}}$ depicted in Theorem 2, we derive an explicit description of the region by applying Fourier-Motzkin elimination [15], [30], [29].

Let $\mathcal{R}(p)$ denote the set of all nonnegative rate triples (R_0, R_1, R_2) such that

$$R_1 \le I(X_1; Y_1|U_0U_2) \tag{12}$$
$$R_2 \le I(X_2; Y_2|U_0U_1) \tag{13}$$
$$R_0 + R_1 \le I(U_0X_1U_2; Y_1) \tag{14}$$
$$R_0 + R_2 \le I(U_0X_2U_1; Y_2) \tag{15}$$
$$R_1 + R_2 \le I(X_1U_2; Y_1|U_0U_1) + I(X_2U_1; Y_2|U_0U_2) \tag{16}$$
$$R_1 + R_2 \le I(X_1; Y_1|U_0U_1U_2) + I(X_2U_1; Y_2|U_0) \tag{17}$$
$$R_0 + R_1 + R_2 \le I(X_1; Y_1|U_0U_1U_2) + I(U_0X_2U_1; Y_2) \tag{18}$$
$$R_1 + R_2 \le I(X_2; Y_2|U_0U_1U_2) + I(X_1U_2; Y_1|U_0) \tag{19}$$
$$R_0 + R_1 + R_2 \le I(X_2; Y_2|U_0U_1U_2) + I(U_0X_1U_2; Y_1) \tag{20}$$
$$2R_1 + R_2 \le I(X_1; Y_1|U_0U_1U_2) + I(X_1U_2; Y_1|U_0) + I(X_2U_1; Y_2|U_0U_2) \tag{21}$$
$$R_0 + 2R_1 + R_2 \le I(X_1; Y_1|U_0U_1U_2) + I(U_0X_1U_2; Y_1) + I(X_2U_1; Y_2|U_0U_2) \tag{22}$$
$$R_1 + 2R_2 \le I(X_2; Y_2|U_0U_1U_2) + I(X_2U_1; Y_2|U_0) + I(X_1U_2; Y_1|U_0U_1) \tag{23}$$
$$R_0 + R_1 + 2R_2 \le I(X_2; Y_2|U_0U_1U_2) + I(U_0X_2U_1; Y_2) + I(X_1U_2; Y_1|U_0U_1) \tag{24}$$

for some fixed joint distribution $p(\cdot) \in \mathcal{P}^*$, and define $\mathcal{R} := \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}(p)$.

Corollary 1: The rate region $\mathcal{R}$ is achievable for the channel $\boldsymbol{C}$, and $\mathcal{R} = \mathcal{R}_{\text{impl}}$.

Remark 8: The proof of this corollary is given in Appendix C. In fact, the explicit rate region obtained by applying Fourier–Motzkin elimination on (2)–(11) contains two extra constraints

$$R_1 \le I(X_1; Y_1|U_0U_1U_2) + I(X_2U_1; Y_2|U_0U_2)$$
$$R_2 \le I(X_2; Y_2|U_0U_1U_2) + I(X_1U_2; Y_1|U_0U_1).$$

However, these two constraints are redundant and thus are excluded. This is shown in the second part of Appendix C by applying the technique introduced in [16].

The close tie between the explicit CMG region and the capacity region of a class of DICs in [11] was pointed out in [30]. Similarly, we will disclose that the explicit region for the ICC is also closely related to the capacity region of a class of DICCs investigated in Section V-B.

IV. RELATIONS BETWEEN $\mathcal{R}_{\text{impl}}$ AND SOME EXISTING RESULTS

In this section, we discuss the relations between the achievable rate region derived in the preceding section and several previously known results [23], [24], [15].

A. The Achievable Rate Region for ICC by Tan [23]

We show that the achievable rate region $\mathcal{R}_{\text{impl}}$ includes the one given in [23, Th. 1] as a subregion. Note that a similar result is presented in [28, Corollary 1].

Let $\mathcal{P}^*_{\text{Tan}}$ denote the set of all the joint distributions $p(\cdot)$ that factors as

$$p(u_0,u_1,u_2,x_1,x_2,y_1,y_2) = p(u_0)p(u_1|u_0)p(u_2|u_0) \cdot p(x_1|u_1)p(x_2|u_2)p(y_1,y_2|x_1,x_2).$$

Let $\mathcal{R}^i_{\text{Tan}}(p)$, $i=1,2,3,4$, denote the set of all nonnegative rate triples (R_0,R_1,R_2) satisfying

$$\begin{aligned} R_1 &\le I(X_1;Y_1|U_1U_2)+s_i \\ R_2 &\le I(X_2;Y_2|U_1U_2)+t_i \\ R_0+R_1+\frac{t_i}{I(X_2;Y_2|U_1U_2)+t_i}R_2 &\le I(U_2X_1;Y_1) \\ R_0+R_2+\frac{s_i}{I(X_1;Y_1|U_1U_2)+s_i}R_1 &\le I(U_1X_2;Y_2) \end{aligned}$$

where s_i and t_i are computed as

$$\begin{aligned} s_1 &= \min\{I(U_1;Y_1|U_0), I(U_1;Y_2|U_0)\} \\ t_1 &= \min\{I(U_2;Y_1|U_0U_1), I(U_2;Y_2|U_0U_1)\} \\ s_2 &= \min\{I(U_1;Y_1|U_0U_2), I(U_1;Y_2|U_0U_2)\} \\ t_2 &= \min\{I(U_2;Y_1|U_0), I(U_2;Y_2|U_0)\} \\ s_3 &= \min\{I(U_1;Y_1|U_0), I(U_1;Y_2|U_0U_2)\} \\ t_3 &= \min\{I(U_2;Y_1|U_0U_1), I(U_2;Y_2|U_0)\} \\ s_4 &= \min\{I(U_1;Y_1|U_0U_2), I(U_1;Y_2|U_0)\} \\ t_4 &= \min\{I(U_2;Y_1|U_0), I(U_2;Y_2|U_0U_1)\} \end{aligned}$$

for a joint distribution $p(\cdot)\in\mathcal{P}^*_{\text{Tan}}$.

Denote the closed convex hull operation by $\text{co}(\cdot)$, and define

$$\mathcal{R}_{\text{Tan}}(p) := \text{co}\left(\bigcup_{i=1}^{4}\mathcal{R}^i_{\text{Tan}}(p)\right).$$

In the following, we restate the achievable result obtained by Tan [23, Th. 1], and further show that our achievable rate region includes this result as a subregion.

Corollary 2 ([23, Th. 1]): Any rate triple $(R_0,R_1,R_2)\in\mathcal{R}_{\text{Tan}} := \bigcup_{p(\cdot)\in\mathcal{P}^*_{\text{Tan}}}\mathcal{R}_{\text{Tan}}(p)$ is achievable for the ICC, i.e., $\mathcal{R}_{\text{Tan}}\subseteq\mathcal{R}_{\text{impl}}$.

Proof: It suffices to show that each $\mathcal{R}^i_{\text{Tan}}(p)$, $i=1,2,3,4$, is achievable for any joint distribution $p(\cdot)\in P^*_{\text{Tan}}$. Let $\mathcal{R}^i_{\text{sub}}(p)$ be the set of all rate triples (R_0,R_1,R_2) such that $R_1=R_{12}+R_{11}$ and $R_2=R_{21}+R_{22}$ with nonnegative rate quadruples $(R_{12},R_{11},R_{21},R_{22})$ satisfying

$$\begin{aligned} R_{11} &\le I(X_1;Y_1|U_0U_1U_2) \\ R_{22} &\le I(X_2;Y_2|U_0U_1U_2) \\ R_{12} &\le s_i \\ R_{21} &\le t_i \\ R_0+R_{12}+R_{11}+R_{21} &\le I(U_0X_1U_2;Y_1) \\ R_0+R_{21}+R_{22}+R_{12} &\le I(U_0X_2U_1;Y_2) \end{aligned}$$

for a joint distribution $p(\cdot)\in P^*$. It is easy to check that for each $i\in\{1,2,3,4\}$, the rate region $\mathcal{R}^i_{\text{sub}}(p)$ is a subset of our achievable rate region $\mathcal{R}_{\text{impl}}(p)$. Note that $\mathcal{P}^*_{\text{Tan}}\subseteq\mathcal{P}^*$. For a distribution $p(\cdot)\in P^*_{\text{Tan}}$, the rate region $\mathcal{R}^i_{\text{sub}}(p)$ reduces to the region with $(R_{12},R_{11},R_{21},R_{22})$ satisfying

$$\begin{aligned} R_{11} &\le I(X_1;Y_1|U_1U_2) \\ R_{22} &\le I(X_2;Y_2|U_1U_2) \\ R_{12} &\le s_i \\ R_{21} &\le t_i \\ R_0+R_{12}+R_{11}+R_{21} &\le I(X_1U_2;Y_1) \\ R_0+R_{21}+R_{22}+R_{12} &\le I(X_2U_1;Y_2). \end{aligned}$$

This is due to the fact that $p(\cdot)\in P^*_{\text{Tan}}$ induces a Markov chain $U_0\to(U_1,U_2)\to(X_1,X_2)\to(Y_1,Y_2)$. It is now clear that $\mathcal{R}^i_{\text{sub}}(p)=\mathcal{R}^i_{\text{Tan}}(p)$ for any joint distribution $p(\cdot)\in P^*_{\text{Tan}}$. Therefore, the rate region $\mathcal{R}_{\text{Tan}}$ is achievable, and $\mathcal{R}_{\text{Tan}}\subseteq\mathcal{R}_{\text{impl}}$. □

It should be noted that the corollary does not indicate that the inclusion, $\mathcal{R}_{\text{Tan}}\subseteq\mathcal{R}_{\text{impl}}$, is strict. Whether this inclusion is strict deserves further investigation. However under some specific setting, the region $\mathcal{R}_{\text{impl}}$ strictly contains $\mathcal{R}_{\text{Tan}}$, which can be justified as follows. In the case of no common information, $\mathcal{R}_{\text{Tan}}(p)$ reduces to $\mathcal{R}_0(Z)$ in [9, Corollary 3.1], while $\mathcal{R}_{\text{impl}}(p)$ reduces to the CMG region (or the HK region). When the channel is Gaussian and the time-sharing variable is fixed as a constant, the HK region demonstrates strict inclusion over $\mathcal{R}_0(Z)$ in [9]. We will show that under this setting, $\mathcal{R}_{\text{impl}}$ improves $\mathcal{R}_{\text{Tan}}$ similarly in Section VI-B.

B. Strong Interference Channel With Common Information

Let $\mathcal{P}_s$ denote the set of all joint distributions $p(u_0,x_1,x_2,y_1,y_2)$ that factor as

$$p(u_0,x_1,x_2,y_1,y_2) = p(u_0)p(x_1|u_0)p(x_2|u_0)p(y_1,y_2|x_1,x_2).$$

As defined in [24], an ICC is considered as a SICC if

$$\begin{aligned} I(X_1;Y_1|X_2U_0) &\le I(X_1;Y_2|X_2U_0) \\ I(X_2;Y_2|X_1U_0) &\le I(X_2;Y_1|X_1U_0) \end{aligned}$$

for all joint probability distributions $p(\cdot)\in\mathcal{P}_s$.

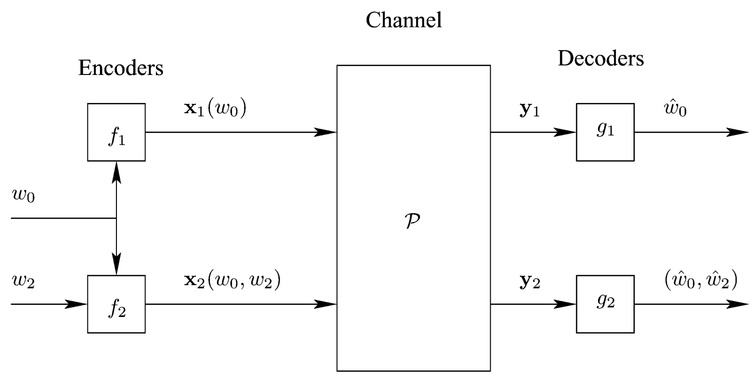

Fig. 3. The asymmetric interference channel with common information.

Let $\mathcal{R}_s(p)$ denote the set of all nonnegative rate triples (R_0, R_1, R_2) such that

$$R_1 \le I(X_1;Y_1|X_2U_0) \tag{25}$$
$$R_2 \le I(X_2;Y_2|X_1U_0) \tag{26}$$
$$R_1 + R_2 \le \min\{I(X_1X_2;Y_1|U_0), I(X_2X_1;Y_2|U_0)\} \tag{27}$$
$$R_0 + R_1 + R_2 \le \min\{I(X_1X_2;Y_1), I(X_2X_1;Y_2)\} \tag{28}$$

for a fixed joint distribution $p(\cdot) \in \mathcal{P}_s$.

Corollary 3 ([24, Achievability of Theorem 1]): Any rate triple $(R_0, R_1, R_2) \in \bigcup_{p(\cdot)\in\mathcal{P}_s} \mathcal{R}_s(p)$ is achievable for the SICC.

Remark 9: By setting $U_t = X_t, t = 1, 2$, and $R_{11} = R_{22} = 0$ in (2)–(11), and removing two redundant ones from the resulting inequalities due to the channel assumptions of the SICC, we can easily obtain (25)–(28).

Remark 10: By letting $U_t = X_t, t = 1, 2$, we treat the private information at each sender as a whole instead of two parts. This differs from what was mentioned earlier in Remark 5. In this case the full private information at each sender is allowed to be crossly observed by the respective nonpairing receivers due to the strong interference.

C. Interference Channel Without Common Information

We now consider the general IC (without common information) as a special case of the ICC, and demonstrate that our achievable rate region for the ICC reduces to the CMG region [15] for the IC.

Let Q denote the time-sharing random variable, and $\mathcal{P}_o$ denote the set of all joint distributions that factor as

$$p(q,u_1,u_2,x_1,x_2,y_1,y_2) = p(q)p(u_1|q)p(u_2|q) \cdot p(x_1|u_1,q)p(x_2|u_2,q)p(y_1,y_2|x_1,x_2).$$

Define $\mathcal{R}_o(p)$ as the set of all rate pairs (R_1, R_2) such that $R_1 = R_{12} + R_{11}$ and $R_2 = R_{21} + R_{22}$ with any nonnegative rate quadruple $(R_{12}, R_{11}, R_{21}, R_{22})$ satisfying

$$R_{11} \le I(X_1;Y_1|U_1U_2Q) \tag{29}$$
$$R_{12} + R_{11} \le I(X_1;Y_1|U_2Q) \tag{30}$$
$$R_{11} + R_{21} \le I(X_1U_2;Y_1|U_1Q) \tag{31}$$
$$R_{12} + R_{11} + R_{21} \le I(X_1U_2;Y_1|Q) \tag{32}$$
$$R_{22} \le I(X_2;Y_2|U_2U_1Q) \tag{33}$$
$$R_{21} + R_{22} \le I(X_2;Y_2|U_1Q) \tag{34}$$
$$R_{22} + R_{12} \le I(X_2U_1;Y_2|U_2Q) \tag{35}$$
$$R_{21} + R_{22} + R_{12} \le I(X_2U_1;Y_2|Q) \tag{36}$$

for a fixed joint distribution $p(\cdot) \in \mathcal{P}_o$, and define $\mathcal{R}_o := \bigcup_{p(\cdot)\in\mathcal{P}_o} \mathcal{R}_o(p)$.

Corollary 4 ([15, Th. 3]): $\mathcal{R}_o$ is an achievable rate region for the IC.

Remark 11: Since no common information is involved, we can choose $U_0 = Q$ and $R_0 = 0$ in (2)–(11), and obtain (29)–(36). On the other hand, one can readily obtain the explicit CMG region ([29, Th. D] and [15, Th. 4]) by setting $U_0 = Q$ and $R_0 = 0$ in (12)–(24).

V. Two Special Cases of the ICC

In this section, we specialize our achievable results in Section III to the following two cases.

A. Asymmetric Interference Channel With Common Information

We first introduce the channel model of this class of the ICCs, namely the asymmetric interference channel with common information (AICC), where one sender does not have private information to transmit. Without loss of generality, we assume that sender 1 only has the common message w_0 to be transmitted to receiver 1, while sender 2 needs transmit both the common message w_0 and the private message w_2 to receiver 2. Fig. 3 depicts the channel model for the AICC, which we denote by $\boldsymbol{C}_a$. We follow the definitions introduced in Section II, and define the capacity region of the channel $\boldsymbol{C}_a$ is a set of all achievable rate pairs (R_0, R_2) for this channel.

Let $\mathcal{P}_a$ denote the set of all joint distributions

$$p(u_0,x_1,u_2,x_2,y_1,y_2) = p(u_0,x_1,u_2,x_2)p(y_1,y_2|x_1,x_2).$$

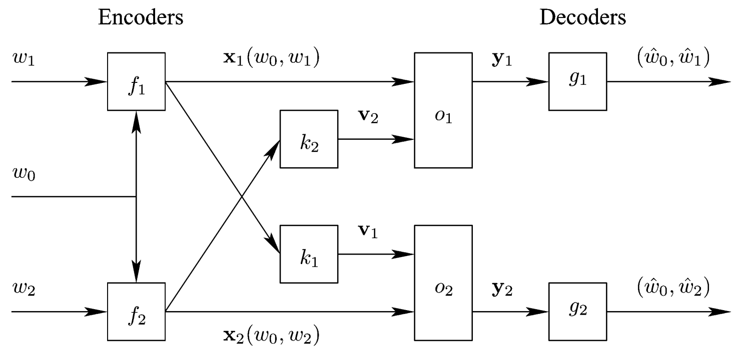

Fig. 4. The class of deterministic interference channels with common information.

Theorem 3: $\mathcal{R}_a := \bigcup_{p(\cdot)\in\mathcal{P}_a} \mathcal{R}_a(p)$ is an achievable rate region for the channel $\boldsymbol{C}_a$, where $\mathcal{R}_a(p)$ is the set of all nonnegative rate pairs (R_0, R_2) such that

$$\begin{aligned} R_2 &\le \min\{I(X_2;Y_2|U_0), I(U_2;Y_1|U_0X_1) \\ &\qquad + I(X_2;Y_2|U_0U_2)\} \\ R_0+R_2 &\le \min\{I(U_0X_2;Y_2), I(U_0X_1U_2;Y_1) \\ &\qquad + I(X_2;Y_2|U_2U_0)\} \end{aligned}$$

for some fixed joint distribution $p(\cdot)\in\mathcal{P}_a$.

Remark 12: 1) It is straightforward to obtain Theorem 3 from Corollary 1 by letting $R_1 = 0$ and $U_1 = X_1$. 2) The coding strategy for this channel remains basically the same as the one for the general ICC: both senders first need cooperate to transmit the common information, while sender 2 treats the private information as two parts, among which only one part is crossly observable to receiver 1. 3) Although the description of this rate region appears simple, establishing the converse remains challenging.

In addition, by letting $U_0 = X_1$ and $U_2 = X_2$, the rate region $\mathcal{R}_a$ reduces to the capacity region for the strong interference channel with unidirectional cooperation [22], [31].

B. A Class of Deterministic Interference Channels With Common Information

We next investigate a class of discrete memoryless DICCs as depicted in Fig. 4. The major attributes of DICCs remain the same as those of an ICC, i.e., the source messages (w_0, w_1, w_2), the channel input and output alphabets $\mathcal{X}_t$ and $\mathcal{Y}_t$, $t = 1, 2$, the encoding functions ($f_1(\cdot)$ and $f_2(\cdot)$) and decoding functions ($g_1(\cdot)$ and $g_2(\cdot)$), the existence of codes, and the achievable rates are defined as the same as those for the general ICC. The distinction lies on the channel transition, which is governed by the following deterministic functions:

$$\begin{aligned} V_t &= k_t(X_t), \quad t = 1,2; \\ Y_1 &= o_1(X_1,V_2), \text{ and } Y_2 = o_2(X_2,V_1) \end{aligned}$$

where V_1 and V_2 represent the interference signals caused by X_1 and X_2 at the corresponding receivers. Furthermore, we assume that there exist two more deterministic functions, $V_2 = h_1(Y_1, X_1)$ and $V_1 = h_2(Y_2, X_2)$. We denote this class of DICCs by $\boldsymbol{C}_d$.

The channel defined above is similar to the one investigated in [11], but there is a slight difference. In [11], it is assumed that $H(Y_1|X_1) = H(V_2)$ and $H(Y_2|X_2) = H(V_1)$ for all product distributions of X_1X_2. It has also been pointed out in [11] that this assumption is equivalent to assuming the existence of $V_2 = h_1(Y_1, X_1)$ and $V_1 = h_2(Y_2, X_2)$. Nevertheless, we assume the latter rather than the former since the former is not satisfied in our case. We will demonstrate that $V_2 = h_1(Y_1, X_1)$ and $V_1 = h_2(Y_2, X_2)$ are the actual governing conditions for this class of DICCs.

Let $\mathcal{P}_d$ denote the set of all joint distributions $p(\cdot)$ that factor as

$$p(v_0,x_1,x_2) = p(v_0)p(x_1|v_0)p(x_2|v_0) \tag{37}$$

where v_0 is the realization of an auxiliary random variable V_0 defined over an arbitrary finite set $\mathcal{V}_0$. Let $\mathcal{R}_d(p)$ denote the set of all nonnegative rate triples (R_0, R_1, R_2) such that

$$R_1 \le H(Y_1|V_0V_2) \tag{38}$$
$$R_2 \le H(Y_2|V_0V_1) \tag{39}$$
$$R_0+R_1 \le H(Y_1) \tag{40}$$
$$R_0+R_2 \le H(Y_2) \tag{41}$$
$$R_1+R_2 \le H(Y_1|V_0V_1)+H(Y_2|V_0V_2) \tag{42}$$
$$R_1+R_2 \le H(Y_1|V_0)+H(Y_2|V_0V_1V_2) \tag{43}$$
$$R_0+R_1+R_2 \le H(Y_1)+H(Y_2|V_0V_1V_2) \tag{44}$$
$$R_1+R_2 \le H(Y_1|V_0V_1V_2)+H(Y_2|V_0) \tag{45}$$
$$R_0+R_1+R_2 \le H(Y_1|V_0V_1V_2)+H(Y_2) \tag{46}$$
$$\begin{aligned} 2R_1+R_2 &\le H(Y_1|V_0)+H(Y_1|V_0V_1V_2) \\ &\qquad + H(Y_2|V_0V_2) \end{aligned} \tag{47}$$
$$\begin{aligned} R_0+2R_1+R_2 &\le H(Y_1)+H(Y_1|V_0V_1V_2) \\ &\qquad + H(Y_2|V_0V_2) \end{aligned} \tag{48}$$
$$\begin{aligned} R_1+2R_2 &\le H(Y_2|V_0)+H(Y_2|V_0V_1V_2) \\ &\qquad + H(Y_1|V_0V_1) \end{aligned} \tag{49}$$
$$\begin{aligned} R_0+R_1+2R_2 &\le H(Y_2)+H(Y_2|V_0V_1V_2) \\ &\qquad + H(Y_1|V_0V_1) \end{aligned} \tag{50}$$

for some fixed joint distribution $p(\cdot)\in\mathcal{P}_d$.

Theorem 4: The capacity region of the channel $\boldsymbol{C}_d$ is the closure of $\bigcup_{p(\cdot)\in\mathcal{P}_d} \mathcal{R}_d(p)$.

Proof:

1) [**Achievability**] It suffices to show that $\mathcal{R}_d(p)$ is achievable for the channel $\boldsymbol{C}_d$ for a fixed joint distribution $p(\cdot) \in \mathcal{P}_d$. As the joint distribution $p(\cdot) \in \mathcal{P}_d$ does not involve V_1 and V_2, it appears difficult to directly apply the superposition coding strategy developed for the general ICC to this channel. Nevertheless, because the interferences V_1 and V_2 are determined by the channel inputs X_1 and X_2, we can extend the joint distribution in the form of (37) to one containing V_1 and V_2 as

$$p(v_0, x_1, x_2, v_1, v_2) = p(v_0)p(x_1|v_0)p(x_2|v_0) \cdot\delta(v_1 - k_1(x_1))\delta(v_2 - k_1(x_2)) \quad (51)$$

where $\delta(\cdot)$ is the Kronecker delta function. Since X_1 and X_2 are conditionally independent given V_0, the interferences V_1 and V_2 also become conditionally independent given V_0. Therefore, the extended joint distribution (51) can be factored as

$$p(v_0, x_1, x_2, v_1, v_2) = p(v_0)p(v_1|v_0)p(v_2|v_0) \cdot p(x_1|v_1, v_0)p(x_2|v_2, v_0)$$

and the achievability of the region $\mathcal{R}_d(p)$ follows readily from Corollary 1.

2) [**Converse**] We first prove that for any nondeterministic (stochastic) $(M_0, M_1, M_2, n, P_e^*)$ code for the channel, there exists a deterministic (M_0, M_1, M_2, n, P_e) code such that $P_e \le P_e^*$. We then upper bound the rates of any deterministic code having $P_e \to 0$ as $n \to \infty$. The detailed steps of the derivations are presented in Appendix D.

The upper bound meets with the inner bound (or the achievable rate region), and thus the theorem follows. □

As mentioned earlier, we assume that there exist $h_1(\cdot,\cdot)$ and $h_2(\cdot,\cdot)$ such that $V_2 = h_1(Y_1, X_1)$ and $V_1 = h_2(Y_2, X_2)$. With this assumption, we have two equalities, $H(V_2^n|W_0) = H(Y_1^n|W_0W_1)$ and $H(V_1^n|W_0) = H(Y_2^n|W_0W_2)$. As can be observed from the converse part of the proof in Appendix D, these two equalities are crucial for us to establish the converse. Moreover, in the absence of common information our assumptions reduce to those made in [11]. In this sense, our assumption is slightly more general compared with the one made in [11]. It is also noteworthy that in the case of no common information, the capacity region of this class of DICCs reduces to the one obtained in [11].

VI. Gaussian Interference Channel With Common Information

In this section, we show how to extend the achievable rate region $\mathcal{R}$ derived for the discrete memoryless ICC to the Gaussian case. We also present a numerical example to illustrate to what extent our region improves the Tan region in [23, Th. 1].

A. Channel Model for the Gaussian ICC

We consider a Gaussian ICC (GICC) in the standard form since any GICC can be transformed to one in the standard form with the capacity region unchanged [7], [30], [19]. As depicted in Fig. 5, a GICC in the standard form can be mathematically expressed as

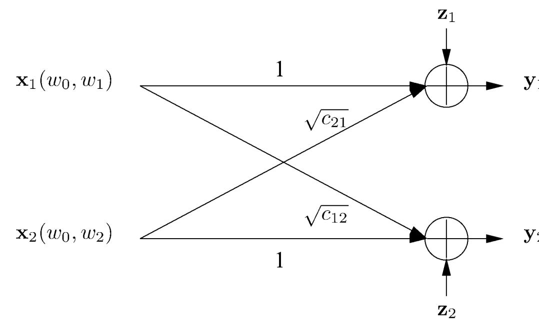

Fig. 5. A Gaussian interference channel with common information.

$$Y_1 = X_1 + \sqrt{c_{21}}X_2 + Z_1 \quad (52)$$
$$Y_2 = X_2 + \sqrt{c_{12}}X_1 + Z_2 \quad (53)$$

where Z_i, $i = 1, 2$, is the additive white Gaussian noise with zero mean and unit variance, and $\sqrt{c_{21}}$ and $\sqrt{c_{12}}$ are the normalized channel gains of the respective interference links.

In addition, the codewords used for this channel are subject to the average power constraint given by $\sum_{t=1}^{n} \|x_{it}\|^2/n \le P_i$, $i = 1, 2$. We only consider Gaussian codewords X_i^n, $i = 1, 2$, for the GICC, since it is shown in the Maximum-Entropy Theorem [32] that Gaussian inputs are optimal for Gaussian channels. Furthermore, we also fix the time-sharing random variable Q as a constant. Regarding how the choice of Q affects the size of the achievable rate region, interested readers are referred to [13] for a detailed exposition.

B. Achievable Rate Region for the GICC

We first define the following mappings of random variables with respect to the joint probability distribution (1):

M1) W_0, W_{12}, W_{11}, W_{21}, and W_{22}, distributed according to $\mathcal{N}(0, 1)$;
M2) $X_1 = \sqrt{\alpha_1 P_1}W_0 + \sqrt{\bar{\alpha}_1\beta_1 P_1}W_{12} + \sqrt{\bar{\alpha}_1\bar{\beta}_1 P_1}W_{11}$;
M3) $X_2 = \sqrt{\alpha_2 P_2}W_0 + \sqrt{\bar{\alpha}_2\beta_2 P_2}W_{21} + \sqrt{\bar{\alpha}_2\bar{\beta}_2 P_2}W_{22}$;
M4) $U_0 = (\sqrt{\alpha_1 P_1} + \sqrt{\alpha_2 P_2})W_0$;
M5) $U_1 = \sqrt{\alpha_1 P_1}W_0 + \sqrt{\bar{\alpha}_1\beta_1 P_1}W_{12}$;
M6) $U_2 = \sqrt{\alpha_2 P_2}W_0 + \sqrt{\bar{\alpha}_2\beta_2 P_2}W_{21}$;

where $\alpha_i, \beta_i \in [0, 1]$, $\bar{\alpha}_i = 1 - \alpha_i$, and $\bar{\beta}_i = 1 - \beta_i$, for $i = 1, 2$. Based on these mappings, and the channel model described by (52) and (53), we obtain

$$\begin{aligned} Y_1 = &\left(\sqrt{\alpha_1 P_1} + \sqrt{c_{21}\alpha_2 P_2}\right)W_0 + \sqrt{\bar{\alpha}_1\beta_1 P_1}W_{12} \\ &+ \sqrt{\bar{\alpha}_1\bar{\beta}_1 P_1}W_{11} + \sqrt{c_{21}\bar{\alpha}_2\beta_2 P_2}W_{21} \\ &+ \sqrt{c_{21}\bar{\alpha}_2\bar{\beta}_2 P_2}W_{22} + Z_1 \end{aligned} \quad (54)$$

$$\begin{aligned} Y_2 = &\left(\sqrt{\alpha_2 P_2} + \sqrt{c_{12}\alpha_1 P_1}\right)W_0 + \sqrt{\bar{\alpha}_2\beta_2 P_2}W_{21} \\ &+ \sqrt{\bar{\alpha}_2\bar{\beta}_2 P_2}W_{22} + \sqrt{c_{12}\bar{\alpha}_1\beta_1 P_1}W_{12} \\ &+ \sqrt{c_{12}\bar{\alpha}_1\bar{\beta}_1 P_1}W_{11} + Z_2. \end{aligned} \quad (55)$$

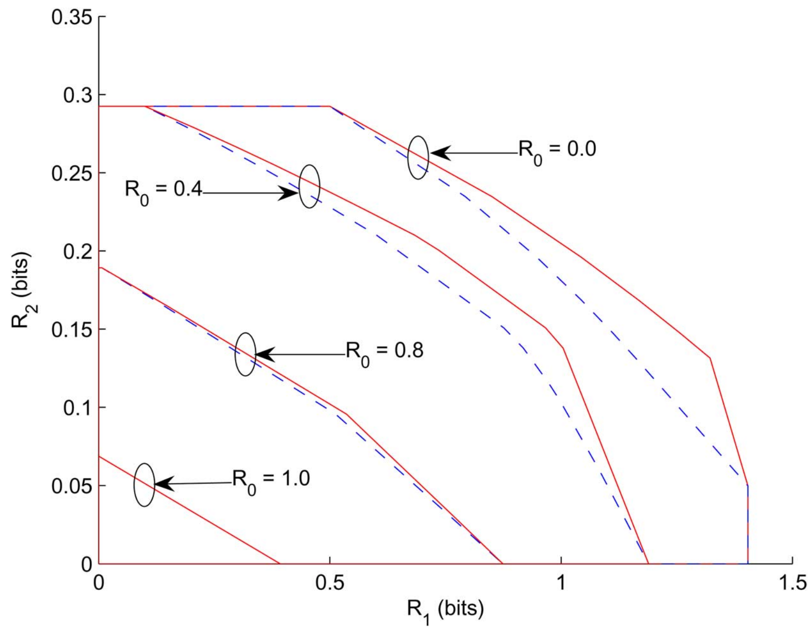

Fig. 6. $P_1 = 6, P_2 = 0.5, c_{21} = 1, c_{12} = 0.25$. The dashed lines characterize the rate regions of $\mathcal{G}_{\text{Tan}}$ sliced at $R_0 = 0, 0.4, 0.8, 1$, respectively, and the solid lines characterize the rate regions of $\mathcal{G}$ sliced at $R_0 = 0, 0.4, 0.8, 1$, respectively.

With the relations between the random variables defined by the mappings, M1)–M6), and the channel input–output relations described by (54) and (55), we evaluate the mutual information terms $I(\cdot)$ in (12)–(24) as shown in (56)–(65) at the bottom of the page where $\gamma(x) := \frac{1}{2}\log_2(1+x)$.

Replacing each mutual information term in (12)–(24) with its corresponding one from (56)–(65), we can obtain the Gaussian counterpart of $\mathcal{R}$, namely $\mathcal{G}$. Since the resulting description of the region $\mathcal{G}$ is rather lengthy, we do not explicitly include it in this paper.

We next compare the obtained achievable rate region $\mathcal{G}$, with $\mathcal{G}_{\text{Tan}}$, the Gaussian counterpart of $\mathcal{R}_{\text{Tan}}$, in Fig. 6. It is difficult to show the comparison in a three-dimensional (3-D) plot. Thus, we slice the 3-D rate regions $\mathcal{G}$ and $\mathcal{G}_{\text{Tan}}$ at different values of R_0, and obtain a number of sliced views as shown in Fig. 6. As can be seen from Fig. 6, the improvement of $\mathcal{G}$ over $\mathcal{G}_{\text{Tan}}$ for $R_0 = 0.0$ is significant, which matches exactly with the result presented in [9, Fig. 10]. It can also be observed that when R_0 is relatively high (e.g., $R_0 = 1.0$), the two regions coincide with each other. This is because most of the power of the two senders is allocated to transmit the high rate common information, while the remaining power for the private information becomes relatively small such that the improvement, primarily gained from allowing cross observation of the private information, diminishes. Note that a similar example has also been given in [28].

VII. Conclusion

We derived in this paper a new achievable rate region for the two user discrete memoryless ICC. We have shown that the derived achievable rate region contains the one established in [23], and reduces to some other existing results developed for the ICC or IC. We also investigated two special cases of the ICC. For the

$$I(X_1;Y_1|U_0U_2) = \gamma(\bar{\alpha}_1 P_1/(c_{21}\bar{\alpha}_2\bar{\beta}_2 P_2 + 1)) \tag{56}$$
$$I(X_2;Y_2|U_0U_1) = \gamma(\bar{\alpha}_2 P_2/(c_{12}\bar{\alpha}_1\bar{\beta}_1 P_1 + 1)) \tag{57}$$
$$I(U_0X_1U_2;Y_1) = \gamma\left(\frac{\left(\sqrt{\alpha_1 P_1} + \sqrt{c_{21}\alpha_2 P_2}\right)^2 + \bar{\alpha}_1 P_1 + c_{21}\bar{\alpha}_2 P_2}{c_{21}\bar{\alpha}_2\bar{\beta}_2 P_2 + 1}\right) \tag{58}$$
$$I(U_0X_2U_1;Y_2) = \gamma\left(\frac{\left(\sqrt{\alpha_2 P_2} + \sqrt{c_{12}\alpha_1 P_1}\right)^2 + \bar{\alpha}_2 P_2 + c_{12}\bar{\alpha}_1 P_1}{c_{12}\bar{\alpha}_1\bar{\beta}_1 P_1 + 1}\right) \tag{59}$$
$$I(X_1U_2;Y_1|U_0U_1) = \gamma((\bar{\alpha}_1\bar{\beta}_1 P_1 + c_{21}\bar{\alpha}_2\beta_2 P_2)/(c_{21}\bar{\alpha}_2\bar{\beta}_2 P_2 + 1)) \tag{60}$$
$$I(X_2U_1;Y_2|U_0U_2) = \gamma((\bar{\alpha}_2\bar{\beta}_2 P_2 + c_{12}\bar{\alpha}_1\beta_1 P_1)/(c_{12}\bar{\alpha}_1\bar{\beta}_1 P_1 + 1)) \tag{61}$$
$$I(X_1;Y_1|U_0U_1U_2) = \gamma(\bar{\alpha}_1\bar{\beta}_1 P_1/(c_{21}\bar{\alpha}_2\bar{\beta}_2 P_2 + 1)) \tag{62}$$
$$I(X_2;Y_2|U_0U_1U_2) = \gamma(\bar{\alpha}_2\bar{\beta}_2 P_2/(c_{12}\bar{\alpha}_1\bar{\beta}_1 P_1 + 1)) \tag{63}$$
$$I(X_1U_2;Y_1|U_0) = \gamma((\bar{\alpha}_1 P_1 + c_{21}\bar{\alpha}_2\beta_2 P_2)/(c_{21}\bar{\alpha}_2\bar{\beta}_2 P_2 + 1)) \tag{64}$$
$$I(X_2U_1;Y_2|U_0) = \gamma((\bar{\alpha}_2 P_2 + c_{12}\bar{\alpha}_1\beta_1 P_1)/(c_{12}\bar{\alpha}_1\bar{\beta}_1 P_1 + 1)) \tag{65}$$

first case in which only one sender has private information to send, we obtained an achievable rate region with a fairly simple description; while for the second case, a class of DICCs, we show that our achievable region is the capacity region. Nevertheless, in a general ICC setting, the tightness of our achievable rate region as an inner bound of the capacity region is unknown.

APPENDIX A
PROOF OF LEMMA 2

As the following lemma will be frequently used, we state it before the proof of Lemma 2.

Lemma 3 [32, Th. 14.2.3]: Let $A_\epsilon^{(n)}$ denote the typical set for the probability distribution $p(s_1, s_2, s_3)$, and let

$$P(\mathbf{S}_1' = \mathbf{s}_1, \mathbf{S}_2' = \mathbf{s}_2, \mathbf{S}_3' = \mathbf{s}_3) = \prod_{i=1}^{n} p(s_{1i}|s_{3i})p(s_{2i}|s_{3i})p(s_{3i})$$

then $P\left\{(\mathbf{S}_1', \mathbf{S}_2', \mathbf{S}_3') \in A_\epsilon^{(n)}\right\} \doteq 2^{-n(I(S_1;S_2|S_3)\pm 6\epsilon)}$.

Proof of Lemma 2: [**Codebook Generation**] Let us fix a joint distribution $p(\cdot)$ that factors in the form of (1). We first generate 2^{nR_0} independent codewords $\mathbf{u}_0(i)$, $i \in \{1, \ldots, 2^{nR_0}\}$, according to $\prod_{t=1}^{n} p(u_{0t})$. At encoder 1, for each codeword $\mathbf{u}_0(i)$, generate $2^{nR_{12}}$ independent codewords $\mathbf{u}_1(i,j)$, $j \in \{1, \ldots, 2^{nR_{12}}\}$, according to $\prod_{t=1}^{n} p(u_{1t}|u_{0t})$. Subsequently, for each pair of codewords $(\mathbf{u}_0(i), \mathbf{u}_1(i,j))$, generate $2^{nR_{11}}$ independent codewords $\mathbf{x}_1(i,j,k)$, $k \in \{1, \ldots, 2^{nR_{11}}\}$, according to $\prod_{t=1}^{n} p(x_{1t}|u_{1t}u_{0t})$. Similarly at encoder 2, for each codeword $\mathbf{u}_0(i)$, generate $2^{nR_{21}}$ independent codewords $\mathbf{u}_2(i,l)$, $l \in \{1, \ldots, 2^{nR_{21}}\}$, according to $\prod_{t=1}^{n} p(u_{2t}|u_{0t})$. Subsequently, for each codeword pair $(\mathbf{u}_0(i), \mathbf{u}_2(i,l))$, generate $2^{nR_{22}}$ independent codewords $\mathbf{x}_2(i,l,m)$, $m \in \{1, \ldots, 2^{nR_{22}}\}$, according to $\prod_{t=1}^{n} p(x_{2t}|u_{2t}u_{0t})$. The entire codebook consisting of all the codewords $\mathbf{u}_0(i)$, $\mathbf{u}_1(i,j)$, $\mathbf{x}_1(i,j,k)$, $\mathbf{u}_2(i,l)$ and, $\mathbf{x}_2(i,l,m)$ with $i \in \{1, \ldots, 2^{nR_0}\}$, $j \in \{1, \ldots, 2^{nR_{12}}\}$, $k \in \{1, \ldots, 2^{nR_{11}}\}$, $l \in \{1, \ldots, 2^{nR_{21}}\}$, and, $m \in \{1, \ldots, 2^{nR_{22}}\}$ is revealed to both receivers.

[**Encoding & Transmission**] Suppose that the source message vector generated at the two senders is $(n_0, n_1, l_1, n_2, l_2) = (i,j,k,l,m)$. Sender 1 transmits the codeword $\mathbf{x}_1(i,j,k)$ with n channel uses, while sender 2 transmits the codeword $\mathbf{x}_2(i,l,m)$ with n channel uses. The transmissions are assumed to be perfectly synchronized.

[**Decoding**] Each receiver accumulates an n-length channel output sequence, $\mathbf{y}_1$ (receiver 1) or $\mathbf{y}_2$ (receiver 2). Let $A_\epsilon^{(n)}$ denote the typical sets of the respective joint distributions. Decoder 1 declares that $(\hat{i}, \hat{j}, \hat{k})$ is received, if $(\hat{i}, \hat{j}, \hat{k})$ is the unique message vector such that $(\mathbf{u}_0(\hat{i}), \mathbf{u}_1(\hat{i},\hat{j}), \mathbf{x}_1(\hat{i},\hat{j},\hat{k}), \mathbf{u}_2(\hat{i},l), \mathbf{y}_1) \in A_\epsilon^{(n)}$ for some l; otherwise, it declares an error. Similarly, decoder 2 looks for a unique message vector $(\hat{i}, \hat{l}, \hat{m})$ such that $(\mathbf{u}_0(\hat{i}), \mathbf{u}_2(\hat{i},\hat{l}), \mathbf{x}_2(\hat{i},\hat{l},\hat{m}), \mathbf{u}_1(\hat{i},j), \mathbf{y}_2) \in A_\epsilon^{(n)}$ for some j; otherwise, it declares an error.

[**Analysis of the Probability of Decoding Error**] Because of the symmetry of the codebook generation, the probability of error does not depend on which message vector is encoded and transmitted. Since the messages are uniformly generated over their respective ranges, the average error probability over all the possible messages is equal to the probability of error incurred when any message vector is encoded and transmitted. We hence only analyze the probability of error at decoder 1 in details, since the same analysis can be performed for decoder 2. Without loss of generality, we assume that a source message vector $(n_0, n_l, l_1, n_2, l_2) = (1,1,1,1,1)$ is encoded and transmitted for the subsequent analysis. We first define the event

$$E_{ijkl} := \left\{(\mathbf{U}_0(i), \mathbf{U}_1(i,j), \mathbf{X}_1(i,j,k), \mathbf{U}_2(i,l), \mathbf{Y}_1) \in A_\epsilon^{(n)}\right\}.$$

The possible error events can be grouped into two classes: 1) the codewords transmitted are not jointly typical, i.e., E_{1111}^c happens; 2) there exist some $(i,j,k) \neq (1,1,1)$ such that E_{ijkl} happens (l may not be 1). Thus the probability of error at decoder 1 can be expressed as

$$P_{e,1}^{(n)} = P\left(E_{1111}^c \bigcup \cup_{(i,j,k)\neq(1,1,1)} E_{ijkl}\right). \tag{66}$$

By applying the union bound, we can upper bound (66) as

$$\begin{aligned} P_{e,1}^{(n)} &\leq P(E_{1111}^c) + P(\cup_{(i,j,k)\neq(1,1,1)} E_{ijkl}) \\ &\leq P(E_{1111}^c) + \sum_{i\neq 1} P(E_{i111}) + \sum_{i\neq 1, l\neq 1} P(E_{i11l}) \\ &\quad + \sum_{j\neq 1} P(E_{1j11}) + \sum_{j\neq 1, l\neq 1} P(E_{1j1l}) + \sum_{k\neq 1} P(E_{11k1}) \\ &\quad + \sum_{k\neq 1, l\neq 1} P(E_{11kl}) + \sum_{i\neq 1, j\neq 1} P(E_{ij11}) \\ &\quad + \sum_{i\neq 1, j\neq 1, l\neq 1} P(E_{ij1l}) + \sum_{i\neq 1, k\neq 1} P(E_{i1k1}) \\ &\quad + \sum_{i\neq 1, k\neq 1, l\neq 1} P(E_{i1kl}) + \sum_{j\neq 1, k\neq 1} P(E_{1jk1}) \\ &\quad + \sum_{j\neq 1, k\neq 1, l\neq 1} P(E_{1jkl}) + \sum_{i\neq 1, j\neq 1, k\neq 1} P(E_{ijk1}) \\ &\quad + \sum_{i\neq 1, j\neq 1, k\neq 1, l\neq 1} P(E_{ijkl}). \end{aligned} \tag{67}$$

Due to the asymptotic equipartition property (AEP) [32], $P(E_{1111}^c)$ in (67) can be made arbitrarily small as long as n is sufficiently large. The rest of the fourteen probability terms in (67) can be evaluated through a standard procedure, which is demonstrated as follows. To evaluate $P(E_{i111})$, we apply Lemma 3 by letting $\mathbf{S}_1' = (\mathbf{U}_0(i), \mathbf{U}_1(i,1), \mathbf{X}_1(i,1,1), \mathbf{U}_2(i,1))$, $\mathbf{S}_2' = \mathbf{Y}_1$, and $\mathbf{S}_3' = \emptyset$ with $\emptyset$ denoting the empty set. Since the assumption of Lemma 3 on the joint distribution of $\mathbf{S}_1'$, $\mathbf{S}_2'$, and $\mathbf{S}_3'$ is satisfied, we have

$$\begin{aligned} P(E_{i111}) &\leq 2^{-n(I(U_0U_1X_1U_2;Y_1)-6\epsilon)} \\ &\stackrel{(a)}{=} 2^{-n(I(U_0X_1U_2;Y_1)-6\epsilon)}. \end{aligned}$$

Note that (a) follows from the fact that $I(U_1;Y_1|U_0U_2X_2) = 0$, which is because U_1, (U_0, U_2, X_1), and Y_1 form a Markov chain $U_1 \rightarrow (U_0, U_2, X_1) \rightarrow Y_1$. Since the case with $\mathbf{S}_3' = \emptyset$ seems not archetypal, we evaluate one more probability term, $P(E_{1jk1})$. Again, we use Lemma 3 by letting $\mathbf{S}'_1 = (\mathbf{U}_1(1,j), \mathbf{X}_1(1,j,k))$, $\mathbf{S}'_2 = \mathbf{Y}_1$, and $\mathbf{S}'_3 = (\mathbf{U}_0(1), \mathbf{U}_2(1,1))$ to obtain

$$P(E_{1jk1}) \le 2^{-n(I(U_1X_1;Y_1|U_0U_2)-6\epsilon)}.$$

By repeatedly applying Lemma 3, we obtain upper bounds of the remaining twelve probability terms. Further, we employ these bounds to derive an upper bound of the probability of error at decoder 1 as

$$\begin{aligned} P_{e,1}^{(n)} \le\, & \epsilon + 2^{nR_0}2^{-n(I(U_0X_1U_2;Y_1)-6\epsilon)} \\ & + 2^{n(R_0+R_{21})}2^{-n(I(U_0X_1U_2;Y_1)-6\epsilon)} \\ & + 2^{nR_{12}}2^{-n(I(X_1;Y_1|U_0U_2)-6\epsilon)} \\ & + 2^{n(R_{12}+R_{21})}2^{-n(I(X_1U_2;Y_1|U_0)-6\epsilon)} \\ & + 2^{nR_{11}}2^{-n(I(X_1;Y_1|U_0U_1U_2)-6\epsilon)} \\ & + 2^{n(R_{11}+R_{21})}2^{-n(I(X_1U_2;Y_1|U_0U_1)-6\epsilon)} \\ & + 2^{n(R_0+R_{12})}2^{-n(I(U_0X_1U_2;Y_1)-6\epsilon)} \\ & + 2^{n(R_0+R_{12}+R_{21})}2^{-n(I(U_0X_1U_2;Y_1)-6\epsilon)} \\ & + 2^{n(R_0+R_{11})}2^{-n(I(U_0X_1U_2;Y_1)-6\epsilon)} \\ & + 2^{n(R_0+R_{11}+R_{21})}2^{-n(I(U_0X_1U_2;Y_1)-6\epsilon)} \\ & + 2^{n(R_{12}+R_{11})}2^{-n(I(X_1;Y_1|U_0U_2)-6\epsilon)} \\ & + 2^{n(R_{12}+R_{11}+R_{21})}2^{-n(I(X_1U_2;Y_1|U_0)-6\epsilon)} \\ & + 2^{n(R_0+R_{12}+R_{11})}2^{-n(I(U_0X_1U_2;Y_1)-6\epsilon)} \\ & + 2^{n(R_0+R_{12}+R_{11}+R_{21})}2^{-n(I(U_0X_1U_2;Y_1)-6\epsilon)}. \end{aligned} \quad (68)$$

It is now easy to check that when (2)–(6) hold and n is sufficiently large, we have $P_{e,1}^{(n)} \le 15\epsilon$. By symmetry, we have $P_{e,2}^{(n)} \le 15\epsilon$ for decoder 2, when (7)–(11) hold and n is sufficiently large. Hence, $\max\left\{P_{e,1}^{(n)}, P_{e,2}^{(n)}\right\} \le 15\epsilon$, and Lemma 2 readily follows. □

Appendix B
Proof of the Convexity of $\mathcal{R}_m$

Let $(R_0^1, R_{12}^1, R_{11}^1, R_{21}^1, R_{22}^1)$ and $(R_0^2, R_{12}^2, R_{11}^2, R_{21}^2, R_{22}^2)$ be two arbitrary rate quintuples belonging to $\mathcal{R}_m$. It suffices to show that for given any $\alpha \in [0,1]$, $(\alpha R_0^1 + (1-\alpha)R_0^2, \alpha R_{12}^1 + (1-\alpha)R_{12}^2, \alpha R_{11}^1 + (1-\alpha)R_{11}^2, \alpha R_{21}^1 + (1-\alpha)R_{21}^2, \alpha R_{22}^1 + (1-\alpha)R_{22}^2) \in \mathcal{R}_m$. Note that the rate region $\mathcal{R}_m$ is the union of regions $\mathcal{R}_m(p)$ over all $p(\cdot) \in \mathcal{P}^*$. Thus, there must exist two sets of auxiliary random variables, namely $\left(U_0^1, U_1^1, U_2^1\right)$ and $\left(U_0^2, U_1^2, U_2^2\right)$ such that their joint distributions $p_1(\cdot)$ and $p_2(\cdot)$ factor as

$$\begin{aligned} & p_1\left(u_0^1, u_1^1, u_2^1, x_1, x_2, y_1, y_2\right) \\ & \quad = p\left(u_0^1\right) p\left(u_1^1|u_0^1\right) p\left(u_2^1|u_0^1\right) \\ & \qquad \cdot p\left(x_1|u_1^1, u_0^1\right) p\left(x_2|u_2^1, u_0^1\right) p\left(y_1, y_2|x_1, x_2\right) \\ & p_2\left(u_0^2, u_1^2, u_2^2, x_1, x_2, y_1, y_2\right) \\ & \quad = p\left(u_0^2\right) p\left(u_1^2|u_0^2\right) p\left(u_2^2|u_0^2\right) \\ & \qquad \cdot p\left(x_1|u_1^2, u_0^2\right) p\left(x_2|u_2^2, u_0^2\right) p\left(y_1, y_2|x_1, x_2\right). \end{aligned}$$

Let T be the independent random variable, taking the value 1 with probability α and 2 with probability $1-\alpha$. We define a new set of auxiliary random variables (U_0, U_1, U_2) such that $U_0 = \left(U_0^T, T\right)$, $U_1 = U_1^T$, and $U_2 = U_2^T$, and then their joint distribution $p_3(\cdot)$ can factor

$$\begin{aligned} p_3(u_0, u_1, u_2, x_1, x_2, y_1, y_2) = p(u_0)p(u_1|u_0)p(u_2|u_0) \\ \cdot p(x_1|u_1, u_0)p(x_2|u_2, u_0)p(y_1, y_2|x_1, x_2). \end{aligned}$$

Since $p_3(\cdot) \in \mathcal{P}^*$, we have $\mathcal{R}_m(p_3) \subseteq \mathcal{R}_m$. It is easy to show that $\left(\alpha R_0^1 + (1-\alpha)R_0^2, \alpha R_{12}^1 + (1-\alpha)R_{12}^2, \alpha R_{11}^1 + (1-\alpha)R_{11}^2, \alpha R_{21}^1 + (1-\alpha)R_{21}^2, \alpha R_{22}^1 + (1-\alpha)R_{22}^2\right) \in \mathcal{R}_m(p_3)$ by following the steps used to prove the convexity of the capacity region for the MACC in Appendix A of [27]. Therefore, we conclude $\left(\alpha R_0^1 + (1-\alpha)R_0^2, \alpha R_{12}^1 + (1-\alpha)R_{12}^2, \alpha R_{11}^1 + (1-\alpha)R_{11}^2, \alpha R_{21}^1 + (1-\alpha)R_{21}^2, \alpha R_{22}^1 + (1-\alpha)R_{22}^2\right) \in \mathcal{R}_m(p_3) \subseteq \mathcal{R}_m$, which proves the convexity.

Appendix C
Proof of Corollary 1

A. Fourier–Motzkin Elimination

We show in detail, how to apply Fourier–Motzkin elimination to obtain the explicit rate region depicted by (12)–(24).

Step 1) By substituting R_{11} with $R_1 - R_{12}$ and R_{22} with $R_2 - R_{21}$, we can rewrite the implicit rate region (2)–(11) for a fixed joint distribution $p(\cdot) \in \mathcal{P}^*$ as

$$R_1 - R_{12} \le a_1 \quad (69)$$
$$R_1 \le b_1 \quad (70)$$
$$R_1 - R_{12} + R_{21} \le c_1 \quad (71)$$
$$R_1 + R_{21} \le d_1 \quad (72)$$
$$R_0 + R_1 + R_{21} \le e_1 \quad (73)$$
$$R_2 - R_{21} \le a_2 \quad (74)$$
$$R_2 \le b_2 \quad (75)$$
$$R_2 - R_{21} + R_{12} \le c_2 \quad (76)$$
$$R_2 + R_{12} \le d_2 \quad (77)$$
$$R_0 + R_2 + R_{12} \le e_2 \quad (78)$$
$$-R_{12} \le 0 \quad (79)$$
$$-R_1 + R_{12} \le 0 \quad (80)$$
$$-R_{21} \le 0 \quad (81)$$
$$-R_2 + R_{21} \le 0 \quad (82)$$

where $a_1 := I(X_1;Y_1|U_0U_1U_2)$, $b_1 := I(X_1;Y_1|U_0U_2)$, $c_1 := I(X_1U_2;Y_1|U_0U_1)$, $d_1 := I(X_1U_2;Y_1|U_0)$, $e_1 := I(U_0X_1U_2;Y_1)$, $a_2 := I(X_2;Y_2|U_0U_2U_1)$, $b_2 := I(X_2;Y_2|U_0U_1)$, $c_2 := I(X_2U_1;Y_2|U_0U_2)$, $d_2 := I(X_2U_1;Y_2|U_0)$, and $e_2 := I(U_0X_2U_1;Y_2)$.

Categorize (69)–(82) into the following three groups such that the inequalities in group 1 do not contain the R_{12} term, those in group 2 contain the negative R_{12} term, and those in group 3 contain the positive R_{12} term.

$$R_1 \le b_1 \quad (83)$$
$$R_1 + R_{21} \le d_1 \quad (84)$$
$$R_0 + R_1 + R_{21} \le e_1 \quad (85)$$

$$R_2 - R_{21} \le a_2 \quad (86)$$
$$R_2 \le b_2 \quad (87)$$
$$-R_{21} \le 0 \quad (88)$$
$$-R_2 + R_{21} \le 0 \quad (89)$$
and
$$R_1 - R_{12} \le a_1 \quad (90)$$
$$R_1 - R_{12} + R_{21} \le c_1 \quad (91)$$
$$-R_{12} \le 0 \quad (92)$$
and
$$R_2 - R_{21} + R_{12} \le c_2 \quad (93)$$
$$R_2 + R_{12} \le d_2 \quad (94)$$
$$R_0 + R_2 + R_{12} \le e_2 \quad (95)$$
$$-R_1 + R_{12} \le 0. \quad (96)$$

Step 2) By adding each inequality from (90)–(92) and each one from (93)–(96), we eliminate R_{12} and obtain the following new inequalities:

$$R_1 + R_2 - R_{21} \le a_1 + c_2 \quad (97)$$
$$R_1 + R_2 \le a_1 + d_2 \quad (98)$$
$$R_0 + R_1 + R_2 \le a_1 + e_2 \quad (99)$$
$$0 \le a_1 \quad (100)$$
$$R_1 + R_2 \le c_1 + c_2 \quad (101)$$
$$R_1 + R_2 + R_{21} \le c_1 + d_2 \quad (102)$$
$$R_0 + R_1 + R_2 + R_{21} \le c_1 + e_2 \quad (103)$$
$$R_{21} \le c_1 \quad (104)$$
$$R_2 - R_{21} \le c_2 \quad (105)$$
$$R_2 \le d_2 \quad (106)$$
$$R_0 + R_2 \le e_2 \quad (107)$$
$$-R_1 \le 0. \quad (108)$$

Observing that (100) always holds, we exclude it first. It is straightforward to verify that (106) is implied by (87), and (105) is implied by (86). We therefore also exclude both (106) and (105). We then categorize the remaining inequalities in (83)–(89) and (97)–(108) into the following three groups according to the different involvement of R_{21}:

$$R_1 \le b_1 \quad (109)$$
$$R_2 \le b_2 \quad (110)$$
$$R_1 + R_2 \le a_1 + d_2 \quad (111)$$
$$R_1 + R_2 \le c_1 + c_2 \quad (112)$$
$$R_0 + R_1 + R_2 \le a_1 + e_2 \quad (113)$$
$$-R_1 \le 0 \quad (114)$$
$$R_0 + R_2 \le e_2 \quad (115)$$
and
$$R_2 - R_{21} \le a_2 \quad (116)$$
$$-R_{21} \le 0 \quad (117)$$
$$R_1 + R_2 - R_{21} \le a_1 + c_2 \quad (118)$$
and
$$R_1 + R_{21} \le d_1 \quad (119)$$
$$R_0 + R_1 + R_{21} \le e_1 \quad (120)$$
$$-R_2 + R_{21} \le 0 \quad (121)$$
$$R_1 + R_2 + R_{21} \le c_1 + d_2 \quad (122)$$
$$R_0 + R_1 + R_2 + R_{21} \le c_1 + e_2 \quad (123)$$
$$R_{21} \le c_1. \quad (124)$$

Step 3: By adding each inequality from (116)–(118) and each one from (119)–(124), we eliminate R_{21} and obtain the following new inequalities:

$$R_1 + R_2 \le a_2 + d_1 \quad (125)$$
$$R_0 + R_1 + R_2 \le a_2 + e_1 \quad (126)$$
$$0 \le a_2 \quad (127)$$
$$R_1 + 2R_2 \le a_2 + c_1 + d_2 \quad (128)$$
$$R_0 + R_1 + 2R_2 \le a_2 + c_1 + e_2 \quad (129)$$
$$R_2 \le a_2 + c_1 \quad (130)$$
$$R_1 \le d_1 \quad (131)$$
$$R_0 + R_1 \le e_1 \quad (132)$$
$$-R_2 \le 0 \quad (133)$$
$$R_1 + R_2 \le c_1 + d_2 \quad (134)$$
$$R_0 + R_1 + R_2 \le c_1 + e_2 \quad (135)$$
$$0 \le c_1 \quad (136)$$
$$2R_1 + R_2 \le a_1 + c_2 + d_1 \quad (137)$$
$$R_0 + 2R_1 + R_2 \le a_1 + c_2 + e_1 \quad (138)$$
$$R_1 \le a_1 + c_2 \quad (139)$$
$$2R_1 + 2R_2 \le a_1 + c_2 + c_1 + d_2 \quad (140)$$
$$R_0 + 2R_1 + 2R_2 \le a_1 + c_2 + c_1 + e_2 \quad (141)$$
$$R_1 + R_2 \le a_1 + c_2 + c_1. \quad (142)$$

We now group (109)–(115) and (125)–(142) together, and we can observe that: i) (127) and (136) always hold, ii) (131) is implied by (109), iii) (134) is implied by (111), iv) (142) is implied by (112), v) (140) is implied by (111) and (112), vi) (135) is implied by (113), and vii) (141) is implied by (112) and (113). By removing the redundant inequalities and reordering the remaining ones, we have

$$R_1 \le b_1 \quad (143)$$
$$R_1 \le a_1 + c_2 \quad (144)$$
$$R_2 \le b_2 \quad (145)$$
$$R_2 \le a_2 + c_1 \quad (146)$$
$$R_0 + R_1 \le e_1 \quad (147)$$
$$R_0 + R_2 \le e_2 \quad (148)$$
$$R_1 + R_2 \le c_1 + c_2 \quad (149)$$
$$R_1 + R_2 \le a_1 + d_2 \quad (150)$$
$$R_0 + R_1 + R_2 \le a_1 + e_2 \quad (151)$$
$$R_1 + R_2 \le a_2 + d_1 \quad (152)$$
$$R_0 + R_1 + R_2 \le a_2 + e_1 \quad (153)$$
$$2R_1 + R_2 \le a_1 + c_2 + d_1 \quad (154)$$
$$R_0 + 2R_1 + R_2 \le a_1 + c_2 + e_1 \quad (155)$$
$$R_1 + 2R_2 \le a_2 + c_1 + d_2 \quad (156)$$
$$R_0 + R_1 + 2R_2 \le a_2 + c_1 + e_2 \quad (157)$$
$$-R_1 \le 0 \quad (158)$$
$$-R_2 \le 0. \quad (159)$$

Let $\mathcal{R}^*(p)$ denote the rate region defined by (143)–(159) for a fixed joint distribution $p(\cdot) \in \mathcal{P}^*$, and let $\mathcal{R}^*$ be defined as $\mathcal{R}^* := \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}^*(p)$.

Note that $\mathcal{R}^*(p)$ has two additional rate constraints (144) and (146), compared with $\mathcal{R}(p)$ (explicitly given in Corollary 1). We next show that both (144) and (146) are redundant by establishing the following equivalence:

$$\mathcal{R}^* = \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}^*(p) \equiv \mathcal{R} = \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}(p). \tag{160}$$

B. Equivalence Between $\mathcal{R}$ and $\mathcal{R}^*$

For any fixed joint distribution $p(\cdot) \in \mathcal{P}^*$, $\mathcal{R}^*(p)$ involves two additional rate constraints compared to $\mathcal{R}(p)$. It implies that $\mathcal{R}^*(p) \subseteq \mathcal{R}(p)$ and $\bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}^*(p) \subseteq \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}(p)$. To show the equivalence, we need prove $\bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}(p) \subseteq \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}^*(p)$. It is sufficient to show that for any given joint distribution $p(\cdot) \in \mathcal{P}^*$, we have $\mathcal{R}(p) \subseteq \mathcal{R}^*(p) \cup \mathcal{R}^*(p_1) \cup \mathcal{R}^*(p_2)$, where $p_1(\cdot)$ and $p_2(\cdot)$ are defined as

$$p_1(u_0,u_2,x_1,x_2,y_1,y_2) = \sum_{u_1\in\mathcal{U}_1} p(u_0,u_1,u_2,x_1,x_2,y_1,y_2)$$
$$p_2(u_0,u_1,x_1,x_2,y_1,y_2) = \sum_{u_2\in\mathcal{U}_2} p(u_0,u_1,u_2,x_1,x_2,y_1,y_2).$$

Without loss of generality, suppose that $(\tilde{R}_0,\tilde{R}_1,\tilde{R}_2)$ is a rate triple such that $(\tilde{R}_0,\tilde{R}_1,\tilde{R}_2) \in \mathcal{R}(p)$ and $(\tilde{R}_0,\tilde{R}_1,\tilde{R}_2) \notin \mathcal{R}^*(p)$ due to

$$I(X_1;Y_1|U_0U_1U_2) + I(X_2U_1;Y_2|U_0U_2) < \tilde{R}_1, \tag{161}$$

for the same given joint distribution $p(\cdot) \in \mathcal{P}^*$. Since $(\tilde{R}_0,\tilde{R}_1,\tilde{R}_2) \in R(p)$, from (12), we have

$$\tilde{R}_1 \le I(X_1;Y_1|U_0U_2). \tag{162}$$

From (17) and (161), we obtain

$$\begin{aligned}\tilde{R}_2 &< I(X_2U_1;Y_2|U_0) - I(X_2U_1;Y_2|U_0U_2)\\ &= I(U_2X_2U_1;Y_2|U_0) - I(X_2U_1;Y_2|U_0U_2)\\ &= I(U_2;Y_2|U_0)\\ &\le I(U_2X_2;Y_2|U_0)\\ &= I(X_2;Y_2|U_0).\end{aligned} \tag{163}$$

From (16) and (161), we have

$$\begin{aligned}\tilde{R}_2 &< I(X_1U_2;Y_1|U_0U_1) - I(X_1;Y_1|U_0U_1U_2)\\ &= I(U_2;Y_1|U_0U_1)\\ &\le I(X_1U_2;Y_1|U_0U_1) + I(X_2;Y_2|U_0U_2).\end{aligned} \tag{164}$$

From (14), we immediately have

$$\tilde{R}_0 + \tilde{R}_1 \le I(U_0U_2X_1;Y_1). \tag{165}$$

From (18) and (161), we obtain

$$\begin{aligned}\tilde{R}_0 + \tilde{R}_2 &< I(U_0X_1U_2;Y_1) - I(X_2U_1;Y_2|U_0U_2)\\ &= I(U_0U_1X_1U_2;Y_1) - I(X_2U_1;Y_2|U_0U_2)\\ &= I(U_0U_2;Y_1)\\ &\le I(U_0U_2X_2;Y_1)\\ &= I(U_0X_2;Y_1).\end{aligned} \tag{166}$$

From (21) and (161), we obtain

$$\begin{aligned}\tilde{R}_1 + \tilde{R}_2 &< I(X_1U_2;Y_1|U_0)\\ &\le I(X_1U_2;Y_1|U_0) + I(X_2;Y_2|U_0U_2).\end{aligned} \tag{167}$$

Similarly, from (22) and (161), we have

$$\begin{aligned}\tilde{R}_0 + \tilde{R}_1 + \tilde{R}_2 &< I(U_0X_1U_2;Y_1)\\ &\le I(U_0X_1U_2;Y_1) + I(X_2;Y_2|U_0U_2).\end{aligned} \tag{168}$$

Setting $\mathcal{U}_1 = \emptyset$ in (143)–(159), we obtain $\mathcal{R}^*(p_1)$ with (R_0,R_1,R_2) satisfying

$$\begin{aligned}R_1 &\le I(X_1;Y_1|U_0U_2)\\ R_2 &\le I(X_2;Y_2|U_0)\\ R_2 &\le I(X_2;Y_2|U_0U_2) + I(X_1U_2;Y_1|U_0)\\ R_0 + R_1 &\le I(U_0X_1U_2;Y_1)\\ R_0 + R_2 &\le I(U_0X_2;Y_2)\\ R_1 + R_2 &\le I(X_1U_2;Y_1|U_0) + I(X_2;Y_2|U_0U_2)\\ R_0 + R_1 + R_2 &\le I(U_0X_1U_2;Y_1) + I(X_2;Y_2|U_0U_2).\end{aligned}$$

Since the rate triple $(\tilde{R}_0,\tilde{R}_1,\tilde{R}_2)$ satisfies (162)–(168), we have $(\tilde{R}_0,\tilde{R}_1,\tilde{R}_2) \in \mathcal{R}^*(p_1)$.

Similarly, if $(\tilde{R}_0,\tilde{R}_1,\tilde{R}_2) \in \mathcal{R}(p)$ and $I(X_2;Y_2|U_0U_1U_2) + I(X_1U_2;Y_1|U_0U_1) < \tilde{R}_2$, i.e., $(\tilde{R}_0,\tilde{R}_1,\tilde{R}_2) \notin \mathcal{R}^*(p)$, then we have $(\tilde{R}_0,\tilde{R}_1,\tilde{R}_2) \in \mathcal{R}^*(p_2)$.

Hence, we have $\mathcal{R}(p) \subseteq \mathcal{R}^*(p) \cup \mathcal{R}^*(p_1) \cup \mathcal{R}^*(p_2)$ and $\bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}(p) \subseteq \bigcup_{p(\cdot)\in\mathcal{P}^*} \mathcal{R}^*(p)$. The equivalence is thus proven.

Appendix D
Proof of the Converse Part of Theorem 4

A. Nondeterministic Codes and Deterministic Codes

In this part, we show that for any nondeterministic (or stochastic) (M_0,M_1,M_2,n,P_e^*) code for the ICC, there exists a deterministic (M_0,M_1,M_2,n,P_e) code for the same channel such that $P_e \le P_e^*$, by applying the technique introduced in [33].

Assign each codeword $\mathbf{x}_i \in \mathcal{X}_i^n$ an index $\mu_i \in \mathcal{I}_{x_i} = \{1,2,\ldots,|\mathcal{X}_i|^n\}, i = 1,2,$ assign each channel output sequence $\mathbf{y}_i \in \mathcal{Y}_i^n$ an index $\nu_i \in \mathcal{I}_{y_i} = \{1,2,\ldots,|\mathcal{Y}_i|^n\}$, $i = 1, 2$, and assign each message pair (w_0,w_i) an index $\vartheta_i \in \mathcal{I}_{w_i} = \{1,2,\ldots,M_0M_i\}$, $i = 1, 2$.

Consider one nondeterministic (M_0,M_1,M_2,n,P_e^*) code with the encoders and the decoders being defined by the following probability matrices

$$\text{Encoder } i: P_{E_i}(\mu_i|\vartheta_i), \quad i = 1,2$$
$$\text{Decoder } i: P_{D_i}(\hat{\vartheta}_i|\nu_i), \quad i = 1,2.$$

We next show that there exist the following lists of random variables

$$A_{E_i}^{M_0 M_i} = (A_{E_i}(1), A_{E_i}(2), \ldots, A_{E_i}(M_0 M_i)), \quad i=1,2$$
$$A_{D_i}^{|\mathcal{Y}_i|^n} = (A_{D_i}(1), A_{D_i}(2), \ldots, A_{D_i}(|\mathcal{Y}_i|^n)), \quad i=1,2$$

such that the encoding and decoding functions of the nondeterministic $(M_0, M_1, M_2, n, P_e^*)$ code can be expressed as $\mu_i = f_i(\vartheta_i, A_{E_i}(\vartheta_i))$, $i=1,2$, and $\hat{\vartheta}_i = g_i(\nu_i, A_{D_i}(\nu_i))$, $i=1,2$.

Let all the elements of $A_{E_i}^{M_0 M_i}$ and $A_{D_i}^{|\mathcal{Y}_i|^n}$, $i=1,2$, be independent of each other and all other random variables, and uniformly distributed over the interval $[0,1)$.

With respect to encoder 1, for each $\vartheta_1 \in \mathcal{I}_{w_i}$ and each $m \in \mathcal{I}_{x_1}$ we define

$$B_{E_1}(\vartheta_1, m) = \sum_{j=1}^{m} p(j|\vartheta_1), \text{ and } B_{E_1}(\vartheta_1, 0) = 0.$$

Suppose that a message pair $(\tilde{w}_0, \tilde{w}_1)$ indexed by $\tilde{\vartheta}_1$ is to be encoded. We let $f_1(\cdot)$ output $\mu_1 = t$, if $A_{E_1}(\tilde{\vartheta}_1)$ falls into the interval $[B_{E_1}(\tilde{\vartheta}_1, t-1), B_{E_1}(\tilde{\vartheta}_1, t))$. Hence, we have

$$P\left(A_{E_1}(\tilde{\vartheta}_1) \in \left[B_{E_1}(\tilde{\vartheta}_1, t-1), B_{E_1}(\tilde{\vartheta}_1, t)\right)\right) = p(t|\tilde{\vartheta}_1)$$

which means that the constructed encoding function $f_1(\cdot)$ is equivalent to the original encoding probability matrix $P_{E_1}(\mu_1|\vartheta_1)$. Similar constructions can be done for encoder 2 and the two decoders.

We define the random variable $A := \left(A_{E_1}^{M_0 M_1}, A_{E_2}^{M_0 M_2}, A_{D_1}^{|\mathcal{Y}_1|^n}, A_{D_2}^{|\mathcal{Y}_2|^n}\right)$, which has a joint probability distribution $p(a)$ over a range $\mathcal{A}$.

The probabilities of error in decoding the given nondeterministic $(M_0, M_1, M_2, n, P_e^*)$ code can now be expressed as

$$\begin{aligned} P_{e,i} &= P((\hat{W}_0, \hat{W}_i) \neq (W_0, W_i)) \\ &= \int_{a \in \mathcal{A}} P((\hat{W}_0, \hat{W}_i) \neq (W_0, W_i)|a) da, \quad i=1,2. \end{aligned}$$

Therefore, there always exists some $a \in \mathcal{A}$ such that

$$P((\hat{W}_0, \hat{W}_i) \neq (W_0, W_i)|a) \leq \max\left\{P_{e,1}^{(n)}, P_{e,2}^{(n)}\right\} \quad i=1,2. \tag{169}$$

Let $P_e = \max\left\{P_{e,1}^{(n)}, P_{e,2}^{(n)}\right\}$. From (169), we have a deterministic (M_0, M_1, M_2, n, P_e) code. By the definition of the $(M_0, M_1, M_2, n, P_e^*)$ code, we have $\max\left\{P_{e,1}^{(n)}, P_{e,2}^{(n)}\right\} \leq P_e^*$, and thus the claim follows immediately.

B. The Converse Based on Deterministic Codes

Based on the discussion in the previous section, it suffices to show that for any deterministic $\left(2^{nR_0}, 2^{nR_1}, 2^{nR_2}, n, P_e\right)$ code with $P_e \to 0$, the rate triple (R_0, R_1, R_2) must satisfy (38)–(50) for some joint distribution $p(v_0)p(x_1|v_0)p(x_2|v_0)$, to establish the converse.

Consider a deterministic $\left(2^{nR_0}, 2^{nR_1}, 2^{nR_2}, n, P_e\right)$ code with $P_e \to 0$. Note that $P_e \to 0$ implies $P_{e,1}^{(n)} \to 0$ and $P_{e,2}^{(n)} \to 0$. Applying Fano-inequality [32] for decoder 1, we obtain

$$H(W_0 W_1|Y_1^n) \leq n(R_0+R_1)P_{e,1}^{(n)} + h\left(P_{e,1}^{(n)}\right) := n\epsilon_{1n},$$

where $h(\cdot)$ is the binary entropy function. Note that $\epsilon_{1n} \to 0$ as $P_{e,1}^{(n)} \to 0$. It easily follows that

$$H(W_1|Y_1^n W_0) \leq H(W_0 W_1|Y_1^n) \leq n\epsilon_{1n}. \tag{170}$$

By symmetry, we also have

$$H(W_2|Y_2^n W_0) \leq H(W_0 W_2|Y_2^n) \leq n\epsilon_{2n}. \tag{171}$$

We now expand the entropy term $H(Y_1^n, V_2^n|W_0, W_1)$ as

$$\begin{aligned} H(Y_1^n V_2^n|W_0 W_1) &\overset{(a)}{=} H(Y_1^n V_2^n|X_1^n W_0 W_1) \\ &\overset{(b)}{=} H(V_2^n|X_1^n W_0 W_1) \\ &\quad + H(Y_1^n|V_2^n X_1^n W_0 W_1) \\ &\overset{(c)}{=} H(Y_1^n|X_1^n W_0 W_1) \\ &\quad + H(V_2^n|Y_1^n X_1^n W_0 W_1) \end{aligned}$$

where (a) follows from the fact that $X_1^n = f_1(W_0, W_1)$ is a deterministic function of W_0 and W_1 for a given $\left(2^{nR_0}, 2^{nR_1}, 2^{nR_2}, n, P_e\right)$ code, and both (b) and (c) are based on the chain rule. Since Y_1 is a deterministic function of X_1 and V_2, $H(Y_1^n|V_2^n X_1^n W_0 W_1) = 0$. Similarly, due to $V_2 = h_1(Y_1, X_1)$, we have $H(V_2^n|Y_1^n X_1^n W_0 W_1) = 0$. Hence, we obtain the following:

$$\begin{aligned} H(V_2^n|X_1^n W_0 W_1) &= H(Y_1^n|X_1^n W_0 W_1) \\ H(V_2^n|W_0 W_1) &\overset{\text{(a)}}{=} H(Y_1^n|W_0 W_1) \\ H(V_2^n|W_0) &\overset{\text{(b)}}{=} H(Y_1^n|W_0 W_1) \end{aligned} \tag{172}$$

where (a) follows from the deterministic relation between X_1^n and (W_0, W_1), and (b) follows from the conditional independence between V_2^n and W_1 given W_0. Analogously, we have

$$H(V_1^n|W_0) = H(Y_2^n|W_0 W_2). \tag{173}$$

Before proceeding to the main part of the converse, we need show the following two inequalities:

$$I(W_1; Y_1^n|W_0) \leq I(W_1; Y_1^n V_1^n|V_2^n W_0) \tag{174}$$
$$I(W_2; Y_2^n|W_0) \leq I(W_2; Y_2^n V_2^n|V_1^n W_0). \tag{175}$$

Inequality (174) can be derived as follows:

$$\begin{aligned} I(W_1; Y_1^n|W_0) &= H(W_1|W_0) - H(W_1|Y_1^n W_0) \\ &\overset{\text{(a)}}{\leq} H(W_1|V_2^n W_0) - H(W_1|Y_1^n V_2^n W_0) \\ &\overset{\text{(b)}}{\leq} H(W_1|V_2^n W_0) - H(W_1|Y_1^n V_1^n V_2^n W_0) \\ &= I(W_1; Y_1^n V_1^n|V_2^n W_0) \end{aligned}$$

where (a) follows from the facts that $H(W_1|W_0) = H(W_1|V_2^n W_0)$, which is due to the conditional independence between W_1 and V_2^n given W_0, and "conditioning reduces entropy", i.e., $H(W_1|Y_1^n V_2^n W_0) \le H(W_1|Y_1^n W_0)$, and (b) follows from "conditioning reduces entropy" as well. Similarly, we can obtain (175).

We now prove each of (38)–(50) by using (170)– (175).

For (38), we have

$$\begin{aligned} nR_1 &= H(W_1) = H(W_1|W_0) \overset{(a)}{=} H(W_1|W_0 V_2^n) \\ &= I(W_1; Y_1^n|W_0 V_2^n) + H(W_1|Y_1^n W_0 V_2^n) \\ &\overset{(b)}{\le} H(Y_1^n|W_0 V_2^n) - H(Y_1^n|W_0 W_1 V_2^n) + n\epsilon_{1n} \\ &\overset{(c)}{=} H(Y_1^n|W_0 V_2^n) + n\epsilon_{1n} \\ &\le \sum_{i=1}^n H(Y_{1i}|V_{2i} W_0) + n\epsilon_{1n} \end{aligned} \tag{176}$$

where (a) follows from the fact that W_1 and V_2^n are conditionally independent given W_0, (b) follows from $H(W_1|Y_1^n W_0 V_2^n) \le H(W_1|Y_1^n W_0) \le n\epsilon_{1n}$, and (c) follows from $H(Y_1^n|W_0 W_1 V_2^n) = H(Y_1^n|X_1^n V_2^n W_0 W_1) = 0$.

Analogously, for (39) we have

$$nR_2 \le \sum_{i=1}^n H(Y_{2i}|V_{1i} W_0) + n\epsilon_{2n}. \tag{177}$$

Regarding (40), we have

$$\begin{aligned} n(R_0 + R_1) &= H(W_0 W_1) \\ &= I(W_0 W_1; Y_1^n) + I(W_0 W_1|Y_1^n) \\ &\overset{(a)}{\le} I(W_0 W_1; Y_1^n) + n\epsilon_{1n} \\ &\le H(Y_1^n) + n\epsilon_{1n} \\ &\le \sum_{i=1}^n H(Y_{1i}) + n\epsilon_{1n} \end{aligned} \tag{178}$$

where (a) follows from (170).

Similarly, for (41) we have

$$n(R_0 + R_2) \le \sum_{i=1}^n H(Y_{2i}) + n\epsilon_{2n}. \tag{179}$$

With respect to (42), we have

$$\begin{aligned} &n(R_1 + R_2) \\ &\quad = H(W_1) + H(W_2) \\ &\quad = H(W_1|W_0) + H(W_2|W_0) \\ &\quad = I(W_1; Y_1^n|W_0) + H(W_1|Y_1^n W_0) \\ &\qquad + I(W_2; Y_2^n|W_0) + H(W_2|Y_2^n W_0) \\ &\quad \overset{(a)}{\le} H(Y_1^n|W_0) - H(Y_1^n|W_0 W_1) + H(Y_2^n|W_0) \\ &\qquad - H(Y_2^n|W_0 W_2) + n(\epsilon_{1n} + \epsilon_{2n}) \\ &\quad \overset{(b)}{=} H(Y_1^n|W_0) - H(V_2^n|W_0) + H(Y_2^n|W_0) \\ &\qquad - H(V_1^n|W_0) + n(\epsilon_{1n} + \epsilon_{2n}) \\ &\quad \le H(Y_1^n V_1^n|W_0) - H(V_1^n|W_0) + H(Y_2^n V_2^n|W_0) \\ &\qquad - H(V_2^n|W_0) + n(\epsilon_{1n} + \epsilon_{2n}) \\ &\quad = H(Y_1^n|V_1^n W_0) + H(Y_2^n|V_2^n W_0) + n(\epsilon_{1n} + \epsilon_{2n}) \\ &\quad \le \sum_{i=1}^n H(Y_{1i}|V_{1i} W_0) + \sum_{i=1}^n H(Y_{2i}|V_{2i} W_0) \\ &\qquad + n(\epsilon_{1n} + \epsilon_{2n}) \end{aligned} \tag{180}$$

where (a) follows from inequalities (170) and (171), and (b) follows from equalities (172) and (173).

Regarding (43), we have

$$\begin{aligned} n(R_1 + R_2) &= H(W_1|W_0) + H(W_2|W_0) \\ &\overset{(a)}{\le} I(W_1; Y_1^n|W_0) + I(W_2; Y_2^n|W_0) \\ &\quad + n(\epsilon_{1n} + \epsilon_{2n}) \\ &\overset{(b)}{\le} I(W_1; Y_1^n|W_0) + I(W_2; Y_2^n V_2^n|V_1^n W_0) \\ &\quad + n(\epsilon_{1n} + \epsilon_{2n}) \\ &= I(W_1; Y_1^n|W_0) + I(W_2; V_2^n|V_1^n W_0) \\ &\quad + I(W_2; Y_2^n|V_1^n V_2^n W_0) + n(\epsilon_{1n} + \epsilon_{2n}) \\ &\le H(Y_1^n|W_0) - H(Y_1^n|W_0 W_1) \\ &\quad + H(V_2^n|V_1^n W_0) \\ &\quad - H(V_2^n|V_1^n W_2 W_0) + H(Y_2^n|V_1^n V_2^n W_0) \\ &\quad - H(Y_2^n|V_1^n V_2^n W_2 W_0) \\ &\quad + n(\epsilon_{1n} + \epsilon_{2n}) \\ &\overset{(c)}{=} H(Y_1^n|W_0) + H(Y_2^n|V_1^n V_2^n W_0) \\ &\quad + n(\epsilon_{1n} + \epsilon_{2n}) \\ &\le \sum_{i=1}^n H(Y_{1i}|W_0) + \sum_{i=1}^n H(Y_{2i}|V_{1i} V_{2i} W_0) \\ &\quad + n(\epsilon_{1n} + \epsilon_{2n}) \end{aligned} \tag{181}$$

in which (a) follows from (170) and (171), (b) follows from (174), and (c) follows from $H(Y_1^n|W_0 W_1) = H(V_2^n|V_1^n W_0)$, $H(V_2^n|V_1^n W_2 W_0) = 0$ which is because V_2^n is determined by X_2^n and X_2^n is determined by (W_0, W_2), and $H(Y_2^n|V_1^n V_2^n W_2 W_0) = H(Y_2^n|X_2^n V_1^n V_2^n W_2 W_0) = 0$.

Similarly, we have

$$n(R_1 + R_2) \le \sum_{i=1}^n H(Y_{2i}|W_0) + \sum_{i=1}^n H(Y_{1i}|V_{1i} V_{2i} W_0) + n(\epsilon_{1n} + \epsilon_{2n}) \tag{182}$$

which corresponds to (45).

For (44), we obtain

$$\begin{aligned} &n(R_0 + R_1 + R_2) \\ &\quad = H(W_0 W_1) + H(W_2|W_0) \\ &\quad \overset{(a)}{\le} I(W_0 W_1; Y_1^n) + I(W_2; Y_2^n|W_0) + n(\epsilon_{1n} + \epsilon_{2n}) \\ &\quad \overset{(b)}{\le} I(W_0 W_1; Y_1^n) + I(W_2; Y_2^n V_2^n|V_1^n W_0) \\ &\qquad + n(\epsilon_{1n} + \epsilon_{2n}) \\ &\quad = I(W_0 W_1; Y_1^n) + I(W_2; V_2^n|V_1^n W_0) \\ &\qquad + I(W_2; Y_2^n|V_1^n V_2^n W_0) + n(\epsilon_{1n} + \epsilon_{2n}) \\ &\quad \le H(Y_1^n) - H(Y_1^n|W_0 W_1) + H(V_2^n|V_1^n W_0) \\ &\qquad - H(V_2^n|V_1^n W_2 W_0) + H(Y_2^n|V_1^n V_2^n W_0) \end{aligned}$$

$$
\begin{aligned}
&- H\left(Y_2^n|V_1^nV_2^nW_2W_0\right) + n(\epsilon_{1n}+\epsilon_{2n})\\
&\overset{(c)}{=} H\left(Y_1^n\right) + H\left(Y_2^n|V_1^nV_2^nW_0\right) + n(\epsilon_{1n}+\epsilon_{2n})\\
&\leq \sum_{i=1}^{n} H(Y_{1i}) + \sum_{i=1}^{n} H(Y_{2i}|V_{1i}V_{2i}W_0) + n(\epsilon_{1n}+\epsilon_{2n})
\end{aligned}
\tag{183}
$$

where (a), (b), and (c) follow from the same arguments for (181). Note that the proof for (183) and the one for (181) only differ in the first few steps, and the rest follows from the same set of arguments and procedures.

Instead of expressing $n(R_0+R_1+R_2)$ as $H(W_0W_1) + H(W_2|W_0)$, we set $n(R_0+R_1+R_2) = H(W_0|W_1) + H(W_0W_2)$. Following the similar steps used in deriving (183), we obtain

$$
n(R_0+R_1+R_2) \leq \sum_{i=1}^{n} H(Y_{2i}) + \sum_{i=1}^{n} H(Y_{1i}|V_{1i}V_{2i}W_0) + n(\epsilon_{1n}+\epsilon_{2n}) \tag{184}
$$

which corresponds to (46).

Now for (47), we have

$$
\begin{aligned}
&n(2R_1+R_2)\\
&= H(W_1|W_0) + H(W_1|W_0) + H(W_2|W_0)\\
&\overset{(a)}{\leq} I\left(W_1;Y_1^n|W_0\right) + I\left(W_1;Y_1^n|W_0\right) + I\left(W_2;Y_2^n|W_0\right)\\
&\quad + n(2\epsilon_{1n}+\epsilon_{2n})\\
&\overset{(b)}{\leq} I\left(W_1;Y_1^n|W_0\right) + I\left(W_1;Y_1^nV_1^n|V_2^nW_0\right)\\
&\quad + I\left(W_2;Y_2^n|W_0\right) + n(2\epsilon_{1n}+\epsilon_{2n})\\
&= I\left(W_1;Y_1^n|W_0\right) + I\left(W_1;V_1^n|V_2^nW_0\right)\\
&\quad + I\left(W_1;Y_1^n|V_1^nV_2^nW_0\right) + I\left(W_2;Y_2^n|W_0\right)\\
&\quad + n(2\epsilon_{1n}+\epsilon_{2n})\\
&= H\left(Y_1^n|W_0\right) - H\left(Y_1^n|W_0W_1\right) + H\left(V_1^n|V_2^nW_0\right)\\
&\quad - H\left(V_1^n|V_2^nW_0W_1\right) + H\left(Y_1^n|V_1^nV_2^nW_0\right)\\
&\quad - H\left(Y_1^n|V_1^nV_2^nW_0W_1\right) + H\left(Y_2^n|W_0\right)\\
&\quad - H\left(Y_2^n|W_0W_2\right) + n(2\epsilon_{1n}+\epsilon_{2n})\\
&\overset{(c)}{=} H\left(Y_1^n|W_0\right) - H\left(Y_1^n|W_0W_1\right) + H\left(Y_1^n|V_1^nV_2^nW_0\right)\\
&\quad + H\left(Y_2^n|W_0\right) + n(2\epsilon_{1n}+\epsilon_{2n})\\
&\overset{(d)}{=} H\left(Y_1^n|W_0\right) - H\left(V_2^n|W_0\right) + H\left(Y_1^n|V_1^nV_2^nW_0\right)\\
&\quad + H\left(Y_2^n|W_0\right) + n(2\epsilon_{1n}+\epsilon_{2n})\\
&\leq H\left(Y_1^n|W_0\right) - H\left(V_2^n|W_0\right) + H\left(Y_1^n|V_1^nV_2^nW_0\right)\\
&\quad + H\left(Y_2^nV_2^n|W_0\right) + n(2\epsilon_{1n}+\epsilon_{2n})\\
&= H\left(Y_1^n|W_0\right) + H\left(Y_1^n|V_1^nV_2^nW_0\right) + H\left(Y_2^n|V_2^nW_0\right)\\
&\quad + n(2\epsilon_{1n}+\epsilon_{2n})\\
&\leq \sum_{i=1}^{n} H\left(Y_{1i}|W_0\right) + \sum_{i=1}^{n} H(Y_{1i}|V_{1i}V_{2i}W_0)\\
&\quad + \sum_{i=1}^{n} H(Y_{2i}|V_{2i}W_0) + n(2\epsilon_{1n}+\epsilon_{2n})
\end{aligned}
\tag{185}
$$

where (a) follows from (170) and (171), (b) follows from (174), (c) follows from the facts that $H\left(V_1^n|V_2^nW_0\right) = H\left(V_1^n|W_0\right) = H\left(Y_2^n|W_0W_2\right)$, $H\left(V_1^n|V_2^nW_0W_1\right) = H\left(V_1^n|X_1^nV_2^nW_0W_1\right) = 0$, and $H\left(Y_1^n|V_1^nV_2^nW_0W_1\right) = H\left(Y_1^n|V_1^nX_1^nV_2^nW_0W_1\right) = 0$, and (d) follows from $H\left(V_2^n|W_0\right) = H\left(Y_1^n|W_0W_1\right)$. Following similar procedures, we obtain

$$
\begin{aligned}
&n(R_1+2R_2)\\
&\quad\leq \sum_{i=1}^{n} H(Y_{2i}|W_0) + \sum_{i=1}^{n} H(Y_{2i}|V_{1i}V_{2i}W_0)\\
&\quad\quad + \sum_{i=1}^{n} H(Y_{1i}|V_{1i}W_0) + n(\epsilon_{1n}+2\epsilon_{2n})
\end{aligned}
\tag{186}
$$

$$
\begin{aligned}
&n(R_0+2R_1+R_2)\\
&\quad\leq \sum_{i=1}^{n} H(Y_{1i}) + \sum_{i=1}^{n} H(Y_{1i}|V_{1i}V_{2i}W_0)\\
&\quad\quad + \sum_{i=1}^{n} H(Y_{2i}|V_{2i}W_0) + n(2\epsilon_{1n}+\epsilon_{2n})
\end{aligned}
\tag{187}
$$

$$
\begin{aligned}
&n(R_0+R_1+2R_2)\\
&\quad\leq \sum_{i=1}^{n} H(Y_{2i}) + \sum_{i=1}^{n} H(Y_{2i}|V_{1i}V_{2i}W_0)\\
&\quad\quad + \sum_{i=1}^{n} H(Y_{1i}|V_{1i}W_0) + n(\epsilon_{1n}+2\epsilon_{2n})
\end{aligned}
\tag{188}
$$

which correspond to (49)–(50) respectively.

We have derived a number of inequalities (176)– (188) which upper bound the rate triple (R_0, R_1, R_2) of a given code for the DICC channel. We now adopt the technique which was used to prove the converse of the capacity region of the MACC in [26] and [27]. Define $V_0 = W_0$, equivalently $p(v_{0i}) = p(w_0)$, i.e., V_0 or V_{0i} is an auxiliary random variable uniformly distributed over the common message set $\mathcal{W}_0 = \{1, \ldots, M_0\}$. Since X_1 and X_2 are conditionally independent given W_0, i.e., $p(x_{1i}, x_{2i}|w_0) = p(x_{1i}|w_0)p(x_{2i}|w_0)$, we can write $p(x_{1i}, x_{2i}|v_{0i}) = p(x_{1i}|v_{0i})p(x_{2i}|v_{0i})$. Due to the introduction of V_0, the region inherits the convexity from the achievable rate region for the general ICC. We now can conclude that the rate of the given code (R_0, R_1, R_2) is upper bounded by (38)–(50) for some choice of joint distribution $p(v_0)p(x_1|v_0)p(x_2|v_0)$. This completes the proof of the converse.

Acknowledgment

The authors wish to thank the Associate Editor Prof. Andrea Goldsmith and the anonymous reviewers for their many valuable comments and suggestions that greatly helped to improve the manuscript. The authors would also like to thank the anonymous reviewers for pointing out [23] and [28].

References

[1] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math. Stat. Prob.*, Berkeley, CA, 1961, vol. 1, pp. 611–644.

[2] R. Ahlswede, “The capacity region of a channel with two senders and two receivers,” *Ann. Prob.*, vol. 2, no. 5, pp. 805–814, 1974.

[3] A. B. Carleial, “A case where interference does not reduce capacity,” *IEEE Trans. Inf. Theory*, vol. IT-21, pp. 569–570, Sep. 1975.

[4] H. Sato, “Two-user communication channels,” *IEEE Trans. Inf. Theory*, vol. IT-23, pp. 295–304, May 1977.

[5] H. Sato, “On degraded Gaussian two-user channels,” *IEEE Trans. Inf. Theory*, vol. IT-24, pp. 637–640, May 1978.

[6] H. Sato, “On the capacity region of a discrete two-user channel for strong interference,” *IEEE Trans. Inf. Theory*, vol. IT-24, pp. 377–379, May 1978.

[7] A. B. Carleial, “Interference channels,” *IEEE Trans. Inf. Theory*, vol. IT-24, pp. 60–70, Jan. 1978.

[8] R. Benzel, “The capacity region of a class of discrete additive degraded interference channels,” *IEEE Trans. Inf. Theory*, vol. IT-25, pp. 228–231, Mar. 1979.

[9] T. S. Han and K. Kobayashi, “A new achievable rate region for the interference channel,” *IEEE Trans. Inf. Theory*, vol. IT-27, pp. 49–60, Jan. 1981.

[10] H. Sato, “The capacity of the Gaussian interference channel under strong interference,” *IEEE Trans. Inf. Theory*, vol. IT-27, pp. 786–788, Nov. 1981.

[11] A. A. E. Gamal and M. H. M. Costa, “The capacity region of a class of deterministic interference channels,” *IEEE Trans. Inf. Theory*, vol. IT-28, pp. 343–346, Mar. 1982.

[12] M. H. M. Costa and A. A. E. Gamal, “The capacity region of the discrete memoryless interference channel with strong interference,” *IEEE Trans. Inf. Theory*, vol. 35, pp. 710–711, Sep. 1987.

[13] I. Sason, “On achievable rate regions for the Gaussian interference channel,” *IEEE Trans. Inf. Theory*, vol. 50, pp. 1345–1356, Jun. 2004.

[14] G. Kramer, “Genie-aided outer bounds on the capacity of interference channels,” *IEEE Trans. Inf. Theory*, vol. 50, pp. 581–586, Mar. 2004.

[15] H. F. Chong, M. Motani, and H. K. Garg, “A comparison of two achievable rate regions for the interference channel,” in *Proc. ITA Workshop*, San Diego, CA, Feb. 2006.

[16] H. F. Chong, M. Motani, H. K. Garg, and H. E. Gamal, “On a simplification of the Han-Kobayashi rate region for the interference channel,” *IEEE Trans. Inf. Theory*, submitted for publication.

[17] T. M. Cover, “An achievable rate region for the broadcast channel,” *IEEE Trans. Inf. Theory*, vol. IT-21, pp. 399–404, Jul. 1975.

[18] N. Devroye, P. Mitran, and V. Tarokh, “Achievable rates in cognitive channels,” *IEEE Trans. Inf. Theory*, vol. 52, pp. 1813–1827, May 2006.

[19] A. Jovičić and P. Viswanath, Cognitive radio: An Information-Theoretic Perspective 2006 [Online]. Available: http://arxiv.org/abs/cs/0604107

[20] A. Arapostathis, On the Capacity of Interference Channels With Degraded Message Sets 2006 [Online]. Available: http://arxiv.org/abs/cs/0605071

[21] I. Maric, R. D. Yates, and G. Kramer, “The strong interference channel with unidirectional cooperation,” in *Proc. UCSD Workshop Inf. Theory and Its Appl.*, San Diego, CA, USA, Feb. 2006.

[22] I. Maric, R. D. Yates, and G. Kramer, “Capacity of interference channels with partial transmitter cooperation,” *IEEE Trans. Inf. Theory*, to be published.

[23] H. H. Tan, “Two-user interference channels with correlated information sources,” *Inf. Contr.*, vol. 44, no. 1, pp. 77–104, 1980.

[24] I. Maric, R. D. Yates, and G. Kramer, “The capacity region of the strong interference channel with common information,” in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Oct. 2005, pp. 1737–1741.

[25] J. Jiang, Y. Xin, and H. K. Garg, “An achievable rate region for interference channels with common information,” in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Oct. 2006, pp. 448–452.

[26] D. Slepian and J. K. Wolf, “A coding theorem for multiple access channels with correlated sources,” *Bell Syst. Tech. J.*, vol. 52, pp. 1037–1076, Sep. 1973.

[27] F. M. J. Willems, “Information Theoretical Results for the Discrete Memoryless Multiple Access Channel,” Ph.D. dissertation, Katholieke Universiteit Leuven, Leuven, Belgium, 1982.

[28] Y. Cao, B. Chen, and J. Zhang, “A new achievable rate region for interference channels with common information,” in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC’07)*, Hong Kong, Mar. 2007, pp. 2069–2073.

[29] K. Kobayashi and T. S. Han, “A further consideration on the HK and the CMG regions for the interference channel,” in *Proc. ITA Workshop*, San Diego, CA, Jan. 2007.

[30] G. Kramer, “Review of rate regions for interference channels,” in *Proc. IZS Workshop*, Zurich, Switzerland, Feb. 2006.

[31] I. Maric, R. D. Yates, and G. Kramer, “The strong interference channel with common information,” in *Proc. 43th Allerton Conf. Commun., Contr. Comput.*, Monticello, IL, Sep. 2005.

[32] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1994.

[33] F. M. J. Williems, “The discrete memoryless multiple-access channel with cribbing encoders,” *IEEE Trans. Inf. Theory*, vol. IT-31, pp. 313–327, May 1985.