Capacity of Two-Way Channels with Symmetry Properties

Jian-Jia Weng[†], Lin Song[‡], Fady Alajaji[†] and Tamás Linder[†]

Abstract

In this paper, we make use of channel symmetry properties to determine the capacity region of three types of two-way networks: (a) two-user memoryless two-way channels (TWCs), (b) two-user TWCs with memory, and (c) three-user multiaccess/degraded broadcast (MA/DB) TWCs. For each network, symmetry conditions under which a Shannon-type random coding inner bound (under independent non-adaptive inputs) is tight are given. For two-user memoryless TWCs, prior results are substantially generalized by viewing a TWC as two interacting state-dependent one-way channels. The capacity of symmetric TWCs with memory, whose outputs are functions of the inputs and independent stationary and ergodic noise processes, is also obtained. Moreover, various channel symmetry properties under which the Shannon-type inner bound is tight are identified for three-user MA/DB TWCs. The results not only enlarge the class of symmetric TWCs whose capacity region can be exactly determined but also imply that interactive adaptive coding, not improving capacity, is unnecessary for such channels.

Index Terms

Network information theory, two-way channels, capacity region, inner and outer bounds, channel symmetry, multiple access and broadcast channels, channels with memory, adaptive coding.

I. INTRODUCTION

Shannon's two-way channel (TWC) [3], which allows two users to exchange data streams in a full-duplex manner, is a basic component of communication systems. To mitigate the interference
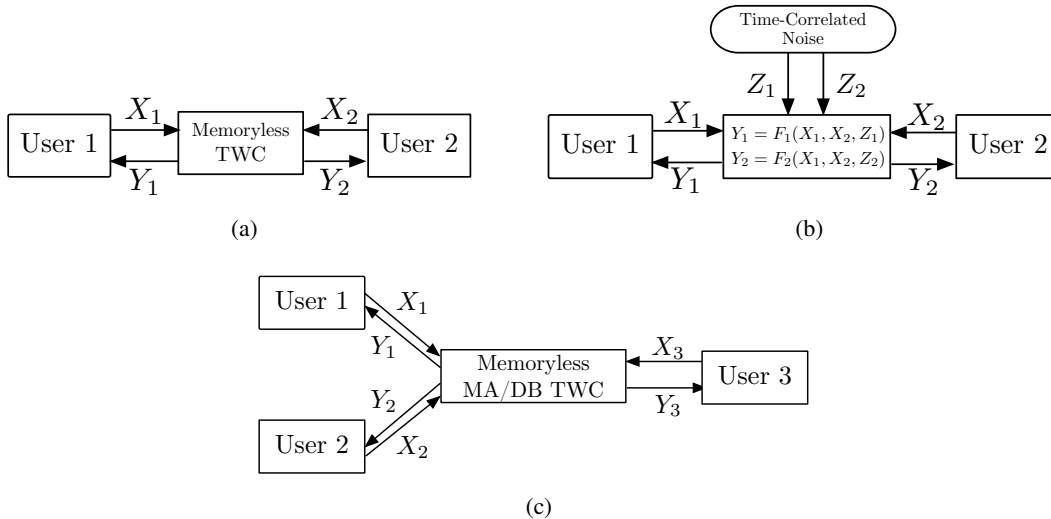
Fig. 1: Block diagrams of the two-way networks considered: (a) point-to-point memoryless TWC with two channel inputs X_1 and X_2 and two channel outputs Y_1 and Y_2; (b) point-to-point TWC with memory, where F_1 and F_2 are deterministic functions and (Z_1, Z_2) is a time-correlated channel noise pair generated from a joint stationary and ergodic process; (c) three-user memoryless MA/DB TWC, where X_i and Y_i respectively denote channel input and output at user j for $j = 1, 2, 3$.

incurred from two-way simultaneous transmission, TWCs are often used in conjunction with orthogonal multiplexing [4]. With increasing demands for fast data transmission, many industrial standards have enabled the use of non-orthogonal multiplexing to accommodate more users [5], [6]. From an information-theoretic viewpoint, the challenge is how each user can effectively maximize its individual transmission rate over the shared channel and concurrently provide sufficient feedback to help the other users' transmissions. These competing objectives impose on each user the challenging task of optimally adapting their inputs to the previously received signals of the other users. As finding such an optimal coding procedure is still elusive, the exact characterization of the capacity region of general TWCs remains open [7], [8, Section 17.5].

This paper revisits this open problem by finding larger classes of TWCs whose capacity region can be exactly obtained. Our approach is to identify channel symmetry properties under which a Shannon-type random coding inner bound (under independent non-adaptive inputs) is tight, thus directly determining the capacity region. As a result, we identify TWCs for which interactive

adaptive coding is useless in terms of improving the users' transmission rates. In particular, we focus on three two-way networks which we depict in Fig. 1. The two-user (point-to-point) memoryless TWC in Fig. 1(a) models device-to-device communication [9]. The simplified TWC with memory in Fig. 1(b), which is a generalization of additive-noise TWC in [1], can capture the effect of time-correlated channel noise which commonly arises in wireless communications. The three-user memoryless multiaccess/degraded broadcasting (MA/DB) TWC [24] in Fig. 1(c) models the communication between two mobile users and one base station, where the shared channel in the users-to-base-station (uplink) direction acts as a multiple-access channel (MAC) while the reverse (downlink) direction acts as a degraded broadcast channel (DBC). For these networks, we derive conditions under which the Shannon-type inner bound is optimal in terms of achieving channel capacity. Such a result also has a practical significance since communication without adaptive coding simplifies system design.

A. Capacity Bounds for TWCs

We briefly review some general results on the capacity of TWCs. In [3], Shannon derived inner and outer capacity bounds in the form of a single-letter expression for two-user memoryless TWCs. The inner bound is obtained via random coding where the users' channel inputs are independent (and non-adaptive), while the inputs are allowed to have arbitrary correlation in the outer bound. In general, the two bounds do not coincide. Follow-up work in [10]-[13] was devoted to improving Shannon's inner bound by using adaptive coding. Two novel outer bounds [14], [15], which restrict the dependency among channel inputs, were proposed to refine Shannon's result. Moreover, methods to efficiently utilize TWCs were investigated by studying the role of feedback [16]. In [17], directed mutual information [18], which is widely used in the study of one-way channels with feedback [19]-[23], was used to characterize the capacity of TWCs, but the obtained multi-letter expressions are often not computable. Recently, the Shannon-type random coding scheme was shown to be optimal in several deterministic multi-user TWC settings [24] such as MA/BC, Z, and interference TWCs, hence finding the channel capacity in these cases. The channel capacity for a variant of these multi-user TWCs, called three-way channels, was also investigated in different network setups such as three-way multi-cast finite-field or phase-fading Gaussian channels [26] and three-way Gaussian channels with multiple unicast sessions [27]. An additional capacity result for deterministic interference TWCs was derived in

[25]. For TWCs with memory, Shannon provided a multi-letter capacity characterization in [3, Section 16] which in general is incalculable.

B. Related Work

Channel symmetry properties, which are extensively investigated to simplify the computation of the capacity of one-way channels, play a key role in determining the capacity region for TWCs. The first channel symmetry property for TWCs was proposed by Shannon [3, Section 12]. Let $[P_{Y_1,Y_2|X_1,X_2}(\cdot,\cdot|\cdot,\cdot)]$ denote the channel transition matrix of a two-user discrete memoryless TWC, where X_j and Y_j denote the channel input and output at user j, respectively. Shannon gave two permutation invariance conditions on $[P_{Y_1,Y_2|X_1,X_2}(\cdot,\cdot|\cdot,\cdot)]$ which guarantee the equality of his inner and outer bounds (see Propositions 1 and 2 in Section II for details). A recent work [28] by Chaaban, Varshney, and Alouini (CVA) presented another tightness condition, where the channel symmetry property is given in terms of conditional entropies for the marginal channel distribution $[P_{Y_j|X_1,X_2}(\cdot|\cdot,\cdot)]$ (see Proposition 3).

The above conditions delineate classes of two-user memoryless TWCs for which Shannon's capacity inner bound is tight, hence exactly yielding their capacity region. Examples include Gaussian TWCs [13], q-ary additive-noise TWCs [1], and more general channel models such as injective semi-deterministic TWCs (ISD-TWCs) [28], Cauchy [28] and exponential family type TWCs [29]. It is worth mentioning that Hekstra and Willems [15] also presented a condition under which Shannon's inner bound is tight. However, their result is only valid for single-output memoryless TWCs.

For three-user MA/BC memoryless TWCs, Cheng and Devroye [24] investigated a class of symmetric TWCs. In particular, they considered deterministic, invertible, and alphabet-restricted MA/BC TWCs, proving that the Shannon-type inner bound is tight for that class of channels. However, to the best of our knowledge, symmetry properties for TWCs beyond these have not been investigated. It is also important to point out that two-user TWCs with memory are not well understood either.

C. A Motivational Example and Proposed Approach

Consider a point-to-point binary-input and binary-output memoryless TWC with transition probability matrix (see Section II-B for the formal description of the channel model)

$$[P_{Y_1,Y_2|X_1,X_2}(\cdot,\cdot|\cdot,\cdot)] = \begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \begin{array}{cccc} 00 & 01 & 10 & 11 \\ \begin{pmatrix} 0.783 & 0.087 & 0.117 & 0.013 \\ 0.0417 & 0.3753 & 0.0583 & 0.5247 \\ 0.261 & 0.609 & 0.039 & 0.091 \\ 0.2919 & 0.1251 & 0.4081 & 0.1749 \end{pmatrix} \end{array},$$

where the rows and columns are indexed by the channel inputs and outputs, respectively. The corresponding marginal channel transition matrices are

$$[P_{Y_2|X_1,X_2}(\cdot|\cdot,0)] = \begin{pmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{pmatrix}, [P_{Y_1|X_1,X_2}(\cdot|0,\cdot)] = \begin{pmatrix} 0.87 & 0.13 \\ 0.417 & 0.583 \end{pmatrix},$$

$$[P_{Y_2|X_1,X_2}(\cdot|\cdot,1)] = \begin{pmatrix} 0.1 & 0.9 \\ 0.7 & 0.3 \end{pmatrix}, [P_{Y_1|X_1,X_2}(\cdot|1,\cdot)] = \begin{pmatrix} 0.87 & 0.13 \\ 0.417 & 0.583 \end{pmatrix}.$$

A thorough examination reveals that for this TWC Shannon's inner bound is actually exact due to the symmetric structures of the channel's marginal transition matrices. However, none of the previously proposed symmetry conditions in the literature are satisfied.

We address this problem by viewing a TWC as two state-dependent one-way channels [3], [30]. Taking the two-user setting as an example, the state-dependent one-way channel from users 1 to 2 has input X_1, output Y_2, state X_2, and transition matrix given by $[P_{Y_2|X_1,X_2}(\cdot|\cdot,\cdot)]$; similarly, the one-way channel $[P_{Y_1|X_1,X_2}(\cdot|\cdot,\cdot)]$ in the reverse direction has input X_2, output Y_1, and channel state X_1. Note that this viewpoint[1] may also be useful for all previously mentioned two-way networks. Another useful tool is the rich set of symmetry concepts for single-user one-way channels.[2] From this perspective, the two one-way channels now interact with each other through the channel states. Clearly, this interaction could improve bi-directional transmission rates by making use of adaptive coding.

[1]Another viewpoint for two-user TWCs is based on compound MACs, see [31, Problem 14.11] and [32].

[2]Channel symmetry properties for single-user one-way memoryless channels can be roughly classified into two types. One type focuses on the structure of the channel transition probability such as Gallager symmetric channels [33], weakly symmetric and symmetric channels [34], and quasi-symmetric channels [35]. The other type aims at the invariance of information quantities including T-symmetric channels [36] and channels with input-invariance symmetry [37].

Our approach is to study symmetry properties for state-dependent one-way channels that imply that the capacity cannot be increased with the availability of channel state information at the transmitter (in addition to the receiver). Such properties can potentially render interactive adaptive coding useless in terms of enlarging TWC capacity. In the two-user memoryless setting, we develop the following two important channel symmetry notions. The *common optimal input distribution* condition identifies a state-dependent one-way channel that has an identical capacity-achieving input distribution for all channel states. The *invariance of input-output mutual information* condition then identifies a state-dependent one-way channel that produces the same input-output mutual information for all channel states under any fixed input distribution. If a TWC satisfies both conditions, one for each direction of the two-way transmission, the optimal transmission scheme of one user is irrelevant to the other user's transmission scheme, implying that the interaction between the users does not increase their transmission rates and hence channel capacity. In fact, the preceding motivational example illustrates this. More formally, we can prove that under certain symmetry properties (identified by the derived conditions), any rate pair inside Shannon's outer bound region is always contained in the inner bound region, implying that the latter bound is tight.

Furthermore, it should be expected that validating generalized channel symmetry properties can be a very complex procedure. However, we show that such a verification can be greatly simplified for some TWCs. For instance, the channel transition matrices $[P_{Y_1|X_1,X_2}(\cdot|\cdot,0)]$ and $[P_{Y_1|X_1,X_2}(\cdot|\cdot,1)]$ in the above example are column permutations of each other and the matrices $[P_{Y_1|X_1,X_2}(\cdot|0,\cdot)]$ and $[P_{Y_1|X_1,X_2}(\cdot|1,\cdot)]$ are identical. It turns out (as we will see later) that these two symmetry properties imply that Shannon's inner bound is tight. Therefore, we not only seek general conditions but also look for conditions which are simple to verify.

D. Summary of Contributions

Most of the conditions that we establish in this paper comprise two parts, one for each direction of the two-way transmission. Our contributions are summarized as follows.

• **Point-to-Point Memoryless TWCs:** six sufficient conditions (Theorems 1-4 and Corollaries 1-2) guaranteeing that Shannon's inner and outer bounds coincide are derived. Three of these are shown to be substantial generalizations of the Shannon and CVA conditions (in Theorems 5-7); our simplest condition can be verified by only observing the channel marginal distributions. Moreover, the capacity region of q-ary additive-noise TWCs with erasures, which subsume
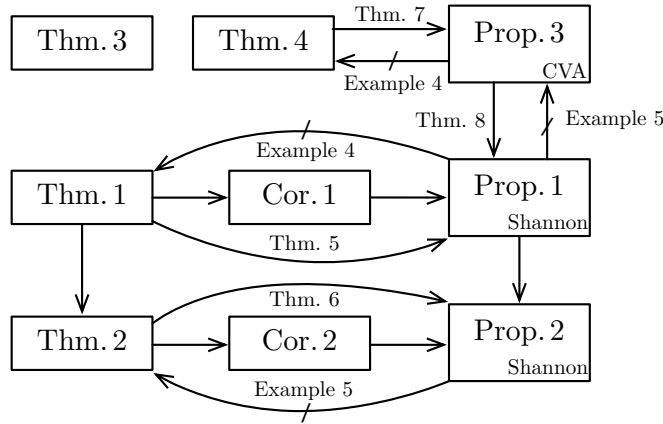
Fig. 2: The relationships between the results yielding the equality of Shannon's capacity bounds in point-to-point memoryless TWCs. Here, $A \to B$ indicates that result A subsumes result B, and $B \not\to A$ indicates that result B does not subsume result A. For example, Prop. 3 $\to$ Prop. 1 and Prop. 1 $\not\to$ Prop. 3 mean that the CVA result in Prop. 3 is more general than the Shannon result in Prop. 1.
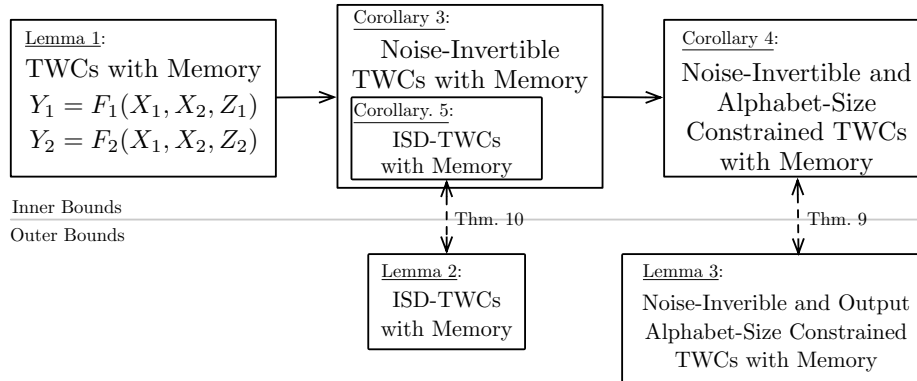

Fig. 3: The relationships between the results for point-to-point TWCs with memory. Here, $A \xleftarrow{\text{Thm. } C} \to B$ indicates that results A and B are combined in Theorem C to determine the capacity region.

several classical TWCs, is fully characterized by our conditions. Several examples illustrating the difference between these conditions are provided. We also refine Shannon's result to show that the CVA condition is a strict generalization of the Shannon condition (Theorem 8), thus answering a question raised in [28]. Implications among our results (and prior results) are depicted in Fig. 2.

- **Point-to-Point TWCs with Memory:** a Shannon-type inner bound and an outer bound for

the capacity of TWCs with memory under certain invertibility and alphabet size constraints are derived (Lemmas 1-2 and Corollaries 3-5). Two sufficient conditions for the tightness of the bounds are given (Theorems 9 and 10). The first condition is derived for TWCs with strict invertibility and alphabet size constraints, characterizing channel capacity in single-letter form. The other condition is specialized for injective semi-deterministic TWCs with memory.[3] The obtained results are related as shown in Fig. 3. We also illustrate via a simple example that when the channel's memory is strong, the Shannon-type random coding scheme does not achieve capacity and adaptive coding is useful.

• **Three-User Memoryless MA/DB TWCs**: we establish a Shannon-type inner bound and an outer bound for the capacity region of MA/DB TWCs (Theorems 11 and 12) where both bounds admit a common rate expression but have different input distribution requirements. Three sufficient conditions (based on different techniques) for these bounds to coincide are established (Theorems 13-15). The first condition involves the existence of independent inputs that can achieve the outer bound (similar to the CVA approach). The second condition is derived from the viewpoint of two interacting state-dependent one-way channels. The last one focuses on the permutation invariance structure of the channel transition matrix (mirroring the Shannon symmetry method). The obtained results extend the results in [24] and readily provide the capacity region for a larger class of MA/DB TWCs. While the channel model here is admittedly simplified, we note that our intention is to illustrate a potential methodology for determining the capacity regions of multi-user two-way channels and to motivate future work in this area.

The rest of the paper is organized as follows. In Section II, point-to-point memoryless TWCs are investigated. TWCs with memory are studied in Section III, and memoryless MA/DB TWCs are examined in Section IV. Concluding remarks are given in Section V.

II. POINT-TO-POINT MEMORYLESS TWCS

In this section, we study two-user memoryless two-way networks. We first formally describe the general model for point-to-point TWCs (not necessarily memoryless) in Section II-A, and then review the prior results for the memoryless case in Section II-B. New symmetry conditions are derived in Section II-C, and we demonstrate how to apply these conditions to finding the

---

[3]ISD-TWC model with memoryless noise were introduced in [28]. Here, we merely extend this setting by allowing noise processes with memory.
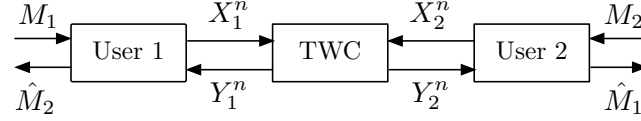
Fig. 4: The information flow of point-to-point two-way transmission.

channel capacity in Section II-D. Comparisons between prior results and our conditions are also presented in Section II-E, and the relationship between Shannon's condition and the CVA condition is examined in Section II-F.

A. General Channel Model

In point-to-point two-way communication as shown in Fig. 4, two users exchange messages M_1 and M_2 via n channel uses. Here, M_1 and M_2 are assumed to be independent and uniformly distributed on the finite sets $\mathcal{M}_1 \triangleq \{1, 2, ..., 2^{nR_1}\}$ and $\mathcal{M}_2 \triangleq \{1, 2, ..., 2^{nR_2}\}$, respectively, for some $R_1, R_2 \geq 0$. Let $\mathcal{X}_j$ and $\mathcal{Y}_j$ be the channel input and output alphabets, respectively for $j = 1, 2$. For $i = 1, 2, \ldots, n$, let $X_{j,i} \in \mathcal{X}_j$ and $Y_{j,i} \in \mathcal{Y}_j$ denote the channel input and output of user j at time i, respectively. The joint probability distribution of all random variables for the entire transmission period is given by

$$P_{M_1,M_2,X_1^n,X_2^n,Y_1^n,Y_2^n} = P_{M_1} P_{M_2} \cdot \left(\prod_{i=1}^{n} P_{X_{1,i}|M_1,Y_1^{i-1}} \right) \cdot$$

$$\left(\prod_{i=1}^{n} P_{X_{2,i}|M_2,Y_2^{i-1}} \right) \cdot \left(\prod_{i=1}^{n} P_{Y_{1,i},Y_{2,i}|X_1^i,X_2^i,Y_1^{i-1},Y_2^{i-1}} \right),$$

where $X_j^i \triangleq (X_{j,1}, X_{j,2}, \ldots, X_{j,i})$ and $Y_j^i \triangleq (Y_{j,1}, Y_{j,2}, \ldots, Y_{j,i})$ for $j = 1, 2$. The n transmissions over a point-to-point TWC can be then described by the sequence of conditional probabilities $\{P_{Y_{1,i},Y_{2,i}|X_1^i,X_2^i,Y_1^{i-1},Y_2^{i-1}}\}_{i=1}^{n}$.

Definition 1: An (n, R_1, R_2) code for a TWC consists of two message sets $\mathcal{M}_1 = \{1, 2, \ldots, 2^{nR_1}\}$ and $\mathcal{M}_2 = \{1, 2, \ldots, 2^{nR_2}\}$, two sequences of encoding functions $f_1^n \triangleq (f_{1,1}, f_{1,2}, \ldots, f_{1,n})$ and $f_2^n \triangleq (f_{2,1}, f_{2,2}, \ldots, f_{2,n})$ such that

$$X_{1,1} = f_{1,1}(M_1), \quad X_{1,i} = f_{1,i}(M_1, Y_1^{i-1})$$
$$X_{2,1} = f_{2,1}(M_2), \quad X_{2,i} = f_{2,i}(M_2, Y_2^{i-1})$$

for $i = 2, 3, \ldots, n$, and two decoding functions g_1 and g_2 such that $\hat{M}_2 = g_1(M_1, Y_1^n)$ and $\hat{M}_1 = g_2(M_2, Y_2^n)$.

When messages M_1 and M_2 are encoded via an (n, R_1, R_2) channel code, the probability of decoding error is defined as $P_{\mathrm{e}}^{(n)}(f_1^n, f_2^n, g_1, g_2) = \Pr\{\hat{M}_1 \neq M_1 \text{ or } \hat{M}_2 \neq M_2\}$.

Definition 2: A rate pair (R_1, R_2) is said to be achievable if there exists a sequence of (n, R_1, R_2) codes with $\lim_{n \to \infty} P_{\mathrm{e}}^{(n)} = 0$.

Definition 3: The capacity region $\mathcal{C}$ of a point-to-point TWC is defined as the closure of the convex hull of all achievable rate pairs.

B. Prior Results for Memoryless TWCs

A point-to-point TWC is said to be *memoryless* if its transition probabilities satisfy

$$P_{Y_{1,i}, Y_{2,i} | X_1^i, X_2^i, Y_1^{i-1}, Y_2^{i-1}} = P_{Y_1, Y_2 | X_1, X_2}$$

for all $i \geq 1$. For a memoryless TWC with transition probability $P_{Y_1, Y_2 | X_1, X_2}$ and input distribution P_{X_1, X_2}, let $\mathcal{R}(P_{X_1, X_2}, P_{Y_1, Y_2 | X_1, X_2})$ denote the set of all rate pairs (R_1, R_2) constrained by

$$R_1 \leq I(X_1; Y_2 | X_2) \text{ and } R_2 \leq I(X_2; Y_1 | X_1). \tag{1}$$

In [3], Shannon showed that the capacity region of a discrete memoryless point-to-point TWC is inner bounded by

$$\mathcal{C}_{\mathrm{I}}(P_{Y_1, Y_2 | X_1, X_2}) \triangleq \overline{\mathrm{co}} \left(\bigcup_{P_{X_1} P_{X_2}} \mathcal{R}(P_{X_1} P_{X_2}, P_{Y_1, Y_2 | X_1, X_2}) \right),$$

and outer bounded by

$$\mathcal{C}_{\mathrm{O}}(P_{Y_1, Y_2 | X_1, X_2}) \triangleq \overline{\mathrm{co}} \left(\bigcup_{P_{X_1, X_2}} \mathcal{R}(P_{X_1, X_2}, P_{Y_1, Y_2 | X_1, X_2}) \right),$$

where $\overline{\mathrm{co}}(\cdot)$ denotes taking the closure of the convex hull. In general, $\mathcal{C}_{\mathrm{I}}$ and $\mathcal{C}_{\mathrm{O}}$ are not matched to each other, but if they coincide, then the exact capacity region is obtained. Our objective is to develop general conditions under which the two bounds coincide.

In the following, the Shannon [3] and CVA [28] conditions that imply the equality of $\mathcal{C}_{\mathrm{I}}$ and $\mathcal{C}_{\mathrm{O}}$ are summarized. In short, the Shannon condition focuses on the permutation invariance structure of the channel transition matrix $[P_{Y_1, Y_2 | X_1, X_2}(\cdot, \cdot | \cdot, \cdot)]$, while the CVA condition involves the existence of independent inputs which can achieve the outer bound. Throughout the paper, we use $I^{(l)}(X_k; Y_j | X_j)$ and $H^{(l)}(Y_j | X_1, X_2)$ to denote the conditional mutual information and the conditional entropy evaluated under input distribution $P_{X_1, X_2}^{(l)}$ for $j, k = 1, 2$ with $j \neq k$. For $P_{X_1, X_2}^{(l)} = P_{X_j}^{(l)} P_{X_k | X_j}^{(l)}$ with $j \neq k$, the conditional entropy $H^{(l)}(Y_j | X_j)$ is evaluated using the

marginal distribution $P^{(l)}_{Y_j|X_j}(y_j|x_j) = \sum_{x_k} P^{(l)}_{X_k|X_j}(x_k|x_j)P_{Y_j|X_j,X_k}(y_j|x_j,x_k)$. Also, for a finite set $\mathcal{A}$, let $\pi^{\mathcal{A}} : \mathcal{A} \to \mathcal{A}$ denote a permutation (bijection), and for any two symbols a' and a'' in $\mathcal{A}$, let $\tau^{\mathcal{A}}_{a',a''} : \mathcal{A} \to \mathcal{A}$ denote the transposition which swaps a' and a'' in $\mathcal{A}$, but leaves the other symbols unaffected. Finally, let $\mathcal{P}(\mathcal{X}_j)$ denote the set of all probability distributions on $\mathcal{X}_j$, and define $P^{\mathrm{U}}_{\mathcal{X}_j}$ as the uniform probability distribution on $\mathcal{X}_j$ for $j = 1, 2$.

Proposition 1 (Shannon's One-Sided Symmetry Condition [3]): For a memoryless TWC with transition probability $P_{Y_1,Y_2|X_1,X_2}$, we have that $\mathcal{C} = \mathcal{C}_{\mathrm{I}} = \mathcal{C}_{\mathrm{O}}$ if for any pair of distinct input symbols $x'_1, x''_1 \in \mathcal{X}_1$, there exists a pair of permutations $(\pi^{\mathcal{Y}_1}[x'_1, x''_1], \pi^{\mathcal{Y}_2}[x'_1, x''_1])$ on $\mathcal{Y}_1$ and $\mathcal{Y}_2$, respectively, (which depend on x'_1 and x''_1) such that for all x_1, x_2, y_1, y_2,

$$P_{Y_1,Y_2|X_1,X_2}(y_1, y_2|x_1, x_2) = P_{Y_1,Y_2|X_1,X_2}(\pi^{\mathcal{Y}_1}[x'_1, x''_1](y_1), \pi^{\mathcal{Y}_2}[x'_1, x''_1](y_2)|\tau^{\mathcal{X}_1}_{x'_1,x''_1}(x_1), x_2). \quad (2)$$

Under this condition, the capacity region is given by

$$\mathcal{C} = \overline{\mathrm{co}} \left(\bigcup_{P_{X_2}} \mathcal{R}\left(P^{\mathrm{U}}_{\mathcal{X}_1} P_{X_2}, P_{Y_1,Y_2|X_1,X_2} \right) \right). \quad (3)$$

In [3], the proof of Proposition 1 is only sketched. To make the paper self-contained and facilitate the understanding of a technique used to derive one of our results (Theorem 15), we provide a full proof in Appendix A. Note that Proposition 1 describes a channel symmetry property with respect to the channel input of user 1, but an analogous condition can be obtained by exchanging the roles of users 1 and 2. The proposition below immediately follows from Proposition 1.

Proposition 2 (Shannon's Two-Sided Symmetry Condition [3]): For a memoryless TWC with transition probability $P_{Y_1,Y_2|X_1,X_2}$, we have that $\mathcal{C} = \mathcal{C}_{\mathrm{I}} = \mathcal{C}_{\mathrm{O}}$ if the TWC satisfies the one-sided symmetry condition with respect to both channel inputs. In this case, the capacity region is rectangular and given by $\mathcal{C} = \mathcal{R}(P^{\mathrm{U}}_{\mathcal{X}_1} P^{\mathrm{U}}_{\mathcal{X}_2}, P_{Y_1,Y_2|X_1,X_2})$.

Proposition 3 (CVA Condition [28]): For a memoryless TWC with transition probability $P_{Y_1,Y_2|X_1,X_2}$, we have that $\mathcal{C} = \mathcal{C}_{\mathrm{I}} = \mathcal{C}_{\mathrm{O}}$ if $H(Y_j|X_1, X_2)$, $j = 1, 2$, does not depend on $P_{X_1|X_2}$ for any fixed P_{X_2} and $P_{Y_j|X_1,X_2}$, and for any $P^{(1)}_{X_1,X_2} = P^{(1)}_{X_2} P^{(1)}_{X_1|X_2}$ there exists $\tilde{P}_{X_1} \in \mathcal{P}(\mathcal{X}_1)$ such that $H^{(1)}(Y_j|X_j) \leq H^{(2)}(Y_j|X_j)$ for $j = 1, 2$, where $P^{(2)}_{X_1,X_2} = \tilde{P}_{X_1} P^{(1)}_{X_2}$.

Thus, if a TWC satisfies any one of the above conditions, the capacity region can be determined by considering independent inputs: $P_{X_1,X_2} = P_{X_1} P_{X_2}$. This result implies that adaptive coding, where channel inputs are generated by interactively adapting to the previously received signals, cannot improve the users' achievable rates and that Shannon's random coding scheme is optimal.

The class of memoryless ISD-TWCs [28] satisfies the CVA condition (but do not necessarily satisfy the Shannon condition) and hence adaptive coding is useless for such channels. A TWC with independent q-ary additive noise [1] is an example of a channel that satisfies both the Shannon and CVA conditions. Although the CVA condition does not require any permutation invariance on the channel marginal distribution $P_{Y_j|X_1,X_2}$, the invariance requirement of $H(Y_j|X_1, X_2)$'s in Proposition 3 does in fact impose a certain symmetry constraint on $P_{Y_j|X_1,X_2}$. More details about these conditions will be provided in the proof of Theorem 7 and Section II-F.

C. Conditions for the Tightness of Shannon's Inner and Outer Bounds

In this section, we present conditions that guarantee the tightness of Shannon's inner bound by considering a TWC as two interacting state-dependent one-way channels. For example, the state-dependent one-way channel from user 1 to user 2 is governed by the marginal distribution $P_{Y_2|X_1,X_2}$ (derived from the channel probability $P_{Y_1,Y_2|X_1,X_2}$), where X_1 and Y_2 are respectively the input and the output of the channel with state X_2.

Let P_X and $P_{Y|X}$ be probability distributions on $\mathcal{X}$ and $\mathcal{Y}$, respectively. To simplify the presentation, we use

$$\mathcal{I}(P_X, P_{Y|X}) = \sum_{x,y} P_X(x) P_{Y|X}(y|x) \log \frac{P_{Y|X}(y|x)}{\sum_{x'} P_X(x') P_{Y|X}(y|x')},$$

as an alternative way of writing the mutual information $I(X; Y)$ between input X (governed by P_X) and corresponding output Y of a channel with transition probability $P_{Y|X}$. A useful fact is that $\mathcal{I}(\cdot, \cdot)$ is concave in the first argument when the second argument is fixed. Moreover, the conditional mutual information $I(X_1; Y_2|X_2 = x_2)$ can be expressed as $\mathcal{I}(P_{X_1|X_2=x_2}, P_{Y_2|X_1,X_2=x_2})$.

Since the TWC is viewed as two state-dependent one-way channels, each of the following theorems consists of two conditions, one for each direction of the two-way transmission. By symmetry, these theorems are valid if the roles of users 1 and 2 are swapped.

Theorem 1: For a memoryless TWC, if conditions (i) and (ii) below are satisfied, then $\mathcal{C}_\mathrm{I} = \mathcal{C}_\mathrm{O}$.

(i) There exists $P_{X_1}^* \in \mathcal{P}(\mathcal{X}_1)$ such that $\arg\max_{P_{X_1|X_2=x_2}} I(X_1; Y_2|X_2 = x_2) = P_{X_1}^*$ for all $x_2 \in \mathcal{X}_2$;

(ii) $\mathcal{I}(P_{X_2}, P_{Y_1|X_1=x_1,X_2})$ does not depend on $x_1 \in \mathcal{X}_1$ for any fixed $P_{X_2} \in \mathcal{P}(\mathcal{X}_2)$.

Proof: For any $P_{X_1,X_2}^{(1)} = P_{X_2}^{(1)} P_{X_1|X_2}^{(1)}$, let $P_{X_1,X_2}^{(2)} = P_{X_1}^* P_{X_2}^{(1)}$, where $P_{X_1}^*$ is given by (i). In light of (i), we have

$$I^{(1)}(X_1; Y_2|X_2) = \sum_{x_2} P_{X_2}^{(1)}(x_2) \cdot I^{(1)}(X_1; Y_2|X_2 = x_2) \tag{4}$$

$$\leq \sum_{x_2} P_{X_2}^{(1)}(x_2) \cdot \left[\max_{P_{X_1|X_2=x_2}} I(X_1; Y_2 | X_2 = x_2) \right] \tag{5}$$

$$= \sum_{x_2} P_{X_2}^{(1)}(x_2) \cdot \mathcal{I}(P_{X_1}^*, P_{Y_2|X_1, X_2=x_2}) \tag{6}$$

$$= \sum_{x_2} P_{X_2}^{(1)}(x_2) \cdot I^{(2)}(X_1; Y_2 | X_2 = x_2) \tag{7}$$

$$= I^{(2)}(X_1; Y_2 | X_2). \tag{8}$$

Moreover,

$$I^{(1)}(X_2; Y_1 | X_1) = \sum_{x_1} P_{X_1}^{(1)}(x_1) \cdot I^{(1)}(X_2; Y_1 | X_1 = x_1)$$

$$= \sum_{x_1} P_{X_1}^{(1)}(x_1) \cdot \mathcal{I}(P_{X_2|X_1=x_1}^{(1)}, P_{Y_1|X_1=x_1, X_2})$$

$$= \sum_{x_1} P_{X_1}^{(1)}(x_1) \cdot \mathcal{I}(P_{X_2|X_1=x_1}^{(1)}, P_{Y_1|X_1=x_1', X_2}) \tag{9}$$

$$\leq \mathcal{I}\left(\sum_{x_1} P_{X_1}^{(1)}(x_1) P_{X_2|X_1}^{(1)}(x_2|x_1), P_{Y_1|X_1=x_1', X_2} \right) \tag{10}$$

$$= \mathcal{I}(P_{X_2}^{(1)}, P_{Y_1|X_1=x_1', X_2})$$

$$= \sum_{x_1'} P_{X_1}^*(x_1') \cdot \mathcal{I}(P_{X_2}^{(1)}, P_{Y_1|X_1=x_1', X_2}) \tag{11}$$

$$= I^{(2)}(X_2; Y_1 | X_1), \tag{12}$$

where (9) holds by the invariance assumption in (ii) and $x_1' \in \mathcal{X}_1$ is arbitrary, (10) holds since the functional $\mathcal{I}(\cdot, \cdot)$ is concave in the first argument, and (11) is obtained from the invariance assumption in (ii). Combining the above yields $\mathcal{R}(P_{X_1,X_2}^{(1)}, P_{Y_1,Y_2|X_1,X_2}) \subseteq \mathcal{R}(P_{X_1}^* P_{X_2}^{(1)}, P_{Y_1,Y_2|X_1,X_2})$, which implies that $C_O \subseteq C_I$ and hence $C_I = C_O$. ∎

Instead of relying on the permutation invariance (row, column, or both) of the channel transition matrix, the symmetry property in the theorem is characterized by a combination of two symmetry properties for state-dependent one-way channels in terms of mutual information: (1) common capacity-achieving input distribution; (2) invariance of input-output mutual information. A special case where condition (i) of Theorem 1 trivially holds is when each one-way channel $P_{Y_2|X_1,X_2=x_2}$, $x_2 \in \mathcal{X}_2$, is T-symmetric[4] [36]; in this case we have $P_{X_1}^* = P_{\mathcal{X}_1}^U$.

We next apply condition (ii) of Theorem 1 for both directions of the two-way transmission.

---

[4]A point-to-point one way channel is called T-symmetric if the optimal input distribution (that maximizes the channels's mutual information) is uniform.

Theorem 2: For a memoryless TWC, if conditions (i) and (ii) below are satisfied, then $\mathcal{C}_\mathrm{I} = \mathcal{C}_\mathrm{O}$.

(i) $\mathcal{I}(P_{X_1}, P_{Y_2|X_1, X_2 = x_2})$ does not depend on $x_2 \in \mathcal{X}_2$ for any fixed $P_{X_1} \in \mathcal{P}(\mathcal{X}_1)$;

(ii) $\mathcal{I}(P_{X_2}, P_{Y_1|X_1 = x_1, X_2})$ does not depend on $x_1 \in \mathcal{X}_1$ for any fixed $P_{X_2} \in \mathcal{P}(\mathcal{X}_2)$.

Proof: From conditions (i) and (ii), we know that $\max_{P_{X_1|X_2 = x_2}} I(X_1; Y_2 | X_2 = x_2)$ has a common maximizer $P_{X_1}^*$ for all $x_2 \in \mathcal{X}_2$ and that $\max_{P_{X_2|X_1 = x_1}} I(X_2; Y_1 | X_1 = x_1)$ has a common maximizer $P_{X_2}^*$ for all $x_1 \in \mathcal{X}_1$. For any $P_{X_1, X_2}^{(1)} = P_{X_1}^{(1)} P_{X_2|X_1}^{(1)}$, let $P_{X_1, X_2}^{(2)} = P_{X_1}^* P_{X_2}^*$. Using the same argument as in (4)-(8), we conclude that $I^{(1)}(X_1; Y_2 | X_2) \leq I^{(2)}(X_1; Y_2 | X_2)$ and $I^{(1)}(X_2; Y_1 | X_1) \leq I^{(2)}(X_2; Y_1 | X_1)$. Thus, $\mathcal{R}(P_{X_1, X_2}^{(1)}, P_{Y_1, Y_2|X_1, X_2}) \subseteq \mathcal{R}(P_{X_1}^* P_{X_2}^*, P_{Y_1, Y_2|X_1, X_2})$, which yields $\mathcal{C}_\mathrm{I} = \mathcal{C}_\mathrm{O}$. ∎

To verify condition (i) in Theorem 1, one should find optimal input distributions for the one-way channel from users 1 to 2 for each state $x_2 \in \mathcal{X}_2$, say, via the Blahut-Arimoto algorithm [38]. This process can sometimes be simplified by testing whether the uniform input distribution is optimal via the Karush-Kuhn-Tucker (KKT) conditions for one-way channel capacity [33]. However, verifying condition (ii) in Theorem 1 may necessitate the evaluation of $\mathcal{I}(P_{X_2}, P_{Y_1|X_1, X_2}(\cdot|x_1, \cdot))$ for all $P_{X_2} \in \mathcal{P}(\mathcal{X}_2)$ and $x_1 \in \mathcal{X}_1$. In practice, such a verification is often complex, especially when the size of the input alphabet is large. Similar difficulties arise when ascertaining the conditions of Theorem 2. In the following results, conditions that are easier to check are presented.

Theorem 3: For a memoryless TWC, if conditions (i) and (ii) below are satisfied, then $\mathcal{C}_\mathrm{I} = \mathcal{C}_\mathrm{O}$.

(i) There exists $P_{X_1}^* \in \mathcal{P}(\mathcal{X}_1)$ such that $\arg\max_{P_{X_1|X_2 = x_2}} I(X_1; Y_2 | X_2 = x_2) = P_{X_1}^*$ for all $x_2 \in \mathcal{X}_2$ and $\mathcal{I}(P_{X_1}^*, P_{Y_2|X_1, X_2 = x_2})$ does not depend on $x_2 \in \mathcal{X}_2$;

(ii) There exists $P_{X_2}^* \in \mathcal{P}(\mathcal{X}_2)$ such that $\arg\max_{P_{X_2|X_1 = x_1}} I(X_2; Y_1 | X_1 = x_1) = P_{X_2}^*$ for all $x_1 \in \mathcal{X}_1$ and $\mathcal{I}(P_{X_2}^*, P_{Y_1|X_1 = x_1, X_2})$ does not depend on $x_1 \in \mathcal{X}_1$.

Proof: For any $P_{X_1, X_2}^{(1)} = P_{X_2}^{(1)} P_{X_1|X_2}^{(1)}$, consider $P_{X_1, X_2}^{(2)} = P_{X_1}^* P_{X_2}^*$, where $P_{X_1}^*$ and $P_{X_2}^*$ are given by (i) and (ii), respectively. Following the same steps as in (4)-(8), we obtain $I^{(1)}(X_1; Y_2 | X_2) \leq I^{(2)}(X_1; Y_2 | X_2)$. By a similar argument, we obtain the inequality $I^{(1)}(X_2; Y_1 | X_1) \leq I^{(2)}(X_2; Y_1 | X_1)$. Hence, $\mathcal{R}(P_{X_1, X_2}^{(1)}, P_{Y_1, Y_2|X_1, X_2}) \subseteq \mathcal{R}(P_{X_1}^* P_{X_2}^*, P_{Y_1, Y_2|X_1, X_2})$ which implies $\mathcal{C}_\mathrm{I} = \mathcal{C}_\mathrm{O}$. ∎

Unlike condition (ii) of Theorem 1 and the conditions in Theorem 2, Theorem 3 only requires checking the existence of a common maximizer and testing whether $\mathcal{I}(P_{X_1}^*, P_{Y_2|X_1, X_2 = x_2})$ is invariant with respect to $x_2 \in \mathcal{X}_2$ and $\mathcal{I}(P_{X_2}^*, P_{Y_1|X_1 = x_1, X_2})$ is invariant with respect to $x_1 \in \mathcal{X}_1$,

thus siginificantly reducing the validation computational complexity vis-a-vis Theorems 1 and 2.

The next two corollaries provide even simpler conditions. Let $[P_{Y_2|X_1,X_2}(\cdot|\cdot, x_2)]$ denote the transition matrix of the channel from users 1 to 2 when the input of user 2 is fixed to be x_2. The matrix $[P_{Y_2|X_1,X_2}(\cdot|\cdot, x_2)]$ has size $|\mathcal{X}_1| \times |\mathcal{Y}_2|$ and its entry at the x_1th row and y_2th column is $P_{Y_2|X_1,X_2}(y_2|x_1, x_2)$. Similarly, let $[P_{Y_2|X_1,X_2}(\cdot|x_1, \cdot)]$ denote the transition matrix of the channel from users 2 to 1 when the input of user 1 is fixed to be x_1.

Corollary 1: For a memoryless TWC, if conditions (i) and (ii) below are satisfied, then $\mathcal{C}_\mathrm{I} = \mathcal{C}_\mathrm{O}$.

(i) The channel with transition matrix $[P_{Y_2|X_1,X_2}(\cdot|\cdot, x_2)]$ is quasi-symmetric[5] for all $x_2 \in \mathcal{X}_2$;

(ii) The matrices $[P_{Y_1|X_1,X_2}(\cdot|x_1, \cdot)]$, $x_1 \in \mathcal{X}_1$, are column permutations of each other.

Proof: It suffices to show that conditions (i) and (ii) imply the conditions of Theorem 1. Under condition (i), we obtain a common maximizer given by $P_{X_1}^* = P_{\mathcal{X}_1}^\mathrm{U}$ since the optimal input distribution for a quasi-symmetric channel is the uniform distribution [35]; this implies condition (i) of Theorem 1. Furthermore, we observe that $\mathcal{I}(P_{X_2}, P_{Y_1|X_1,X_2}(\cdot|x_1, \cdot))$ is invariant with respect to column permutations of the transition matrix $P_{Y_1|X_1,X_2}(\cdot|x_1, \cdot)$ for given P_{X_2}. Since the matrices $[P_{Y_1|X_1,X_2}(\cdot|x_1, \cdot)]$, $x_1 \in \mathcal{X}_1$, are column permutations of each other, we conclude that $\mathcal{I}(P_{X_2}, P_{Y_1|X_1=x_1,X_2})$ does not depend on $x_1 \in \mathcal{X}_1$ for any fixed $P_{X_2} \in \mathcal{P}(\mathcal{X}_2)$, which is the second condition of Theorem 1. ∎

Corollary 2: For a memoryless TWC, if conditions (i) and (ii) below are satisfied, then $\mathcal{C}_\mathrm{I} = \mathcal{C}_\mathrm{O}$.

(i) The matrices $[P_{Y_2|X_1,X_2}(\cdot|\cdot, x_2)]$, $x_2 \in \mathcal{X}_2$, are column permutations of each other;

(ii) The matrices $[P_{Y_1|X_1,X_2}(\cdot|x_1, \cdot)]$, $x_1 \in \mathcal{X}_1$, are column permutations of each other.

Proof: It suffices to show that conditions (i) and (ii) imply the conditions of Theorem 2. This can be done using a similar argument as in the second part of the proof of Corollary 1, and hence the details are omitted. ∎

If the transition probability $P_{Y_1,Y_2|X_1,X_2}$ satisfies conditions (i) and (ii) of Theorem 1, the capacity region is given by

$$\mathcal{C} = \overline{\mathrm{co}} \left(\bigcup_{P_{X_2}} \mathcal{R}(P_{X_1}^* P_{X_2}, P_{Y_1,Y_2|X_1,X_2}) \right), \tag{13}$$

---

[5]A discrete memoryless channel with transition matrix $[P_{Y|X}(\cdot|\cdot)]$ is said to be weakly-symmetric if the rows are permutations of each other and all the column sums are identical [34]. A discrete memoryless channel is said to be quasi-symmetric if its transition matrix $[P_{Y|X}(\cdot|\cdot)]$ can be partitioned along its columns into weakly-symmetric sub-matrices [35].

where $P_{X_1}^*$ is given by condition (i). For example, condition (i) trivially holds when each one-way channel with fixed state $x_2 \in \mathcal{X}_2$ from users 1 to 2 is T-symmetric. In this case, we have $P_{X_1}^* = P_{\mathcal{X}_1}^{\mathsf{U}}$ and the capacity region becomes

$$\mathcal{C} = \overline{\mathsf{co}} \left(\bigcup_{P_{X_2}} \mathcal{R}(P_{\mathcal{X}_1}^{\mathsf{U}} P_{X_2}, P_{Y_1, Y_2 | X_1, X_2}) \right). \tag{14}$$

In fact, this is also the capacity region for memoryless TWCs which satisfy Corollary 1 because condition (ii) of Corollary 1 implies condition (ii) of Theorem 1 (this follows from the proof of Corollary 1). Moreover, the proof of Theorem 2 demonstrates that a common maximizer exists for each direction of the two-way transmission under the conditions of Theorem 2. Let $\arg\max_{P_{X_1 | X_2 = x_2}} I(X_1; Y_2 | X_2 = x_2) = P_{X_1}^*$ for all $x_2 \in \mathcal{X}_2$ and $\arg\max_{P_{X_2 | X_1 = x_1}} I(X_2; Y_1 | X_1 = x_1) = P_{X_2}^*$ for all $x_1 \in \mathcal{X}_1$. A TWC which satisfies the conditions of Theorem 2 has the capacity region

$$\mathcal{C} = \mathcal{R}(P_{X_1}^* P_{X_2}^*, P_{Y_1, Y_2 | X_1, X_2}). \tag{15}$$

The region is rectangular which suggests that such a two-way transmission inherently comprises two independent one-way transmissions. A memoryless TWC that satisfies the conditions in either Theorem 3 or Corollary 2 also has a capacity region given by (15).

To end this section, we remark that it is possible to combine different conditions to determine the capacity region of a broader class of memoryless TWCs as shown below.

Theorem 4: For a memoryless TWC, if both of the following conditions are satisfied, then $\mathcal{C} = \mathcal{C}_{\mathrm{I}} = \mathcal{C}_{\mathrm{O}}$ with $\mathcal{C}$ given by (13):

(i) There exists $P_{X_1}^* \in \mathcal{P}(\mathcal{X}_1)$ such that $\arg\max_{P_{X_1 | X_2 = x_2}} I(X_1; Y_2 | X_2 = x_2) = P_{X_1}^*$ for all $x_2 \in \mathcal{X}_2$;

(ii) $H(Y_1 | X_1, X_2)$ does not depend on $P_{X_1 | X_2}$ given P_{X_2} and $P_{Y_1 | X_1, X_2}$, and $P_{X_1}^*$ given in (i) satisfies $H^{(1)}(Y_1 | X_1) \leq H^{(2)}(Y_1 | X_1)$ for any $P_{X_1, X_2}^{(1)} = P_{X_2}^{(1)} P_{X_1 | X_2}^{(1)}$, where $P_{X_1, X_2}^{(2)} = P_{X_1}^* P_{X_2}^{(1)}$.

Here, condition (i) is directly from Theorem 1; condition (ii) is obtained by extracting the CVA condition related to the channel from user 2 to user 1. In order that the two conditions jointly determine the capacity region, the $\tilde{P}_{X_1}$ required by the CVA condition is forced to be $P_{X_1}^*$.

Proof of Theorem 4: Given any $P_{X_1, X_2}^{(1)} = P_{X_2}^{(1)} P_{X_1 | X_2}^{(1)}$, let $P_{X_1, X_2}^{(2)} = P_{X_1}^* P_{X_2}^{(1)}$. Invoking the same argument as in (4)-(8), we obtain that $I^{(1)}(X_1; Y_2 | X_2) \leq I^{(2)}(X_1; Y_2 | X_2)$ using condition (i). Moreover, condition (ii) implies that $I^{(1)}(X_2; Y_1 | X_1) = H^{(1)}(Y_1 | X_1) - H^{(1)}(Y_1 | X_1, X_2) \leq$

$H^{(2)}(Y_1|X_1) - H^{(2)}(Y_1|X_1, X_2) = I^{(2)}(X_2; Y_1|X_1)$. Combining the above then completes the proof. ∎

D. Examples

We next illustrate the proposed conditions via examples.

Example 1 (Memoryless Binary Additive-Noise TWCs with Erasures): Let $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$ and $\mathcal{Y}_1 = \mathcal{Y}_2 = \mathcal{Z} = \{0, 1, \mathbf{E}\}$, where $\mathbf{E}$ denotes channel erasure. A binary additive noise TWC with erasures is defined by the channel equations

$$Y_{1,i} = (X_{1,i} \oplus_2 X_{2,i} \oplus_2 Z_{1,i}) \cdot \mathbf{1}\{Z_{1,i} \neq \mathbf{E}\} + \mathbf{E} \cdot \mathbf{1}\{Z_{1,i} = \mathbf{E}\},$$

$$Y_{2,i} = (X_{1,i} \oplus_2 X_{2,i} \oplus_2 Z_{2,i}) \cdot \mathbf{1}\{Z_{2,i} \neq \mathbf{E}\} + \mathbf{E} \cdot \mathbf{1}\{Z_{2,i} = \mathbf{E}\},$$

where $\oplus_2$ denotes modulo-2 addition, $\{(Z_{1,i}, Z_{2,i})\}_{i=1}^{\infty}$ is a memoryless joint noise-erasure process that is independent of the users' messages and has components $Z_{1,i}, Z_{2,i} \in \mathcal{Z}$ such that $\Pr(Z_{j,i} = \mathbf{E}) = \varepsilon_j$, $\Pr(Z_{j,i} = 1) = \alpha_j$, where $0 \leq \varepsilon_j + \alpha_j \leq 1$ for $j = 1, 2$, and $\mathbf{1}\{\cdot\}$ denotes the indicator function. Here, we adopt the convention $\mathbf{E} \cdot 0 = 0$ and $\mathbf{E} \cdot 1 = \mathbf{E}$ to simplify the representation of the channel equations.[6] The channel equations yield the following transition matrices for the one-way channels:

$$[P_{Y_2|X_1,X_2}(\cdot|\cdot, 0)] = \begin{pmatrix} 1 - \varepsilon_2 - \alpha_2 & \alpha_2 & \varepsilon_2 \\ \alpha_2 & 1 - \varepsilon_2 - \alpha_2 & \varepsilon_2 \end{pmatrix},$$

$$[P_{Y_2|X_1,X_2}(\cdot|\cdot, 1)] = \begin{pmatrix} \alpha_2 & 1 - \varepsilon_2 - \alpha_2 & \varepsilon_2 \\ 1 - \varepsilon_2 - \alpha_2 & \alpha_2 & \varepsilon_2 \end{pmatrix},$$

$$[P_{Y_1|X_1,X_2}(\cdot|0, \cdot)] = \begin{pmatrix} 1 - \varepsilon_1 - \alpha_1 & \alpha_1 & \varepsilon_1 \\ \alpha_1 & 1 - \varepsilon_1 - \alpha_1 & \varepsilon_1 \end{pmatrix},$$

$$[P_{Y_1|X_1,X_2}(\cdot|1, \cdot)] = \begin{pmatrix} \alpha_1 & 1 - \varepsilon_1 - \alpha_1 & \varepsilon_1 \\ 1 - \varepsilon_1 - \alpha_1 & \alpha_1 & \varepsilon_1 \end{pmatrix}.$$

As all our proposed conditions are only based on the marginal transition probabilities, the relationship between $Z_{1,i}$ and $Z_{2,i}$ can be arbitrary. By Corollary 2, we obtain that the optimal channel input distribution is $P_{X_1}^* P_{X_2}^* = P_{\mathcal{X}_1}^{\mathsf{U}} P_{\mathcal{X}_2}^{\mathsf{U}}$ since the marginal channel transition matrices

---

[6]Strictly speaking, $X_{1,i} \oplus_2 X_{2,i} \oplus_2 Z_{j,i}$ is undefined when $Z_{j,i} = \mathbf{E}$, but we set $(X_{1,i} \oplus_2 X_{2,i} \oplus_2 \mathbf{E}) \cdot 0 = 0$.

not only exhibit column permutation properties but also are quasi-symmetric. The capacity region is given by

$$\mathcal{C} = \left\{ (R_1, R_2) : R_1 \le (1 - \varepsilon_2)\left(1 - H_{\mathsf{b}}\left(\frac{\alpha_2}{1 - \varepsilon_2}\right)\right), \ R_2 \le (1 - \varepsilon_1)\left(1 - H_{\mathsf{b}}\left(\frac{\alpha_1}{1 - \varepsilon_1}\right)\right) \right\},$$

where $H_{\mathsf{b}}(\cdot)$ denotes the binary entropy function. One can verify that this TWC also satisfies the conditions of Theorems 1-3 and Corollary 1.

Remark 1: Various TWCs are special cases of this TWC model:

1) If $\alpha_1 = \alpha_2 = 0$, then the memoryless binary additive TWC with erasures is recovered:

$$Y_{1,i} = (X_{1,i} \oplus_2 X_{2,i}) \cdot \mathbf{1}\{Z_{1,i} \ne \mathbf{E}\} + \mathbf{E} \cdot \mathbf{1}\{Z_{1,i} = \mathbf{E}\},$$

$$Y_{2,i} = (X_{1,i} \oplus_2 X_{2,i}) \cdot \mathbf{1}\{Z_{2,i} \ne \mathbf{E}\} + \mathbf{E} \cdot \mathbf{1}\{Z_{2,i} = \mathbf{E}\}.$$

The capacity region is given by

$$\mathcal{C} = \{(R_1, R_2) : R_1 \le 1 - \varepsilon_2, R_2 \le 1 - \varepsilon_1\}.$$

2) If $\varepsilon_1 = \varepsilon_2 = 0$, then the memoryless binary additive-noise TWC is obtained:

$$Y_{1,i} = X_{1,i} \oplus_2 X_{2,i} \oplus Z_{1,i},$$

$$Y_{2,i} = X_{1,i} \oplus_2 X_{2,i} \oplus Z_{2,i}.$$

The capacity region of this channel is given by

$$\mathcal{C} = \{(R_1, R_2) : R_1 \le 1 - H_{\mathsf{b}}(\alpha_2), R_2 \le 1 - H_{\mathsf{b}}(\alpha_1)\}.$$

3) If $\varepsilon_1 = \varepsilon_2 = 0$ and $\alpha_1 = \alpha_2 = 0$, then we obtain the memoryless binary additive TWC given by $Y_{1,i} = X_{1,i} \oplus_2 X_{2,i}$ and $Y_{2,i} = X_{1,i} \oplus_2 X_{2,i}$. The capacity region is given by $\mathcal{C} = \{(R_1, R_2) : R_1 \le 1, R_2 \le 1\}$ [3], [24].

Remark 2: Example 1 can be generalized to a non-binary setting: for some integer $q > 2$, $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1, \dots, q - 1\}$ and $\mathcal{Y}_1 = \mathcal{Y}_2 = \mathcal{Z} = \{0, 1, \dots, q - 1, \mathbf{E}\}$, the q-ary channel model obeys the same equations as in Example 1 with modulo-2 addition replaced with the modulo-q operation $\oplus_q$. Furthermore, the channel noise-erasure variables have marginal distributions given by $\Pr(Z_{j,i} = \mathbf{E}) = \varepsilon_j$ and $\Pr(Z_{j,i} = z) = \alpha_j/(q - 1)$ for $z = 0, 1, \dots, q - 1$, where $0 \le \alpha_j + \varepsilon_j \le 1$ for $j = 1, 2$. By Corollary 2, we directly have that $\mathcal{C}_{\mathsf{I}} = \mathcal{C}_{\mathsf{O}}$, and

$$\mathcal{C} = \left\{ (R_1, R_2) : R_1 \le (1 - \varepsilon_2)\left(\log_2 q - H_q\left(\frac{\alpha_2}{(q - 1)(1 - \varepsilon_2)}\right)\right), \right.$$

$$R_2 \leq (1 - \varepsilon_1) \left(\log_2 q - H_q \left(\frac{\alpha_1}{(q-1)(1-\varepsilon_1)} \right) \right) \bigg\},$$

where $H_q(x) \triangleq x \log_2(q-1) - x \log_2 x - (1-x) \log_2(1-x)$.

Example 2 (Data Access TWCs): Let $q = 2^m$ for some integer $m \geq 1$ and consider the alphabets $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X} = \{0, 1, \ldots, q-1\}$, $\mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1, \ldots, q-1, \mathbf{E}\}$, and $\mathcal{Z} = \{0, 1, 2\}$. A data access TWC linking two storage devices is described by

$$Y_{1,i} = (X_{1,i} \boxplus_q X_{2,i}) \cdot \mathbf{1}\{Z_{1,i} = 0\} + ((q-1) \boxplus_q X_{1,i} \boxplus_q X_{2,i}) \cdot \mathbf{1}\{Z_{1,i} = 1\} + \mathbf{E} \cdot \mathbf{1}\{Z_{1,i} = 2\},$$

$$Y_{2,i} = (X_{1,i} \boxplus_q X_{2,i}) \cdot \mathbf{1}\{Z_{2,i} = 0\} + ((q-1) \boxplus_q X_{1,i} \boxplus_q X_{2,i}) \cdot \mathbf{1}\{Z_{2,i} = 1\} + \mathbf{E} \cdot \mathbf{1}\{Z_{2,i} = 2\},$$

where $a \boxplus_q b$ denotes bit-wise addition for the length-q standard binary representation of $a, b \in \mathcal{X}$, and $\{(Z_{1,i}, Z_{2,i})\}_{i=1}^{\infty}$ is a memoryless joint noise-erasure process that is independent of the stored messages and has components $Z_{1,i}, Z_{2,i} \in \mathcal{Z}$ such that $\Pr(Z_{j,i} = 1) = \alpha_j$, $\Pr(Z_{j,i} = \mathbf{E}) = \varepsilon_j$, where $0 \leq \alpha_j + \varepsilon_j \leq 1$ for $j = 1, 2$. This channel model can capture the effect of user signal superpositions (when $Z_{j,i} = 0$), bit-level burst errors which flip all bits of $X_{1,i} \boxplus_q X_{2,i}$ (when $Z_{j,i} = 1$), and data package losses (when $Z_{j,i} = 2$).

For this channel, an application of Corollary 2 immediately gives the capacity region:

$$\mathcal{C} = \left\{ (R_1, R_2) : R_1 \leq (1 - \varepsilon_2) \left(m - H_{\mathsf{b}} \left(\frac{\alpha_2}{1 - \varepsilon_2} \right) \right), \ R_2 \leq (1 - \varepsilon_1) \left(m - H_{\mathsf{b}} \left(\frac{\alpha_1}{1 - \varepsilon_1} \right) \right) \right\}.$$

The next example redervies a known result in [28] based on our approach.

Example 3 (Memoryless Injective Semi-Deterministic TWCs [28]): Let $\mathcal{T}_j$ and $\mathcal{Z}_j$ denote finite sets. A memoryless ISD-TWC is defined in [28] by the channel equations

$$Y_{j,i} = h_j(X_{j,i}, T_{j,i}), \text{ and } T_{j,i} = \tilde{h}_j(X_{k,i}, Z_{j,i}), \text{ for } j, k \in \{1, 2\}, \ j \neq k, \tag{16}$$

where $h_j : \mathcal{X}_j \times \mathcal{T}_j \to \mathcal{Y}_j$ is invertible in $\mathcal{T}_j$ and $\tilde{h}_j : \mathcal{X}_k \times \mathcal{Z}_j \to \mathcal{T}_j$ is invertible in $\mathcal{Z}_j$, i.e., for every $x_j \in \mathcal{X}_j$, $h_j(x_j, t_j)$ is one-to-one in $t_j \in \mathcal{T}_j$ and for every $x_k \in \mathcal{X}_k$, $\tilde{h}_j(x_k, z_j)$ is one-to-one in $z_j \in \mathcal{Z}_j$. Here, $\{(Z_{1,i}, Z_{2,i})\}_{i=1}^{\infty}$ is a memoryless joint noise process that is independent of users' messages. For this channel, we have [28]

$$I(X_1; Y_2 | X_2 = x_2) \leq \max_{P_{X_1}} H(\tilde{h}_2(X_1, Z_2)) - H(Z_2).$$

This upper bound does not depend on X_2, and hence a common maximizer exists, i.e., $P_{X_1}^* = \arg\max_{P_{X_1}} H(\tilde{h}_2(X_1, Z_2))$. Moreover, the value of $\max_{P_{X_1}} I(X_1; Y_2 | X_2 = x_2)$ is identical for all $x_2 \in \mathcal{X}_2$. We immediately observe that condition (i) in Theorem 3 holds. By a similar argument,
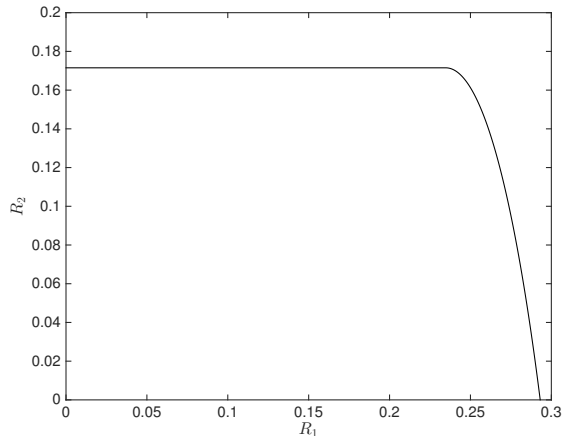
Fig. 5: The capacity region of the point-to-point memoryless TWC in Example 4.

condition (ii) in Theorem 3 also holds, implying that Shannon's inner and outer bounds coincide. The capacity region is given by

$$\mathcal{C} = \left\{ (R_1, R_2) : R_1 \leq \max_{P_{X_1}} H(\tilde{h}_2(X_1, Z_2)) - H(Z_2), R_2 \leq \max_{P_{X_2}} H(\tilde{h}_1(X_2, Z_1)) - H(Z_1) \right\}.$$

Example 4: Consider the TWC with $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1\}$ and transition probability

$$[P_{Y_1,Y_2|X_1,X_2}] = \begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \begin{array}{cccc} 00 & 01 & 10 & 11 \\ \left(\begin{array}{cccc} 0.783 & 0.087 & 0.117 & 0.013 \\ 0.36279 & 0.05421 & 0.50721 & 0.07579 \\ 0.261 & 0.609 & 0.039 & 0.091 \\ 0.173889 & 0.243111 & 0.243111 & 0.339889 \end{array} \right) \end{array}.$$

The one-way channel marginal distributions are

$$[P_{Y_2|X_1,X_2}(\cdot|\cdot, 0)] = \begin{pmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{pmatrix}, [P_{Y_2|X_1,X_2}(\cdot|\cdot, 1)] = \begin{pmatrix} 0.87 & 0.13 \\ 0.417 & 0.583 \end{pmatrix},$$

with $[P_{Y_1|X_1,X_2}(\cdot|0, \cdot)] = [P_{Y_1|X_1,X_2}(\cdot|1, \cdot)] = [P_{Y_2|X_1,X_2}(\cdot|\cdot, 1)]$.

Shannon's symmetry condition in Proposition 1 does not hold for this channel since there are no permutations of $\mathcal{Y}_1$ and $\mathcal{Y}_2$ which can result in (2). Furthermore, since $H(Y_2|X_1 = 0, X_2 = 0) = H_b(0.1)$ and $H(Y_2|X_1 = 1, X_2 = 0) = H_b(0.3)$, $H(Y_2|X_1, X_2)$ depends on $P_{X_1|X_2}$ for fixed P_{X_2}. Thus, the CVA condition in Proposition 3 does not hold either. However, the conditions of Theorem 1 are satisfied since a common maximizer exists for the one-way channel from

users 1 to 2 given by $P_{X_1}^*(0) = 0.471$, and condition (ii) trivially holds. By considering all input distributions of the form $P_{X_1,X_2} = P_{X_1}^* P_{X_2}$, where $P_{X_2} \in \mathcal{P}(\mathcal{X}_2)$, one can compute the capacity region as shown in Fig. 5. We note that, with some extra effort, one can show that the conditions of Theorem 4 also hold [2].

Finally, we point out (without proof) that the channels in the examples in [3, Fig. 2 & Tab. II] and [28, Section IV-B] satisfy the conditions of Theorem 1.

E. Comparison with Prior Results

In this section, we show that Theorems 1 and 2 generalize the Shannon results in Propositions 1 and 2, respectively, and that Theorem 4 subsumes the CVA result in Proposition 3 as a special case.

Theorem 5: A TWC that satisfies the Shannon's one-sided symmetry condition of Proposition 1 must satisfy the conditions of Theorem 1.

Proof: If a TWC satisfies the Shannon condition in Proposition 1, the capacity-achieving input distribution is of the form $P_{X_1,X_2} = P_{\mathcal{X}_1}^{\mathsf{U}} P_{X_2}$ for some $P_{X_2} \in \mathcal{P}(\mathcal{X}_2)$ [3]. This implies that condition (i) of Theorem 1 is satisfied because a common maximizer exists for all $x_2 \in \mathcal{X}$ and is given by $P_{X_1}^* = P_{\mathcal{X}_1}^{\mathsf{U}}$. To prove that condition (ii) is also satisfied, we consider the transition matrices $[P_{Y_1|X_1,X_2}(\cdot|x_1', \cdot)]$ and $[P_{Y_1|X_1,X_2}(\cdot|x_1'', \cdot)]$ for arbitrary $x_1', x_1'' \in \mathcal{X}_1$ and show that these are column permutations of each other and hence $\mathcal{I}(P_{X_2}, P_{Y_1|X_1=x_1',X_2}) = \mathcal{I}(P_{X_2}, P_{Y_1|X_1=x_1'',X_2})$. The first claim is true because

$$P_{Y_1|X_1,X_2}(y_1|x_1', x_2) = P_{Y_1|X_1,X_2}(\pi^{\mathcal{Y}_1}[x_1', x_1''](y_1)|\tau_{x_1',x_1''}^{\mathcal{X}_1}(x_1'), x_2) \tag{17}$$

$$= P_{Y_1|X_1,X_2}(\pi^{\mathcal{Y}_1}[x_1', x_1''](y_1)|x_1'', x_2),$$

where (17) is obtained by marginalizing over Y_2 on both sides of (2). For the second claim, we have

$$\mathcal{I}(P_{X_2}, P_{Y_1|X_1=x_1',X_2})$$

$$= \sum_{x_2,y_1} P_{X_2}(x_2) P_{Y_1|X_1,X_2}(y_1|x_1', x_2) \log \frac{P_{Y_1|X_1,X_2}(y_1|x_1', x_2)}{\sum_{\tilde{x}_2} P_{X_2}(\tilde{x}_2) P_{Y_1|X_1,X_2}(y_1|x_1', \tilde{x}_2)}$$

$$= \sum_{x_2,y_1} P_{X_2}(x_2) P_{Y_1|X_1,X_2}(\pi^{\mathcal{Y}_1}[x_1', x_1''](y_1)|x_1'', x_2) \log \frac{P_{Y_1|X_1,X_2}(\pi^{\mathcal{Y}_1}[x_1', x_1''](y_1)|x_1'', x_2)}{\sum_{\tilde{x}_2} P_{X_2}(\tilde{x}_2) P_{Y_1|X_1,X_2}(\pi^{\mathcal{Y}_1}[x_1', x_1''](y_1)|x_1'', \tilde{x}_2)}$$

$$\tag{18}$$

$$= \sum_{x_2, \tilde{y}_1} P_{X_2}(x_2) P_{Y_1|X_1,X_2}(\tilde{y}_1|x_1'', x_2) \log \frac{P_{Y_1|X_1,X_2}(\tilde{y}_1|x_1'', x_2)}{\sum_{\tilde{x}_2} P_{X_2}(\tilde{x}_2) P_{Y_1|X_1,X_2}(\tilde{y}_1|x_1'', \tilde{x}_2)}$$

$$= \mathcal{I}(P_{X_2}, P_{Y_1|X_1=x_1'',X_2}),$$

where (18) holds by the first claim. ∎

Remark 3: Since the optimal input distribution of user 1 in Theorem 1 is not necessarily uniform as illustrated in Example 4, Theorem 1 is more general than Proposition 1.

Theorem 6: A TWC that satisfies the Shannon two-sided symmetry condition of Proposition 2 must satisfy the conditions of Theorem 2.

This theorem is immediate, and hence the proof is omitted. Together with Example 5 given in the next section, Theorem 2 is shown to be more general than Proposition 2. We next show that the symmetry properties identified by the conditions of Theorem 4 are more general than those in the CVA condition.

Theorem 7: A TWC that satisfies the CVA condition in Proposition 3 must satisfy the conditions in Theorem 4.

Proof: Suppose that the condition of Proposition 3 is satisfied. To prove the theorem, we show that for $j = 1, 2$, $H(Y_j|X_1 = x_1', X_2 = x_2) = H(Y_j|X_1 = x_1'', X_2 = x_2)$ for all $x_1', x_1'' \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$. Given arbitrary pairs (x_1', x_2) and (x_1'', x_2), consider the probability distributions

$$P^{(1)}_{X_1,X_2}(a, b) = \begin{cases} 1, & \text{if } a = x_1' \text{ and } b = x_2, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$P^{(2)}_{X_1,X_2}(a, b) = \begin{cases} 1, & \text{if } a = x_1'' \text{ and } b = x_2, \\ 0, & \text{otherwise.} \end{cases}$$

Noting that $P^{(1)}_{X_2} = P^{(2)}_{X_2}$, we have $H(Y_j|X_1 = x_1', X_2 = x_2) = H^{(1)}(Y_j|X_1, X_2) = H^{(2)}(Y_j|X_1, X_2) = H(Y_j|X_1 = x_1'', X_2 = x_2)$, where the first and last equality are due to the definitions of $P^{(1)}_{X_1,X_2}$ and $P^{(2)}_{X_1,X_2}$, respectively, and the second equality follows from the CVA condition since $P^{(1)}_{X_2} = P^{(2)}_{X_2}$. Thus $H(Y_j|X_1 = x_1, X_2 = x_2)$ does not depend on x_1 for fixed x_2 as claimed. Also, since $H(Y_j|X_1, X_2 = x_2) = \sum_{x_1} P_{X_1|X_2}(x_1|x_2) \cdot H(Y_j|X_1 = x_1, X_2 = x_2)$, $H(Y_j|X_1, X_2 = x_2)$ does not depend on $P_{X_1|X_2=x_2}$.

Next, we show that condition (i) of Theorem 4 holds by constructing a common maximizer from the CVA condition. For fixed $x_2 \in \mathcal{X}_2$, let $P^*_{X_1|X_2=x_2} = \arg\max_{P_{X_1|X_2=x_2}} I(X_1; Y_2|X_2 = x_2) = \arg\max_{P_{X_1|X_2=x_2}} [H(Y_2|X_2 = x_2) - H(Y_2|X_1, X_2 = x_2)]$ and define $P^{(1)}_{X_1,X_2} = P^{(1)}_{X_2} P^*_{X_1|X_2}$

for some $P_{X_2}^{(1)} \in \mathcal{P}(\mathcal{X}_2)$. Since $H(Y_j|X_1, X_2 = x_2)$ does not depend on $P_{X_1|X_2=x_2}$, $P_{X_1|X_2=x_2}^*$ is in fact a maximizer of $H(Y_2|X_2 = x_2)$. Note that the maximizer $P_{X_1|X_2=x_2}^*$ is not necessarily unique, but any choice works for our purposes. Now for $P_{X_1,X_2}^{(1)}$, by the CVA condition, there exists $\tilde{P}_{X_1} \in \mathcal{P}(\mathcal{X}_1)$ such that $H^{(1)}(Y_2|X_2) \leq H^{(2)}(Y_2|X_2)$, where $P_{X_1,X_2}^{(2)} = \tilde{P}_{X_1} P_{X_2}^{(1)}$. Since $P_{X_1|X_2=x_2}^*$ is the maximizer for $H(Y_2|X_2 = x_2)$, we have

$$
\begin{aligned}
H^{(1)}(Y_2|X_2) &= \sum_{x_2} P_{X_2}^{(1)}(x_2) \cdot H^{(1)}(Y_2|X_2 = x_2) \\
&= \sum_{x_2} P_{X_2}^{(1)}(x_2) \cdot \left[\max_{P_{X_1|X_2=x_2}} H(Y_2|X_2 = x_2) \right] \\
&\geq \sum_{x_2} P_{X_2}^{(1)}(x_2) \cdot H^{(2)}(Y_2|X_2 = x_2) \\
&= H^{(2)}(Y_2|X_2)
\end{aligned}
$$

Thus, $H^{(1)}(Y_2|X_2) = H^{(2)}(Y_2|X_2)$, i.e.,

$$
\sum_{x_2} P_{X_2}^{(1)}(x_2) \cdot H^{(1)}(Y_2|X_2 = x_2) = \sum_{x_2} P_{X_2}^{(1)}(x_2) \cdot H^{(2)}(Y_2|X_2 = x_2).
$$

Since $H^{(2)}(Y_2|X_2 = x_2) \leq H^{(1)}(Y_2|X_2 = x_2)$ for each $x_2 \in \mathcal{X}_2$, we obtain $H^{(1)}(Y_2|X_2 = x_2) = H^{(2)}(Y_2|X_2 = x_2)$, i.e., $\tilde{P}_{X_1}$ achieves the same value for $H(Y_2|X_2 = x_2)$ as $P_{X_1|X_2=x_2}^*$ for all $x_2 \in \mathcal{X}_2$. Consequently, $\tilde{P}_{X_1}$ is a common maximizer and thus condition (i) of Theorem 4 is satisfied. Moreover, since the common maximizer $\tilde{P}_{X_1}$ is from the CVA condition, we have that $H^{(1)}(Y_1|X_1) \leq H^{(2)}(Y_1|X_1)$, which together with the fact that $H(Y_1|X_1, X_2)$ does not depend on $P_{X_1|X_2}$ given P_{X_2} and $P_{Y_1|X_1,X_2}$ (guaranteed by the CVA condition) implies that condition (ii) of Theorem 4 holds. ∎

Remark 4: As illustrated by Example 4, a TWC that satisfies the conditions of Theorem 4 does not necessarily satisfy the CVA condition in Proposition 3. Therefore, Theorem 4 is a more general result than Proposition 3. We note that the main difference between Theorem 4 and Proposition 3 lies in the fact that we allow $H(Y_2|X_1, X_2)$ to depend on $P_{X_1|X_2}$, given P_{X_2}.

F. Connection Between the Shannon and CVA Conditions

In this section, we connect Shannon's result to the CVA condition. First, the proof in Appendix A shows that Shannon's symmetry conditions are more than sufficient for $\mathcal{C}_I$ and $\mathcal{C}_O$ to coincide. In fact, assume that the marginal channels $P_{Y_j|X_1,X_2}$'s (derived from $P_{Y_1,Y_2|X_1,X_2}$) satisfy the following extended Shannon's symmetry condition: for any pair of distinct input symbols x_1',

$x_1'' \in \mathcal{X}_1$, there exists a pair of permutations $(\pi^{\mathcal{Y}_1}[x_1', x_1''], \pi^{\mathcal{Y}_2}[x_1', x_1''])$ on $\mathcal{Y}_1$ and $\mathcal{Y}_2$, respectively, (which depend on x_1' and x_1'') such that for all x_1, x_2, y_1, y_2,

$$\begin{cases} P_{Y_1|X_1,X_2}(y_1|x_1,x_2) = P_{Y_1|X_1,X_2}(\pi^{\mathcal{Y}_1}[x_1', x_1''](y_1)|\tau_{x_1',x_1''}^{\mathcal{X}_1}(x_1), x_2), \\ P_{Y_2|X_1,X_2}(y_2|x_1,x_2) = P_{Y_2|X_1,X_2}(\pi^{\mathcal{Y}_2}[x_1', x_1''](y_2)|\tau_{x_1',x_1''}^{\mathcal{X}_1}(x_1), x_2), \end{cases} \tag{19}$$

then $\mathcal{C}_\mathrm{I} = \mathcal{C}_\mathrm{O} = \mathcal{C}$ with $\mathcal{C}$ given by (3).

The extended Shannon's symmetry conditions are more general than their original versions since (2) implies (19) but the reverse implication is not true as shown below.

Example 5: Consider the TWC with $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1\}$ and transition probability

$$[P_{Y_1,Y_2|X_1,X_2}] = \begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \begin{array}{cccc} 00 & 01 & 10 & 11 \\ \left(\begin{array}{cccc} 0.25 & 0.5 & 0.25 & 0 \\ 0.375 & 0375 & 0.125 & 0.125 \\ 0.125 & 0.125 & 0.375 & 0.375 \\ 0.125 & 0.125 & 0.375 & 0.375 \end{array} \right) \end{array}.$$

The marginal distributions are

$$[P_{Y_1|X_1,X_2}] = \begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \begin{array}{cc} 0 & 1 \\ \left(\begin{array}{cc} 0.75 & 0.25 \\ 0.75 & 0.25 \\ 0.25 & 0.75 \\ 0.25 & 0.75 \end{array} \right) \end{array}, \ [P_{Y_2|X_1,X_2}] = \begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \begin{array}{cc} 0 & 1 \\ \left(\begin{array}{cc} 0.5 & 0.5 \\ 0.5 & 0.5 \\ 0.5 & 0.5 \\ 0.5 & 0.5 \end{array} \right) \end{array}.$$

Clearly, neither of the Shannon conditions in Proposition 1 or 2 holds, but the extended condition in (19) holds.

We now show that the above extended symmetry condition implies the CVA condition.

Theorem 8: A TWC that satisfies the condition in (19) must satisfy the CVA condition of Proposition 3.

Proof: If the marginal channels $P_{Y_1|X_1,X_2}$ and $P_{Y_2|X_1,X_2}$ satisfy the extended one-sided symmetry condition, then $H(Y_j|X_1 = x_1, X_2 = x_2)$ does not depend on $x_1 \in \mathcal{X}_1$ for any fixed $x_2 \in \mathcal{X}_2$ since the rows of $[P_{Y_j|X_1,X_2}(\cdot|\cdot, x_2)]$ are permutations of each other. Hence, $H(Y_j|X_1, X_2)$ does not depend on $P_{X_1|X_2}$ given $P_{X_2} \in \mathcal{P}(\mathcal{X}_2)$ as required by the CVA condition.

Next, for any given $P_{X_1,X_2}^{(1)} = P_{X_2}^{(1)} P_{X_1|X_2}^{(1)}$, we show that $P_{X_1,X_2}^{(2)} = \tilde{P}_{X_1} P_{X_2}^{(1)}$ with the choice $\tilde{P}_{X_1} = P_{\mathcal{X}_1}^\mathsf{U}$ meets the remaining requirements of the CVA condition in Proposition 3. Since

the TWC satisfies the extended Shannon condition, Lemma 6 in Appendix A gives the two inequalities: $I^{(1)}(X_1; Y_2|X_2) \leq I^{(2)}(X_1; Y_2|X_2)$ and $I^{(1)}(X_2; Y_1|X_1) \leq I^{(2)}(X_2; Y_1|X_1)$. Observing that $I^{(1)}(X_1; Y_2|X_2) = H^{(1)}(Y_2|X_2) - H^{(1)}(Y_2|X_1, X_2) = H^{(1)}(Y_2|X_2) - H^{(2)}(Y_2|X_1, X_2)$, we immediately obtain that $H^{(1)}(Y_2|X_2) \leq H^{(2)}(Y_2|X_2)$ since $I^{(1)}(X_1; Y_2|X_2) \leq I^{(2)}(X_1; Y_2|X_2)$. Moreover, since $H^{(1)}(Y_1|X_1, X_2) = H^{(2)}(Y_1|X_1, X_2)$ and $I^{(1)}(X_2; Y_1|X_1) \leq I^{(2)}(X_2; Y_1|X_1)$, we have that $H^{(1)}(Y_1|X_1) \leq H^{(2)}(Y_1|X_1)$. Thus, the CVA condition is fulfilled. ∎

Remark 5: In [28], the existence of examples showing that the Shannon and CVA results are not equivalent was posed as an open question. The example below shows that the CVA condition is more general than the extended (one-sided) Shannon's symmetry condition (19). Together with Example 5, we conclude that the CVA result is more general than the Shannon result.

Example 6: Consider the TWC with $\mathcal{X}_1 = \mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1, 2\}$ and $\mathcal{X}_2 = \{0, 1\}$ and marginal distributions given by

$$[P_{Y_1|X_1,X_2}(\cdot|\cdot, 0)] = [P_{Y_1|X_1,X_2}(\cdot|\cdot, 1)] = [P_{Y_2|X_1,X_2}(\cdot|\cdot, 0)] = [P_{Y_2|X_1,X_2}(\cdot|\cdot, 1)] = \begin{pmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{pmatrix}.$$

Clearly, there are no relabeling functions for $\mathcal{Y}_1$ and $\mathcal{Y}_2$ which recover $[P_{Y_1|X_1,X_2}(\cdot|\cdot, 0)]$ after exchanging the labels of $X_1 = 0$ and $X_1 = 1$, so that the extended one-sided symmetry condition does not hold. To check the CVA condition, we first observe that $H(Y_j|X_1 = x_1, X_2 = x_2)$ does not depend on $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$; thus $H(Y_j|X_1, X_2)$ does not depend on P_{X_1, X_2} for $j = 1, 2$. Furthermore, for any given $P_{X_1,X_2}^{(1)} = P_{X_2}^{(1)} P_{X_1|X_2}^{(1)}$, consider $P_{X_1,X_2}^{(2)} = \tilde{P}_{X_1} P_{X_2}^{(1)}$ with $\tilde{P}_{X_1} = P_{\mathcal{X}_1}^{\mathrm{U}}$. Then, we have $I^{(1)}(X_1; Y_2|X_2) = \sum_{x_2} P_{X_2}^{(1)}(x_2) \cdot I^{(1)}(X_1; Y_2|X_2 = x_2) \leq \sum_{x_2} P_{X_2}^{(1)}(x_2) \cdot I^{(2)}(X_1; Y_2|X_2 = x_2) = I^{(2)}(X_1; Y_2|X_2)$, where the inequality follows from the fact that $P_{\mathcal{X}_1}^{\mathrm{U}}$ is the capacity-achieving input distribution for all one-way channels from users 1 to 2. On the other hand, since the matrices $[P_{Y_1|X_1,X_2}(\cdot|x_1, \cdot)]$, $x_1 \in \mathcal{X}_1$, are column permutations of each other, $\mathcal{I}(P_{X_2}, P_{Y_1|X_1=x_1,X_2})$ does not depend on $x_1 \in \mathcal{X}_1$ for any fixed $P_{X_2} \in \mathcal{P}(\mathcal{X}_2)$. One can then follow the proof of Theorem 1 to obtain that $I^{(1)}(X_2; Y_1|X_1) \leq I^{(2)}(X_2; Y_1|X_1)$. Now, since $H(Y_j|X_1, X_2)$ does not depend on the input distribution, we conclude that $H^{(1)}(Y_j|X_j) \leq H^{(2)}(Y_j|X_j)$ for $j = 1, 2$, and thus the CVA condition is satisfied.

Remark 6: The channel in the above example in fact also satisfies the conditions of Theorem 1. Nevertheless, the connection between the conditions of Theorem 1 and the CVA condition is still unclear.

We close this section by noting that the symmetry properties induced by our proposed conditions are not necessarily specific to two-user memoryless TWCs as we will see in Section IV. It is also worth mentioning that the proposed conditions can be used to investigate whether or not Shannon-type random coding schemes (under independent and non-adaptive inputs) provide tight bounds for other classical communication scenarios such as MACs with feedback and one-way compound channels. In particular, our conditions can be used to identify compound channels where the availability of channel state information at the transmitter (in addition to the receiver) cannot improve capacity.

III. Two-Way Symmetric Channels with Memory

We here consider point-to-point TWCs with memory whose inputs and outputs are related via functions F_1 and F_2 as follows:

$$Y_{1,i} = F_1(X_{1,i}, X_{2,i}, Z_{1,i}), \tag{20}$$

$$Y_{2,i} = F_2(X_{1,i}, X_{2,i}, Z_{2,i}), \tag{21}$$

where $\{(Z_{1,i}, Z_{2,i})\}_{i=1}^{\infty}$ is a stationary and ergodic noise process which is independent of the users' messages M_1 and M_2. Note that this model is a special case of the general model introduced in Section II-A; it is also a generalization of the discrete additive-noise TWC considered in [1].

We first state (without proof) an inner bound for arbitrary (time-invariant) functions F_1 and F_2. The bound can be proved via Shannon's standard random coding scheme (under non-adaptive independent inputs) for information stable one-way channels with memory, applied in each direction of the two-way transmission.

Lemma 1 (Inner Bound): For the channel described in (20) and (21), a rate pair (R_1, R_2) is achievable if there exist two sequences of codes (f_1^n, g_1) and (f_2^n, g_2) with message sets $\mathcal{M}_1 = \{1, 2, \ldots, 2^{nR_1}\}$ and $\mathcal{M}_2 = \{1, 2, \ldots, 2^{nR_2}\}$, respectively, such that

$$R_1 \leq \lim_{n \to \infty} \frac{1}{n} I(X_1^n; Y_2^n | X_2^n),$$

$$R_2 \leq \lim_{n \to \infty} \frac{1}{n} I(X_2^n; Y_1^n | X_1^n),$$

where the mutual information terms are evaluated under a sequence of product input probability distributions $\{P_{X_1^n} P_{X_2^n}\}_{n=1}^{\infty}$ and the inputs X_j^n are independent of $\{(Z_{1,i}, Z_{2,i})\}_{i=1}^{n}$, $j = 1, 2$.

We say that $F_j(X_1, X_2, Z_j)$ is invertible in Z_j if $F_j(x_1, x_2, \cdot)$ is one-to-one for any fixed $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$. Under this invertibility condition, we obtain the following corollary.

Corollary 3: If F_j is invertible in Z_j for $j = 1, 2$, a rate pair (R_1, R_2) is achievable if

$$R_1 \leq \lim_{n \to \infty} \frac{1}{n} H(Y_2^n | X_2^n) - \bar{H}(Z_2), \tag{22}$$

$$R_2 \leq \lim_{n \to \infty} \frac{1}{n} H(Y_1^n | X_1^n) - \bar{H}(Z_1), \tag{23}$$

for product distributions $\{P_{X_1^n} P_{X_2^n}\}_{n=1}^{\infty}$, where $\bar{H}(Z_j)$ denotes the entropy rate of the noise process $\{Z_{j,i}\}_{i=1}^{\infty}$ and the inputs X_j^n are independent of $\{(Z_{1,i}, Z_{2,i})\}_{i=1}^{n}$, $j = 1, 2$.

Proof: The proof follows from the fact that

$$I(X_1^n; Y_2^n | X_2^n) = H(Y_2^n | X_2^n) - H(Y_2^n | X_1^n, X_2^n)$$

$$= H(Y_2^n | X_2^n) - H(Z_2^n | X_1^n, X_2^n)$$

$$= H(Y_2^n | X_2^n) - H(Z_2^n),$$

where the second equality holds since F_2 is invertible in Z_2 and the last equality holds since the channel inputs are generated independently of the noise process $\{(Z_{2,1}, Z_{2,i})\}_{i=1}^{\infty}$. Applying a similar argument to $I(X_1^n; Y_2^n | X_2^n)$ completes the proof. ∎

If we further impose cardinality constraints on the alphabets, we can simplify the expressions in (22) and (23) as follows.

Corollary 4: Suppose that $|\mathcal{X}_2| = |\mathcal{Y}_1| = |\mathcal{Z}_1| = q_1$ and that $|\mathcal{X}_1| = |\mathcal{Y}_2| = |\mathcal{Z}_2| = q_2$ for some integers $q_1, q_2 \geq 2$. Then, a rate pair (R_1, R_2) is achievable if

$$R_1 \leq \log q_2 - \bar{H}(Z_2),$$

$$R_2 \leq \log q_1 - \bar{H}(Z_1).$$

Proof: The proof hinges on noting that $H(Y_j^n | X_j^n) \leq n \cdot \log q_j$ and that the uniform input distribution $P_{X_1^n, X_2^n} = (P_{\mathcal{X}_1}^{\mathsf{U}} P_{\mathcal{X}_2}^{\mathsf{U}})^n$ achieves the upper bound. More specifically, we have to show that if $P_{X_1^n, X_2^n}$ is the uniform distribution, then $P_{Y_{j^n} | X_j^n}(y_j^n | x_j^n)$ is uniform on $\mathcal{Y}_j^n$ for any given $X_j^n = x_j^n$, and hence $H(Y_j^n | X_j^n = x_j^n) = n \cdot \log q_j$. By symmetry, we only provide the details for $H(Y_2^n | X_2^n)$. Suppose that $P_{X_1^n, X_2^n}$ is the uniform distribution on $\mathcal{X}_1^n \times \mathcal{X}_2^n$. Let F_2^{-1} denote the inverse of F_2 for fixed (x_1, x_2) so that $z_2 = F_2^{-1}(x_1, x_2, y_2)$. Then, for any x_2^n we have

$$P_{Y_2^n | X_2^n}(y_2^n | x_2^n) = \sum_{x_1^n} P_{Y_2^n | X_1^n, X_2^n}(y_2^n | x_1^n, x_2^n) P_{X_1^n | X_2^n}(x_1^n | x_2^n)$$

$$= \left(\frac{1}{q_2}\right)^n \sum_{x_1^n} P_{Y_2^n | X_1^n, X_2^n}(F_2(x_1^n, x_2^n, z_2^n) | x_1^n, x_2^n)$$

$$= \left(\frac{1}{q_2}\right)^n \sum_{x_1^n} P_{Z_2^n|X_1^n,X_2^n}(F_2^{-1}(x_1^n, x_2^n, y_2^n)|x_1^n, x_2^n)$$

$$= \left(\frac{1}{q_2}\right)^n \sum_{z_2^n} P_{Z_2^n}(z_2^n) \qquad (24)$$

$$= \left(\frac{1}{q_2}\right)^n,$$

where (24) holds since (X_1^n, X_2^n) is independent of Z_2^n and F_2 is onto in Z_2 due to the cardinality constraint. Clearly, $P_{Y_2^n|X_2^n=x_2^n}$ is the uniform distribution for any x_2^n, implying that $H(Y_2^n|X_2^n) = n \log q_2$. ∎

Next we consider ISD-TWCs as in Example 3 and [28], but with the assumption that the noise process $\{(Z_{1,i}, Z_{2,i})\}_{i=1}^\infty$ can have memory. Note that any ISD-TWC with memory is a special case of the system model in (20) and (21) satisfying the invertibility condition in Z_1 and Z_2. Thus, Corollary 3 applies to ISD-TWCs with memory to obtain the following result.

Corollary 5: For the ISD-TWC with memory, a rate pair (R_1, R_2) is achievable if

$$R_1 \leq \lim_{n\to\infty} \frac{1}{n} \max_{P_{X_1^n}} H(\tilde{h}_2(X_1^n, Z_2^n)) - \bar{H}(Z_2),$$

$$R_2 \leq \lim_{n\to\infty} \frac{1}{n} \max_{P_{X_2^n}} H(\tilde{h}_1(X_2^n, Z_1^n)) - \bar{H}(Z_1),$$

where $\bar{H}(Z_j)$ denotes the entropy rate of the process $\{Z_{j,i}\}_{i=1}^\infty$ for $j = 1, 2$.

We note that Corollary 4 also applies to ISD-TWCs with memory under identical alphabet size constraints so that any rate pair in $\{(R_1, R_2 : R_1 \leq \log q_2 - \bar{H}(Z_2), R_2 \leq \log q_1 - \bar{H}(Z_1))\}$ is achievable for ISD-TWCs with memory. We next derive converses to Corollaries 4 and 5.

Lemma 2 (Outer Bound for Noise-Invertible TWCs with Memory): Suppose that $|\mathcal{Y}_j| = q_j$ for some integer $q_j \geq 2$. If F_j is invertible in Z_j for $j = 1, 2$, any achievable rate pair (R_1, R_2) must satisfy

$$R_1 \leq \log q_2 - \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n H(Z_{2,i}|Z_1^{i-1}, Z_2^{i-1}),$$

$$R_2 \leq \log q_1 - \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n H(Z_{1,i}|Z_1^{i-1}, Z_2^{i-1}),$$

where the limits exist because $\{(Z_{1,i}, Z_{2,i})\}_{i=1}^\infty$ is stationary.

Proof: For an achievable rate pair (R_1, R_2), we have

$$nR_1 = H(M_1|M_2)$$

$$= I(M_1; Y_2^n|M_2) + H(M_1|Y_2^n, M_2)$$

$$\leq I(M_1; Y_2^n|M_2) + n\epsilon_n \tag{25}$$

$$= \sum_{i=1}^{n} \left[H(Y_{2,i}|M_2, Y_2^{i-1}) - H(Y_{2,i}|M_1, M_2, Y_2^{i-1}) \right] + n\epsilon_n \tag{26}$$

$$\leq \sum_{i=1}^{n} \left[\log q_2 - H(Y_{2,i}|M_1, M_2, Y_2^{i-1}) \right] + n\epsilon_n \tag{27}$$

$$\leq \sum_{i=1}^{n} \left[\log q_2 - H(Y_{2,i}|M_1, M_2, Y_1^{i-1}, Y_2^{i-1}, X_{1,i}, X_{2,i}) \right] + n\epsilon_n$$

$$= \sum_{i=1}^{n} \left[\log q_2 - H(Z_{2,i}|M_1, M_2, Y_1^{i-1}, Y_2^{i-1}, X_1^i, X_2^i) \right] + n\epsilon_n \tag{28}$$

$$= \sum_{i=1}^{n} \left[\log q_2 - H(Z_{2,i}|M_1, M_2, Y_1^{i-1}, Y_2^{i-1}, X_1^i, X_2^i, Z_1^{i-1}, Z_2^{i-1}) \right] + n\epsilon_n \tag{29}$$

$$= \sum_{i=1}^{n} \left[\log q_2 - H(Z_{2,i}|Z_1^{i-1}, Z_2^{i-1}) \right] + n\epsilon_n \tag{30}$$

$$= n \log q_2 - \sum_{i=1}^{n} H(Z_{2,i}|Z_1^{i-1}, Z_2^{i-1}) + n\epsilon_n, \tag{31}$$

where (25) is due to Fano's inequality with $\epsilon_n \to 0$ as $n \to \infty$, (27) follows from $|\mathcal{Y}_2| = q_2$, (28) and (29) hold since F_j is invertible in Z_j given $(X_{1,i}, X_{2,i})$, and (30) holds since

$$H(Z_{2,i}|Z_1^{i-1}, Z_2^{i-1}) = H(Z_{2,i}|M_1, M_2, Z_1^{i-1}, Z_2^{i-1}) \tag{32}$$

$$= H(Z_{2,i}|M_1, M_2, Z_1^{i-1}, Z_2^{i-1}, X_{1,1}, X_{2,1}) \tag{33}$$

$$= H(Z_{2,i}|M_1, M_2, Z_1^{i-1}, Z_2^{i-1}, X_{1,1}, X_{2,1}, Y_{1,1}, Y_{2,1}) \tag{34}$$

$$= H(Z_{2,i}|M_1, M_2, Z_1^{i-1}, Z_2^{i-1}, X_1^2, X_2^2, Y_{1,1}, Y_{2,1}) \tag{35}$$

$$= H(Z_{2,i}|M_1, M_2, Z_1^{i-1}, Z_2^{i-1}, X_1^i, X_2^i, Y_1^{i-1}, Y_2^{i-1}) \tag{36}$$

where (32) is due to the fact that $\{(Z_{1,i}, Z_{2,i})\}_{i=1}^{\infty}$ is independent of (M_1, M_2), (33) and (35) hold since $X_{j,i} = f_{j,i}(M_j, Y_j^{i-1})$ for $j = 1, 2$, (34) follows from the identity $Y_{j,i} = F_j(X_{1,i}, X_{2,i}, Z_{j,i})$, and (36) is obtained by recursively using the same argument as in (33)-(35). Similarly, we have

$$nR_2 \leq n \log q_1 - \sum_{i=1}^{n} H(Z_{1,i}|Z_1^{i-1}, Z_2^{i-1}) + n\hat{\epsilon}_n. \tag{37}$$

The proof is completed by dividing both sides of (31) and (37) by n and letting $n \to \infty$. $\blacksquare$

Lemma 3 (Outer Bound for ISD-TWCs with Memory): For the ISD-TWC with memory, any achievable rate pair (R_1, R_2) must satisfy

$$R_1 \leq \lim_{n \to \infty} \frac{1}{n} \left[\max_{P_{X_1^n}} H(\tilde{h}_2(X_1^n, Z_2^n)) - \sum_{i=1}^{n} H(Z_{2,i}|Z_1^{i-1}, Z_2^{i-1}) \right],$$

$$R_2 \leq \lim_{n \to \infty} \frac{1}{n} \left[\max_{P_{X_2^n}} H(\tilde{h}_1(X_2^n, Z_1^n)) - \sum_{i=1}^{n} H(Z_{1,i}|Z_1^{i-1}, Z_2^{i-1}) \right].$$

Proof: The proof is similar to the proof of the previous lemma and hence the details are omitted. The main difference is that the first term in (26) is now upper bounded as follows

$$\sum_{i=1}^{n} H(Y_{2,i}|M_2, Y_2^{i-1}) = \sum_{i=1}^{n} H(h_2(X_{2,i}, T_{2,i})|M_2, Y_2^{i-1}, X_2^i, T_2^{i-1})$$

$$\leq \sum_{i=1}^{n} H(T_{2,i}|T_2^{i-1})$$

$$= H(T_2^n)$$

$$\leq \max_{P_{X_1^n}} H(\tilde{h}_2(X_1^n, Z_2^n)),$$

where the first equality holds since X_2^i is a function of M_2 and Y_2^{i-1} and $Y_2 = h_2(X_2, T_2)$ is invertible in T_2 given X_2. ∎

Based on the preceding inner and outer bounds, the capacity region for two classes of TWCs with memory (whose component noise processes are independent of each other) can be exactly determined as follows.

Theorem 9: For a TWC with memory such that $\{Z_{1,i}\}_{i=1}^{\infty}$ and $\{Z_{2,i}\}_{i=1}^{\infty}$ are stationary ergodic and mutually independent, F_j is invertible in Z_j for $j = 1, 2$, and $|\mathcal{X}_2| = |\mathcal{Y}_1| = |\mathcal{Z}_1| = q_1$ and $|\mathcal{X}_1| = |\mathcal{Y}_2| = |\mathcal{Z}_2| = q_2$ for some integers $q_1, q_2 \geq 2$, the capacity region is given by

$$\mathcal{C} = \left\{ (R_1, R_2) : R_1 \leq \log q_2 - \bar{H}(Z_2), R_2 \leq \log q_1 - \bar{H}(Z_1) \right\}. \tag{38}$$

Theorem 10: For a ISD-TWC with memory such that $\{Z_{1,i}\}_{i=1}^{\infty}$ and $\{Z_{2,i}\}_{i=1}^{\infty}$ are stationary ergodic and mutually independent, the capacity region is given by

$$\mathcal{C} = \left\{ (R_1, R_2) : R_1 \leq \lim_{n \to \infty} \frac{1}{n} \max_{P_{X_1^n}} H(\tilde{h}_2(X_1^n, Z_2^n)) - \bar{H}(Z_2), \right.$$

$$\left. R_2 \leq \lim_{n \to \infty} \frac{1}{n} \max_{P_{X_2^n}} H(\tilde{h}_1(X_2^n, Z_1^n)) - \bar{H}(Z_1) \right\}. \tag{39}$$

Remark 7: Theorem 10 generalizes [28, Corollary 1] for memoryless ISD-TWCs. If one further has $|\mathcal{X}_2| = |\mathcal{T}_1| = |\mathcal{Z}_1| = q_1$ and $|\mathcal{X}_1| = |\mathcal{T}_2| = |\mathcal{Z}_2| = q_2$ for some integers $q_1, q_2 \geq 2$,

then $\lim_{n\to\infty} \frac{1}{n} \max_{P_{X_1^n}} H(\tilde{h}_2(X_1^n, Z_2^n)) = \log q_1$ and that $\lim_{n\to\infty} \frac{1}{n} \max_{P_{X_2^n}} H(\tilde{h}_1(X_2^n, Z_1^n)) = \log q_2$.

The next example shows that if the noise processes are *dependent*, then Shannon's random coding scheme is not optimal.

Example 7 (Adaptation is Useful): Let $q_1 = q_2 = 2$ and suppose that the channel is given by

$$Y_{1,i} = F_1(X_{1,i}, X_{2,i}, Z_{1,i}) = X_{1,i} \oplus_2 X_{2,i} \oplus_2 Z_{1,i},$$

$$Y_{2,i} = F_2(X_{1,i}, X_{2,i}, Z_{2,i}) = X_{1,i} \oplus_2 X_{2,i} \oplus_2 Z_{2,i},$$

where $\{Z_{1,i}\}_{i=1}^n$ is assumed to be memoryless with $Z_{1,i}$ uniformly distributed on $\mathcal{Z}_1 = \{0,1\}$ for $i = 1, 2, \ldots, n$, and $\{Z_{2,i}\}_{i=1}^n$ is given by $Z_{2,1} = 0$ and $Z_{2,i} = Z_{1,i-1}$ for $i = 2, 3, \ldots, n$. Since the functions F_1 and F_2 are invertible in Z_1 and Z_2, the outer bound in Lemma 2 indicates that

$$R_1 \le \log 2 - \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n H(Z_{2,i}|Z_1^{i-1}, Z_2^{i-1}) = 1 - 0 = 1$$

$$R_2 \le \log 2 - \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n H(Z_{1,i}|Z_1^{i-1}, Z_2^{i-1}) = 1 - H(Z_{1,i}) = 0.$$

We claim that the rate pair $(R_1, R_2) = (1,0)$ can be achieved by an adaptive coding scheme. Let $\{M_{1,i}\}_{i=1}^n$ denote the binary messages to be sent from users 1 to 2. For $i = 1, 2, \ldots, n$, set the encoding function of user 1 as $X_{1,i} = f_{1,i}(\{M_{1,i}\}_{i=1}^n, Y_1^{i-1}) \triangleq M_{1,i} \oplus_2 X_{1,i-1} \oplus_2 Y_{1,i-1}$ with initial conditions $X_{1,0} = X_{2,0} = Y_{1,0} = 0$, and set the encoder output of user 2 to be zero, i.e., $X_{2,i} = 0$ for all i. With this coding scheme, the received signal at user 2 is given by

$$Y_{2,i} = X_{1,i} \oplus_2 X_{2,i} \oplus_2 Z_{2,i}$$

$$= M_{1,i} \oplus_2 X_{1,i-1} \oplus_2 Y_{1,i-1} \oplus_2 Z_{2,i}$$

$$= M_{1,i} \oplus_2 X_{1,i-1} \oplus_2 X_{1,i-1} \oplus_2 Z_{1,i-1} \oplus_2 Z_{2,i} = M_{1,i},$$

and thus the rate pair $(1,0)$ is achievable. This achievability result together with the outer bound imply that the channel capacity is given by $\mathcal{C} = \{(R_1, R_2) : R_1 \le 1, R_2 = 0\}$. However, the Shannon-type random coding scheme only provides $R_1 \le 1 - \bar{H}(Z_2) = 0$ and $R_2 \le 1 - \bar{H}(Z_2) = 0$ by Corollary 4.

IV. MULTIPLE ACCESS/DEGRADED BROADCAST TWCS

This section considers a three-user two-way communication scenario combining multiaccess and broadcasting. We first introduce the channel model and derive inner and outer bounds for
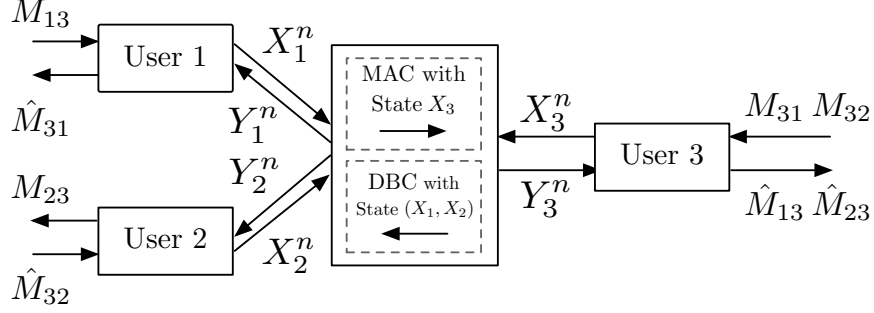
Fig. 6: The information flow of MA/DB TWCs.

the capacity region. Then, sufficient conditions for the two bounds to coincide are provided, along with illustrative examples.

A. Channel Model

Two-way communication over a discrete additive-noise MA/DB TWC comprises three users as depicted in Fig. 6. Users 1 and 2 want to transmit messages M_{13} and M_{23}, respectively, to user 3 through the TWC that acts as a MAC in the forward direction. User 3 wishes to broadcast messages M_{31} and M_{32} to users 1 and 2, respectively, through the TWC that acts as a DBC in the reverse direction. The messages are assumed to be independent of each other and uniformly distributed over their alphabets. The joint distribution of all the variables for n channel uses is given by

$$P_{M_{\{13,23,31,32\}},X^n_{\{1,2,3\}},Y^n_{\{1,2,3\}}} = P_{M_{13}}P_{M_{23}}P_{M_{31}}P_{M_{32}} \cdot \left(\prod_{i=1}^n P_{X_{1,i}|M_{13},Y_1^{i-1}}\right) \cdot \left(\prod_{i=1}^n P_{X_{2,i}|M_{23},Y_2^{i-1}}\right) \cdot$$
$$\left(\prod_{i=1}^n P_{X_{3,i}|M_{\{31,32\}},Y_3^{i-1}}\right) \cdot \left(\prod_{i=1}^n P_{Y_{1,i},Y_{2,i},Y_{3,i}|X^i_{\{1,2,3\}},Y^{i-1}_{\{1,2,3\}}}\right),$$

where $M_{\{13,23,31,32\}} \triangleq \{M_{13}, M_{23}, M_{31}, M_{32}\}$, $X^n_{\{1,2,3\}} \triangleq \{X_1^n, X_2^n, X_3^n\}$, and $Y^n_{\{1,2,3\}} \triangleq \{Y_1^n, Y_2^n, Y_3^n\}$. Thus, the n transmissions can be described by the sequence of input-output conditional probabilities $\{P_{Y_{1,i},Y_{2,i},Y_{3,i}|X^i_{\{1,2,3\}},Y^{i-1}_{\{1,2,3\}}}\}_{i=1}^n$.

To simplify our analysis, we assume that the channel is memoryless in the sense that given current channel inputs, the current channel outputs are independent of past signals, i.e., $P_{Y_{1,i},Y_{2,i},Y_{3,i}|X^i_{\{1,2,3\}},Y^{i-1}_{\{1,2,3\}}} = P_{Y_{1,i},Y_{2,i},Y_{3,i}|X_{1,i},X_{2,i},X_{3,i}}$ for all i. Furthermore, the two directions of transmission are assumed to interact in a way such that $P_{Y_{1,i},Y_{2,i},Y_{3,i}|X_{1,i},X_{2,i},X_{3,i}} = P_{Y_{1,i},Y_{2,i}|X_{1,i},X_{2,i},X_{3,i}}P_{Y_{3,i}|X_{1,i},X_{2,i},X_{3,i}}$. Let all channel input and output alphabets other than $\mathcal{Y}_3$

equal $\mathcal{Q} \triangleq \{0, 1, ..., q-1\}$ for some $q \geq 2$. The MA/DB TWC is defined by the transition probability $P_{Y_3|X_1,X_2,X_3}$ in the MA direction and the transmission equations in the DB direction are given by

$$Y_{1,i} = X_{1,i} \oplus_q X_{3,i} \oplus_q Z_{1,i}, \tag{40}$$

$$Y_{2,i} = X_{2,i} \oplus_q X_{3,i} \oplus_q Z_{1,i} \oplus_q Z_{2,i}, \tag{41}$$

for $i = 1, 2, ..., n$, where $Z_{1,i}, Z_{2,i} \in \mathcal{Q}$ denote additive noise variables, the components of the memoryless and independent noise processes $\{Z_{1,i}\}_{i=1}^n$ and $\{Z_{2,i}\}_{i=1}^n$, respectively. We also assume that the channel noise processes are independent of all users' messages. Thus, the channel transition probability of this MA/DB TWC at time i can be written as

$$P_{Y_{1,i},Y_{2,i},Y_{3,i}|X_1^i,X_2^i,X_3^i,Y_1^{i-1},Y_2^{i-1},Y_3^{i-1}}(y_{1,i}, y_{2,i}, y_{3,i}|x_1^i, x_2^i, x_3^i, y_1^{i-1}, y_2^{i-1}, y_3^{i-1})$$

$$= P_{Y_{1,i},Y_{2,i},Y_{3,i}|X_{1,i},X_{2,i},X_{3,i}}(y_{1,i}, y_{2,i}, y_{3,i}|x_{1,i}, x_{2,i}, x_{3,i})$$

$$= P_{Y_{3,i}|X_{1,i},X_{2,i},X_{3,i}}(y_{3,i}|x_{1,i}, x_{2,i}, x_{3,i}) P_{Y_{1,i}|X_{1,i},X_{2,i},X_{3,i},Y_{3,i}}(y_{1,i}|x_{1,i}, x_{2,i}, x_{3,i}, y_{3,i})$$

$$\quad \cdot P_{Y_{2,i}|X_{1,i},X_{2,i},X_{3,i},Y_{1,i},Y_{3,i}}(y_{2,i}|x_{1,i}, x_{2,i}, x_{3,i}, y_{1,i}, y_{3,i})$$

$$= P_{Y_3|X_1,X_2,X_3}(y_{3,i}|x_{1,i}, x_{2,i}, x_{3,i}) \cdot \Pr\{Z_{1,i} = y_{1,i} \ominus_q x_{1,i} \ominus_q x_{3,i}\}$$

$$\quad \cdot \Pr\{Z_{2,i} = y_{2,i} \ominus_q x_{2,i} \ominus_q x_{3,i} \ominus_q (y_{1,i} \ominus_q x_{1,i} \ominus_q x_{3,i})\}$$

$$= P_{Y_3|X_1,X_2,X_3}(y_{3,i}|x_{1,i}, x_{2,i}, x_{3,i}) P_{Z_1}(y_{1,i} \ominus_q x_{1,i} \ominus_q x_{3,i}) P_{Z_2}(y_{2,i} \ominus_q x_{2,i} \ominus_q y_{1,i} \oplus_q x_{1,i}),$$

where $\ominus_q$ denotes modulo-q subtraction.

We next define channel codes, achievable rates, and channel capacity for the MA/DB TWC.

Definition 4: An $(n, R_{13}, R_{23}, R_{31}, R_{32})$ channel code for the memoryless MA/DB TWC consists of four message sets $\mathcal{M}_{13} = \{1, 2, ..., 2^{nR_{13}}\}$, $\mathcal{M}_{23} = \{1, 2, ..., 2^{nR_{23}}\}$, $\mathcal{M}_{31} = \{1, 2, ..., 2^{nR_{31}}\}$, $\mathcal{M}_{32} = \{1, 2, ..., 2^{nR_{32}}\}$, three sequences of encoding functions: $f_1^n = (f_{1,1}, f_{1,2}, ..., f_{1,n})$, $f_2^n = (f_{2,1}, f_{2,2}, ..., f_{2,n})$, $f_3^n = (f_{3,1}, f_{3,2}, ..., f_{3,n})$ such that

$$X_{1,1} = f_{1,1}(M_{13}), \qquad X_{1,i} = f_{1,i}(M_{13}, Y_1^{i-1}), \tag{42}$$

$$X_{2,1} = f_{2,1}(M_{23}), \qquad X_{2,i} = f_{2,i}(M_{23}, Y_2^{i-1}), \tag{43}$$

$$X_{3,1} = f_{3,1}(M_{31}, M_{32}), \quad X_{3,i} = f_{3,i}(M_{31}, M_{32}, Y_3^{i-1}), \tag{44}$$

for $i = 2, ..., n$, and three decoding functions g_1, g_2, and g_3, such that $\hat{M}_{31} = g_1(M_{13}, Y_1^n)$, $\hat{M}_{32} = g_2(M_{23}, Y_2^n)$, and $(\hat{M}_{13}, \hat{M}_{23}) = g_3(M_{31}, M_{32}, Y_3^n)$.

When messages are encoded via the channel code, the probability of decoding error is defined as $P_{\mathrm{e}}^{(n)}(f_1^n, f_2^n, f_3^n, g_1, g_2, g_3) = \Pr\{\hat{M}_{13} \neq M_{13} \text{ or } \hat{M}_{23} \neq M_{23} \text{ or } \hat{M}_{31} \neq M_{31} \text{ or } \hat{M}_{32} \neq M_{32}\}$.

Definition 5: A rate quadruple $(R_{13}, R_{23}, R_{31}, R_{32})$ is said to be achievable for the memoryless MA/DB TWC if there exists a sequence of $(n, R_{13}, R_{23}, R_{31}, R_{32})$ codes with $\lim_{n \to \infty} P_{\mathrm{e}}^{(n)} = 0$.

Definition 6: The capacity region $\mathcal{C}^{\mathrm{MA\text{-}DBC}}$ of the memoryless MA/DB TWC is the closure of the convex hull of all achievable rate quadruples $(R_{13}, R_{23}, R_{31}, R_{32})$.

B. Capacity Inner and Outer Bounds for the Memoryless MA/DB TWCs

Let $\mathcal{R}^{\mathrm{MA\text{-}DBC}}(P_{X_1,X_2,X_3,V}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2})$ denote the set of rate quadruples $(R_{13}, R_{23}, R_{31}, R_{32})$ which satisfy the constraints

$$R_{13} \leq I(X_1; Y_3 | X_2, X_3),$$

$$R_{23} \leq I(X_2; Y_3 | X_1, X_3),$$

$$R_{13} + R_{23} \leq I(X_1, X_2; Y_3 | X_3),$$

$$R_{31} \leq I(X_3; X_3 \oplus_q Z_1 | V),$$

$$R_{32} \leq I(V; X_3 \oplus_q Z_1 \oplus_q Z_2),$$

where V is an auxiliary random variable with alphabet $\mathcal{V}$ such that $|\mathcal{V}| \leq q + 1$ and the mutual information terms are evaluated according to the joint probability distribution $P_{X_1,X_2,X_3,V,Y_3,Z_1,Z_2} = P_{X_1,X_2,X_3,V} P_{Y_3|X_1,X_2,X_3} P_{Z_1} P_{Z_2}$. We next establish a Shannon-type inner bound and an outer bound for the capacity of MA/DB TWCs in Theorems 11 and 12, respectively. Note that the achievable scheme in Theorem 11 is given by combining Shannon's standard (non-adaptive) coding schemes for the MAC [8, Theorem 4.2] and the DBC [8, Theorem 5.2], and hence the proof is omitted here. The derivation for the outer bound is given in Appendix B.

Theorem 11 (Inner Bound): For a memoryless MA/DB TWC with MA transition probability $P_{Y_3|X_1,X_2,X_3}$ and DB noise distributions P_{Z_1} and P_{Z_2}, any rate quadruple $(R_{13}, R_{23}, R_{31}, R_{32}) \in \mathcal{C}_{\mathrm{I}}^{\mathrm{MA\text{-}DBC}}(P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2})$ is achievable, where

$$\mathcal{C}_{\mathrm{I}}^{\mathrm{MA\text{-}DBC}}(P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2})$$

$$\triangleq \overline{\mathrm{co}} \left(\bigcup_{P_{X_1} P_{X_2} P_{V,X_3}} \mathcal{R}^{\mathrm{MA\text{-}DBC}}(P_{X_1} P_{X_2} P_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2}) \right).$$

Theorem 12 (Outer Bound): For a memoryless MA/DB TWC with MA transition probability $P_{Y_3|X_1,X_2}$ and DB noise distributions P_{Z_1} and P_{Z_2}, all achievable rate quadruples $(R_{13}, R_{23}, R_{31}, R_{32})$ belong to $\mathcal{C}_{\text{O}}^{\text{MA-DBC}}(P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2})$, where

$$\mathcal{C}_{\text{O}}^{\text{MA-DBC}}(P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2})$$
$$\triangleq \overline{\text{co}} \left(\bigcup_{P_{X_1,X_2,X_3,V}} \mathcal{R}^{\text{MA-DBC}}(P_{X_1,X_2,X_3,V}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2}) \right).$$

C. Conditions for the Tightness of the Inner and Outer Bounds

The inner and outer bounds derived in the previous section are of the same form but have different restrictions on the joint distribution $P_{X_1,X_2,X_3,V}$, and hence they do not match. Here, we establish conditions under which the two bounds have matching input distributions, implying that they coincide and yield the capacity region. The proofs of Theorems 13-15 are given in Appendices C-E, respectively.

Theorem 13: The inner and outer capacity bounds in Theorems 11 and 12 coincide if for every conditional input distribution $P_{X_1,X_2|X_3}^{(1)}$, there exists a product input distribution $P_{X_1,X_2|X_3}^{(2)} = \tilde{P}_{X_1}\tilde{P}_{X_2}$ (which depends on $P_{X_1,X_2|X_3}^{(1)}$) such that

$$I^{(1)}(X_1; Y_3|X_2, X_3 = x_3) \leq I^{(2)}(X_1; Y_3|X_2, X_3 = x_3) \tag{45}$$

$$I^{(1)}(X_2; Y_3|X_1, X_3 = x_3) \leq I^{(2)}(X_2; Y_3|X_1, X_3 = x_3) \tag{46}$$

$$I^{(1)}(X_1, X_2; Y_3|X_3 = x_3) \leq I^{(2)}(X_1, X_2; Y_3|X_3 = x_3) \tag{47}$$

hold for all $x_3 \in \mathcal{X}_3$. Under this condition, the capacity region is given by

$$\mathcal{C}^{\text{MA-DBC}} = \overline{\text{co}} \left(\bigcup_{P_{X_1} P_{X_2} P_{V,X_3}} \mathcal{R}^{\text{MA-DBC}}\left(P_{X_1} P_{X_2} P_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2} \right) \right).$$

A special case of the above theorem is when $\tilde{P}_{X_1}\tilde{P}_{X_2}$ does not depend on $P_{X_1,X_2|X_3}$. This case may happen when $P_{Y_3|X_1,X_2,X_3}$ has a strong symmetry property.

Corollary 6: The inner and outer capacity bounds in Theorems 11 and 12 coincide if there exists an input distributions $P_{X_1,X_2}^{(2)} = P_{X_1}^* P_{X_2}^*$ such that for all $P_{X_1,X_2|X_3}^{(1)}$ and $x_3 \in \mathcal{X}_3$ the inequalities given in (45)-(47) hold. In this case, the capacity region is given by

$$\mathcal{C}^{\text{MA-DBC}} = \overline{\text{co}} \left(\bigcup_{P_{V,X_3}} \mathcal{R}^{\text{MA-DBC}}\left(P_{X_1}^* P_{X_2}^* P_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2} \right) \right).$$

The next result is derived by treating the channel as a composition of state-dependent one-way channels.

Theorem 14: The inner and outer capacity bounds in Theorems 11 and 12 coincide if the following conditions hold:

(i) There exists $P_{X_1}^* \in \mathcal{P}(\mathcal{X}_1)$ such that $\arg\max_{P_{X_1|X_2=x_2,X_3=x_3}} I(X_1; Y_3|X_2 = x_2, X_3 = x_3) = P_{X_1}^*$ for all $x_2 \in \mathcal{X}_2$ and $x_3 \in \mathcal{X}_3$ and $\mathcal{I}(P_{X_1}^*, P_{Y_3|X_1,X_2=x_2,X_3=x_3})$ does not depend on x_2 for every fixed x_3;

(ii) For any $P_{X_2} \in \mathcal{P}(\mathcal{X}_2)$, $\mathcal{I}(P_{X_2}, P_{Y_3|X_1=x_1,X_2,X_3=x_3})$ does not depend on $x_1 \in \mathcal{X}_1$ and $x_3 \in \mathcal{X}_3$;

(iii) For any fixed P_{X_1,X_2}, we have $\mathcal{I}(P_{X_1,X_2}, P_{Y_3|X_1,X_2,X_3=x_3})$ does not depend on $x_3 \in \mathcal{X}_3$, and for each $x_3 \in \mathcal{X}_3$ we have $\mathcal{I}(P_{X_1,X_2}, P_{Y_3|X_1,X_2,X_3=x_3}) \leq \mathcal{I}(P_{X_1}^* P_{X_2}, P_{Y_3|X_1,X_2,X_3=x_3})$, where $P_{X_1}^*$ is given by condition (i) and $P_{X_2}(x_2) = \sum_{x_1} P_{X_1,X_2}(x_1, x_2)$ for $x_2 \in \mathcal{X}_2$.

Under this condition, the capacity region is given by

$$\mathcal{C}^{\text{MA-DBC}} = \overline{\text{co}} \left(\bigcup_{P_{X_2} P_{V,X_3}} \mathcal{R}^{\text{MA-DBC}} \left(P_{X_1}^* P_{X_2} P_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2} \right) \right).$$

Next, we derive our last sufficient condition by generalizing Shannon's condition (in Proposition 1) to the three-user setting. This new condition is easier to verify than the previous ones.

Theorem 15: The inner and outer capacity bounds in Theorems 11 and 12 coincide if the following conditions hold:

(i) For any relabeling $\tau_{x_1',x_1''}^{\mathcal{X}_1}$ on $\mathcal{X}_1$, there exists a permutation $\pi^{\mathcal{Y}_3}[x_1', x_1'']$ on $\mathcal{Y}_3$ such that for all x_1, x_2, x_3, and y_3, we have

$$P_{Y_3|X_1,X_2,X_3}(y_3|x_1,x_2,x_3) = P_{Y_3|X_1,X_2,X_3}\left(\pi^{\mathcal{Y}_3}[x_1', x_1''](y_3) \big| \tau_{x_1',x_1''}^{\mathcal{X}_1}(x_1), x_2, x_3\right) \tag{48}$$

(ii) For any relabeling $\tau_{x_2',x_2''}^{\mathcal{X}_2}$ on $\mathcal{X}_2$, there exists a permutation on $\pi^{\mathcal{Y}_3}[x_2', x_2'']$ on $\mathcal{Y}_3$ such that for all x_1, x_2, x_3, and y_3, we have

$$P_{Y_3|X_1,X_2,X_3}(y_3|x_1,x_2,x_3) = P_{Y_3|X_1,X_2,X_3}\left(\pi^{\mathcal{Y}_3}[x_1', x_1''](y_3) \big| x_1, \tau_{x_2',x_2''}^{\mathcal{X}_2}(x_2), x_3\right). \tag{49}$$

Under these conditions, the capacity region is given by

$$\mathcal{C}^{\text{MA-DBC}} = \overline{\text{co}} \left(\bigcup_{P_{V,X_3}} \mathcal{R}^{\text{MA-DBC}} \left(P_{\mathcal{X}_1}^U P_{\mathcal{X}_2}^U P_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2} \right) \right), \tag{50}$$

where $P_{\mathcal{X}_i}^U$ denotes uniform probability distribution on $\mathcal{X}_i$ for $i = 1, 2$.

D. Examples

We next illustrate Theorems 13-15 via three examples.

Example 8 (Additive-Noise MA/DB TWC): Consider a discrete memoryless additive-noise MA/DB TWC in which the inputs and outputs of the DBC are described by (40) and (41) and the inputs and outputs of MAC are related via

$$Y_{3,i} = X_{1,i} \oplus_q X_{2,i} \oplus_q X_{3,i} \oplus_q Z_{3,i}, \tag{51}$$

where $\{Z_{3,i}\}_{i=1}^{\infty}$ with $Z_{3,i} \in \mathcal{Q}$ is a discrete memoryless noise process which is independent of all user messages and the noise processes $\{Z_{1,i}\}_{i=1}^{\infty}$ and $\{Z_{2,i}\}_{i=1}^{\infty}$. For any $x_3 \in \mathcal{X}_3$, we have the following bounds:

$$I(X_1; Y_3 | X_2, X_3 = x_3) = H(Y_3 | X_2, X_3 = x_3) - H(Y_3 | X_1, X_2, X_3 = x_3) \le \log_2 q - H_{\mathsf{b}}(Z_3),$$

$$I(X_2; Y_3 | X_1, X_3 = x_3) = H(Y_3 | X_1, X_3 = x_3) - H(Y_3 | X_1, X_2, X_3 = x_3) \le \log_2 q - H_{\mathsf{b}}(Z_3),$$

$$I(X_1, X_2; Y_3 | X_3 = x_3) = H(Y_3 | X_3 = x_3) - H(Y_3 | X_1, X_2, X_3 = x_3) \le \log_2 q - H_{\mathsf{b}}(Z_3),$$

where equalities hold when $P_{X_1, X_2} = P_{\mathcal{X}_1}^{\mathsf{U}} P_{\mathcal{X}_2}^{\mathsf{U}}$. Choosing $\tilde{P}_{X_1} = P_{\mathcal{X}_1}^{\mathsf{U}}$ and $\tilde{P}_{X_2} = P_{\mathcal{X}_2}^{\mathsf{U}}$, it is clear that (45)-(47) in Theorem 13 hold, and hence the capacity region given by

$$\mathcal{C}^{\text{MA-DBC}} = \overline{\text{co}} \left(\bigcup_{P_{V,X_3}} \mathcal{R}^{\text{MA-DBC}} \Big(P_{\mathcal{X}_1}^{\mathsf{U}} P_{\mathcal{X}_2}^{\mathsf{U}} P_{U,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2} \Big) \right)$$

$$= \overline{\text{co}} \Bigg(\bigcup_{P_{V,X_3}} \Big\{ (R_{13}, R_{23}, R_{31}, R_{32}) : R_{13} + R_{23} \le \log_2 q - H_{\mathsf{b}}(Z_3), $$

$$R_{31} \le I(X_1; X_3 \oplus_2 Z_1 | V), R_{32} \le I(X_2 \oplus Z_1 \oplus Z_2; V) \Big\} \Bigg).$$

Example 9: Suppose that $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}_3 = \{0,1\}$, $\mathcal{Y}_1 = \mathcal{Y}_2 = \{0,1\}$, and $\mathcal{Y}_3 = \{0,1,2\}$. We consider a discrete memoryless MA/DB TWC in which the DB direction is described by (40) and (41) and the channel transition matrix $[P_{Y_3|X_1,X_2,X_3}(\cdot|\cdot,\cdot,\cdot)]$ for the MA direction is given

by

$$
\begin{array}{c}
\quad\ \ 0 \qquad\ \ 1 \qquad\ \ 2 \\
\begin{array}{c}
000 \\
100 \\
010 \\
110 \\
001 \\
101 \\
011 \\
111
\end{array}
\left(\begin{array}{ccc}
1-\varepsilon & 0 & \varepsilon \\
1-\varepsilon & 0 & \varepsilon \\
0 & 1-\varepsilon & \varepsilon \\
0 & 1-\varepsilon & \varepsilon \\
0 & \varepsilon & 1-\varepsilon \\
0 & \varepsilon & 1-\varepsilon \\
1-\varepsilon & \varepsilon & 0 \\
1-\varepsilon & \varepsilon & 0
\end{array}\right)
\end{array}
$$

where $0 \leq \varepsilon \leq 1$. Since each marginal channel governed by the transition matrix $[P_{Y_3|X_1,X_2,X_3}(\cdot|\cdot,x_2,x_3)]$ is quasi-symmetric, we immediately have that $P_{X_1}^* = P_{\mathcal{X}_1}^{\mathrm{U}}$. Also, since $[P_{Y_3|X_1,X_2,X_3}(\cdot|\cdot,x_2,x_3)]$, $x_2 \in \mathcal{X}_2$ and $x_3 \in \mathcal{X}_3$, are column permutations of each other, for any fixed $x_3 \in \mathcal{X}_3$, $\mathcal{I}(P_{X_1}^*, P_{Y_3|X_1,X_2=x_2,X_3=x_3})$ does not depend on $x_2 \in \mathcal{X}_2$. Thus, condition (i) of Theorem 14 holds. Moreover, condition (ii) holds since the matrices $[P_{Y_3|X_1,X_2,X_3}(\cdot|x_1,\cdot,x_3)]$, $x_1 \in \mathcal{X}_1$ and $x_3 \in \mathcal{X}_3$, are column permutations of each other.

Verifying condition (iii) involves several steps. We first observe that $\mathcal{I}(P_{X_1,X_2}, P_{Y_3|X_1,X_2,X_3=x_3})$ does not depend on $x_3 \in \mathcal{X}_3$ for any fixed P_{X_1,X_2} since the matrices $[P_{Y_3|X_1,X_2,X_3}(\cdot|\cdot,\cdot,x_3)]$, $x_3 \in \mathcal{X}_3$, are column permutations of each other. From (97) and (98) in Appendix D, it suffices to consider input distributions of this form: $P_{X_1,X_2,X_3,V} = P_{X_1,X_2}P_{X_3,V}$. Thus, given any $P_{X_1,X_2,X_3,V}^{(1)} = P_{X_1,X_2}^{(1)}P_{X_3,V}^{(1)}$, we define $P_{X_1,X_2,X_3,V}^{(2)}(x_1,x_2,x_3,v) = P_{X_1,X_2,X_3,V}^{(1)}(x_1 \oplus_2 1, x_2, x_3, v)$ for all x_1,x_2,x_3,v. Also, let $P_{X_1,X_2,X_3,V}^{(3)} = (P_{X_1,X_2,X_3,V}^{(1)} + P_{X_1,X_2,X_3,V}^{(2)})/2$ so that we have $P_{X_1,X_2,X_3,V}^{(3)} = P_{X_1}^{(3)}P_{X_2}^{(1)}P_{X_3,V}^{(1)}$ with $P_{X_1}^{(3)} = P_{\mathcal{X}_1}^{\mathrm{U}} = P_{X_1}^*$. Now, since (48) holds in this example, one can directly obtain that $I^{(1)}(X_1,X_2;Y_3|X_3=x_3) \leq I^{(3)}(X_1,X_2;Y_3|X_3=x_3)$ from the proof of Lemma 7. As a result, this TWC satisfies all conditions of Theorem 14 and has capacity region given by

$$
\mathcal{C}^{\text{MA-DBC}} = \overline{\mathrm{co}}\left(\bigcup_{P_{X_2}P_{V,X_3}} \mathcal{R}^{\text{MA-DBC}}\Big(P_{\mathcal{X}_1}^{\mathrm{U}}P_{X_2}P_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2}\Big)\right).
$$

Example 10 (Binary MA/DB TWC with Erasures): Suppose that $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}_3 = \{0,1\}$, $\mathcal{Y}_1 = \mathcal{Y}_2 = \{0,1\}$, and $\mathcal{Y}_3 = \{0,1,\mathbf{E}\}$, where $\mathbf{E}$ denotes erasure symbol. We consider a discrete

memoryless MA/DB TWC in which the DBC direction is described by (40) and (41) and the MAC direction is described by

$$Y_{3,i} = (X_{1,i} \oplus_2 X_{2,i} \oplus_2 X_{3,i}) \cdot \mathbf{1}(Z_{3,i} \neq \mathbf{E}) + \mathbf{E} \cdot \mathbf{1}(Z_{3,i} = \mathbf{E}), \tag{52}$$

where $\{Z_{3,i}\}_{i=1}^{\infty}$ with $Z_{3,i} \in \{0, \mathbf{E}\}$ is a discrete memoryless noise process which is independent of all users' messages and the noise processes $\{Z_{1,i}\}_{i=1}^{\infty}$ and $\{Z_{2,i}\}_{i=1}^{\infty}$. Also, we assume that $\Pr(Z_{3,i} = \mathbf{E}) = \varepsilon$ for all i, thereby obtaining the channel transition matrix $[P_{Y_3|X_1,X_2,X_3}(\cdot|\cdot,\cdot,\cdot)]$:

$$
\begin{array}{c}
\begin{array}{ccc} 0 & 1 & \mathbf{E} \end{array} \\
\begin{array}{c} 000 \\ 100 \\ 010 \\ 110 \\ 001 \\ 101 \\ 011 \\ 111 \end{array}
\left(
\begin{array}{ccc}
1-\varepsilon & 0 & \varepsilon \\
0 & 1-\varepsilon & \varepsilon \\
0 & 1-\varepsilon & \varepsilon \\
1-\varepsilon & 0 & \varepsilon \\
0 & 1-\varepsilon & \varepsilon \\
1-\varepsilon & 0 & \varepsilon \\
1-\varepsilon & 0 & \varepsilon \\
0 & 1-\varepsilon & \varepsilon
\end{array}
\right).
\end{array}
$$

It can be directly verified that (48) and (49) in Theorem 15 hold. Hence, the inner and outer bounds coincide and the capacity region is given by

$$\mathcal{C}^{\text{MA-DBC}} = \overline{\text{co}}\left(\bigcup_{P_{V,X_3}} \mathcal{R}^{\text{MA-DBC}}\left(P_{\mathcal{X}_1}^{\text{U}} P_{\mathcal{X}_2}^{\text{U}} P_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2} \right) \right)$$

$$= \overline{\text{co}}\Big(\bigcup_{P_{V,X_3}} \big\{ (R_{13}, R_{23}, R_{31}, R_{32}) : R_{13} + R_{23} \leq 1 - H_{\mathsf{b}}(\epsilon), $$

$$R_{31} \leq I(X_1; X_3 \oplus_2 Z_1|V), R_{32} \leq I(X_2 \oplus_2 Z_1 \oplus_2 Z_2; V) \big\} \Big).$$

Remark 8: Examples 9 and 10 also satisfy Theorem 13 since the product distribution $\tilde{P}_{X_1} \tilde{P}_{X_2}$ required by Theorem 13 are explicitly given in these examples. Moreover, it is straightforward to show that Examples 9 and 10 do not satisfy the conditions of Theorems 15 and 14, respectively. In other words, Theorems 14 and 15 are neither equivalent nor special cases of each other.

V. Conclusion

We have identified salient symmetry conditions for three types of two-way noisy networks: two-user TWCs with and without memory, and three-user MA/DB TWCs, under which Shannon-type random coding inner bounds exactly yield channel capacity. These tightness results, which

subsume previously established symmetry properties as special cases, delineate large families of TWCs for which user interactive adaptive coding is not beneficial in terms of improving capacity. Future research directions include identifying necessary conditions for the tightness of Shannon-type inner bounds and deriving conditions under which Han's adaptive coding inner bound [13] is tight. An additional interesting avenue of investigation is to examine whether adaptive coding is useful for the (almost) lossless and lossy transmission of correlated sources over TWCs whose capacity are achievable by the Shannon-type random coding scheme.

APPENDIX

A. Proof of Proposition 1 (Shannon's One-sided Symmetry Condition)

The proof of Proposition 1 is based on the following lemmas.

Lemma 4: If a memoryless TWC satisfies the conditions in Proposition 1, then for any input distribution $P_{X_1,X_2}^{(1)}$, any $x_1', x_1'' \in \mathcal{X}_1$, and $P_{X_1,X_2}^{(2)}(\cdot,\cdot) \triangleq P_{X_1,X_2}^{(1)}(\tau_{x_1',x_1''}^{\mathcal{X}_1}(\cdot),\cdot)$, the following hold:

$$I^{(1)}(X_1;Y_2|X_2) = I^{(2)}(X_1;Y_2|X_2), \tag{53}$$

$$I^{(1)}(X_2;Y_1|X_1) = I^{(2)}(X_2;Y_1|X_1), \tag{54}$$

$$\mathcal{R}(P_{X_1,X_2}^{(1)}, P_{Y_1,Y_2|X_1,X_2}) = \mathcal{R}(P_{X_1,X_2}^{(2)}, P_{Y_1,Y_2|X_1,X_2}). \tag{55}$$

Proof: For any $P_{X_1,X_2}^{(1)}$ and $P_{X_1,X_2}^{(2)}(\cdot,\cdot) \triangleq P_{X_1,X_2}^{(1)}(\tau_{x_1',x_1''}^{\mathcal{X}_1}(\cdot),\cdot)$, we have

$$I^{(2)}(X_1;Y_2|X_2) = \sum_{x_2} P_{X_2}^{(2)}(x_2) I^{(2)}(X_1;Y_2|X_2 = x_2)$$

$$= \sum_{x_2} P_{X_2}^{(2)}(x_2) \sum_{x_1,y_2} P_{X_1|X_2}^{(2)}(x_1|x_2) P_{Y_2|X_1,X_2}(y_2|x_1,x_2) \log \frac{P_{Y_2|X_1,X_2}(y_2|x_1,x_2)}{P_{Y_2|X_2}^{(2)}(y_2|x_2)}$$

$$= \sum_{x_1,x_2,y_2} P_{X_1,X_2}^{(2)}(x_1,x_2) P_{Y_2|X_1,X_2}(y_2|x_1,x_2)$$

$$\cdot \log \frac{P_{Y_2|X_1,X_2}(y_2|x_1,x_2)}{\sum_{\tilde{x}_1} P_{X_1|X_2}^{(2)}(\tilde{x}_1|x_2) P_{Y_2|X_1,X_2}(y_2|\tilde{x}_1,x_2)}$$

$$= \sum_{x_1,x_2,y_2} P_{X_1,X_2}^{(1)}(\tau_{x_1',x_1''}^{\mathcal{X}_1}(x_1),x_2) P_{Y_2|X_1,X_2}(\pi^{\mathcal{Y}_2}[x_1',x_1''](y_2)|\tau_{x_1',x_1''}^{\mathcal{X}_1}(x_1),x_2)$$

$$\cdot \log \frac{P_{Y_2|X_1,X_2}(\pi^{\mathcal{Y}_2}[x_1',x_1''](y_2)|\tau_{x_1',x_1''}^{\mathcal{X}_1}(x_1),x_2)}{\sum_{\tilde{x}_1} P_{X_1|X_2}^{(1)}(\tau_{x_1',x_1''}^{\mathcal{X}_1}(\tilde{x}_1)|x_2) P_{Y_2|X_1,X_2}(\pi^{\mathcal{Y}_2}[x_1',x_1''](y_2)|\tau_{x_1',x_1''}^{\mathcal{X}_1}(\tilde{x}_1),x_2)}$$

$$\tag{56}$$

$$= \sum_{x_1,x_2,y_2} P^{(1)}_{X_1,X_2}(\tau^{\mathcal{X}_1}_{x'_1,x''_1}(x_1),x_2) P_{Y_2|X_1,X_2}(\pi^{\mathcal{Y}_2}_2[x'_1,x''_1](y_2)|\tau^{\mathcal{X}_1}_{x'_1,x''_1}(x_1),x_2)$$

$$\cdot \log \frac{P_{Y_2|X_1,X_2}(\pi^{\mathcal{Y}_2}[x'_1,x''_1](y_2)|\tau^{\mathcal{X}_1}_{x'_1,x''_1}(x_1),x_2)}{\sum_{\tilde{x}_1} P^{(1)}_{X_1|X_2}(\tilde{x}_1|x_2) P_{Y_2|X_1,X_2}(\pi^{\mathcal{Y}_2}[x'_1,x''_1](y_2)|\tilde{x}_1,x_2)} \tag{57}$$

$$= \sum_{x_1,x_2,y_2} P^{(1)}_{X_1,X_2}(\tau^{\mathcal{X}_1}_{x'_1,x''_1}(x_1),x_2) P_{Y_2|X_1,X_2}(\pi^{\mathcal{Y}_2}_2[x'_1,x''_1](y_2)|\tau^{\mathcal{X}_1}_{x'_1,x''_1}(x_1),x_2)$$

$$\cdot \log \frac{P_{Y_2|X_1,X_2}(\pi^{\mathcal{Y}_2}[x'_1,x''_1](y_2)|\tau^{\mathcal{X}_1}_{x'_1,x''_1}(x_1),x_2)}{P^{(1)}_{Y_2|X_2}(\pi^{\mathcal{Y}_2}[x'_1,x''_1](y_2)|x_2)}$$

$$= \sum_{x_1,x_2,\tilde{y}_2} P^{(1)}_{X_1,X_2}(\tau^{\mathcal{X}_1}_{x'_1,x''_1}(x_1),x_2) P_{Y_2|X_1,X_2}(\tilde{y}_2|\tau^{\mathcal{X}_1}_{x'_1,x''_1}(x_1),x_2)$$

$$\cdot \log \frac{P_{Y_2|X_1,X_2}(\tilde{y}_2|\tau^{\mathcal{X}_1}_{x'_1,x''_1}(x_1),x_2)}{P^{(1)}_{Y_2|X_2}(\tilde{y}_2|x_2)} \tag{58}$$

$$= \sum_{\tilde{x}_1,x_2,\tilde{y}_2} P^{(1)}_{X_1,X_2}(\tilde{x}_1,x_2) P_{Y_2|X_1,X_2}(\tilde{y}_2|\tilde{x}_1,x_2) \log \frac{P_{Y_2|X_1,X_2}(\tilde{y}_2|\tilde{x}_1,x_2)}{P^{(1)}_{Y_2|X_2}(\tilde{y}_2|x_2)} \tag{59}$$

$$= I^{(1)}(X_1;Y_2|X_2), \tag{60}$$

where (56) holds by the definition of $P^{(2)}_{X_1,X_2}(x_1,x_2)$ and the fact that $P_{Y_2|X_1,X_2}(y_2|x_1,x_2) = P_{Y_2|X_1,X_2}(\pi^{\mathcal{Y}_2}[x'_1,x''_1](y_2)|\tau^{\mathcal{X}_1}_{x'_1,x''_1}(x_1),x_2)$ due to the Shannon condition in (2), (57) and (59) hold since $\tau^{\mathcal{X}_1}_{x'_1,x''_1}$ is a bijection, and (58) holds since $\pi^{\mathcal{Y}_2}[x'_1,x''_1]$ is a bijection.

By a similar argument, we can verify that $I^{(1)}(X_2;Y_1|X_1) = I^{(2)}(X_2;Y_1|X_1)$. The proof is then completed by noting that the identity $\mathcal{R}(P^{(1)}_{X_1,X_2}, P_{Y_1,Y_2|X_1,X_2}) = \mathcal{R}(P^{(2)}_{X_1,X_2}, P_{Y_1,Y_2|X_1,X_2})$ follows from the definition of $\mathcal{R}$ in (1). ∎

Lemma 5: If a memoryless TWC satisfies the condition in Proposition 1, then for any input distribution $P^{(1)}_{X_1,X_2}$, any $x'_1, x''_1 \in \mathcal{X}_1$, and $P^{(2)}_{X_1,X_2}(\cdot,\cdot) \triangleq P^{(1)}_{X_1,X_2}(\tau^{\mathcal{X}_1}_{x'_1,x''_1}(\cdot),\cdot)$, we have

$$\mathcal{R}(P^{(1)}_{X_1,X_2}, P_{Y_1,Y_2|X_1,X_2}) \subseteq \mathcal{R}(P^{(3)}_{X_1,X_2}, P_{Y_1,Y_2|X_1,X_2}) \tag{61}$$

where $P^{(3)}_{X_1,X_2}(x_1,x_2) \triangleq \frac{1}{2}(P^{(1)}_{X_1,X_2}(x_1,x_2) + P^{(2)}_{X_1,X_2}(x_1,x_2))$.

Proof: The proof relies on the concavity of $I(X_1;Y_2|X_2)$ and $I(X_2;Y_1|X_1)$ in P_{X_1,X_2} [3]. For any given $P^{(1)}_{X_1,X_2}$ and $P^{(2)}_{X_1,X_2}(\cdot,\cdot) = P^{(1)}_{X_1,X_2}(\tau^{\mathcal{X}_1}_{x'_1,x''_1}(\cdot),\cdot)$, let $P^{(3)}_{X_1,X_2} = \frac{1}{2}(P^{(1)}_{X_1,X_2} + P^{(2)}_{X_1,X_2})$. The concavity property then implies that

$$I^{(3)}(X_1;Y_2|X_2) \geq \frac{1}{2}I^{(1)}(X_1;Y_2|X_2) + \frac{1}{2}I^{(2)}(X_1;Y_2|X_2) \tag{62}$$

$$= I^{(1)}(X_1; Y_2 | X_2), \tag{63}$$

and that

$$I^{(3)}(X_2; Y_1 | X_1) \geq \frac{1}{2} I^{(1)}(X_2; Y_1 | X_1) + \frac{1}{2} I^{(2)}(X_2; Y_1 | X_1) \tag{64}$$

$$= I^{(1)}(X_2; Y_1 | X_1), \tag{65}$$

where (63) and (65) follow from Lemma 4. The proof is completed by invoking the definition of $\mathcal{R}$ in (1). ∎

Lemma 6: If a memoryless TWC satisfies the condition in Proposition 1, then for any given input distribution $P_{X_1,X_2} = P_{X_1|X_2} P_{X_2}$, we have

$$\mathcal{R}(P_{X_1,X_2}, P_{Y_1,Y_2|X_1,X_2}) \subseteq \mathcal{R}\left(P_{\mathcal{X}_1}^{\mathsf{U}} P_{X_2}, P_{Y_1,Y_2|X_1,X_2}\right), \tag{66}$$

where $P_{\mathcal{X}_1}^{\mathsf{U}}$ denotes the uniform probability distribution on $\mathcal{X}_1$.

Proof: Without loss of generality, we assume that $\mathcal{X}_1 \triangleq \{1, 2, ..., \kappa\}$. Define $\mathcal{P}_m = \{P_{X_1,X_2} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2) : P_{X_1,X_2}(1, x_2) = P_{X_1,X_2}(2, x_2) = \cdots = P_{X_1,X_2}(m, x_2) \text{ for all } x_2 \in \mathcal{X}_2\}$, where $1 \leq m \leq \kappa$. Lemma 5 shows that for any $P_{X_1,X_2}^{(1)} \in \mathcal{P}_1$, one can construct $P_{X_1,X_2}^{(3)} \in \mathcal{P}_2$ in such a way that (61) holds. We now extend this result by induction on m showing that for any $P_{X_1,X_2}^{(1)} \in \mathcal{P}_m$ with $2 \leq m < \kappa$, there exists a $P_{X_1,X_2}^{(m+2)} \in \mathcal{P}_{m+1}$ such that $\mathcal{R}(P_{X_1,X_2}^{(1)}, P_{Y_1,Y_2|X_1,X_2}) \subseteq \mathcal{R}(P_{X_1,X_2}^{(m+2)}, P_{Y_1,Y_2|X_1,X_2})$.

Suppose that the above claim is true up to m for some $1 \leq m < \kappa$, where the base case $m = 1$ was proved in Lemma 5. We next prove the claim for $m + 1$. For any $P_{X_1,X_2}^{(1)} \in \mathcal{P}_m$, define

$$P_{X_1,X_2}^{(m+2)}(x_1, x_2) \triangleq \frac{1}{m+1} \sum_{i=1}^{m+1} P_{X_1,X_2}^{(i)}(x_1, x_2),$$

where $P_{X_1,X_2}^{(i)}(\cdot, \cdot) \triangleq P_{X_1,X_2}^{(1)}(\tau_{i-1,m+1}^{\mathcal{X}_1}(\cdot), \cdot)$ for $2 \leq i \leq m+1$. Due to the Shannon's one-sided symmetry condition and Lemma 4, we have that $I^{(i)}(X_1; Y_2 | X_2) = I^{(1)}(X_1; Y_2 | X_2)$ and that $I^{(i)}(X_2; Y_1 | X_1) = I^{(1)}(X_2; Y_1 | X_1)$ for $2 \leq i \leq m+1$. Concavity then implies that

$$I^{(m+2)}(X_1; Y_2 | X_2) \geq \frac{1}{m+1} \sum_{i=1}^{m+1} I^{(i)}(X_1; Y_2 | X_2)$$

$$= I^{(1)}(X_1; Y_2 | X_2).$$

Similarly, we obtain that $I^{(m+2)}(X_2; Y_1 | X_1) \geq I^{(1)}(X_2; Y_1 | X_1)$. Moreover, since $P_{X_1,X_2}^{(1)} \in \mathcal{P}_m$, we have that $P_{X_1,X_2}^{(m+2)}(x_1, x_2) = 1/(m+1) \cdot (m \cdot P_{X_1,X_2}^{(1)}(1, x_2) + P_{X_1,X_2}^{(1)}(m+1, x_2))$ for $1 \leq x_1 \leq m+1$ and all $x_2 \in \mathcal{X}_2$, i.e., $P_{X_1,X_2}^{(m+2)} \in \mathcal{P}_{m+1}$, thereby proving the claim.

Since any $P_{X_1,X_2} = P_{X_1|X_2}P_{X_2} \in \mathcal{P}_\kappa$ can be expressed as $P_{\mathcal{X}_1}^{\mathrm{U}}P_{X_2}$, in view of the definition of $\mathcal{R}$ the proof is complete. ∎

We are now ready to prove Proposition 1.

Proof of Proposition 1: Note that

$$\mathcal{C}_{\mathrm{O}}(P_{Y_1,Y_2|X_1,X_2}) = \overline{\mathrm{co}}\left(\bigcup_{P_{X_1,X_2}} \mathcal{R}(P_{X_1,X_2}, P_{Y_1,Y_2|X_1,X_2})\right)$$

$$\subseteq \overline{\mathrm{co}}\left(\bigcup_{P_{X_2}} \mathcal{R}\left(P_{\mathcal{X}_1}^{\mathrm{U}}P_{X_2}, P_{Y_1,Y_2|X_1,X_2}\right)\right) \tag{67}$$

$$\subseteq \mathcal{C}_{\mathrm{I}}(P_{Y_1,Y_2|X_1,X_2}), \tag{68}$$

where (67) follows from Lemma 6. Together with $\mathcal{C}_{\mathrm{I}}(P_{Y_1,Y_2|X_1,X_2}) \subseteq \mathcal{C}_{\mathrm{O}}(P_{Y_1,Y_2|X_1,X_2})$, this gives:

$$\mathcal{C} = \mathcal{C}_{\mathrm{I}}(P_{Y_1,Y_2|X_1,X_2}) = \mathcal{C}_{\mathrm{O}}(P_{Y_1,Y_2|X_1,X_2}) = \overline{\mathrm{co}}\left(\bigcup_{P_{X_2}} \mathcal{R}\left(P_{\mathcal{X}_1}^{\mathrm{U}}P_{X_2}, P_{Y_1,Y_2|X_1,X_2}\right)\right). \tag{69}$$

∎

We remark that, based on the proof of Proposition 1, it is straightforward to prove Shannon's two-sided symmetry condition in Proposition 2.

B. Proof of Theorem 12

Proof: Suppose that $(R_{13}, R_{23}, R_{31}, R_{32})$ is an achievable quadruple. We derive the necessary conditions for those rates by the standard converse method. For R_{13}, we have

$$nR_{13} = H(M_{13}|M_{23}, M_{31}, M_{32})$$

$$= I(M_{13}; Y_3^n|M_{23}, M_{31}, M_{32}) + H(M_{13}|Y_3^n, M_{23}, M_{31}, M_{32})$$

$$\leq I(M_{13}; Y_3^n|M_{23}, M_{31}, M_{32}) + n\epsilon_n \tag{70}$$

$$\leq I(M_{13}; Y_2^n, Y_3^n|M_{23}, M_{31}, M_{32}) + n\epsilon_n$$

$$= \sum_{i=1}^{n} I(M_{13}; Y_{2,i}, Y_{3,i}|Y_2^{i-1}, Y_3^{i-1}, M_{23}, M_{31}, M_{32}) + n\epsilon_n$$

$$= \sum_{i=1}^{n} \Big(H(Y_{2,i}, Y_{3,i}|X_{2,i}, X_{3,i}, Y_2^{i-1}, Y_3^{i-1}, M_{23}, M_{31}, M_{32})$$

$$\qquad - H(Y_{2,i}, Y_{3,i}|X_{2,i}, X_{3,i}, Y_2^{i-1}, Y_3^{i-1}, M_{23}, M_{31}, M_{32}, M_{13}) \Big) + n\epsilon_n \tag{71}$$

$$\leq \sum_{i=1}^{n} \Big(H(Y_{2,i}, Y_{3,i}|X_{2,i}, X_{3,i}) - H(Y_{2,i}, Y_{3,i}|X_{1,i}, X_{2,i}, X_{3,i}) \Big) + n\epsilon_n \tag{72}$$

$$= \sum_{i=1}^{n} I(X_{1,i}; Y_{2,i}, Y_{3,i}|X_{2,i}, X_{3,i}) + n\epsilon_n$$

$$= \sum_{i=1}^{n} I(X_{1,i}; X_{2,i} \oplus_q X_{3,i} \oplus_q Z_{1,i} \oplus_q Z_{2,i}, Y_{3,i}|X_{2,i}, X_{3,i}) + n\epsilon_n$$

$$= \sum_{i=1}^{n} I(X_{1,i}; Y_{3,i}|X_{2,i}, X_{3,i}) + I(X_{1,i}; Z_{1,i} \oplus_q Z_{2,i}|Y_{3,i}, X_{2,i}, X_{3,i}) + n\epsilon_n$$

$$= \sum_{i=1}^{n} I(X_{1,i}; Y_{3,i}|X_{2,i}, X_{3,i}) + n\epsilon_n, \tag{73}$$

where (70) follows from Fano's inequality with $\epsilon_n \to 0$ as $n \to \infty$, (71) holds since $X_{2,i} = f_{2,i}(M_{23}, Y_2^{i-1})$ and $X_{3,i} = f_{3,i}(M_{31}, M_{32}, Y_3^{i-1})$, (72) follows since the channel is memoryless, and (73) follows since $(Z_{1,i}, Z_{2,i})$ is independent of $(Y_{3,i}, X_{1,i}, X_{2,i}, X_{3,i})$. By symmetry, we also have

$$nR_{23} \le \sum_{i=1}^{n} I(X_{2,i}; Y_{3,i}|X_{1,i}, X_{3,i}) + n\epsilon_n. \tag{74}$$

For the sum rate $R_{13} + R_{23}$, we have

$$n(R_{13} + R_{23}) = H(M_{13}, M_{23}|M_{31}, M_{32})$$

$$\le I(M_{13}, M_{23}; Y_3^n|M_{31}, M_{32}) + n\epsilon_n$$

$$= \sum_{i=1}^{n} \left(H(Y_{3,i}|X_{3,i}, Y_3^{i-1}, M_{31}, M_{32}) - H(Y_{3,i}|Y_3^{i-1}, M_{31}, M_{32}, M_{13}, M_{23}) \right) + n\epsilon_n$$

$$\le \sum_{i=1}^{n} \left(H(Y_{3,i}|X_{3,i}) - H(Y_{3,i}|Y_3^{i-1}, M_{31}, M_{32}, M_{13}, M_{23}) \right) + n\epsilon_n$$

$$\le \sum_{i=1}^{n} \left(H(Y_{3,i}|X_{3,i}) - H(Y_{3,i}|X_{1,i}, X_{2,i}, X_{3,i}) \right) + n\epsilon_n$$

$$= \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; Y_{3,i}|X_{3,i}) + n\epsilon_n,$$

where $\epsilon_n \to 0$ as $n \to \infty$ by Fano's inequality. Therefore, for the rates in the MAC direction, we have

$$R_{13} \le \frac{1}{n} \sum_{i=1}^{n} I(X_{1,i}; Y_{3,i}|X_{2,i}, X_{3,i}) + \epsilon_n \le I(X_1; Y_3|X_2, X_3) + \epsilon_n$$

$$R_{23} \le \frac{1}{n} \sum_{i=1}^{n} I(X_{2,i}; Y_{3,i}|X_{1,i}, X_{3,i}) + \epsilon_n \le I(X_2; Y_3|X_1, X_3) + \epsilon_n$$

$$R_{13} + R_{23} \le \frac{1}{n} \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; Y_{3,i}|X_{3,i}) + \epsilon_n \le I(X_1, X_2; Y_3|X_3) + \epsilon_n$$

where the inequalities hold since $I(X_1; Y_3|X_2, X_3)$, $I(X_2; Y_3|X_1, X_3)$, and $I(X_1, X_2; Y_3|X_3)$ are concave[7] in the joint input distribution P_{X_1,X_2,X_3}, where $P_{X_1,X_2,X_3} = \frac{1}{n}\sum_{i=1}^n P_{X_{1,i},X_{2,i},X_{3,i}}$.

For the achievable rate R_{32} in the DB direction, we have

$$nR_{32} = H(M_{32}|M_{23})$$

$$\leq I(M_{32}; Y_2^n|M_{23}) + n\epsilon_n$$

$$= \sum_{i=1}^n I(M_{32}; Y_{2,i}|Y_2^{i-1}, M_{23}, X_2^i) + n\epsilon_n$$

$$= \sum_{i=1}^n I(M_{32}; X_{3,i} \oplus_q Z_{1,i} \oplus_q Z_{2,i}|X_3^{i-1} \oplus_q Z_1^{i-1} \oplus_q Z_2^{i-1}, M_{23}, X_2^i) + n\epsilon_n$$

$$= \sum_{i=1}^n I(M_{32}; X_{3,i} \oplus_q Z_{1,i} \oplus_q Z_{2,i}|X_3^{i-1} \oplus_q Z_1^{i-1} \oplus_q Z_2^{i-1}, M_{23}) + n\epsilon_n \tag{75}$$

$$\leq \sum_{i=1}^n I(M_{32}, X_3^{i-1} \oplus_q Z_1^{i-1} \oplus_q Z_2^{i-1}, M_{23}; X_{3,i} \oplus_q Z_{1,i} \oplus_q Z_{2,i}) + n\epsilon_n \tag{76}$$

$$\leq \sum_{i=1}^n I(M_{32}, M_{23}, M_{13}, X_3^{i-1} \oplus_q Z_1^{i-1} \oplus_q Z_2^{i-1}, X_3^{i-1} \oplus_q Z_1^{i-1}; X_{3,i} \oplus_q Z_{1,i} \oplus_q Z_{2,i}) + n\epsilon_n$$

$$= \sum_{i=1}^n I(M_{\{32,23,13\}}, \tilde{Y}_1^{i-1}, \tilde{Y}_2^{i-1}; \tilde{Y}_{2,i}) + n\epsilon_n \tag{77}$$

where (75) holds since X_2^i is a function of $(X_3^{i-1} \oplus_q Z_1^{i-1} \oplus_q Z_2^{i-1}, M_{23})$, (76) follows from the chain rule and the non-negativity of mutual information, and (77) is expressed in terms of $\tilde{Y}_{1,i} \triangleq X_{3,i} \oplus_q Z_{1,i}$, and $\tilde{Y}_{2,i} \triangleq X_{3,i} \oplus_q Z_{1,i} \oplus_q Z_{2,i} = \tilde{Y}_{1,i} \oplus_q Z_{2,i}$.

For R_{31}, we have

$$nR_{31} = H(M_{31}|M_{\{32,23,13\}})$$

$$\leq I(M_{31}; Y_1^n, Y_2^n|M_{\{32,23,13\}}) + n\epsilon_n$$

$$= \sum_{i=1}^n I(M_{31}; Y_{1,i}, Y_{2,i}|Y_1^{i-1}, Y_2^{i-1}, M_{\{32,23,13\}}) + n\epsilon_n$$

$$\leq \sum_{i=1}^n I(M_{31}, X_{3,i}; Y_{1,i}, Y_{2,i}|Y_1^{i-1}, Y_2^{i-1}, M_{\{32,23,13\}}) + n\epsilon_n$$

$$= \sum_{i=1}^n I(M_{31}, X_{3,i}; Y_{1,i}, Y_{2,i}|Y_1^{i-1}, Y_2^{i-1}, M_{\{32,23,13\}}, X_1^i, X_2^i) + n\epsilon_n \tag{78}$$

[7]This follows from the fact that $I(A; C|B)$ is concave in $P_{A,B}$ for fixed $P_{C|A,B}$ [3].

$$= \sum_{i=1}^{n} I(M_{31}, X_{3,i}; \tilde{Y}_{1,i}, \tilde{Y}_{2,i} | Y_1^{i-1}, Y_2^{i-1}, M_{\{32,23,13\}}, X_1^i, X_2^i) + n\epsilon_n$$

$$= \sum_{i=1}^{n} I(M_{31}, X_{3,i}; \tilde{Y}_{1,i}, \tilde{Y}_{2,i} | \tilde{Y}_1^{i-1}, \tilde{Y}_2^{i-1}, M_{\{32,13,23\}}) + n\epsilon_n \tag{79}$$

$$= \sum_{i=1}^{n} I(X_{3,i}; \tilde{Y}_{1,i}, \tilde{Y}_{2,i} | \tilde{Y}_1^{i-1}, \tilde{Y}_2^{i-1}, M_{\{32,13,23\}})$$

$$\quad + \sum_{i=1}^{n} I(M_{31}; \tilde{Y}_{1,i}, \tilde{Y}_{2,i} | \tilde{Y}_1^{i-1}, \tilde{Y}_2^{i-1}, M_{\{32,13,23\}}, X_{3,i}) + n\epsilon_n$$

$$= \sum_{i=1}^{n} I(X_{3,i}; \tilde{Y}_{1,i}, \tilde{Y}_{2,i} | \tilde{Y}_1^{i-1}, \tilde{Y}_2^{i-1}, M_{\{32,13,23\}}) + n\epsilon_n \tag{80}$$

$$= \sum_{i=1}^{n} I(X_{3,i}; \tilde{Y}_{1,i} | \tilde{Y}_1^{i-1}, \tilde{Y}_2^{i-1}, M_{\{32,13,23\}}) + n\epsilon_n \tag{81}$$

where (78) holds since $X_{1,i} = f_{1,i}(M_{13}, Y_1^{i-1})$ and $X_{2,i} = f_{2,i}(M_{23}, Y_2^{i-1})$, (79) holds since $(Y_1^{i-1}, Y_2^{i-1}, X_1^i, X_2^i)$ can be generated knowing $(M_{13}, M_{23}, \tilde{Y}_1^{i-1}, \tilde{Y}_2^{i-1})$, (80) holds because $M_{31} \,\multimap\!\!-\!\!\multimap\, (\tilde{Y}_1^{i-1}, \tilde{Y}_2^{i-1}, M_{\{32,13,23\}}, X_{3,i}) \,\multimap\!\!-\!\!\multimap\, (\tilde{Y}_{1,i}, \tilde{Y}_{2,i})$ form a Markov chain, and (81) holds since $\tilde{Y}_{2,i} \,\multimap\!\!-\!\!\multimap\, (\tilde{Y}_{1,i}, \tilde{Y}_1^{i-1}, \tilde{Y}_2^{i-1}, M_{\{32,13,23\}}) \,\multimap\!\!-\!\!\multimap\, X_{3,i}$ form a Markov chain. Note that these Markov chain properties hold since $\{Z_{1,i}\}_{i=1}^{n}$ and $\{Z_{2,i}\}_{i=1}^{n}$ are independent memoryless processes and are independent of all user messages.

Setting $V_i = (\tilde{Y}_1^{i-1}, \tilde{Y}_2^{i-1}, W_{\{32,13,23\}})$, we have that $V_i \,\multimap\!\!-\!\!\multimap\, X_{3,i} \,\multimap\!\!-\!\!\multimap\, (\tilde{Y}_{1,i}, \tilde{Y}_{2,i})$ form a Markov chain. From (77) and (81), we obtain that $nR_{32} \le \sum_{i=1}^{n} I(V_i; \tilde{Y}_{2,i}) + n\epsilon_n$ and $nR_{31} \le \sum_{i=1}^{n} I(X_{3,i}; \tilde{Y}_{1,i} | V_i) + n\epsilon_n$. Let K be a time-sharing random variable that is uniform over $\{1, 2, ..., n\}$ and independent of all messages, inputs, and outputs. Setting $V = (K, V_K)$, $X_3 = X_{3,K}$, $Z_1 = Z_{1,K}$, $Z_2 = Z_{2,K}$ $\tilde{Y}_1 = X_3 \oplus_q Z_1 = \tilde{Y}_{1,K}$, $\tilde{Y}_2 = X_3 \oplus_q Z_1 \oplus_q Z_2 = \tilde{Y}_{2,K}$, we have

$$nR_{32} \le \sum_{i=1}^{n} I(V_i; \tilde{Y}_{2,i}) + n\epsilon_n = nI(V_K; \tilde{Y}_{2,K} | K) + n\epsilon_n \le nI(V; \tilde{Y}_2) + n\epsilon_n$$
$$= nI(V; X_3 \oplus_q Z_1 \oplus_q Z_2) + n\epsilon_n,$$

and

$$nR_{31} \le \sum_{i=1}^{n} I(X_{3,i}; \tilde{Y}_{1,i} | V_i) + n\epsilon_n = nI(X_3; \tilde{Y}_1 | V) + n\tilde{\epsilon}_n$$
$$= nI(X_3; X_3 \oplus_q Z_1 | V) + n\epsilon_n$$

for some $P_{Z_1,Z_2,X_3,V} = P_{X_3,V}P_{Z_1}P_{Z_2}$. Combining the obtained bounds for rates R_{13} and R_{23}, the proof is completed by letting $n \to \infty$. The bound on the alphabet size of V can be established by the convex cover method [8]. ∎

C. Proof of Theorem 13

Proof: Consider a MA-DB TWC governed by $P_{Y_3|X_1,X_2,X_3}$, P_{Z_1}, and P_{Z_2}. Recall that

$$\mathcal{R}^{\text{MA-DBC}}(P_{X_1,X_2,X_3,V}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2}) = \{(R_{13}, R_{23}, R_{31}, R_{32}) :$$

$$R_{13} \leq I(X_1; Y_3|X_2, X_3), \tag{82}$$

$$R_{23} \leq I(X_2; Y_3|X_1, X_3), \tag{83}$$

$$R_{13} + R_{23} \leq I(X_1, X_2; Y_3|X_3), \tag{84}$$

$$R_{31} \leq I(X_3; X_3 \oplus_q Z_1|V), \tag{85}$$

$$R_{32} \leq I(V; X_3 \oplus_q Z_1 \oplus_q Z_2)\}. \tag{86}$$

Since (82)-(84) do not depend on V and (85) and (86) do not depend on (X_1, X_2), we have

$$\mathcal{R}^{\text{MA-DBC}}(P_{X_1,X_2,X_3,V}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2}) = \mathcal{R}^{\text{MA-DBC}}(P_{X_1,X_2|X_3}P_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2}).$$

$$\tag{87}$$

To complete the proof, it suffices to show that for every $P_{X_1,X_2|X_3}$ and the corresponding $\tilde{P}_{X_1}\tilde{P}_{X_2}$ (which depends on $P_{X_1,X_2|X_3}$) given by our assumption, satisfies

$$\mathcal{R}^{\text{MA-DBC}}(P_{X_1,X_2|X_3}P_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2}) \subseteq \mathcal{R}^{\text{MA-DBC}}(\tilde{P}_{X_1}\tilde{P}_{X_2}P_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2}),$$

$$\tag{88}$$

since then we clearly have $\mathcal{C}_O^{\text{MA-DBC}}(P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2}) \subseteq \mathcal{C}_I^{\text{MA-DBC}}(P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2})$. To show (88), consider two input distributions $P_{X_1,X_2,X_3,V}^{(1)} \triangleq P_{X_1,X_2|X_3}^{(1)} P_{V,X_3}^{(1)}$ and $P_{X_1,X_2,X_3,V}^{(2)} \triangleq \tilde{P}_{X_1}\tilde{P}_{X_2}P_{V,X_3}^{(1)}$, where $\tilde{P}_{X_1}\tilde{P}_{X_2}$ is given by the assumption. Then,

$$I^{(1)}(X_3; X_3 \oplus_q Z_1|V) = I^{(2)}(X_3; X_3 \oplus_q Z_1|V) \tag{89}$$

$$I^{(1)}(V; X_3 \oplus_q Z_1 \oplus_q Z_2) = I^{(2)}(V; X_3 \oplus_q Z_1 \oplus_q Z_2) \tag{90}$$

since $P_{X_1,X_2,X_3,U}^{(1)}$ and $P_{X_1,X_2,X_3,U}^{(2)}$ have the same marginal $P_{U,X_3}^{(1)}$. Furthermore,

$$I^{(1)}(X_1; Y_3|X_2, X_3) = \sum_{x_3} P_{X_3}^{(1)}(x_3) \cdot I^{(1)}(X_1; Y_3|X_2, X_3 = x_3)$$

$$\leq \sum_{x_3} P_{X_3}^{(1)}(x_3) \cdot I^{(2)}(X_1; Y_3 | X_2, X_3 = x_3)$$

$$= I^{(2)}(X_1; Y_3 | X_2, X_3),$$

where the inequality follows from (45) and the last equality holds since $P_{X_1,X_2,X_3,V}^{(1)}$ and $P_{X_1,X_2,X_3,V}^{(2)}$ have the same marginal $P_{X_3}^{(1)}$. Similarly, we obtain that $I^{(1)}(X_2; Y_3 | X_1, X_3) \leq I^{(2)}(X_2; Y_3 | X_1, X_3)$ and $I^{(1)}(X_1, X_2; Y_3 | X_3) \leq I^{(2)}(X_1, X_2; Y_3 | X_3)$. Consequently, (88) holds.

∎

D. Proof of Theorem 14

Proof: Similar to the proof in Theorem 13, for any $P_{X_1,X_2|X_3} P_{V,X_3} = P_{X_2|X_3} P_{X_1|X_2,X_3} P_{V,X_3}$, it suffices to show that

$$\mathcal{R}^{\text{MA-DBC}}(P_{X_1,X_2|X_3} P_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2})$$

$$\subseteq \mathcal{R}^{\text{MA-DBC}}(P_{X_1}^* P_{X_2|X_3} P_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2}), \tag{91}$$

where $P_{X_1}^*$ is given by conditions (i).

For any $P_{X_1,X_2,X_3,V}^{(1)} = P_{X_1,X_2|X_3}^{(1)} P_{V,X_3}^{(1)}$, let $P_{X_1,X_2,X_3,V}^{(2)} = P_{X_1}^* P_{X_2}^{(1)} P_{V,X_3}^{(1)}$, where $P_{X_1}^*$ is given by condition (i) and $P_{X_2}^{(1)}$ denotes the marginal distribution of X_2 derived from $P_{X_1,X_2,X_3,V}^{(1)}$. For the rate constraints in the DB direction, the same identities as in (89)-(90) can be obtained because $P_{X_1,X_2,X_3,V}^{(1)}$ and $P_{X_1,X_2,X_3,V}^{(2)}$ share a common marginal distribution given by $P_{V,X_3}^{(1)}$. For R_{13} in the MA direction, we have

$$I^{(1)}(X_1; Y_3 | X_2, X_3)$$

$$= \sum_{x_2,x_3} P_{X_2,X_3}^{(1)}(x_2, x_3) \cdot I^{(1)}(X_1; Y_3 | X_2 = x_2, X_3 = x_3)$$

$$= \sum_{x_2,x_3} P_{X_2,X_3}^{(1)}(x_2, x_3) \cdot \mathcal{I}(P_{X_1|X_2=x_2,X_3=x_3}^{(1)}, P_{Y_3|X_1,X_2=x_2,X_3=x_3})$$

$$\leq \sum_{x_2,x_3} P_{X_2,X_3}^{(1)}(x_2, x_3) \cdot \max_{P_{X_1|X_2=x_2,X_3=x_3}} \mathcal{I}(P_{X_1|X_2=x_2,X_3=x_3}, P_{Y_3|X_1,X_2=x_2,X_3=x_3})$$

$$= \sum_{x_2,x_3} P_{X_2,X_3}^{(1)}(x_2, x_3) \cdot \mathcal{I}(P_{X_1}^*, P_{Y_3|X_1,X_2=x_2,X_3=x_3}) \tag{92}$$

$$= \sum_{x_3} P_{X_3}^{(1)}(x_3) \sum_{x_2} P_{X_2|X_3}^{(1)}(x_2|x_3) \cdot \mathcal{I}(P_{X_1}^*, P_{Y_3|X_1,X_2=x_2,X_3=x_3})$$

$$= \sum_{x_3} P_{X_3}^{(1)}(x_3) \left(\sum_{x_2} P_{X_2|X_3}^{(1)}(x_2|x_3) \right) \cdot \mathcal{I}(P_{X_1}^*, P_{Y_3|X_1,X_2=x_2',X_3=x_3}) \tag{93}$$

$$= \sum_{x'_2} P^{(1)}_{X_2}(x'_2) \sum_{x_3} P^{(1)}_{X_3}(x_3) \cdot \mathcal{I}(P^*_{X_1}, P_{Y_3|X_1,X_2=x'_2,X_3=x_3})$$

$$= I^{(2)}(X_1; Y_3 | X_2, X_3),$$

where (92) and (93) directly follow from condition (i).

For R_{23}, we have

$$I^{(1)}(X_2; Y_3 | X_1, X_3)$$

$$= \sum_{x_1,x_3} P^{(1)}_{X_1,X_3}(x_1, x_3) \cdot I^{(1)}(X_2; Y_3 | X_1 = x_1, X_3 = x_3)$$

$$= \sum_{x_1,x_3} P^{(1)}_{X_1,X_3}(x_1, x_3) \cdot \mathcal{I}(P^{(1)}_{X_2|X_1=x_1,X_3=x_3}, P_{Y_3|X_1=x_1,X_2,X_3=x_3})$$

$$= \sum_{x_1,x_3} P^{(1)}_{X_1,X_3}(x_1, x_3) \cdot \mathcal{I}(P^{(1)}_{X_2|X_1=x_1,X_3=x_3}, P_{Y_3|X_1=x'_1,X_2,X_3=x'_3}) \tag{94}$$

$$\leq \mathcal{I}\left(\sum_{x_1,x_3} P^{(1)}_{X_1,X_3}(x_1, x_3) P^{(1)}_{X_2|X_1,X_3}(x_2|x_1, x_3), P_{Y_3|X_1=x'_1,X_2,X_3=x'_3} \right) \tag{95}$$

$$= \mathcal{I}(P^{(1)}_{X_2}, P_{Y_3|X_1=x'_1,X_2,X_3=x'_3})$$

$$= \sum_{x'_1,x'_3} P^*_{X_1}(x'_1) P^{(1)}_{X_3}(x'_3) \cdot \mathcal{I}(P^{(1)}_{X_2}, P_{Y_3|X_1=x'_1,X_2,X_3=x'_3}) \tag{96}$$

$$= I^{(2)}(X_2; Y_3 | X_2, X_3),$$

where (94) and (96) follow from condition (ii) and (95) is due to convexity of $\mathcal{I}(\cdot, \cdot)$ in its first argument.

Moreover, for the sum rate $R_{13} + R_{23}$, we have

$$I^{(1)}(X_1, X_2; Y_3 | X_3) = \sum_{x_3} P^{(1)}_{X_3}(x_3) \cdot I^{(1)}(X_1, X_2; Y_3 | X_3 = x_3)$$

$$= \sum_{x_3} P^{(1)}_{X_3}(x_3) \cdot \mathcal{I}\left(P^{(1)}_{X_1,X_2|X_3=x_3}, P_{Y_3|X_1,X_2,X_3=x_3} \right)$$

$$= \sum_{x_3} P^{(1)}_{X_3}(x_3) \cdot \mathcal{I}\left(P^{(1)}_{X_1,X_2|X_3=x_3}, P_{Y_3|X_1,X_2,X_3=x'_3} \right) \tag{97}$$

$$\leq \mathcal{I}\left(\sum_{x_3} P^{(1)}_{X_3}(x_3) P^{(1)}_{X_1,X_2|X_3}(x_1, x_2|x_3), P_{Y_3|X_1,X_2,X_3=x'_3} \right) \tag{98}$$

$$= \mathcal{I}\left(P^{(1)}_{X_1,X_2}, P_{Y_3|X_1,X_2,X_3=x'_3} \right)$$

$$\leq \mathcal{I}\left(P^*_{X_1} P^{(1)}_{X_2}, P_{Y_3|X_1,X_2,X_3=x'_3} \right) \tag{99}$$

$$= \sum_{x'_3} P^{(1)}_{X_3}(x'_3) \cdot \mathcal{I}\left(P^*_{X_1} P^{(1)}_{X_2}, P_{Y_3|X_1,X_2,X_3=x'_3} \right)$$

$$= I^{(2)}(X_1, X_2; Y_3|X_3),$$

where (97) and (99) follow from condition (iii) and (98) is due to convexity of $\mathcal{I}(\cdot, \cdot)$ in its first argument. Therefore, (91) holds under conditions (i)-(iii). ∎

E. Proof of Theorem 15

It suffices to show that

$$\mathcal{R}^{\text{MA-DBC}}(P_{X_1,X_2|X_3}P_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2})$$

$$\subseteq \mathcal{R}^{\text{MA-DBC}}(P^{\text{U}}_{\mathcal{X}_1} P^{\text{U}}_{\mathcal{X}_2} P_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2}) \tag{100}$$

for any $P_{X_1,X_2|X_3}P_{V,X_3}$. We first give a proof sketch. Analogous to Shannon's proof for point-to-point TWCs (see Appendix A), we want to show that for any input distribution $P^{(1)}_{X_1,X_2,X_3,V} = P^{(1)}_{X_1,X_2|X_3}P^{(1)}_{V,X_3}$, if we set $P^{(2)}_{X_1,X_2,X_3,V} = P^{(2)}_{X_1,X_2|X_3}P^{(1)}_{V,X_3}$ and $P^{(3)}_{X_1,X_2,X_3,V} = P^{(3)}_{X_1,X_2|X_3}P^{(1)}_{V,X_3}$, where

$$P^{(2)}_{X_1,X_2|X_3}(\cdot, \cdot|\cdot) \triangleq P^{(1)}_{X_1,X_2|X_3}(\tau^{\mathcal{X}_1}_{x'_1,x''_1}(\cdot), \cdot|\cdot), \tag{101}$$

$$P^{(3)}_{X_1,X_2|X_3}(\cdot, \cdot|\cdot) \triangleq \frac{1}{2}\left(P^{(1)}_{X_1,X_2|X_3}(\cdot, \cdot|\cdot) + P^{(2)}_{X_1,X_2|X_3}(\cdot, \cdot|\cdot)\right), \tag{102}$$

and $x'_1, x''_1 \in \mathcal{X}_1$, we have

$$\mathcal{R}^{\text{MA-DBC}}(P^{(1)}_{X_1,X_2|X_3}P^{(1)}_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2}) = \mathcal{R}^{\text{MA-DBC}}(P^{(2)}_{X_1,X_2|X_3}P^{(1)}_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2}) \tag{103}$$

$$\subseteq \mathcal{R}^{\text{MA-DBC}}(P^{(3)}_{X_1,X_2|X_3}P^{(1)}_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2}), \tag{104}$$

where the last inclusion is shown using (48) and extending Lemma 5 to the MA/DBC setup. Then, we use an induction argument as in the proof of Lemma 6 to obtain

$$\mathcal{R}^{\text{MA-DBC}}(P_{X_1,X_2|X_3}P_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2}) \subseteq \mathcal{R}^{\text{MA-DBC}}(P^{\text{U}}_{\mathcal{X}_1} P_{X_2|X_3}P_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2}). \tag{105}$$

Next, we consider input distributions of the form $P^{(1)}_{X_1,X_2,X_3,V} = P^{\text{U}}_{\mathcal{X}_1} P^{(1)}_{X_2|X_3} P^{(1)}_{X_3,V}$ and set $P^{(2)}_{X_1,X_2,X_3,V} = P^{(2)}_{X_1,X_2|X_3}P^{(1)}_{V,X_3}$ and $P^{(3)}_{X_1,X_2,X_3,V} = P^{(3)}_{X_1,X_2|X_3}P^{(1)}_{V,X_3}$, where

$$P^{(2)}_{X_1,X_2|X_3}(\cdot, \cdot|\cdot) \triangleq P^{(1)}_{X_1,X_2|X_3}(\cdot, \tau^{\mathcal{X}_2}_{x'_2,x''_2}(\cdot)|\cdot),$$

$$P^{(3)}_{X_1,X_2|X_3}(\cdot, \cdot|\cdot) \triangleq \frac{1}{2}\left(P^{(1)}_{X_1,X_2|X_3}(\cdot, \cdot|\cdot) + P^{(2)}_{X_1,X_2|X_3}(\cdot, \cdot|\cdot)\right),$$

and $x'_2, x''_2 \in \mathcal{X}_2$. It can be shown via (49) that (103)-(104) also hold, and thus applying an induction argument again yields

$$\mathcal{R}^{\text{MA-DBC}}(P^{\text{U}}_{\mathcal{X}_1} P_{X_2|X_3}P_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2}) \subseteq \mathcal{R}^{\text{MA-DBC}}(P^{\text{U}}_{\mathcal{X}_1} P^{\text{U}}_{\mathcal{X}_2} P_{V,X_3}, P_{Y_3|X_1,X_2,X_3}, P_{Z_1}, P_{Z_2}).$$

$$(106)$$

Combining (105) and (106) then proves our claim. Due to symmetry, we only prove (105).

Lemma 7: For any $P^{(1)}_{X_1,X_2,X_3,V} = P^{(1)}_{X_1,X_2|X_3} P^{(1)}_{V,X_3}$, let $P^{(2)}_{X_1,X_2,X_3,V} = P^{(2)}_{X_1,X_2|X_3} P^{(1)}_{V,X_3}$ and $P^{(3)}_{X_1,X_2,X_3,V} = P^{(3)}_{X_1,X_2|X_3} P^{(1)}_{V,X_3}$, where $P^{(2)}_{X_1,X_2|X_3}$ and $P^{(3)}_{X_1,X_2|X_3}$ are given by (101) and (102), respectively. Then, (103)-(104) hold.

Proof: We have

$$I^{(2)}(X_1; Y_3 | X_2, X_3 = x_3)$$

$$= \sum_{x_1,x_2,y_3} P^{(2)}_{X_1,X_2|X_3}(x_1, x_2 | x_3) P_{Y_3|X_1,X_2,X_3}(y_3 | x_1, x_2, x_3)$$

$$\cdot \log \frac{P_{Y_3|X_1,X_2,X_3}(y_3 | x_1, x_2, x_3)}{\sum_{\tilde{x}_1} P^{(2)}_{X_1|X_2,X_3}(\tilde{x}_1 | x_2, x_3) P_{Y_3|X_1,X_2,X_3}(y_3 | \tilde{x}_1, x_2, x_3)}$$

$$= \sum_{x_1,x_2,y_3} P^{(1)}_{X_1,X_2|X_3}(\tau^{\mathcal{X}_1}_{x_1',x_1''}(x_1), x_2 | x_3) P_{Y_3|X_1,X_2,X_3}(\pi^{\mathcal{Y}_3}[x_1',x_1''](y_3) | \tau^{\mathcal{X}_1}_{x_1',x_1''}(x_1), x_2, x_3)$$

$$\cdot \log \frac{P_{Y_3|X_1,X_2,X_3}(\pi^{\mathcal{Y}_3}[x_1',x_1''](y_3) | \tau^{\mathcal{X}_1}_{x_1',x_1''}(x_1), x_2, x_3)}{\sum_{\tilde{x}_1} P^{(1)}_{X_1|X_2,X_3}(\tau^{\mathcal{X}_1}_{x_1',x_1''}(\tilde{x}_1) | x_2, x_3) P_{Y_3|X_1,X_2,X_3}(\pi^{\mathcal{Y}_3}[x_1',x_1''](y_3) | \tau^{\mathcal{X}_1}_{x_1',x_1''}(\tilde{x}_1), x_2, x_3)}$$

$$(107)$$

$$= \sum_{x_1,x_2,y_3} P^{(1)}_{X_1,X_2|X_3}(x_1, x_2 | x_3) P_{Y_3|X_1,X_2,X_3}(y_3 | x_1, x_2, x_3)$$

$$\cdot \log \frac{P_{Y_3|X_1,X_2,X_3}(y_3 | x_1, x_2, x_3)}{\sum_{\tilde{x}_1} P^{(1)}_{X_1|X_2,X_3}(\tilde{x}_1 | x_2, x_3) P_{Y_3|X_1,X_2,X_3}(y_3 | \tilde{x}_1, x_2, x_3)} \qquad (108)$$

$$= I^{(1)}(X_1; Y_2 | X_2, X_3 = x_3),$$

where (107) follows from (48) and (101), (108) holds since $\pi^{\mathcal{Y}_3}[x_1', x_1'']$ and $\tau^{\mathcal{X}_1}_{x_1',x_1''}$ are bijections. By a similar argument, we have that $I^{(2)}(X_2; Y_3 | X_1, X_3 = x_3) = I^{(1)}(X_2; Y_3 | X_1, X_3 = x_3)$ and that $I^{(2)}(X_1, X_2; Y_3 | X_3 = x_3) = I^{(1)}(X_1, X_2; Y_3 | X_3 = x_3)$. Next, using the concavity of $I(X_1; Y_3 | X_2, X_3 = x_3)$, $I(X_2; Y_3 | X_1, X_3 = x_3)$, and $I(X_1, X_2; Y_3 | X_3 = x_3)$ in $P_{X_1,X_2|X_3}(\cdot, \cdot | x_3)$[8] yields that

$$I^{(3)}(X_1; Y_3 | X_2, X_3 = x_3) \geq \frac{1}{2} I^{(1)}(X_1; Y_3 | X_2, X_3 = x_3) + \frac{1}{2} I^{(2)}(X_1; Y_3 | X_2, X_3 = x_3)$$

$$= I^{(1)}(X_1; Y_3 | X_2, X_3 = x_3),$$

---

[8]$I(X_1; Y_3 | X_2, X_3 = x_3)$ and $I(X_2; Y_3 | X_1, X_3 = x_3)$ are concave function of $P_{X_1,X_2|X_3}(\cdot, \cdot | x_3)$ since $I(X_1; Y_2 | X_2)$ and $I(X_2; Y_1 | X_1)$ are both concave in the input distribution P_{X_1,X_2} [3].

$$I^{(3)}(X_2; Y_3 | X_1, X_3 = x_3) \geq \frac{1}{2} I^{(1)}(X_2; Y_3 | X_1, X_3 = x_3) + \frac{1}{2} I^{(2)}(X_2; Y_3 | X_1, X_3 = x_3)$$

$$= I^{(1)}(X_2; Y_3 | X_1, X_3 = x_3),$$

$$I^{(3)}(X_1, X_2; Y_3 | X_3 = x_3) \geq \frac{1}{2} I^{(1)}(X_1, X_2; Y_3 | X_3 = x_3) + \frac{1}{2} I^{(2)}(X_1, X_2; Y_3 | X_3 = x_3)$$

$$= I^{(1)}(X_1, X_2; Y_3 | X_3 = x_3),$$

and hence

$$I^{(3)}(X_1; Y_3 | X_2, X_3) \geq I^{(1)}(X_1; Y_3 | X_2, X_3),$$

$$I^{(3)}(X_2; Y_3 | X_1, X_3) \geq I^{(1)}(X_2; Y_3 | X_1, X_3),$$

$$I^{(3)}(X_1, X_2; Y_3 | X_3) \geq I^{(1)}(X_1, X_2; Y_3 | X_3),$$

since $P_{X_3}^{(1)} = P_{X_3}^{(3)}$. Together with the definition of $\mathcal{R}^{\text{MA-DBC}}$ given in Section IV-B, the inclusions in (103)-(104) are proved. ∎

Now, without loss of generality, suppose that $\mathcal{X}_1 = \{1, 2, ..., \kappa\}$. For $1 \leq m \leq \kappa$, define Λ_m as the set of all conditional probability distributions $P_{X_1, X_2 | X_3}$ satisfying $P_{X_1, X_2 | X_3}(1, x_2 | x_3) = P_{X_1, X_2 | X_3}(2, x_2 | x_3) = \cdots = P_{X_1, X_2 | X_3}(m, x_2 | x_3)$ for any fixed $x_2 \in \mathcal{X}_2$ and $x_3 \in \mathcal{X}_3$. As in the proof of Lemma 6, it can be shown by induction on m that

$$\mathcal{R}^{\text{MA-DBC}}(P_{X_1, X_2 | X_3} P_{V, X_3}, P_{Y_3 | X_1, X_2, X_3}, P_{Z_1}, P_{Z_2}) \subseteq \mathcal{R}^{\text{MA-DBC}}(\tilde{P}_{X_1, X_2 | X_3} P_{V, X_3}, P_{Y_3 | X_1, X_2, X_3}, P_{Z_1}, P_{Z_2})$$

$$(109)$$

where $P_{X_1, X_2 | X_3} \in \Lambda_m$ and $\tilde{P}_{X_1, X_2 | X_3} \in \Lambda_{m+1}$ for $1 \leq m < \kappa$. Note that the base case $m = 1$ was proved in Lemma 7. Since $P_{X_1, X_2 | X_3} \in \Lambda_\kappa$ can be expressed as $P_{X_1, X_2 | X_3} = P_{\mathcal{X}_1}^{\text{U}} P_{X_2 | X_3}$, (105) holds. To show (106), we consider input probability distributions of the form $P_{X_1, X_2, X_3, V} = P_{\mathcal{X}_1}^{\text{U}} P_{X_2 | X_3} P_{X_3, V}$. By changing the roles of X_1 and X_2 in the above derivation, the rest of the proof is straightforward. ∎

REFERENCES

[1] L. Song, F. Alajaji, and T. Linder, "Adaptation is useless for two discrete additive-noise two-way channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, Jul. 2016, pp. 1854-1858.

[2] J. J. Weng, L. Song, F. Alajaji, and T. Linder, "Sufficient conditions for the tightness of Shannon's capacity bounds for two-way channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Vail, CO, Jun. 2018, pp. 1410-1414.

[3] C. E. Shannon, "Two-way communications channels," in *Proc. 4th Berkeley Symp. Math. Stat. Probab.,* Chicago, IL, Jun. 1961, pp. 611-644.

[4] A. Jamalipour, T. Wada, T. Yamazato, "A tutorial on multiple access technologies for beyond 3G mobile networks", *IEEE Commun. Mag.*, vol. 43, no. 2, pp. 110-117, Feb. 2005.

[5] Y. Yuan Z. Yuan, G. Yu, C.-H. Hwang, P. K. Liao, A. Li, and K. Takeda, "Non-orthogonal transmission technology in LTE evolution," *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 68-74, July 2016.

[6] Z. Ding, X. Lei, G. K. Karaginnidis, R. Schober, J. Yuan, and V. K. Bhargava,, "A survey on non-orthogonal multiple access for 5G networks: research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181-2195, Oct. 2017.

[7] E. C. van der Meulen, "A survey of multi-way channels in information theory: 1961-1976", *IEEE Trans. Inf. Theory*, vol. IT-23, no. 2, pp. 1-37, Jan. 1977.

[8] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[9] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surv. Tuts.,* vol. 16, no. 4, pp. 18011819, Fourth Quart. 2014.

[10] G. Dueck, "The capacity region of the two-way channel can exceed the inner bound," *Inform. Contr.*, vol. 40, pp. 258-266, Mar. 1979.

[11] J. Schalkwijk, "The binary multiplying channel - a coding scheme that operates beyond Shannons inner bound region," *IEEE Trans. Inf. Theory*, vol. IT-28, pp. 107-110, Jan. 1982.

[12] J. Schalkwijk, "On an extension of an achievable rate region for the binary multiplying channel," *IEEE Trans. Inf. Theory*, vol. IT-29, pp. 445-448, May 1983.

[13] T. S. Han, "A general coding scheme for the two-way channel," *IEEE Trans. Inf. Theory*, vol. IT-30, pp. 35-44, Jan. 1984.

[14] Z. Zhang, T. Berger, and J. Schalkwijk, "New outer bounds to capacity regions of two-way channels," *IEEE Trans. Inf. Theory*, vol. IT-32, pp. 383-386, May 1986.

[15] A. Hekstra and F. Willems, "Dependence balance bounds for single output two-way channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 44-53, Jan. 1989.

[16] H. B. Meeuwissen. "Information Theoretical Aspects of Two-Way Communication," Ph.D. Dissertation, TU Eindhoven, 1998.

[17] G. Kramer, "Directed Information for Channels with Feedback," Ph.D. Dissertation, Swiss Federal Institute of Technology Zurich, 1998.

[18] J. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Information Theory Its Applic. (ISITA-90)*, Waikiki, HI, Nov. 1990, pp. 303-305.

[19] S. Tatikonda, "Control Under Communication Constraints," Ph.D. Dissertation, MIT, Cambridge, MA, 2000.

[20] J. Chen and T. Berger, "The capacity of finite-state Markov channels with feedback," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 780-789, Mar. 2005.

[21] Y. H. Kim, "A coding theorem for a class of stationary channels with feedback," *IEEE Trans. Inf. Theory*, vol. 54, no. 4, pp. 1488-1499, Apr. 2008.

[22] H. H. Permuter, T. Weissman, and A. J. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 644-662, Feb. 2009.

[23] H. H. Permuter, T. Weissman, and J. Chen, "Capacity region of the finite-state multiple-access channel with and without feedback", *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2455-2477, Jun. 2009.

[24] Z. Cheng and N. Devroye, "Two-way networks: when adaptation is useless," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1793-1813, Mar. 2014.

[25] C. Suh, J. Cho, and D. Tse, "Two-way interference channel capacity: how to have the cake and eat it too," *IEEE Trans. Inf. Theory*, vol. 64, no. 6, pp. 4259-4281, Mar. 2018.

[26] L. Ong, "Capacity results for two classes of three-way channels," in *Proc. IEEE Int. Symp. Commun. Inf. Technol.*, Gold Coast, Oct. 2012, pp. 471-476.

[27] A. Chaaban, H. Maier, A. Sezgin, and R. Mathar, "Three-way channels with multiple unicast sessions: capacity approximation via network transformation," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7086-7102, Dec. 2016.

[28] A. Chaaban, L. R. Varshney, and M. Alouini, "The capacity of injective semi-deterministic two-way channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Jun. 2017, pp. 431-436.

[29] L. R. Varshney, "Two way communications over exponential family type channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Jun. 2013, pp. 2795-2799.

[30] F. Jelinek, "Coding for and decomposition of two-way channels," *IEEE Trans. Inf. Theory,* vol. IT-10, pp. 5-17, Jan. 1964.

[31] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed., New York, U.S.: Cambridge Univ. Press, 2011.

[32] D. Gündüz, E. Erkip, A. Goldsmith, and H. V. Poor, "Source and channel coding for correlated sources over multiuser channels," *IEEE Trans. Inf. Theory,* vol. 55, no. 9, pp. 3927-3944, Sep. 2009.

[33] R. G. Gallager, *Information Theory and Reliable Comminication*, New York, U.S.: John Wiley & Sons, 1968.

[34] T. M. Cover and J. Thomas, *Elements of Information Theory,* 2nd ed., New York, U.S.: John Wiley & Sons, 2006.

[35] F. Alajaji and P.-N. Chen, *An Introduction to Single-User Information Theory,* Singapore: Springer, 2018.

[36] M. Rezaeian and A. Grant, "Computation of total capacity for discrete memoryless multiple-access channels," *IEEE Trans. Inf. Theory,* vol. 50, no. 11, pp. 2779-2784, Nov. 2004.

[37] B. Xie and R. Wesel, "A mutual information invariance approach to symmetry in discrete memoryless channels," in *Proc. Inf. Theory Appl. Workshop (ITA'08)*, San Diego, CA, Feb. 2008, pp. 444-448.

[38] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. IT-18, pp. 460-473, Jul. 1972.