The Capacity Region of a Class of Three-Receiver Broadcast Channels With Degraded Message Sets

Chandra Nair, *Member, IEEE*, and Abbas El Gamal, *Fellow, IEEE*

***Abstract*—Körner and Marton established the capacity region for the two-receiver broadcast channel with degraded message sets. Recent results and conjectures suggest that a straightforward extension of the Körner–Marton region to more than two receivers is optimal. This paper shows that this is not the case. We establish the capacity region for a class of three-receiver broadcast channels with two-degraded message sets and show that it can be strictly larger than the straightforward extension of the Körner–Marton region. The idea is to split the private message into two parts, superimpose one part onto the "cloud center" representing the common message, and superimpose the second part onto the resulting "satellite codeword." One of the receivers finds the common message directly by decoding the "cloud center," a second receiver finds it *indirectly* by decoding a satellite codeword, and a third receiver finds it by jointly decoding the transmitted codeword. This idea is then used to establish new inner and outer bounds on the capacity region of the general three-receiver broadcast channel with two and three degraded message sets. We show that these bounds are tight for some nontrivial cases. The results suggest that finding the capacity region of the three-receiver broadcast channel with degraded message sets is at least as hard finding as the capacity region of the general two-receiver broadcast channel with common and private message.**

***Index Terms*—Broadcast channel, capacity, degraded message sets.**

I. Introduction

A BROADCAST channel with degraded message sets is a model for communication scenarios where a sender wishes to communicate a common message to *all* receivers, a first private message to a first subset of the receivers, a second private message to a second subset of the first subset, and so on. Such scenario can arise, for example, in video or music broadcasting over a wireless network at varying levels of quality. The common message represents the lowest quality version to be sent to all receivers, the first private message represents the additional information needed for the first subset of receivers to decode the second lowest quality version, and so on. What is the set of simultaneously achievable rates for communicating such degraded message sets over the network?

This question was first studied by Körner and Marton in 1977 [1]. They considered a general two-receiver discrete-memoryless broadcast channel with sender X and receivers Y_1 and Y_2. A common message $M_0 \in [1 : 2^{nR_0}]$ is to be sent to both receivers and a private message $M_1 \in [1 : 2^{nR_1}]$ is to be sent only to receiver Y_1. They showed that the capacity region is given by the set of all rate pairs (R_0, R_1) such that[1]

$$\begin{aligned} R_0 &\le \min\{I(U;Y_1), I(U;Y_2)\} \\ R_1 &\le I(X;Y_1|U) \end{aligned} \tag{1}$$

for some $p(u,x)$. These rates are achieved using superposition coding [2]. The common message is represented by the auxiliary random variable U and the private message is superimposed to generate X. The main contribution of [1] is proving a strong converse using the technique of images-of-a-set [3].

Extending the Körner–Marton result to more than two receivers has remained open even for the simple case of three receivers Y_1, Y_2, Y_3 with two-degraded message sets, where a common message M_0 is to be sent to all receivers and a private message M_1 is to be sent only to receiver Y_1. The straightforward extension of the Körner–Marton region to this case yields the achievable rate region consisting of the set of all rate pairs (R_0, R_1) such that

$$\begin{aligned} R_0 &\le \min\{I(U;Y_1), I(U;Y_2), I(U;Y_3)\} \\ R_1 &\le I(X;Y_1|U) \end{aligned} \tag{2}$$

for some $p(u,x)$. Is this region optimal?

In [4], it was shown that the above region (and its natural extension to $k > 3$ receivers) is optimal for a class of product discrete-memoryless and Gaussian broadcast channels, where each of the receivers who decodes only the common message is a degraded version of the unique receiver that also decodes the private message. In [5], it was shown that a straightforward extension of Körner–Marton region is optimal for the class of linear deterministic broadcast channels, where the operations are performed in a finite field. In addition to establishing the degraded message set capacity for this class, the authors gave an explicit characterization of the optimal auxiliary random variables. In a recent paper, Borade *et al.* [6] introduced *multilevel* broadcast channels, which combine aspects of degraded broadcast channels and broadcast channels with degraded message sets. They established an achievable rate region as well as a "mirror-

Manuscript received December 20, 2007; revised April 14, 2009. Current version published September 23, 2009. The work of C. Nair was supported in part by the Direct Grant for research at the Chinese University of Hong Kong. The material in this paper was presented in part at the International Symposium of Information Theory, Toronto, ON, Canada, June 2008.

C. Nair is with the Information Engineering Department, Chinese University of Hong Kong, Shatin, N. T., Hong Kong (e-mail: chandra.nair@gmail.com).

A. El Gamal is with the Electrical Engineering Department, Stanford University, Stanford, CA 94305 USA (e-mail: abbas@ee.stanford.edu).

Communicated by G. Kramer, Associate Editor for Shannon Theory.

Color version of Figure 3 in this paper is available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2009.2027512

[1]The Körner–Marton characterization does not include the second term inside the min in the first inequality $I(U;Y_1)$. Instead it includes the bound $R_0 + R_1 \le I(X;Y_1)$. It can be easily shown that the two characterizations are equivalent.

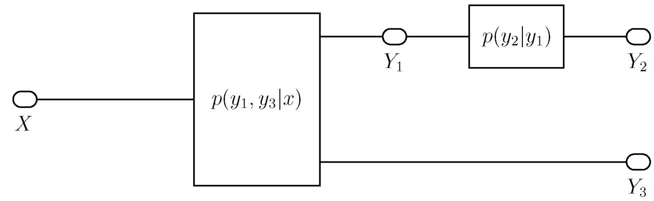

Fig. 1. Multilevel three-receiver broadcast channels. Message M_0 is to be sent to all receivers and message M_1 is to be sent only to Y_1.

image" outer bound for these channels. Their achievable rate region is again a straightforward extension of the Körner–Marton region to k-receiver multilevel broadcast channels. In particular, Conjecture 5 of [6] states that the capacity region for the three-receiver multilevel broadcast channels depicted in Fig. 1 is the set of all rate pairs (R_0, R_1) such that

$$\begin{aligned} R_0 &\leq \min\{I(U;Y_2), I(U;Y_3)\} \\ R_1 &\leq I(X;Y_1|U) \end{aligned} \tag{3}$$

for some $p(u,x)$. Note that this region, henceforth referred to as the BZT region, is the same as (2) because in the multilevel broadcast channel Y_2 is a degraded version of Y_1 and therefore $I(U;Y_2) \leq I(U;Y_1)$.

In this paper, we show that the straightforward extension of the Körner–Marton region to more than two receivers is not in general optimal. We establish the capacity region of the multilevel broadcast channels depicted in Fig. 1 as the set of rate pairs (R_0, R_1) such that

$$\begin{aligned} R_0 &\leq \min\{I(U;Y_2), I(V;Y_3)\} \\ R_1 &\leq I(X;Y_1|U) \\ R_0 + R_1 &\leq I(V;Y_3) + I(X;Y_1|V) \end{aligned}$$

for some $p(u)p(v|u)p(x|v)$ (i.e., $U \to V \to X$ forms a Markov chain), and show that it can be strictly larger than the BZT region. In our coding scheme, the common message M_0 is represented by U (the cloud centers), part of M_1 is superimposed on U to obtain V (satellite codewords), and the rest of M_1 is superimposed on V to yield X. Receiver Y_1 finds M_0, M_1 by decoding X. Receiver Y_2 finds M_0 by decoding U, whereas receiver Y_3 finds M_0 *indirectly* by decoding a satellite codeword V.

Although it seems surprising that higher rates can be achieved by having Y_3 decode more than it needs to, this result can be explained by the fact that for a general two-receiver broadcast channel $X \to (Y_1, Y_2)$, one can have the conditions $I(U;Y_1) < I(U;Y_2)$ and $I(X;Y_1) > I(X;Y_2)$ hold simultaneously [13]. Now, considering our three-receiver broadcast channel scenario, suppose we have a choice of U such that $I(U;Y_3) < I(U;Y_2)$. In this case, requiring both Y_2 and Y_3 to directly decode U necessitates that the rate of the common message be less than $I(U;Y_3)$. From the above fact, a V may exist such that $U \to V \to X$ and $I(V;Y_3) > I(V;Y_2)$, in which case the rate of the common message can be increased to $I(U;Y_2)$ and Y_3 can still find U indirectly by decoding V. Thus, although the additional "degree of freedom" resulting from the introduction of V comes at the expense of having Y_3 decode more than it is required to, it can yield higher achievable rates.

The rest of this paper is organized as follows. In Section II, we provide needed definitions. In Section III, we establish the capacity region for the multilevel broadcast channel in Fig. 1 (Theorem 1). In Section IV, we show through an example that the capacity region for the multilevel broadcast channel can be strictly larger than the BZT region. In Section V, we extend the results on the multilevel broadcast channel to establish inner and outer bounds on the capacity region of the general three-receiver broadcast channel with two-degraded message sets (Propositions 5 and 6). We show that these bounds are tight when Y_1 is less noisy than Y_2 (Proposition 7), which is a more relaxed condition than the degradedness condition of the multilevel broadcast channel model. We then extend the inner bound to three-degraded message sets (Theorem 2). Although Proposition 5 is a special case of Theorem 2, it is presented earlier for clarity of exposition. Finally, we show that the inner bound of Theorem 2 when specialized to the case of two-degraded message sets, where M_0 is to be sent to all receivers and M_1 is to be sent to Y_1 and Y_2, reduces to the straightforward extension of the Körner–Marton region (Corollary 1). We show that this inner bound is tight for deterministic broadcast channels (Proposition 10) and when Y_1 is less noisy than Y_3 and Y_2 is less noisy than Y_3 (Proposition 11).

II. Definitions

Consider a discrete-memoryless three-receiver broadcast channel consisting of an input alphabet $\mathcal{X}$, output alphabets $\mathcal{Y}_1$, $\mathcal{Y}_2$, and $\mathcal{Y}_3$, and a probability transition function $p(y_1, y_2, y_3|x)$.

A $(2^{nR_0}, 2^{nR_1}, n)$ two-degraded message set code for a three-receiver broadcast channel consists of: 1) a pair of messages (M_0, M_1) uniformly distributed over $[1:2^{nR_0}] \times [1:2^{nR_1}]$, 2) an encoder that assigns a codeword $x^n(m_0, m_1)$, for each message pair $(m_0, m_1) \in [1:2^{nR_0}] \times [1:2^{nR_1}]$, and 3) three decoders, one that maps each received y_1^n sequence into an estimate $(\hat{m}_{01}, \hat{m}_1) \in [1:2^{nR_0}] \times [1:2^{nR_1}]$, a second that maps each received Y_3^n sequence into an estimate $\hat{m}_{02} \in [1:2^{nR_0}]$, and a third that maps each received Y_2^n sequence into an estimate $\hat{m}_{03} \in [1:2^{nR_0}]$.

The probability of error is defined as

$$P_e^{(n)} = \mathrm{P}\{\hat{M}_1 \neq M_1 \text{ or } \hat{M}_{0k} \neq M_0 \text{ for } k = 1, 2, \text{ or } 3\}.$$

A rate tuple (R_0, R_1) is said to be achievable if there exists a sequence of $(2^{nR_0}, 2^{nR_1}, n)$ 2-degraded message set codes with $P_e^{(n)} \to 0$. The capacity region of the broadcast channel is the closure of the set of achievable rates.

A three-receiver *multilevel* broadcast channel [6] is a three-receiver broadcast channel with two-degraded message sets where $p(y_1, y_2, y_3|x) = p(y_1, y_3|x)p(y_2|y_1)$ for every $(x, y_1, y_2, y_3) \in \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Y}_3$ (see Fig. 1).

In addition to considering the multilevel three-receiver broadcast channel and the general three-receiver broadcast channel with two-degraded message sets, we will also consider the following two scenarios.

1) Three-receiver broadcast channel with three message sets, where M_0 is to be sent to all receivers, M_1 is to be sent to Y_1 and Y_3, and M_2 is to be sent only to Y_1.
2) Three-receiver broadcast channel with two-degraded message sets, where M_0 is to be sent to all receivers and M_1 is to be sent to Y_1 and Y_3.

Definitions of codes, achievability, and capacity regions for these cases are straightforward extensions of the above definitions. Clearly, the two-degraded message set scenarios are special cases of the three-degraded message set. Nevertheless, we will begin with the special class of multilevel broadcast channel because we are able to establish its capacity region.

III. Capacity of three-Receiver Multilevel Broadcast Channel

A key result of this paper is given in the following theorem.

Theorem 1: The capacity region of the three-receiver multilevel broadcast channel in Fig. 1 is the set of rate pairs (R_0, R_1) such that

$$\begin{aligned} R_0 &\leq \min\{I(U;Y_2), I(V;Y_3)\} \\ R_1 &\leq I(X;Y_1|U) \\ R_0 + R_1 &\leq I(V;Y_3) + I(X;Y_1|V) \end{aligned} \tag{4}$$

for some $p(u)p(v|u)p(x|v)$, where the cardinalities of the auxiliary random variables satisfy $\|\mathcal{U}\| \leq \|\mathcal{X}\| + 4$ and $\|V\| \leq \|\mathcal{X}\|^2 + 5\|\mathcal{X}\| + 4$.

Remark 3.1: It is straightforward to show by setting $U = V$ in the above theorem that the BZT region (3) is contained in the capacity region (4). We show in the next section that the capacity region (4) can be strictly larger than the BZT region.

Remark 3.2: It is straightforward to show that the above region is convex and therefore there is no need to use a time-sharing auxiliary random variable.

The proof of Theorem 1 is given in the following subsections. We first prove the converse. In Section III-C, we prove achievability, and in Section III-D, we establish the bounds on the cardinalities of the auxiliary random variables.

A. Converse of Theorem 1

We show that the region in Theorem 1 forms an outer bound to the capacity region. The proof is quite similar to previous weak converse and outer bound proofs for two-receiver broadcast channels (e.g., see [7]–[9]). Suppose we are given a sequence of codes for the multilevel broadcast channel with $P_e^{(n)} \to 0$. For each code, we form the empirical distribution for M_0, M_1, X^n.

Since $X \to Y_1 \to Y_2$ forms a *physically degraded* broadcast channel, it follows that the code rate pair (R_0, R_1) must satisfy the inequalities

$$\begin{aligned} R_0 &\leq I(U;Y_2) \\ R_1 &\leq I(X;Y_1|U) \end{aligned} \tag{5}$$

for some $p(u,x)$, where U, X, Y_1, Y_2 are defined as follows [7], [12]. Let $U_i = (M_0, Y_1^{i-1})$, $i = 1, \ldots, n$, and let Q be a time-sharing random variable uniformly distributed over the set $\{1, 2, \ldots, n\}$ and independent of X^n, Y_1^n, Y_3^n, Y_2^n. We then set $U = (Q, U_Q)$ and $X = X_Q$, $Y_1 = Y_{1Q}$, and $Y_2 = Y_{2Q}$. Thus, we have established the bounds in five.

Next, since the decoding requirements of the broadcast channel $X \to (Y_1, Y_3)$ makes it a broadcast channel with degraded message sets, the code rate pair must satisfy the inequalities

$$\begin{aligned} R_0 &\leq \min\{I(V;Y_3), I(V,Y_1)\} \\ R_0 + R_1 &\leq I(V;Y_3) + I(X;Y_1|V) \end{aligned}$$

for some $p(v,x)$[8], where U_2 is identified as follows. Let $V_i = (M_0, Y_1^{i-1}, Y_{3\,i+1}^{\;n})$, $i = 1, \ldots, n$, then we set $V = (Q, V_Q)$.

Combining the above two outer bounds, we see that $U \to V \to X$ forms a Markov chain. Observe that this Markov nature of the auxiliary random variables along with the degraded nature of $X \to Y_1 \to Y_2$ implies that $I(V;Y_1) \geq I(V;Y_2) \geq I(U;Y_2)$. Thus, we have shown that the code rate pair (R_0, R_1) must be in region (4). This establishes the converse to Theorem 1.

B. Achievability of Theorem 1

The interesting part of the proof of Theorem 1 is achievability. We split the rate of the private message M_1 into two parts M_{11}, M_{12} with rates S_1, S_2, respectively. Thus, $R_1 = S_1 + S_2$. The common message M_0 is represented by U, (M_0, M_{11}) is represented by V, and (M_0, M_1) is represented by X. Receiver Y_1 finds (M_0, M_1) by decoding X, receiver Y_2 finds M_0 by decoding U, and receiver Y_3 finds M_0 indirectly by decoding V. We now provide details of the proof.

Code Generation: Fix a distribution $p(u)p(v|u)p(x|v)$. Randomly and independently generate 2^{nR_0} sequences $u^n(m_0)$, $m_0 \in \{1, 2, \ldots, 2^{nR_0}\} := [1 : 2^{nR_0}]$, each distributed uniformly over the set of ϵ-typical[2] u^n sequences. For each $u^n(m_0)$, randomly and independently generate 2^{nS_1} sequences $v^n(m_0, s_1)$, $s_1 \in [1 : 2^{nS_1}]$, each distributed uniformly over the set of conditionally ϵ-typical v^n sequences given $u^n(m_0)$. For each $v^n(m_0, s_1)$, randomly and independently generate 2^{nS_2} sequences $x^n(m_0, s_1, s_2)$, $s_2 \in [1 : 2^{nS_2}]$, each distributed uniformly over the set of conditionally ϵ-typical x^n sequences given $v^n(m_0, s_1)$.

Encoding: To send the message pair $(m_0, m_1) \in [1 : 2^{nR_0}] \times [1 : 2^{nR_1}]$, the sender expresses m_1 by the pair $(s_1, s_2) \in [1 : 2^{nS_1}] \times [1 : 2^{nS_2}]$ and sends $x^n(m_0, s_1, s_2)$.

Decoding and Analysis of Error Probability:

1) Receiver Y_2 declares that m_0 is sent if it is the unique message such that $u^n(m_0)$ and y_2^n are jointly ϵ-typical. It is easy to see that this can be achieved with arbitrarily small probability of error if

$$R_0 < I(U;Y_2). \tag{6}$$

[2]We assume strong typicality [10] throughout this paper.

2) Receiver Y_1 declares that (m_0, s_1, s_2) is sent if it is the unique triple such that $x^n(m_0, s_1, s_2)$ and y_1^n are jointly ϵ-typical. It is easy to see using joint decoding that this decoding succeeds with high probability as long as

$$\begin{aligned} R_0 + S_1 + S_2 &< I(X; Y_1) \\ S_1 + S_2 &< I(X; Y_1|U) \\ S_2 &< I(X; Y_1|V). \end{aligned} \tag{7}$$

3) Receiver Y_3 finds m_0 as follows. It declares that $m_0 \in [1 : 2^{nR_0}]$ is sent if it is the unique index such that $v^n(m_0, s_1)$ and y_3^n are jointly ϵ-typical for some $s_1 \in [1 : 2^{nS_1}]$.
We claim that receiver Y_3 can correctly decode m_0 with arbitrarily small probability of error if

$$R_0 + S_1 < I(V; Y_3). \tag{8}$$

Since $R_0 + S_1 < I(V; Y_3)$, there exists $\delta > 0$ such that $R_0 + S_1 \le I(V; Y_3) - 2\delta$. Suppose $(1, 1) \in [1 : 2^{nR_0}] \times [1 : 2^{nS_1}]$ is the message pair sent, then the probability of error averaged over the choice of codebooks can be upper bounded as follows:

$$\begin{aligned} P_e^{(n)} &\le \mathrm{P}\{(V^n(1,1), Y_3^n) \text{ are not jointly } \epsilon\text{-typical}\} \\ &\quad + \mathrm{P}\{(V^n(m_0, s_1), Y_3^n) \text{ are jointly } \epsilon\text{-typical} \\ &\qquad \text{for some } m_0 \neq 1\} \\ &\overset{(a)}{<} \delta' + \sum_{m_0 \neq 1} \sum_{s_1} \mathrm{P}\{(V^n(m_0, s_1), Y_3^n) \\ &\qquad \text{jointly } \epsilon\text{-typical}\} \\ &\overset{(b)}{\le} \delta' + 2^{n(R_0+S_1)} 2^{-n(I(V;Y_3)-\delta)} \\ &\overset{(c)}{\le} \delta' + 2^{-n\delta} \end{aligned}$$

where (a) follows by the union of events bound, (b) follows by the fact that for $m_0 \neq 1$, $V^n(m_0, s_1)$ and Y_3^n are generated completely independently and thus each probability term under the sum is upper bounded by $2^{-n(I(V;Y_3)-\delta)}$ [10] as $n \to \infty$, (c) follows from $R_0 + S_1 \le I(V; Y_3) - 2\delta$. We know that $\delta' \to 0$ as $\epsilon \to 0$ and therefore with arbitrarily high probability, any $V^n(m_0, s_1)$ jointly ϵ-typical with the received Y_3^n sequence must be of the form $V^n(1, s_1)$. Hence, receiver Y_3 can correctly decode M_0 with arbitrarily small probability of error if

$$R_0 + S_1 < I(V; Y_3).$$

Thus, from (6)–(8), all receivers can decode their intended messages with arbitrarily small probability of error if

$$\begin{aligned} R_0 &< I(U; Y_2) \\ R_0 + S_1 + S_2 &< I(X; Y_1) \\ S_1 + S_2 &< I(X; Y_1|U) \\ S_2 &< I(X; Y_1|V) \\ R_0 + S_1 &< I(V; Y_3). \end{aligned}$$

Substituting $S_1 + S_2 = R_1$ and using the Fourier–Motzkin procedure [17] to eliminate S_1 and S_2 shows that any rate pair (R_0, R_1), satisfying the conditions in four, is achievable. This completes the proof of achievability of Theorem 1.

We will refer to the decoding step of Y_3 as *indirect decoding*, since the receiver decodes U indirectly by decoding V. Do we achieve the same region by having Y_3 *jointly decode* M_0, M_{11}? To answer this question, note that for the joint decoder, the probability of error can be made arbitrarily small if

$$\begin{aligned} R_0 + S_1 &< I(V; Y_3) \\ S_1 &< I(V; Y_3|U). \end{aligned}$$

Since bounding the probability of error for the indirect decoder requires only the first inequality, it is in general less restrictive than the joint decoder. Now, combining the conditions for the joint decoder to succeed with (6) and (7) and performing Fourier–Motzkin to eliminate S_1 and S_2, we obtain the set of rate pairs (R_0, R_1) satisfying

$$\begin{aligned} R_0 &< \min\{I(U; Y_2), I(V; Y_3)\} \\ R_1 &< I(X; Y_1|U) \\ R_0 + R_1 &< I(V; Y_3) + I(X; Y_1|V) \\ R_1 &< I(V; Y_3|U) + I(X; Y_1|V) \end{aligned}$$

for some $p(u)p(v|u)p(x|v)$.

Note that this region involves one more inequality than the capacity region given by (4). However, by optimizing the choice of V for each given U, we can show that this inequality is not necessary. There are two cases.

Case 1) $I(U; Y_2) < I(U; Y_3)$: In this case, it is easy to see that the optimal choice is to set $V = U$. Thus, indirect decoding and joint decoding yield the same region.

Case 2) $I(U; Y_2) > I(U; Y_3)$: In this case, for any V, at the corner point of the indirect decoding region prescribed by the pair of random variables (U, V), we have

$$\begin{aligned} R_1^* = I(X; Y_1|V) + \min\{&I(V; Y_1|U), I(V; Y_3) \\ &- \min[I(U; Y_2), I(V; Y_3)]\}. \end{aligned}$$

Clearly $\min\{I(U; Y_2), I(V; Y_3)\} \ge I(U; Y_3)$, which implies that

$$R_1^* \le I(V; Y_3|U) + I(X; Y_1|V)$$

i.e., R_1^* satisfies the additional constraint that joint decoding imposes and hence the corner point is in the joint decoding region. Thus, the regions obtained via indirect decoding and those obtained via joint decoding are equal.

Remark 3.3: In spite of this equivalence, indirect decoding offers some advantages over joint decoding.

1) Indirect decoding yields less inequalities than joint decoding, and thus results in simpler achievable rate region descriptions. This is akin to the equivalent but simpler description of the Han–Kobayashi achievable rate region for the interference channel in [15].
2) Proving the converse for the joint decoding region directly seems very difficult. Using indirect decoding (which shows

that the extra inequality in the description of the joint decoding region is superfluous) makes proving the converse quite straightforward.

3) As we generalize achievability to broadcast channels with various message set requirements, it is not clear that the extra inequalities imposed by joint decoding would still be redundant. Hence, it is conceivable that indirect decoding can outperform joint decoding in general.

C. *Proof of Cardinality Bounds in Theorem 1*

The bounds on the cardinality of the auxiliary random variables are based on a strengthened version of Carathéodory's theorem by Fenchel and Eggleston stated in [11]. The strengthened Carathéodory theorem along with standard arguments [12] imply that for any choice of the auxiliary random variable U, there exists a random variable U_1 with cardinality bounded by $\|\mathcal{X}\| + 1$ such that $I(U;Y_2) = I(U_1;Y_2)$ and $I(X;Y_1|U) = I(X;Y_1|U_1)$. Similarly for any choice of V, one can obtain a random variable V_1 with cardinality bounded by $\|\mathcal{X}\| + 1$ such that $I(V;Y_3) = I(V_1;Y_3)$ and $I(X;Y_1|V) = I(X;Y_1|V_1)$. While these cardinality-bounded random variables do not change the numerical value of the bounds in (4), it is not clear that they preserve the Markov condition $U_1 \rightarrow V_1 \rightarrow X$. To circumvent this problem and preserve the Markov chain, we adapt arguments from [11], where the authors dealt with the same issue, to establish the cardinality bounds stated in Theorem 1. For completeness, we provide an outline of the argument.

This argument is proved in two steps. In the first step, a random variable U_1 and transition probabilities $p(v|u_1)$ are constructed such that the following are held constant: $p(x)$, the marginal probability $p(X)$ (and hence $p(Y_1), p(Y_2), p(Y_3)$), $H(Y_1|U)$, $H(Y_2|U)$, $H(Y_3|U)$, $H(Y_3|V,U)$, and $H(Y_1|V,U)$. Using standard arguments [11], [12], there exists a random variable U_1 (with cardinality of U_1 bounded by $\|\mathcal{X}\| + 4$) and transition probabilities $p(v|u_1)$ that satisfy the above constraints. Note that the distribution of V is not necessarily preserved and hence denotes the resulting random variable as V'.

We thus have random variables $U_1 \rightarrow V' \rightarrow X$ such that

$$\begin{aligned} I(U;Y_2) &= I(U_1;Y_2) \\ I(U;Y_3) &= I(U_1;Y_3) \\ I(X;Y_1|U) &= I(X;Y_1|U_1) \\ I(V;Y_1|U) &= I(V';Y_1|U_1). \end{aligned} \tag{9}$$

In the second step, for each $U_1 = u_1$, a new random variable $V_1(u_1)$ is found such that the following are held constant: $p(x|u_1)$, the marginal distribution of X conditioned on $U_1 = u_1$, $H(Y_1|V', U_1 = u_1)$, and $H(Y_3|V', U_1 = u_1)$. Again standard arguments imply that there exists a random variable $V_1(u_1)$ (with cardinality of V_1 bounded by $\|\mathcal{X}\|+1$) and transition probabilities $p(x|v_1(u_1))$ that satisfy the above constraints. This in particular implies that

$$\begin{aligned} I(V_1(U_1);Y_3|U_1) &= I(V';Y_3|U_1) = I(V;Y_3|U) \\ I(V_1(U_1);Y_1|U_1) &= I(V';Y_1|U_1) = I(V;Y_1|U). \end{aligned} \tag{10}$$

Now, set $V_1 = (U_1, V_1(U_1))$ and observe the following as a consequence of (9) and (10):

$$\begin{aligned} I(V_1;Y_3) &= I(U_1;Y_3) + I(V_1(U_1);Y_3|U_1) \\ &= I(U;Y_3) + I(V;Y_3|U) = I(V;Y_3) \\ I(X;Y_1|V_1) &= I(X;Y_1|U_1) - I(V_1(U_1);Y_1|U_1) \\ &= I(X;Y_1|U) - I(V;Y_1|U) = I(X:Y_1|V). \end{aligned}$$

We thus have the required random variables U_1, V_1 satisfying the cardinality bounds $\|\mathcal{X}\| + 4$ and $(\|\mathcal{X}\| + 4)(\|\mathcal{X}\| + 1)$, respectively, as desired. Furthermore, observe that $U_1 = f(V_1)$ and hence $U_1 \rightarrow V_1 \rightarrow X$ forms a Markov chain.

IV. Multilevel Product Broadcast Channel

In this section, we show that the BZT region can be strictly smaller than the capacity region in Theorem 1.

Consider the product of two three-receiver broadcast channels given by the Markov relationships

$$\begin{aligned} &X_1 \rightarrow Y_{31} \rightarrow Y_{11} \rightarrow Y_{21} \\ &X_2 \rightarrow Y_{12} \rightarrow Y_{22}. \end{aligned} \tag{11}$$

In the Appendix, we derive the following simplified characterizations for the capacity and the BZT regions.

Proposition 1: The BZT region for the above product channel reduces to the set of rate pairs (R_0, R_1) such that

$$R_0 \le I(U_1;Y_{21}) + I(U_2;Y_{22}) \tag{12a}$$
$$R_0 \le I(U_1;Y_{31}) \tag{12b}$$
$$R_1 \le I(X_1;Y_{11}|U_1) + I(X_2;Y_{12}|U_2) \tag{12c}$$

for some $p(u_1)p(u_2)p(x_1|u_1)p(x_2|u_2)$.

Proposition 2: The capacity region for the product channel reduces to the set of rate pairs (R_0, R_1) such that

$$R_0 \le I(U_1;Y_{21}) + I(U_2;Y_{22}) \tag{13a}$$
$$R_0 \le I(V_1;Y_{31}) \tag{13b}$$
$$R_1 \le I(X_1;Y_{11}|U_1) + I(X_2;Y_{12}|U_2) \tag{13c}$$
$$R_0 + R_1 \le I(V_1;Y_{31}) + I(X_1;Y_{11}|V_1) + I(X_2;Y_{12}|U_2) \tag{13d}$$

for some $p(u_1)p(v_1|u_1)p(x_1|v_1)p(u_2)p(x_2|u_2)$.

Now we compare these two regions via the following examples.

A. *Discrete-Memoryless Example*

Consider the multilevel product broadcast channel example in Fig. 2, where $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y}_{12} = \mathcal{Y}_{21} = \{0, 1\}$, and $\mathcal{Y}_{11} = \mathcal{Y}_{31} = \mathcal{Y}_{32} = \{0, E, 1\}$, $Y_{31} = X_1$, $Y_{12} = X_2$, the channels $Y_{31} \rightarrow Y_{11}$ and $Y_{12} \rightarrow Y_{22}$ are binary erasure channels (BEC) with erasure probability $\frac{1}{2}$, and the channel $Y_{11} \rightarrow Y_{21}$ is given by the transition probabilities: $\mathrm{P}\{Y_{21} = E|Y_{11} = E\} = 1$, $\mathrm{P}\{Y_{21} = E|Y_{11} = 0\} = \mathrm{P}\{Y_{21} = E|Y_{11} = 1\} = 2/3$,

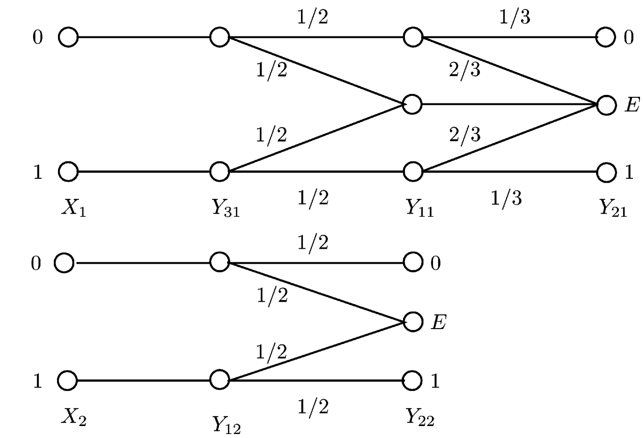

Fig. 2. Product multilevel broadcast channel example.

$\mathrm{P}(Y_{21} = 0|Y_{11} = 0\} = \mathrm{P}\{Y_{21} = 1|Y_{11} = 1\} = 1/3$. Therefore, the channel $X_1 \to Y_{21}$ is effectively a BEC with erasure probability $5/6$.

The BZT region can be simplified to the following.

Proposition 3: The BZT region for the above example reduces to the set of rate pairs (R_0, R_1) satisfying

$$\begin{aligned} R_0 &\leq \min\left\{\frac{p}{6} + \frac{q}{2}, p\right\} \\ R_1 &\leq \frac{1-p}{2} + 1 - q \end{aligned} \tag{14}$$

for some $0 \leq p, q \leq 1$.

The proof of this proposition is given in the Appendix. It is quite straightforward to see that $(R_0, R_1) = (\frac{1}{2}, \frac{5}{12})$ lies on the boundary of this region.

The capacity region can be simplified to the following

Proposition 4: The capacity region for the channel in Fig. 2 reduces to set of rate pairs (R_0, R_1) satisfying

$$\begin{aligned} R_0 &\leq \min\left\{\frac{r}{6} + \frac{s}{2}, t\right\} \\ R_1 &\leq \frac{1-r}{2} + 1 - s \\ R_0 + R_1 &\leq t + \frac{1-t}{2} + 1 - s \end{aligned} \tag{15}$$

for some $0 \leq r \leq t \leq 1, 0 \leq s \leq 1$.

The proof of this proposition is also given in the Appendix. Note that substituting $r = t$ yields the BZT region. By setting $r = 0$, $s = 1$, and $t = 1$, it is easy to see that $(R_0, R_1) = (1/2, 1/2)$ lies on the boundary of the capacity region. On the other hand, for $R_0 = 1/2$, the maximum achievable R_1 in the BZT region is 5/12. Thus, the capacity region is strictly larger than the BZT region.

Fig. 3 plots the BZT region and the capacity region for the example channel. Both regions are specified by two line segments. The boundary of the BZT regions consists of the line segments: $(0, 3/2)$ to $(0.6, 0.2)$ and $(0.6, 0.2)$ to $(2/3, 0)$. The capacity region on the other hand is formed by the pair of line segments: $(0, 3/2)$ to $(1/2, 1/2)$ and $(1/2, 1/2)$ to $(2/3, 0)$. Note that the boundaries of the two regions coincide on the line segment joining $(0.6, 0.2)$ to $(2/3, 0)$.

B. Gaussian Example

Consider a three-receiver Gaussian product multilevel broadcast channel, where

$$\begin{aligned} &Y_{31} = X_1 + Z_1, \quad Y_{11} = Y_{31} + Z_2, \quad Y_{21} = Y_{11} + Z_3 \\ &Y_{12} = X_2 + Z_4, \quad Y_{22} = Y_{12} + Z_5. \end{aligned}$$

The power of noise component Z_i is N_i for $i = 1, 2, \ldots, 5$. We assume a total average power constraint P on $X = (X_1, X_2)$.

Using Gaussian signaling, it can be easily shown that the BZT region is the set of all (R_0, R_1) such that

$$\begin{aligned} R_0 \leq\; &\mathcal{C}\left(\frac{\alpha P_1}{\bar{\alpha} P_1 + N_1 + N_2 + N_3}\right) \\ &+ \mathcal{C}\left(\frac{\beta(P - P_1)}{\bar{\beta}(P - P_1) + N_4 + N_5}\right) \\ R_0 \leq\; &\mathcal{C}\left(\frac{\alpha P_1}{\bar{\alpha} P_1 + N_1}\right) \\ R_1 \leq\; &\mathcal{C}\left(\frac{\bar{\alpha} P_1}{N_1 + N_2}\right) + \mathcal{C}\left(\frac{\bar{\beta}(P - P_1)}{N_4}\right) \end{aligned} \tag{16}$$

for some $0 \leq P_1 \leq P$, $0 \leq \alpha, \beta \leq 1$. Now using Gaussian signaling to evaluate region (13), we obtain the achievable rate region consisting of the set of all (R_0, R_1) such that

$$\begin{aligned} R_0 \leq\; &\mathcal{C}\left(\frac{a P_1}{\bar{a} P_1 + N_1 + N_2 + N_3}\right) \\ &+ \mathcal{C}\left(\frac{b(P - P_1)}{\bar{b}(P - P_1) + N_4 + N_5}\right) \\ R_0 \leq\; &\mathcal{C}\left(\frac{(a + a_1) P_1}{(1 - a - a_1) P_1 + N_1}\right) \\ R_0 + R_1 \leq\; &\mathcal{C}\left(\frac{a P_1}{\bar{a} P_1 + N_1 + N_2 + N_3}\right) \\ &+ \mathcal{C}\left(\frac{b(P - P_1)}{\bar{b}(P - P_1) + N_4 + N_5}\right) \\ &+ \mathcal{C}\left(\frac{\bar{a} P_1}{N_1 + N_2}\right) + \mathcal{C}\left(\frac{\bar{b}(P - P_1)}{N_4}\right) \\ R_0 + R_1 \leq\; &\mathcal{C}\left(\frac{(a + a_1) P_1}{(1 - a - a_1) P_1 + N_1}\right) \\ &+ \mathcal{C}\left(\frac{(1 - a - a_1) P_1}{N_1 + N_2}\right) + \mathcal{C}\left(\frac{\bar{b}(P - P_1)}{N_4}\right) \end{aligned} \tag{17}$$

for some $0 \leq P_1 \leq P$ and $0 \leq a, a_1, b, a + a_1 \leq 1$.

Now consider the above regions with the parameters values: $P = 1, N_1 = 0.4, N_2 = N_3 = 0.1, N_4 = 0.5, N_5 = 0.1$. Fixing $R_1 = 0.5 \log(0.49/0.3)$, we can show that the maximum achievable R_0 in the Gaussian BZT region is $0.5 \log(2.2033957\ldots)$. This is attained using the values $P_1 = 0.5254962\ldots$, $\bar{\alpha} P_1 = 0.02003176\ldots$, and $\bar{\beta} \bar{P}_1 = 0.2852085\ldots$

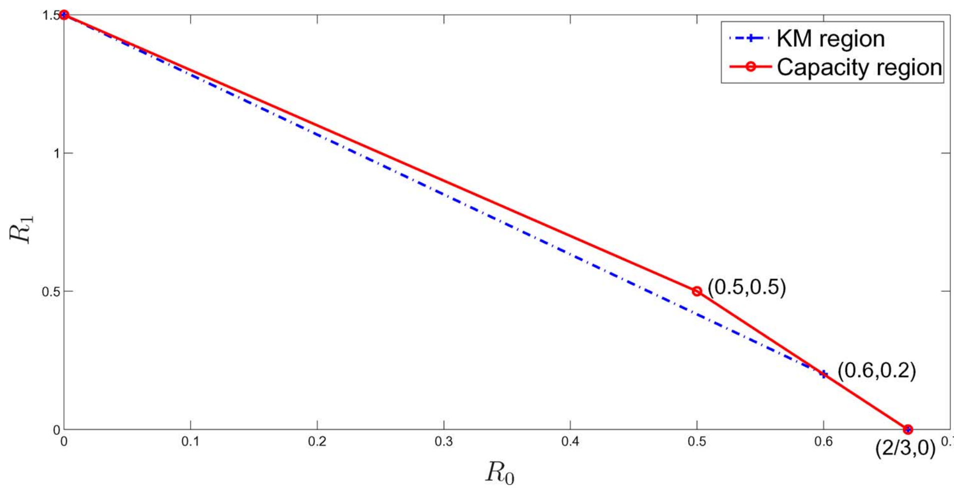

Fig. 3. BZT and the capacity regions for the channel in Fig. 2.

On the other hand, setting $a_1 = 0.03$, $P_1 = 0.5204962$, and retaining the values $\bar{a}P_1 = \bar{\alpha}P_1 = 0.02003176\ldots$, $\bar{b}\bar{P}_1 = \bar{\beta}\bar{P}_1 = 0.2852085\ldots$, the inequalities for region 17 reduce to

$$\begin{aligned} R_0 &\le 0.5\log 2.2038147\ldots \\ R_0 &\le 0.5\log 2.2761073\ldots \\ R_0 + R_1 &\le 0.5\log 2.2038138\ldots + 0.5\log(0.49/0.3) \\ R_0 + R_1 &\le 0.5\log 2.276102975\ldots + 0.5\log 1.5842896\ldots \\ &= 0.5\log 2.2077631\ldots + 0.5\log(0.49/0.3). \end{aligned}$$

Therefore, the rate pairs $(R_0, R_1) = (0.5\log 2.2038147\ldots, 0.5\log(0.49/0.3))$ are achievable (which is outside the BZT region).

Remark 4.1: Note that the BZT region can be viewed as a restriction of the capacity region onto $a_1 = 0$ plane. At the above extreme point of the BZT region, it can be shown that, if we keep the products $\bar{\alpha}P_1$ and $\bar{\beta}\bar{P}_1$ constant, then any small perturbation $\Delta P_1 < 0$, $\Delta a_1 > 0$, $0.1P_1(P_1 + 0.4)/(x(x-0.1))\Delta a_1 > -\Delta P_1$, where $x = \bar{\alpha}P_1 + 0.5$, leads to a strict increase in R_0 for a fixed R_1. The improvement presented is obtained by taking $\Delta P_1 = -0.005$, and $\Delta a_1 = 0.03$, respectively.

Thus, restricted to Gaussian signaling, the BZT region (12) is strictly contained in region (13). However, we have not been able to prove that Gaussian signaling is optimal for either the BZT region or the capacity region.

Remark 4.2: The reader may ask why we did not consider the more general product channel

$$\begin{aligned} X_1 &\to Y_{31} \to Y_{11} \to Y_{21} \\ X_2 &\to Y_{12} \to Y_{22} \to Y_{32}. \end{aligned}$$

In fact, we considered this more general class at first but were unable to show that the capacity region conditions reduce to the separated form

$$\begin{aligned} R_0 &\le I(U_1; Y_{21}) + I(U_2; Y_{22}) \\ R_0 &\le I(V_1; Y_{31}) + I(V_2; Y_{32}) \\ R_0 + R_1 &\le I(U_1; Y_{21}) + I(U_2; Y_{22}) \\ &\quad + I(X_1; Y_{11}|U_1) + I(X_2; Y_{12}|U_2) \\ R_0 + R_1 &\le I(V_1; Y_{31}) + I(V_2; Y_{32}) \\ &\quad + I(X_1; Y_{11}|V_1) + I(X_2; Y_{12}|V_2) \end{aligned}$$

for some $p(u_1)p(v_1|u_1)p(x_1|v_1)p(u_2)p(v_2|u_2)p(x_2|v_2)$.

V. General Three-Receiver Broadcast Channel With Degraded Message Sets

In this section, we extend the results in Section III to obtain inner and outer bounds on the capacity region of general three-receiver broadcast channel with degraded message sets. We first consider the same two-degraded message set scenario as in Section III but without the condition that $X \to Y_1 \to Y_2$ form a degraded broadcast channel. We establish inner and outer bounds for this case and show that they are tight when the channel $X \to Y_1$ is *less noisy* than the channel $X \to Y_2$, which is a more general class than degraded broadcast channels [13]. We then extend our results to the case of three-degraded message sets, where M_0 is to be sent to all receivers, M_1 is to be sent to receivers Y_1 and Y_2, and M_2 is to be sent only to receiver Y_1. A special case of this inner bound gives an inner bound to the capacity of the two-degraded message set scenario where M_0 is to be sent to all receivers and M_1 is to be sent to receivers Y_1 and Y_2 only.

A. Inner and Outer Bounds for Two-Degraded Message Sets

We use rate splitting, superposition coding, indirect decoding, and the Marton achievability scheme for the general two-receiver broadcast channels [14] to establish the following inner bound.

Proposition 5: A rate pair (R_0, R_1) is achievable in a general three-receiver broadcast channel with two-degraded message sets if it satisfies the following inequalities:

$$\begin{aligned} R_0 &\le \min\{I(V_2;Y_2), I(V_3;Y_3)\} \\ 2R_0 &\le I(V_2;Y_2) + I(V_3;Y_3) - I(V_2;V_3|U) \\ R_0 + R_1 &\le \min\{I(X;Y_1), I(V_2;Y_2) + I(X;Y_1|V_2), \\ &\qquad I(V_3;Y_3) + I(X;Y_1|V_3)\} \\ 2R_0 + R_1 &\le I(V_2;Y_2) + I(V_3;Y_3) \\ &\qquad + I(X;Y_1|V_2,V_3) - I(V_2;V_3|U) \\ 2R_0 + 2R_1 &\le I(V_2;Y_2) + I(X;Y_1|V_2) + I(V_3;Y_3) \\ &\qquad + I(X;Y_1|V_3) - I(V_2;V_3|U) \\ 2R_0 + 2R_1 &\le I(V_2;Y_2) + I(V_3;Y_3) + I(X;Y_1|U) \\ &\qquad + I(X;Y_1|V_2,V_3) - I(V_2;V_3|U) \end{aligned} \tag{18}$$

for some

$$\begin{aligned} p(u,v_2,v_3,x) &= p(u)p(v_2|u)p(x,v_3|v_2) \\ &= p(u)p(v_3|u)p(x,v_2|v_3) \end{aligned}$$

(or in other words, both $U \to V_2 \to (V_3, X)$ and $U \to V_3 \to (V_2, X)$ form Markov chains).

Proof: The general idea is to split M_1 into four independent parts M_{10}, M_{11}, M_{12}, and M_{13}. The message pair (M_0, M_{10}) is represented by U. Using superposition and Marton coding, the message triple (M_0, M_{10}, M_{12}) is represented by V_2 and the message triple (M_0, M_{10}, M_{13}) is represented by V_3. Finally, using superposition coding, the message pair (M_0, M_1) is represented by X. Receiver Y_1 decodes U, V_2, V_3, X, and receivers Y_2 and Y_3 find M_0 via indirect decoding of V_2 and V_3, respectively, as in Theorem 1.

We now provide a more detailed outline of the proof.

Code Generation: Let $R_1 = S_0 + S_1 + S_2 + S_3$, where $S_i \ge 0$, $i = 0,1,2,3$, and $T_2 \ge S_2$ and $T_3 \ge S_3$. Fix a probability mass function of the required form

$$\begin{aligned} p(u,v_2,v_3,x) &= p(u)p(v_2|u)p(x,v_3|v_2) \\ &= p(u)p(v_3|u)p(x,v_2|v_3). \end{aligned}$$

Randomly and independently generate $2^{n(R_0+S_0)}$ sequences $u_1^n(m_0,s_0)$, $m_0 \in [1:2^{nR_0}], s_0 \in [1:2^{nS_0}]$, each distributed uniformly over the set of typical u_1^n sequences. For each $u^n(m_0,s_0)$, randomly and independently generate (a) 2^{nT_2} sequences $v_2^n(m_0,s_0,t_2)$, $t_2 \in [1:2^{nT_2}]$, each distributed uniformly over the set of conditionally typical v_2^n sequences, and (b) 2^{nT_3} sequences $v_3^n(m_0,s_0,t_3)$, $t_3 \in [1:2^{nT_3}]$, each distributed uniformly over the set of conditionally typical v_3^n sequences. Randomly partition the 2^{nT_2} sequences $v_2^n(m_0,s_0,t_2)$ into 2^{nS_2} equal size bins and the 2^{nT_3} $v_3^n(m_0,s_0,t_3)$ sequences into 2^{nS_3} equal size bins. To ensure that each product bin contains a jointly typical pair $(v_2^n(m_0,s_0,t_2), v_3^n(m_0,s_0,t_3))$ with arbitrarily high probability, we require that (see [16] for the proof)

$$S_2 + S_3 < T_2 + T_3 - I(V_2;V_3|U). \tag{19}$$

Finally, for each chosen jointly typical pair $(v_2^n(m_0,s_0,t_2), v_3^n(m_0,s_0,t_3))$ in each product bin (s_2,s_3), randomly and conditionally independently generate 2^{nS_1} sequences $x^n(m_0,s_0,s_2,s_3,s_1)$, $s_1 \in [1:2^{nS_1}]$, each distributed uniformly over the set of conditionally typical x^n sequences.

Encoding: To send the message pair (m_0,m_1), we express m_1 by the quadruple (s_0,s_1,s_2,s_3) and send the codeword $X^n(m_0,s_0,s_2,s_3,s_1)$.

Decoding:

1) Receiver Y_1 declares that (m_0,s_0,s_2,s_3,s_1) is sent if it is the unique rate tuple such that y_1^n is jointly typical with

$$\begin{aligned} ((u^n(m_0,s_0), v_2^n(m_0,s_0,t_2), \\ v_3^n(m_0,s_0,t_3), x^n(m_0,s_0,s_2,s_3,s_1)) \end{aligned}$$

and s_2 is the product bin number of $v_2^n(m_0,s_0,t_2)$ and s_3 is the product bin number of $v_2^n(m_0,s_0,t_3)$. Assuming $(m_0,s_0,s_1,s_2,s_3) = (1,1,1,1,1)$ is sent, we partition the error event into the following events.

a) Error event corresponding to $(m_0,s_0) \ne (1,1)$ occurs with arbitrarily small probability provided

$$R_0 + S_0 + S_1 + S_2 + S_3 < I(X;Y_1). \tag{20}$$

b) Error event corresponding to $m_0 = 1, s_0 = 1, s_2 \ne 1, s_3 \ne 1$ occurs with arbitrarily small probability provided

$$S_1 + S_2 + S_3 < I(X;Y_1|U). \tag{21}$$

c) Error event corresponding to $m_0 = 1, s_0 = 1, s_2 = 1, s_3 \ne 1$ occurs with arbitrarily small probability provided

$$S_1 + S_3 < I(X;Y_1|U,V_2) = I(X;Y_1|V_2). \tag{22}$$

The equality follows from the fact that $U \to V_2 \to (V_3, X)$ form a Markov chain.

d) Error event corresponding to $m_0 = 1, s_0 = 1, s_2 \ne 1, s_3 = 1$ occurs with arbitrarily small probability provided

$$S_1 + S_2 < I(X;Y_1|U,V_3) = I(X;Y_1|V_3). \tag{23}$$

The above equality uses the fact that $U \to V_3 \to (V_2, X)$ forms a Markov chain.

e) Error event corresponding to $m_0 = 1, s_0 = 1, s_2 = 1, s_3 = 1, s_1 \ne 1$ occurs with arbitrarily small probability provided

$$S_1 < I(X;Y_1|U,V_2,V_3) = I(X;Y_1|V_2,V_3). \tag{24}$$

Note that the equality here uses a weaker Markov structure $U \to (V_2,V_3) \to X$.

Thus, receiver Y_1 decodes (m_0,s_0,s_2,s_3,s_1) with arbitrarily small probability of error provided (20)–(24) hold.

2) Receiver Y_2 decodes (m_0,s_0) (and hence m_0) via indirect decoding using $v_2^n(m_0,s_0,t_2)$ (as in Theorem 1). This can

be achieved with arbitrarily small probability of error provided

$$R_0 + S_0 + T_2 < I(V_2; Y_2). \quad (25)$$

3) Receiver Y_3 decodes (m_0, s_0) (and hence m_0) via indirect decoding using $v_3^n(m_0, s_0, t_3)$ (as in Theorem 1). This can be achieved with arbitrarily small probability of error provided

$$R_0 + S_0 + T_3 < I(V_3; Y_3). \quad (26)$$

Combining (19)(20)–(24)(26), we obtain the following:

$$\begin{aligned} S_2 &\le T_2 \\ S_3 &\le T_3 \\ S_2 + S_3 &\le T_2 + T_3 - I(V_2; V_3|U) \\ R_0 + R_1 &\le I(X; Y_1) \\ S_1 + S_2 + S_3 &\le I(X; Y_1|U) \\ S_1 + S_3 &\le I(X; Y_1|V_2) \\ S_1 + S_2 &\le I(X; Y_1|V_3) \\ S_1 &\le I(X; Y_1|V_2, V_3) \\ R_0 + S_0 + T_2 &\le I(V_2; Y_2) \\ R_0 + S_0 + T_3 &\le I(V_3; Y_3) \end{aligned} \quad (27)$$

for some

$$\begin{aligned} p(u, v_2, v_3, x) &= p(u)p(v_2|v_1)p(x, v_3|v_2) \\ &= p(u)p(v_3|v_1)p(x, v_2|v_3). \end{aligned}$$

Using the Fourier–Motzkin procedure to eliminate T_2, T_3, S_1, S_2, and S_3, we obtain the inequalities in (18). □

Remark 5.1: The above achievability scheme can be adapted to any joint distribution $p(u, v_2, v_3, x)$. However, by letting $\tilde{V}_2 = (V_2, U)$ and letting $\tilde{V}_3 = (V_3, U)$, we observe that the region remains unchanged. Hence, without loss of generality, we assume the structure of the auxiliary random variables as described in the proposition. Further, using the construction of $\tilde{V}_2, \tilde{V}_3$, observe that one can restrict to triples (U, V_2, V_3), where $U = f(V_2) = g(V_3)$, and f and g are two deterministic mappings. Note that the auxiliary random variables in the outer bound described in the next subsection also possess the same structure.

Remark 5.2: A special choice of the auxiliary random variables is to set V_2 or V_3 equal to U (i.e., only one of the receivers tries to indirectly decode M_0), say let $V_2 = U$. This reduces the inequalities in Proposition 5 (after removing the redundant ones) to

$$\begin{aligned} R_0 &\le \min\{I(U; Y_2), I(V_3; Y_3)\} \\ R_0 + R_1 &\le \min\{I(X; Y_1), I(V_3; Y_3) + I(X; Y_1|V_3), \\ &\qquad I(U; Y_2) + I(X; Y_1|U)\} \end{aligned} \quad (28)$$

where $U \to V_3 \to X$ form a Markov chain.

This region includes the capacity region of the multilevel case in Theorem 1 and hence is tight in this setting.

Remark 5.3: Note that the rate splitting scheme we used in the proof of the proposition includes *rate transfer*, where part of the split message M_{10} is combined with M_0 and encoded using U. This rate transfer can be used also in the Körner–Marton two-receiver broadcast channel with degraded message sets. Recall that without rate splitting, we obtain the decoding constraints

$$\begin{aligned} R_0 &< I(U; Y_2) \\ R_0 + R_1 &< I(X; Y_1) \\ R_1 &< I(X; Y_1|U). \end{aligned} \quad (29)$$

Using rate splitting, we divide M_1 into two independent parts at rates R_{10} and R_{11}, and set $S_1 = R_0 + R_{10}$, $S_2 = R_{11}$. This yields the decoding constraints

$$\begin{aligned} R_0 + R_{10} &< I(U; Y_2) \\ R_0 + R_{10} + R_{11} &< I(X; Y_1) \\ R_{11} &< I(X; Y_1|U). \end{aligned}$$

Performing Fourier–Motzkin procedure, we obtain

$$\begin{aligned} R_0 &< I(U; Y_2) \\ R_0 + R_1 &< I(X; Y_1) \\ R_0 + R_1 &< I(U; Y_3) + I(X; Y_1|U). \end{aligned} \quad (30)$$

It is easy to see that the region given by the new rate splitting arguments is identical to the original region. However, the form of the new region is more conducive to the establishment of the weak converse. The same equivalence holds for the three-receiver broadcast channel with two-degraded message sets discussed in Section III.

Similar rate transfer arguments have been used before. For instance, Liang [19] used it for the two-receiver broadcast channels to obtain a region that is at least as large as the Marton region. The equivalence of the region obtained by Liang to the original Marton region was later established in [18].

We now establish the following outer bound.

Proposition 6: Any achievable rate pair (R_0, R_1) for the general three-receiver broadcast channel with two-degraded message sets must satisfy the conditions

$$\begin{aligned} R_0 &\le \min\{I(U; Y_1), I(V_2; Y_2) - I(V_2; Y_1|U), \\ &\qquad I(V_3; Y_3) - I(V_3; Y_1|U)\} \\ R_1 &\le I(X; Y_1|U) \end{aligned}$$

for some

$$\begin{aligned} p(u, v_2, v_3, x) &= p(u)p(v_2|u)p(x, v_3|u) \\ &= p(u)p(v_3|u)p(x, v_2|v_3) \end{aligned}$$

i.e., the same structure of the auxiliary random variables as in Proposition 5. Further, one can restrict the cardinalities of U, V_2, V_3 to $\|\mathcal{U}\| \le \|\mathcal{X}\| + 6$, $\|\mathcal{V}_2\| \le (\|\mathcal{X}\| + 1)(\|\mathcal{X}\| + 6)$, and $\|\mathcal{V}_3\| \le (\|\mathcal{X}\| + 1)(\|\mathcal{X}\| + 6)$.

Proof: The proof follows largely standard arguments. The auxiliary random variables are identified as $U_i = (M_0, Y_1^{i-1})$,

$V_{2i} = (U_i, Y_{2\,i+1}^{\,n})$, and $V_{3i} = (U_i, Y_{3\,i+1}^{\,n})$. With this identification inequalities, $R_0 \le I(U;Y_1)$ and $R_1 \le I(X;Y_1|U)$ are immediate. The other two inequalities also follow from standard arguments and are briefly outlined here

$$\begin{aligned} nR_0 &\le n\epsilon_n + \sum_i I(M_0; Y_{2i}|Y_{2\,i+1}^{\,n}) \\ &\le n\epsilon_n + \sum_i I(M_0, Y_{2\,i+1}^{\,n}, Y_1^{i-1}; Y_{2i}) \\ &\quad - I(Y_1^{i-1}; Y_{2i}|M_0, Y_{2\,i+1}^{\,n}) \\ &\overset{(a)}{=} n\epsilon_n + \sum_i I(M_0, Y_{2\,i+1}^{\,n}, Y_1^{i-1}; Y_{2i}) \\ &\quad - I(Y_{2\,i+1}^{\,n}; Y_{1i}|M_0, Y_1^{i-1}) \\ &= n\epsilon_n + \sum_i I(U_{2i}; Y_{2i}) - I(U_{2i}; Y_{1i}|U_i) \end{aligned}$$

where $\epsilon_n \to 0$ as n approaches infinity, and (a) follows by the Csiszár sum equality. The cardinality bounds are established using a similar argument as in Section III-D. To create a set of new auxiliary random variables with the bounds of Proposition 6, we first replace V_2 by (V_2, U) and V_3 by (V_3, U). It is easy to see from the Markov chain relationships $U \to V_2 \to (V_3, X)$ and $U \to V_3 \to (V_2, X)$ that the following region is same as the that of Proposition 6:

$$\begin{aligned} R_0 \le \min\{&I(U;Y_1), I(U,V_2;Y_2) + I(X;Y_1|U,V_2) \\ &- I(X:Y_1|U), \\ &I(U,V_3;Y_3) + I(X;Y_1|U,V_3) - I(X:Y_1|U)\} \\ R_1 \le I&(X;Y_1|U). \end{aligned} \tag{31}$$

Then, using standard arguments, one can replace U by U^* satisfying $\|\mathcal{U}^*\| \le \|\mathcal{X}\| + 6$, such that the distribution of X and $H(Y_1|U)$, $H(Y_1|U,V_2)$, $H(Y_1|U,V_3)$, $H(Y_3|U)$, $H(Y_3|U,V_2)$, $H(Y_2|U)$, and $H(Y_2|U,V_3)$ is preserved. Now for each $U^* = u$, one can find $V_2^*(u)$ with cardinality less than $\|\mathcal{X}\| + 1$ each, such that the distribution of X conditioned on $U^* = u$, $H(Y_1|U^* = u, V_2)$, and $H(Y_2|U^* = u, V_2)$ is preserved. Similarly, one can find for each $U^* = u$ a random variable $V_3^*(u)$ with cardinality less than $\|\mathcal{X}\| + 1$ each, such that the distribution of X conditioned on $U^* = u$, $H(Y_1|U^* = u, V_3)$, and $H(Y_3|U_1^* = u, V_3)$ is preserved. This yields random variables U^*, V_2^*, V_3^* that preserve the region in (31). (Note that as the distribution of X conditioned on $U = u$ is preserved by both $V_2^*(u)$ and $V_3^*(u)$, it is possible to get a consistent triple of random variables U^*, V_2^*, V_3^*.) Finally, setting $\tilde{U} = U^*$, $\tilde{V}_2 = (U^*, V_2^*(U^*))$, and $\tilde{V}_3 = (U^*, V_3^*(U^*))$ gives the desired bounds on cardinality as well as the desired Markov relations. □

Remark 5.4: The above outer bound appears to be very different from the inner bound of Proposition 5. However, by taking appropriate sums of the inequalities defining the region of Proposition 6, we arrive at the conditions

$$\begin{aligned} R_0 &\le \min\{I(V_2;Y_2) - I(V_2;Y_1|U), \\ &\qquad I(V_3;Y_3) - I(V_3;Y_1|U)\} \\ R_0 + R_1 &\le \min\{I(X;Y_1), I(V_2;Y_2) + I(X;Y_1|V_2), \\ &\qquad I(V_3;Y_3) + I(X;Y_1|V_3)\} \\ 2R_0 + R_1 &\le I(V_2;Y_2) + I(V_3;Y_3) + I(X;Y_1|V_2,V_3) \\ &\quad - I(V_2;V_3|U_1) + I(V_2;V_3|Y_1,U) \\ 2R_0 + 2R_1 &\le I(V_2;Y_2) + I(V_3;Y_3) + I(X;Y_1|U) \\ &\quad + I(X;Y_1|V_2,V_3) - I(V_2;V_3|U_1) \\ &\quad + I(V_2;V_3|Y_1,U). \end{aligned}$$

These conditions, which include some redundancy, are closer in structure to the inequalities defining the inner bound of Proposition 5.

Remark 5.5: The outer bound in Proposition 6 reduces to the capacity region for the multilevel case in Theorem 1. To see this observe that when $X \to Y_1 \to Y_2$ form a Markov chain, then

$$\begin{aligned} R_0 &\le I(V_2;Y_2) - I(V_2;Y_1|U) \\ &\le I(V_2;Y_2) - I(V_2;Y_2|U) = I(U;Y_2). \end{aligned} \tag{32}$$

Thus, any rate pair (R_0, R_1) satisfying the constraints of Proposition 6 must satisfy

$$R_0 \le \min\{I(U;Y_2), I(V_3;Y_3)\} \tag{33}$$
$$R_1 \le I(X;Y_1|U)$$
$$R_0 + R_1 \le I(V_3;Y_3) + I(X;Y_1|V_3). \tag{34}$$

However, any rate pair satisfying these constraints is achievable as shown in Theorem 1, and hence, the outer bound of Proposition 6 is tight for this setting.

The inner and outer bounds match if Y_1 is less noisy than Y_2[13], that is, if $I(U;Y_2) \le I(U;Y_1)$ for all $p(u)p(x|u)$. As shown in [13], this condition is more general than degradedness. As such, it defines a larger class than multilevel broadcast channels.

Proposition 7: The capacity region for the three-receiver broadcast channel with two-degraded message sets when Y_1 is a *less noisy* receiver than Y_2 is given by the set of rate pairs (R_0, R_1) such that

$$R_0 \le \min\{I(U;Y_2), I(V;Y_3)\} \tag{35}$$
$$R_1 \le I(X;Y_1|U)$$
$$R_0 + R_1 \le I(V;Y_3) + I(X;Y_1|V) \tag{36}$$

for some $p(u)p(v|u)p(x|v)$.

From the definition of less noisy receivers [13], we have $I(V;Y_2|U=u) \le I(V;Y_1|U=u)$ for every choice of u, and thus, $I(V;Y_2|U) \le I(V;Y_1|U)$ for every $p(u)p(v|u)p(x|v)$. Using (32), it follows that the general outer bound is contained in (33). Any rate pair satisfying (35) also satisfies (under the less noisy assumption) the constraints in (28) and thus is achievable by setting $V_2 = U$ in the region of Proposition 5.

B. Inner Bound for Three-Degraded Message Sets

We establish an inner bound to the capacity region of the broadcast channel with three-degraded message sets where M_0

is to be sent to all three receivers, M_1 is to be sent only to Y_1 and Y_2, and M_2 is to be sent only to Y_1. We then specialize the result to the case of two-degraded message sets scenario, where M_0 is to be sent to all receivers and M_1 is to be sent to Y_1 and Y_2 and establish optimality for two classes of channels.

The achievability proof of the region for the above scenario is closely related to that of Proposition 5. To explain the connection, consider the more general three-receiver broadcast channel scenario, where message M_0 is to be decoded by all receivers, message M_{12} is to be decoded by receivers Y_1, Y_2, message M_{13} is to be decoded by receivers Y_1, Y_3, and message M_{11} is to be decoded by receiver Y_1. Observe that letting $R_{12} = R_{13} = 0$ yields the two-degraded message set scenario considered in Proposition 5, and letting $R_{13} = 0$ yields the three-degraded message set requirement under consideration. Thus, the region in Proposition 5 and the region for the three-degraded message sets given in Theorem 2 below can be thought of as lower dimensional projections of the region for the more general broadcast channel scenario with message sets in the union of these two message sets. With this motivation, we identify each message set in the superset by the subset of receivers that are required to decode it, and associate with each receiver subset an auxiliary random variable as follows:

$$U : \{Y_1, Y_2, Y_3\}, \quad V_2 : \{Y_1, Y_2\}, \quad V_3 : \{Y_1, Y_3\}, \quad W : Y_1.$$

Since receiver Y_1 is required to decode all messages, one can show that setting $W = X$ is optimal. We also use the *rate transfer* technique alluded to in Remark 5.3 to establish the achievable region.

Let $R_1 = R_{10} + R_{11}$ and $R_2 = S_0 + S_1 + S_2 + S_3$ be the rate splitting as proposed in Proposition 5.

Code generation proceeds similarly to Proposition 5, i.e., we first generate $2^{n(R_0+R_{10}+S_0)}$ u^n sequences. For each u^n sequence, we generate 2^{nT_2} v_2^n sequences and 2^{nT_3} v_3^n sequences and then partition them into $2^{n(R_{11}+S_2)}$ and 2^{nS_3} bins, respectively. We then find a jointly typical (v_2^n, v_3^n) pair in each product bin, and generate 2^{nS_1} x^n sequences for each such pair.

Decoding proceeds in a similar way. Y_1 decodes M_0, M_1, M_2 by decoding X, Y_2 decodes M_0, M_1 by decoding V_2, and Y_3 decodes M_0 by indirectly decoding U from V_3. To ensure that the encoding and the decoding are successful with high probability, we impose the following constraints on the rates:

$$\begin{aligned}
R_{11} + S_2 &\le T_2 \\
S_3 &\le T_3 \\
R_{11} + S_2 + S_3 &\le T_2 + T_3 - I(V_2; V_3|U) \\
R_0 + R_1 + R_2 &\le I(X; Y_1) \\
R_{11} + S_1 + S_2 + S_3 &\le I(X; Y_1|U) \\
S_1 + S_3 &\le I(X; Y_1|U, V_2) = I(X; Y_1|V_2) \\
S_1 + S_2 + R_{11} &\le I(X; Y_1|U, V_3) = I(X; Y_1|V_3) \\
S_1 &\le I(X; Y_1|U, V_2, V_3) = I(X; Y_1|V_2, V_3) \\
R_0 + S_0 + R_{10} + T_2 &\le I(U, V_2; Y_2) = I(V_2; Y_2) \\
T_2 &\le I(V_2; Y_2|U) \\
R_0 + S_0 + R_{10} + T_3 &\le I(U, V_3; Y_3) = I(V_3; Y_3)
\end{aligned} \tag{37}$$

for some

$$\begin{aligned}
p(u, v_2, v_3, x) &= p(u)p(v_2|v_1)p(x, v_3|v_2) \\
&= p(u)p(v_3|v_1)p(x, v_2|v_3).
\end{aligned}$$

Eliminating S_0, S_1, S_2, S_3, R_{10}, R_{11}, T_2, and T_3 via the Fourier–Motzkin procedure with the rate splitting constraints $R_2 = S_0 + S_1 + S_2 + S_3$ and $R_1 = R_{10} + R_{11}$, we obtain the following achievable rate region.

Theorem 2: A rate triple (R_0, R_1, R_2) is achievable in a general three-receiver broadcast channel with three-degraded message sets if it satisfies the conditions

$$\begin{aligned}
R_0 &\le I(V_3; Y_3) \\
R_0 + R_1 &\le \min\{I(V_2; Y_2), I(V_2; Y_2|U) \\
&\qquad + I(V_3; Y_3) - I(V_2; V_3|U)\} \\
2R_0 + R_1 &\le I(V_2; Y_2) + I(V_3; Y_3) - I(V_2; V_3|U) \\
R_0 + R_1 + R_2 &\le \min\{I(X; Y_1), I(V_2; Y_2) \\
&\qquad + I(X; Y_1|V_2), I(V_3; Y_3) \\
&\qquad + I(X; Y_1|V_3), I(V_2; Y_2|U) \\
&\qquad + I(V_3; Y_3) + I(X; Y_1|V_2, V_3) \\
&\qquad - I(V_2; V_3|U)\} \\
2R_0 + R_1 + R_2 &\le I(V_2; Y_2) + I(V_3; Y_3) \\
&\qquad + I(X; Y_1|V_2, V_3) - I(V_2; V_3|U) \\
2R_0 + 2R_1 + R_2 &\le I(V_2; Y_2) + I(V_3; Y_3) \\
&\qquad + I(X; Y_1|V_3) - I(V_2; V_3|U) \\
2R_0 + 2R_1 + 2R_2 &\le \min\{I(V_2; Y_2) + I(V_3; Y_3) \\
&\qquad + I(X; Y_1|V_2) + I(X; Y_1|V_3) \\
&\qquad - I(V_2; V_3|U), I(V_2; Y_2|U) \\
&\qquad + I(V_3; Y_3) + I(X; Y_1|U) \\
&\qquad + I(X; Y_1|V_2, V_3) - I(V_2; V_3|U)\}
\end{aligned} \tag{38}$$

for some

$$\begin{aligned}
p(u_1, u_2, u_3, x) &= p(u_1)p(u_2|u_1)p(x, u_3|u_2) \\
&= p(u_1)p(u_3|u_1)p(x, u_2|u_3)
\end{aligned}$$

i.e., as before both $U_1 \to U_2 \to (U_3, X)$ and $U_1 \to U_3 \to (U_2, X)$ form Markov chains.

Proposition 8: The region of Theorem 2 reduces to the inner bound of Proposition 5 by setting $R_1 = 0$.

Proof: To show this, denote by $\mathcal{R}_A$ the rate region prescribed by the constraints in (27), and $\mathcal{R}_B$ the rate region prescribed by the constraints in (37). Note that in (27) the rate R_2, which corresponds to the rate of the private message to receiver Y_1, is denoted as R_1, i.e., we need to compare the rate pairs (R_0, R_2) from (37) to the rate pairs (R_0, R_1) from (27). We compare the set of constraints in (27) and in (37) when $R_1 = 0$, i.e., $R_{10} = R_{11} = 0$. Observe that (37) has exactly one extra constraint, $T_2 < I(V_2; Y_2|U)$, when compared to the constraints in (27). Therefore, $\mathcal{R}_B \subseteq \mathcal{R}_A$. Hence, it suffices to show that $\mathcal{R}_A \subseteq \mathcal{R}_B$.

Consider any rate pair $(R_0, S_0, S_1, S_2, S_3)$ and random variables U, V_2, V_3 satisfying the constraints in (27). We consider two cases.

Case 1) $R_0 + S_0 > I(U;Y_2)$. Since $R_0 + S_0 + T_2 < I(V_2;Y_2)$, this implies that the rates and the corresponding auxiliary random variables also satisfy $T_2 \le I(V_2;Y_2|U)$, and hence belong to $\mathcal{R}_B$.

Case 2) $R_0 + S_0 \le I(U;Y_2)$. Consider the following identification: $\tilde{R}_0 = R_0$, $\tilde{S}_0 = S_0$, $\tilde{S}_1 = S_1 + S_2$, $\tilde{S}_2 = 0$, $\tilde{S}_3 = S_3$, $\tilde{U} = U$, $\tilde{T}_2 = 0$, $\tilde{T}_3, S_3$, $\tilde{V}_2 = U$, $\tilde{V}_3 = V_3$. It is easy to see that the rate pairs (R_0, R_2) satisfy all the required constraints in (37) and hence belong to $\mathcal{R}_B$. Thus, $\mathcal{R}_A \subseteq \mathcal{R}_B$ as desired. □

Remark 5.6: Indeed a natural extension of this argument implies that the region in Proposition 5 does not change under the addition of the constraints $T_2 < I(V_2;Y_2|U)$, and $T_3 < I(V_3;Y_3|U)$. Therefore, a joint decoding strategy would have resulted in the same region as the indirect decoding strategy. However, as mentioned in part 3 of Remark 3.3, it is not clear to the authors whether this is always the case.

We now consider a two-degraded message set scenario where M_0 is to be sent to all receivers and M_1 is to be sent to receivers Y_1 and Y_2. The following inner bound follows from Theorem 2 by setting $R_2 = 0$.

Corollary 1: A rate pair (R_0, R_1) is achievable in a three-receiver broadcast channel with two-degraded message sets, where M_0 is to be decoded by all three receivers and M_1 is to be decoded only by Y_1 and Y_2 if it satisfies the following conditions:

$$\begin{aligned} R_0 &\le I(U;Y_3) \\ R_0 + R_1 &\le \min\{I(U;Y_3) + I(X;Y_1|U), \\ &\qquad I(U;Y_3) + I(X;Y_2|U)\} \\ R_0 + R_1 &\le \min\{I(X;Y_1), I(X;Y_2)\} \end{aligned} \tag{39}$$

for some $p(u)p(x|u)$.

This region is the straightforward extension of the Körner–Marton scheme to the current scenario.

Proposition 9: The region described by Corollary 1 coincides with the region described by Theorem 2 when $R_2 = 0$.

Proof: By setting $R_2 = 0$, $V_2 = X$, and $V_3 = U$, the region in Theorem 2 reduces to (39). Thus, region in (39) is contained in region (38). There it suffices to show that the projection of the region (38) to the plane $R_2 = 0$ is contained in region (39). To prove this, observe that

$$\begin{aligned} R_0 + R_1 &\le I(V_2;Y_2|U) + I(V_3;Y_3) - I(V_2;V_3|U) \\ &= I(V_3;Y_3) + I(V_3;Y_2|U) + I(V_2;Y_2|V_3) \\ &\quad - I(V_3;Y_2|V_2) - I(V_3;V_2|U) \\ &= I(V_3;Y_3) + I(V_2;Y_2|V_3) - I(V_3;V_2|Y_2,U) \\ &\le I(V_3;Y_3) + I(X;Y_2|V_3). \end{aligned}$$

Thus, the rate pairs must satisfy the following inequalities:

$$\begin{aligned} R_0 &\le I(V_3;Y_3) \\ R_0 + R_1 &\le \min\{I(V_3;Y_3) + I(X;Y_2|V_3), \\ &\qquad I(V_3;Y_3) + I(X;Y_1|V_3)\} \\ R_0 + R_1 &\le \min\{I(X;Y_2), I(X;Y_1)\}. \end{aligned} \tag{40}$$

Clearly, this is contained inside region (39) and hence region (38) reduces to the one in Corollary 1 when $R_2 = 0$. □

Inner bound (Corollary 1) is optimal for the following two special classes of broadcast channels.

Proposition 10: Achievable region (39) is tight for deterministic three-receiver broadcast channels. Indeed it is tight as long as the channel $X \to Y_3$ is deterministic.

Proof: By setting $U = Y_3$ in (39), we see that rate pairs (R_0, R_1) is achievable if

$$\begin{aligned} R_0 &\le H(Y_3) \\ R_0 + R_1 &\le \min\{H(Y_1), H(Y_2)\} \end{aligned}$$

for some $p(x)$. Clearly, these constraints also constitute an outer bound and hence they provide a tight characterization of the capacity region. □

Proposition 11: Achievable region (39) is optimal when Y_1 is a less noisy receiver than Y_2 and Y_3 is a less noisy receiver than Y_2.

Proof: To show optimality, we set $U_i = (M_0, Y_3^{i-1})$, and thus, the only nontrivial inequality in the converse is $R_0 + R_1 \le I(U;Y_3) + \min\{I(X;Y_1|U), I(X;Y_3|U)\}$. To prove this, observe that

$$\begin{aligned} nR_1 &\le \sum_i I(M_1;Y_{1i}|M_0, Y_{1\,i+1}^n) \\ &\le \sum_i I(M_1;Y_{1i}|M_0, Y_{1\,i+1}^n, Y_3^{i-1}) \\ &\quad + \sum_i I(Y_3^{i-1};Y_{1i}|M_0, Y_{1\,i+1}^n)) \\ &\overset{(a)}{=} \sum_i I(M_1, Y_{1\,i+1}^n;Y_{1i}|M_0, Y_3^{i-1}) \\ &\quad - \sum_i I(Y_{1\,i+1}^n;Y_{1i}|M_0, Y_3^{i-1}) \\ &\quad + \sum_i I(Y_{1\,i+1}^n;Y_{3i}|M_0, Y_3^{i-1}) \\ &\overset{(b)}{\le} \sum_i I(X_i;Y_{1i}|M_0, Y_3^{i-1}) \end{aligned}$$

where (a) follows by the Csiszár sum equality and (b) uses the assumption that Y_1 is a less noisy than Y_3, which implies that $I(Y_{1\,i+1}^n;Y_{3i}|M_0, Y_3^{i-1}) \le I(Y_{1\,i+1}^n;Y_{1i}|M_0, Y_3^{i-1})$. The bound $R_1 \le I(X;Y_2|U)$ can be proved similarly. □

Remark 5.7: Note that this result generalizes [4, Th. 3.2], where the authors assume that the receivers Y_2 and Y_1 are degraded versions of Y_3.

VI. Conclusion

Recent results and conjectures on the capacity region of ($k > 2$)-receiver broadcast channels with degraded message sets [4]–[6] have lent support to the general belief that the straightforward extension of the Körner–Marton region for the two-receiver case is optimal. This paper shows that this is not the case. We showed that the capacity region of the three-receiver broadcast channels with two-degraded message sets can be strictly larger than the straightforward extension of the Körner–Marton region. Achievability is proved using rate splitting and superposition coding. We showed that a simpler characterization of the capacity region results using indirect decoding instead of joint decoding. Using these ideas, we devised a new inner bound to the capacity of the general three-receiver broadcast channel with three-degraded message sets and showed that it is tight in some cases.

The results in this paper suggest that the capacity of the $k >$ 2-receiver broadcast channels with degraded message sets is at least as hard to characterize in a single-letter way as the capacity region of the general two-receiver broadcast channel with one common and one private message sets. However, it would be interesting to explore the optimality of our new inner bounds for classes where capacity is known for the general two-receiver case, such as deterministic and vector Gaussian broadcast channels. It would also be interesting to investigate applications of indirect decoding to other problems, for example, the three-receiver broadcast channels with confidential message sets [11].

Our results also show that a straighforward extension of Marton's achievable rate region to more than two receivers is not in general optimal. The structure of the auxiliary random variables in the inner bounds can be naturally extended to three or more receivers with arbitrary message set requirements as will be detailed in a future publication.

Appendix
Proof of Propositions 1–4

To prove Propositions 1 and 2, note that it is straightforward to show that each simplified characterization is contained in the original region as the characterizations are obtained by using the channels independently. So we only prove the other nontrivial direction.

Proof of Proposition 1: We prove that for the product broadcast channel given by (11), the BZT region (3) reduces to the expression (12).

Consider the first term in the BZT region

$$\begin{aligned} R_0 \le I(U;Y_2) &= I(U;Y_{21},Y_{22}) \\ &= I(U;Y_{21}) + I(U;Y_{22}|Y_{21}) \\ &\le I(U;Y_{21}) + I(U,Y_{21};Y_{22}) \\ &\le I(U;Y_{21}) + I(U,Y_{11};Y_{22}). \end{aligned}$$

Now set $U_1 = U$ and $U_2 = (U, Y_{11})$. Thus, the above inequality becomes

$$R_0 \le I(U_1;Y_{21}) + I(U_2;Y_{22}).$$

This inequality is the first term (12a) in (12). To complete the equivalence, we have to show that the remaining constraints of (12) are also satisfied by our choice $U_1 = U$ and $U_2 = (U, Y_{11})$.

Observe that

$$R_0 \le I(U;Y_3) = I(U_1;Y_{31}).$$

Finally, consider the last term

$$\begin{aligned} R_1 &\le I(X;Y|U) = I(X_1,X_2;Y_{11},Y_{12}|U) \\ &= H(Y_{11},Y_{12}|U) - H(Y_{11},Y_{12}|X_1,X_2,U) \\ &= H(Y_{11}|U) + H(Y_{12}|U,Y_{11}) \\ &\quad - H(Y_{11}|X_1,U) - H(Y_{12}|X_2,U) \\ &= I(X_1;Y_{11}|U) + H(Y_{12}|U,Y_{11}) - H(Y_{12}|X_2,U,Y_{11}) \\ &= I(X_1;Y_{11}|U_1) + I(X_2;Y_{12}|U_2). \end{aligned}$$

This implies that all constraints of (12) are satisfied by the choice $U_1 = U$ and $U_2 = (U, Y_{11})$. The fact that $p(u_1)p(u_2)p(x_1|u_1)p(x_2|u_2)$ suffices follows from the structure of the mutual information terms.

Proof of Proposition 2: We prove that for the product broadcast channel (11) the capacity region given by Theorem 1 reduces to the expression (13).

Consider the first term (13a) in the capacity region

$$\begin{aligned} R_0 \le I(U;Y_2) &= I(U;Y_{21},Y_{22}) \\ &= I(U;Y_{21}) + I(U;Y_{22}|Y_{21}) \\ &\le I(U;Y_{21}) + I(U,Y_{21};Y_{22}) \\ &\le I(U;Y_{21}) + I(U,Y_{11};Y_{22}). \end{aligned}$$

Now set $U_1 = U$ and $U_2 = (U, Y_{11})$.

The second term (13b) in the capacity region is $R_0 \le I(V;Y_{31})$. Now set $V_1 = V$ and from $U \to V \to (X_1, X_2)$, we have $U_1 \to V_1 \to X_1$. Thus, the second term can be rewritten as $R_0 \le I(V_1;Y_{31})$

Consider the third term in the capacity region

$$\begin{aligned} R_1 &\le I(X_1,X_2;Y_{11},Y_{12}|U) \\ &= I(X_1;Y_{11}|U) + I(X_2;Y_{12}|U,Y_{11}) \\ &= I(X_1;Y_{11}|U_1) + I(X_2;Y_{12}|U_2). \end{aligned}$$

Finally, consider the last term in the capacity region

$$\begin{aligned} R_0 + R_1 &\le I(V;Y_{31}) + I(X_1,X_2;Y_{11},Y_{12}|V) \\ &= I(V;Y_{31}) + I(X_1;Y_{11}|V) + I(X_2;Y_{12}|V,Y_{11}) \\ &\le I(V;Y_{31}) + I(X_1;Y_{11}|V) + I(X_2;Y_{12}|U,Y_{11}) \\ &= I(V_1;Y_{31}) + I(X_1;Y_{11}|V_1) + I(X_2;Y_{12}|U_2). \end{aligned}$$

The fact that $p(u_1)p(v_1)p(x_1|v_1)p(u_2)p(x_2|u_2)$ suffices follows from the structure of the mutual information terms.

In the proof of Propositions 3 and 4, we will make use of the following simple fact about the entropy function [10]:

$$H(ap, 1-p, (1-a)p) = H(p, 1-p) + pH(a, 1-a).$$

Proof of Proposition 3: We prove that the region given by (12) reduces to (14) for the binary erasure channel described by the example in Section IV-A.

Let $\mathrm{P}\{U_1 = i\} = \alpha_i,\ \mathrm{P}\{X_1 = 0 | U_1 = i\} = \mu_i$. Then

$$\begin{aligned} I(U_1;Y_{21}) &= H\left(\sum_i \frac{\alpha_i \mu_i}{6}, \frac{5}{6}, \sum_i \frac{\alpha_i(1-\mu_i)}{6}\right) \\ &\quad - \sum_i \alpha_i H\left(\frac{\mu_i}{6}, \frac{5}{6}, \frac{1-\mu_i}{6}\right) \\ &= \frac{1}{6} H\left(\sum_i \alpha_i\mu_i, \sum_i \alpha_i(1-\mu_i)\right) \\ &\quad - \frac{1}{6}\sum_i \alpha_i H(\mu_i, 1-\mu_i) \\ I(U_1;Y_{31}) &= H\left(\sum_i \alpha_i\mu_i, \sum_i \alpha_i(1-\mu_i)\right) \\ &\quad - \sum_i \alpha_i H(\mu_i, 1-\mu_i) \\ I(X_1;Y_{11}|U_1) &= \sum_i \alpha_i H\left(\frac{\mu_i}{2}, \frac{1}{2}, \frac{1-\mu_i}{2}\right) \\ &\quad - \sum_i \alpha_i \mu_i H\left(\frac{1}{2}, \frac{1}{2}\right) \\ &\quad - \sum_i \alpha_i(1-\mu_i) H\left(\frac{1}{2}, \frac{1}{2}\right) \\ &= \frac{1}{2}\sum_i \alpha_i H(\mu_i, 1-\mu_i). \end{aligned}$$

Similarly, let $\mathrm{P}\{U_2 = i\} = \beta_i,\ \mathrm{P}\{X_2 = 0 | U_2 = i\} = \nu_i$. Then

$$\begin{aligned} I(U_2;Y_{22}) &= \frac{1}{2} H\left(\sum_i \beta_i\nu_i, \sum_i \beta_i(1-\nu_i)\right) \\ &\quad - \frac{1}{2}\sum_i \beta_i H(\nu_i, 1-\nu_i) \\ I(X_2;Y_{12}|U_2) &= \sum_i \beta_i H(\nu_i, 1-\nu_i). \end{aligned}$$

Now setting $\sum_i \beta_i H(\nu_i, 1-\nu_i) = 1-q$, and $\sum_i \alpha_i H(\mu_i, 1-\mu_i) = 1-p$, we obtain

$$\begin{aligned} I(U_1;Y_{21}) &= \frac{1}{6} H\left(\sum_i \alpha_i\mu_i, \sum_i \alpha_i(1-\mu_i)\right) \\ &\quad - \frac{1}{6}\sum_i \alpha_i H(\mu_i, 1-\mu_i) \\ &\le \frac{1}{6}(1-(1-p)) = \frac{p}{6} \\ I(U_1;Y_{31}) &= H\left(\sum_i \alpha_i\mu_i, \sum_i \alpha_i(1-\mu_i)\right) \\ &\quad - \sum_i \alpha_i H(\mu_i, 1-\mu_i) \\ &\le 1-(1-p) = p \end{aligned}$$

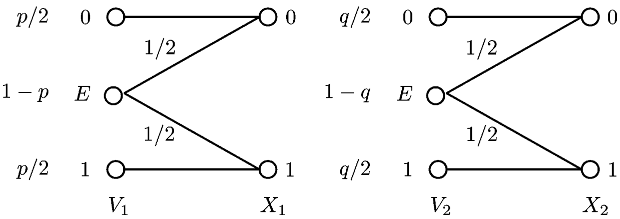

Fig. 4. Auxiliary channels that achieve the boundary of the BZT region.

$$\begin{aligned} I(X_1;Y_{11}|U_1) &= \frac{1-p}{2} \\ I(U_2;Y_{21}) &= \frac{1}{6} H\left(\sum_i \alpha_i\mu_i, \sum_i \alpha_i(1-\mu_i)\right) \\ &\quad - \frac{1}{6}\sum_i \alpha_i H(\mu_i, 1-\mu_i) \\ &\le \frac{1}{2}(1-(1-q)) = \frac{q}{2} \\ I(X_2;Y_{12}|U_2) &= 1-q. \end{aligned}$$

Therefore, any rate pair in the BZT region must satisfy the conditions

$$\begin{aligned} R_0 &\le \min\left\{\frac{p}{6} + \frac{q}{2}, p\right\} \\ R_1 &\le \frac{1-p}{2} + 1 - q \end{aligned}$$

for some $0 \le p, q \le 1$.

It is easy to see that equality is achieved when the marginals of V_1 are given by $\mathrm{P}\{U_1 = 0\} = \mathrm{P}\{U_1 = 1\} = p/2$, $\mathrm{P}\{U_1 = E\} = 1-p$ and the marginals of V_2 are given by $\mathrm{P}\{U_2 = 0\} = \mathrm{P}\{U_2 = 1\} = q/2$, $\mathrm{P}\{U_2 = E\} = 1-q$ (see Fig. 4).

Proof of Proposition 4: We prove that the region (13) reduces to region (15) for the binary erasure channel described by the example in Section IV-A.

Assume that $\mathrm{P}\{U_1 = i\} = \alpha_i$, $\mathrm{P}\{X_1 = 0 | U_1 = i\} = \mu_i$, $\mathrm{P}\{U_2 = i\} = \beta_i$, $\mathrm{P}\{X_2 = 0 | U_2 = i\} = \nu_i$, $\mathrm{P}\{V_1 = i\} = \gamma_i$, and $\mathrm{P}\{X_1 = 0 | V_1 = i\} = \omega_i$. Further, there exist $r, s, t \in [0,1]$ such that

$$\begin{aligned} H(X_1|U_1) &= \sum_i \alpha_i H(\mu_i, 1-\mu_i) = 1-r \\ H(X_2|U_2) &= \sum_i \beta_i H(\nu_i, 1-\nu_i) = 1-s \\ H(X_1|V_1) &= \sum_i \gamma_i H(\omega_i, 1-\omega_i) = 1-t. \end{aligned}$$

Clearly, from the Markov condition $U_1 \to V_1 \to X_1$, we require $1-t \le 1-r$, or equivalently, $r \le t$.

We can also establish the following in a similar fashion:

$$\begin{aligned} I(U_1;Y_{21}) &= \frac{1}{6} H\left(\sum_i \alpha_i\mu_i, \sum_i \alpha_i(1-\mu_i)\right) \\ &\quad - \frac{1}{6}\sum_i \alpha_i H(\mu_i, 1-\mu_i) \le \frac{r}{6} \end{aligned}$$

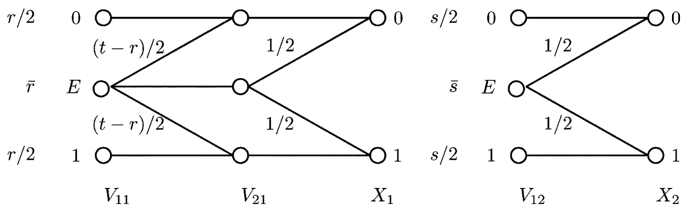

Fig. 5. Auxiliary channels that achieve the boundary of the capacity region.

$$
\begin{aligned}
I(U_2;Y_{22}) &= \frac{1}{2}H\left(\sum_i \beta_i\nu_i, \sum_i \beta_i(1-\nu_i)\right) \\
&\quad - \frac{1}{2}\sum_i \beta_i H(\nu_i, 1-\nu_i) \le \frac{s}{2} \\
I(V_1;Y_{31}) &= H\left(\sum_i \gamma_i\omega_i, \sum_i \gamma_i(1-\omega_i)\right) \\
&\quad - \sum_i \gamma_i H(\omega_i, 1-\omega_i) \le t \\
I(X_1;Y_{11}|U_1) &= \frac{1}{2}\sum_i \alpha_i H(\mu_i, 1-\mu_i) = \frac{1-r}{2} \\
I(X_2;Y_{12}|U_2) &= \sum_i \beta_i H(\nu_i, 1-\nu_i) = 1-s \\
I(X_1;Y_{11}|V_1) &= \frac{1}{2}\sum_i \gamma_i H(\omega_i, 1-\omega_i) = \frac{1-t}{2}.
\end{aligned}
$$

Thus, any rate pair in the capacity region must satisfy

$$
\begin{aligned}
R_0 &\le \min\left\{\frac{r}{6} + \frac{s}{2}, t\right\} \\
R_1 &\le \frac{1-r}{2} + 1 - s \\
R_0 + R_1 &\le t + \frac{1-t}{2} + 1 - s
\end{aligned}
$$

for some $0 \le r \le t \le 1, 0 \le s \le 1$. Note that substituting $r = t$ yields the BZT region.

Equality in the above conditions is achieved by the choices of auxiliary random variables shown in Fig. 5, and thus the above region is the capacity region.

Acknowledgment

The authors would like to thank Y.-H. Kim, G. Kramer, and the anonymous reviewers for valuable suggestions that have greatly improved the presentation of their results.

References

[1] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 1, pp. 60–64, Jan. 1977.

[2] T. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 2–14, Jan. 1972.

[3] J. Körner and K. Marton, "Images of a set via two channels and their role in multi-user communication," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 6, pp. 751–761, Nov. 1977.

[4] S. Diggavi and D. Tse, "On opportunistic codes and broadcast codes with degraded message sets," in *Proc. Inf. Theory Workshop (ITW)*, 2006, pp. 227–231.

[5] V. Prabhakaran, S. Diggavi, and D. Tse, "Broadcasting with degraded message sets: A deterministic approach," in *Proc. 45th Annu. Allerton Conf. Commun. Control Comput.*, 2007.

[6] S. Borade, L. Zheng, and M. Trott, "Multilevel broadcast networks," in *Proc. Int. Symp. Inf. Theory*, 2007, pp. 1151–1155.

[7] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Probl. Peredac. Inf.*, vol. 10, no. 3, pp. 3–14, 1974.

[8] A. El Gamal, "The capacity of a class of broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 2, pp. 166–169, Mar. 1979.

[9] C. Nair and A. El Gamal, "An outer bound to the capacity region of the broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-53, no. 1, pp. 350–355, Jan. 2007.

[10] T. Cover and J. Thomas*, Elements of Information Theory*. New York: Wiley-Interscience, 1991.

[11] I. Csizár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 339–348, May 1978.

[12] R. F. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 6, pp. 629–637, Nov. 1975.

[13] J. Körner and K. Marton, "A source network problem involving the comparison of two channels ii," in *Top. Inf. Theory*, I. Csiszar and P. Elias, Eds., Keszthely, Hungary, Aug. 1975, pp. 411–423.

[14] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 3, pp. 306–311, May 1979.

[15] H.-F. Chong, M. Motani, H. K. Garg, and H. El Gamal, "On the Han-Kobayashi region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3188–3195, Jul. 2008.

[16] A. El Gamal and E. C. van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 120–121, Jan. 1981.

[17] A. Schrijver*, Theory of Integer and Linear Programming*. New York: Wiley, 1986.

[18] Y. Liang, G. Kramer, and H. V. Poor, "Equivalence of two inner bounds on the capacity region of the broadcast channel," in *Proc. 46th Annu. Allerton Conf. Commun. Control Comput.*, Monticello, IL, Sep. 23–26, 2008, pp. 1417–1421.

[19] Y. Liang, "Multiuser communications with relaying and user cooperation," Ph.D. dissertation, Dept. Electr. Comput. Eng., Univ. Illinois Urbana-Champaign, Urbana, IL, 2005.

Chandra Nair (S'02–M'05) received the B.S. degree in electrical engineering from Indian Institute of Technology (IIT), Madras, India, in 1999 and the M.S. and Ph.D. degrees from the Electrical Engineering Department, Stanford University, Stanford, CA, in 2002 and 2005, respectively.

Currently, he is an Assistant Professor at the Information Engineering Department, Chinese University of Hong Kong, Shatin, Hong Kong. Previously, he was a Postdoctoral Researcher at the Theory Group, Microsoft Research, Redmond, VA. His research interests are in discrete optimization problems arising in electrical engineering and computer science, algorithm design, networking, and information theory.

Dr. Nair has received the Stanford and Microsoft Graduate Fellowships (2000–2004, 2005) for his graduate studies, and he was awarded the Philips and Siemens (India) Prizes in 1999 for his undergraduate academic performance.

Abbas El Gamal (S'71–M'73–SM'83–F'00) received the B.Sc. degree in electrical engineering from Cairo University, Cairo, Egypt, in 1972, the M.S. degree in statistics and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1977 and 1978, respectively.

From 1978 to 1980, he was an Assistant Professor of Electrical Engineering at the University of Southern California. He has been on the Stanford faculty since 1981, where he is currently Professor of Electrical Engineering and the Director of the Information Systems Laboratory. He was on leave from Stanford from 1984 to 1988 first as Director of LSI Logic Research Lab, then as cofounder and Chief Scientist of Actel Corporation. In 1990, he cofounded Silicon Architects, which was later acquired by Synopsys. His research has spanned several areas, including information theory, digital imaging, and integrated circuit design and design automation. He has authored or coauthored over 150 papers and 25 patents in these areas.