On an Outer bound and an Inner Bound for the General Broadcast Channel

Amin Aminzadeh Gohari
EECS Department
University of California
Berkeley, CA 94720, USA
Email: aminzade@eecs.berkeley.edu

Abbas El Gamal
EE Department
Stanford University
Stanford, CA 94305, USA
Email: abbas@ee.stanford.edu

Venkat Anantharam
EECS Department
University of California
Berkeley, CA 94720, USA
Email: ananth@eecs.berkeley.edu

***Abstract*—In this paper, we study the Nair-El Gamal outer bound and Marton's inner bound for general two-receiver broadcast channels. We show that the Nair-El Gamal outer bound can be made fully computable. For the inner bound, we show that, unlike in the Gaussian case, for a degraded broadcast channel even without a common message, Marton's coding scheme without a superposition variable is in general insufficient for obtaining the capacity region. Further, we prove various results that help to restrict the search space for computing the sum-rate for Marton's inner bound. We establish the capacity region along certain directions and show that it coincides with Marton's inner bound. Lastly, we discuss an idea that may lead to a larger inner bound.**

I. Introduction

In this paper, we consider the general two-receiver broadcast channel with an input alphabet $\mathcal{X}$, output alphabets $\mathcal{Y}$ and $\mathcal{Z}$, and conditional probability distribution function $q(y,z|x)$. The capacity region of this channel is defined as the set of rate triples (R_0, R_1, R_2) such that the sender X can reliably communicate a common message at rate R_0 to both receivers and two private messages at rates R_1 and R_2 to receivers Y and Z respectively, see [1] or [2]. The capacity region of this channel is known for several special cases but unknown in general. The best known general inner bound to the capacity region is due to Marton [3][6]. A long standing outer bound for the general broadcast channel was due to Körner and Marton. Recently Nair and El Gamal proposed a new outer bound for the general broadcast channel and showed that it is strictly contained in the Körner-Marton outer bound for the binary skew symmetric broadcast channel (BSSC) [8]. Following this work, a series of outer bounds on the capacity region of the general broadcast channel were reported. It is not known whether any of these outer bounds is strictly contained in the Nair-El Gamal outer bound.

In this paper, we study the aforementioned inner and outer bounds. We find bounds on the cardinalities of the auxiliary random variables appearing in the Nair-El Gamal outer bound, thus making it fully computable and generalizing an earlier result by Nair and Zizhou [5] for the case of $R_0 = 0$. We then consider Marton's inner bound and show that, unlike in the Gaussian broadcast channel case, "Marton's coding scheme" alone is not sufficient to achieve the capacity region of the general degraded broadcast channel. Necessity of the "superposition-coding" aspect of the inner bound had previously been observed for a non-degraded broadcast channel [12]. We provide several results that help to restrict the search space for computing the sum rate in Marton's inner bound, and also describe certain directions along which Marton's inner bound equals the capacity region. Lastly, we discuss an idea that may lead to a larger inner bound for the general broadcast channel.

The rest of the paper is organized as follows. In section II, we introduce the basic notation and definitions we use. Section III contains the main results of the paper, and section IV contains some of the proofs. The rest of the proofs can be found in [11].

II. Notation and Definitions

Let $\mathcal{C}(q(y,z|x))$ denote the capacity region of the broadcast channel $q(y,z|x)$. The notation X^i is used to denote the vector $(X_1, X_2, ..., X_i)$, and X_i^n to denote $(X_i, X_{i+1}, ..., X_n)$. Given random variables $K_1, K_2 \in \{0, 1, 2, ..., k-1\}$ for a natural number k, $K_1 \oplus K_2$ denotes $(K_1 + K_2) \mod k$.

Definition 1: [3][2][4][6] Let $\mathcal{C}_M(q(y,z|x))$ denote Marton's inner bound for the channel $q(y,z|x)$, defined as the set of non-negative rate triples (R_0, R_1, R_2) satisfying

$$R_0 + R_1 < I(UW;Y), \quad (1)$$
$$R_0 + R_2 < I(VW;Z), \quad (2)$$
$$R_0 + R_1 + R_2 < I(UW;Y) + I(V;Z|W) - I(U;V|W), \quad (3)$$
$$R_0 + R_1 + R_2 < I(U;Y|W) + I(VW;Z) - I(U;V|W), \quad (4)$$
$$2R_0 + R_1 + R_2 < I(UW;Y) + I(VW;Z) - I(U;V|W) \quad (5)$$

for some random variables $(U,V,W,X,Y,Z) \sim p(u,v,w,x)q(y,z|x)$.

Definition 2: [8] Let $\mathcal{C}_{NE}(q(y,z|x))$ denote the Nair-El Gamal outer bound on the channel $q(y,z|x)$, defined as the set of non-negative rate triples (R_0, R_1, R_2) satisfying

$$\begin{aligned} R_0 &\le \min\{I(W;Y), I(W;Z)\}, \\ R_0 + R_1 &\le I(UW;Y), \\ R_0 + R_2 &\le I(VW;Z), \\ R_0 + R_1 + R_2 &\le I(UW;Y) + I(V;Z|UW), \\ R_0 + R_1 + R_2 &\le I(VW;Z) + I(U;Y|VW) \end{aligned}$$

for some random variables $(U, V, W, X, Y, Z) \sim p(u)p(v)p(w|u,v)p(x|u,v,w)q(y,z|x)$,

Definition 3: [7] Let $\mathcal{C}_{d_1}(q(y,z|x))$ and $\mathcal{C}_{d_2}(q(y,z|x))$ denote the degraded message set capacity regions, i.e. when $R_1 = 0$ and $R_2 = 0$, respectively. The capacity region $\mathcal{C}_{d_1}(q(y,z|x))$ is the set of of non-negative rate pairs (R_0, R_2) satisfying

$$\begin{aligned} R_0 &\le I(W;Y), \\ R_2 &\le I(X;Z|W), \\ R_0 + R_2 &\le I(X;Z) \end{aligned}$$

for some random variables $(W, X, Y, Z) \sim p(w,x)q(y,z|x)$. The capacity region $\mathcal{C}_{d_2}(q(y,z|x))$ is defined similarly.

III. Statement of the results

A. On the Nair-El Gamal outer bound

Theorem 1: For a general broadcast channel $q(y,z|x)$, the Nair-El Gamal outer bound $\mathcal{R}_{NE}$ is the set of non-negative rate triples (R_0, R_1, R_2) satisfying

$$R_0 \le \min\{I(W;Y), I(W;Z)\}, \quad (6)$$

$$R_0 + R_1 \le I(UW;Y), \quad (7)$$

$$R_0 + R_2 \le I(VW;Z), \quad (8)$$

$$R_0 + R_1 + R_2 \le \min\{I(UW;Y) + I(X;Z|UW), I(VW;Z) + I(X;Y|VW)\} \quad (9)$$

for some random variables $(U, V, W, X, Y, Z) \sim p(w,x)p(u|w,x)p(v|w,x)q(y,z|x)$ with $|\mathcal{U}| \le |\mathcal{X}|$, $|\mathcal{V}| \le |\mathcal{X}|$, $|\mathcal{W}| \le |\mathcal{X}| + 6$.

Note that the above result makes the Nair-El Gamal outer bound fully computable.

B. On Marton's Inner Bound

1) Insufficiency of Marton's coding scheme without a superposition variable: In Marton's inner bound the auxiliary random variable W corresponds to the "superposition-coding" aspect of the bound, and the random variables U and V correspond to the "Marton-coding" aspect of the bound. When $R_0 = 0$ (private messages only) and $W = \emptyset$, Marton's inner bound reduces to the the set of non-negative rate pairs (R_1, R_2) satisfying

$$R_1 \le I(U;Y|Q), \quad (10)$$

$$R_2 \le I(V;Z|Q), \quad (11)$$

$$R_1 + R_2 \le I(U;Y|Q) + I(V;Z|Q) - I(U;V|Q) \quad (12)$$

for some random variables $(Q, U, V, X, Y, Z) \sim p(q)p(u,v,x|q)q(y,z|x)$.

It is known that this inner bound is tight for Gaussian broadcast channels (through dirty paper coding), implying that W is unnecessary for achieving the capacity region of this class of degraded broadcast channels. We show through an example that this is not the case in general.

Lemma 1: There are degraded broadcast channels for which Marton's private message inner bound without W is strictly contained in the capacity region of the channel.

2) Computing the sum-rate for Marton's Inner Bound: For any $\lambda \in [0,1]$, let

$$T(\lambda) = \max_{p(u,v,w,x)} \big(\lambda I(W;Y) + (1-\lambda) I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W)\big).$$

Computing the sum-rate for Marton's inner bound is closely related to the following maximization problem for $\lambda \in [0,1]$:

Observation 1: The maximum of the sum-rate for Marton's inner bound is equal to $\min_{\lambda \in [0,1]} T(\lambda)$.

The main theorem of this section restricts the search space for computing $T(\lambda)$. In this section, we only deal with broadcast channels $q(y,z|x)$ with strictly positive transition matrices, i.e. when $q(y|x) > 0, q(z|x) > 0$ for all x, y, z. In order to evaluate $T(\lambda)$ when $q(y|x)$ or $q(z|x)$ become zero for some y or z, one can use the continuity of $T(\lambda)$ in $q(y,z|x)$ and take the limit of $T(\lambda)$ for a sequence of channels with positive entries converging to the desired channel. The reason for dealing with this class of broadcast channels should become clear by the following lemma.

Lemma 2: Take an arbitrary broadcast channel $q(y,z|x)$ with strictly positive transition matrices (i.e. $q(y|x) > 0, q(z|x) > 0$ for all x, y, z). Let $p(u,v,w,x)$ be an arbitrary joint distribution maximizing $T(\lambda)$ for some $\lambda \in [0,1]$. If $p(u,w)$ and $p(v,w)$ are positive for some triple (u,v,w), then it must be the case that $p(u,v,w) > 0$, $p(u,w,y) > 0$ and $p(v,w,z) > 0$ for all y and z.

Theorem 2: Take an arbitrary broadcast channel $q(y,z|x)$ with strictly positive transition matrices. In computing $T(\lambda)$ for some $\lambda \in [0,1]$, it suffices to take the maximum over auxiliary random variables $p(u,v,w,x)q(y,z|x)$ simultaneously satisfying the following constraints

- $|\mathcal{U}| \le \min(|\mathcal{X}|, |\mathcal{Y}|)$, $|\mathcal{V}| \le \min(|\mathcal{X}|, |\mathcal{Z}|)$, $|\mathcal{W}| \le |\mathcal{X}|$,
- $H(X|UVW) = 0$,
- For arbitrary w such that $p(w) > 0$, let the functions

  $$\begin{aligned} &f_{u,w} : \mathcal{X} \to \mathbb{R} \text{ for every } u \in \mathcal{U} \text{ such that } p(u|w) > 0, \\ &g_{v,w} : \mathcal{X} \to \mathbb{R} \text{ for every } v \in \mathcal{V} \text{ such that } p(v|w) > 0, \\ &\text{and } h_w : \mathcal{X} \to \mathbb{R} \end{aligned}$$

  be defined by

  $$f_{u,w}(x) = \textstyle\sum_y q(y|x) \log p(uy|w),$$
  $$g_{v,w}(x) = \textstyle\sum_z q(z|x) \log p(vz|w),$$

  $$h_w(x) = \min_{u',v': p(u'|w)>0, p(v'|w)>0} \Big(\log(p(u'v'|w)) - f_{u',w}(x) - g_{v',w}(x) \Big).$$

These definitions make sense because of Lemma 2. Then, for any (u, v) where $p(u|w) > 0$ and $p(v|w) > 0$, the following two equations hold

$$\log(p(uv|w)) = \max_x f_{u,w}(x) + g_{v,w}(x) + h_w(x),$$

and

$$p(x_0|u, v, w) = 1 \text{ for some } x_0 \in \mathcal{X} \Rightarrow$$
$$x_0 \in argmax_x f_{u,w}(x) + g_{v,w}(x) + h_w(x).$$

- Given any w, random variables U_w, V_w, X_w, Y_w, Z_w distributed according to $p(u, v, x, y, z|w)$ satisfy the following:

$$I(\overline{U}; Y_w) \geq I(\overline{U}; V_w Z_w) \text{ for any } \overline{U} \to U_w \to V_w X_w Y_w Z_w,$$
$$I(\overline{V}; Z_w) \geq I(\overline{V}; U_w Y_w) \text{ for any } \overline{V} \to V_w \to U_w X_w Y_w Z_w.$$

Discussion 1: The first constraint imposes cardinality bounds on $|\mathcal{U}|$ and $|\mathcal{V}|$ that are better than those reported in [9]. *However, we only claim the improved cardinality bounds for $T(\lambda)$ and not the whole capacity region.* The second constraint is not new, and can be found in [9]. The other constraints are useful in restricting the search space due to the constraints imposed on $p(u, v, w, x)$. For instance, take arbitrary w, u_0, u_1, v_0, v_1 where $p(w) > 0, p(u_i|w) > 0, p(v_i|w) > 0$ for $i = 0, 1$. Assume further that $p(x_0|u_0, v_0, w) = p(x_0|u_1, v_1, w) = 1$ for some x_0. Then the third constraint can be used to show that $p(u_0, v_0, w)p(u_1, v_1, w) \leq p(u_1, v_0, w)p(u_0, v_1, w)$.

C. On the capacity region

Lemma 3: For a broadcast channel $q(y, z|x)$ and real numbers λ_0, λ_1 and λ_2 such that $\lambda_0 \geq \lambda_1 + \lambda_2$,

$$\max_{(R_0,R_1,R_2)\in\mathcal{C}(q(y,z|x))} (\lambda_0 R_0 + \lambda_1 R_1 + \lambda_2 R_2) =$$
$$\max\{\max_{(R_0,R_2)\in\mathcal{C}_{d_1}(q(y,z|x))} (\lambda_0 R_0 + \lambda_2 R_2),$$
$$\max_{(R_0,R_1)\in\mathcal{C}_{d_2}(q(y,z|x))} (\lambda_0 R_0 + \lambda_1 R_1)\},$$

where $\mathcal{C}_{d_1}(q(y, z|x))$ and $\mathcal{C}_{d_2}(q(y, z|x))$ are the degraded message set capacity regions for the given channel.

Corollary 1: The above observation essentially says that if $\lambda_0 \geq \lambda_1 + \lambda_2$, then *a* maximum of $\lambda_0 R_0 + \lambda_1 R_1 + \lambda_2 R_2$ over triples (R_0, R_1, R_2) in the capacity region occurs when either $R_1 = 0$ or $R_2 = 0$.

Remark 1: Since $\mathcal{C}_{d_1}(q(y, z|x)) \cup \mathcal{C}_{d_1}(q(y, z|x)) \subset \mathcal{C}_M(q(y, z|x)) \subset \mathcal{C}(q(y, z|x))$, the above lemma implies that Marton's inner bound is tight along the direction of such $(\lambda_0, \lambda_1, \lambda_2)$, i.e.

$$\max_{(R_0,R_1,R_2)\in\mathcal{C}(q(y,z|x))} (\lambda_0 R_0 + \lambda_1 R_1 + \lambda_2 R_2) =$$
$$\max_{(R_0,R_1,R_2)\in\mathcal{C}_M(q(y,z|x))} (\lambda_0 R_0 + \lambda_1 R_1 + \lambda_2 R_2),$$

whenever $\lambda_0 \geq \lambda_1 + \lambda_2$.

Based on numerical simulations for certain broadcast channels, we conjecture that the Nair-El Gamal outer bound is also tight along the direction of any such $(\lambda_0, \lambda_1, \lambda_2)$. However if this conjecture turns out to be false, it would imply that the Nair-El Gamal outer bound is not tight.

D. An achievable region

Since capacity is defined in the limit of large block length, it is natural to expect that it has an invariant structure with respect to shifts in time. This suggests that it should be expressed via a formula that has a fixed-point character, namely it should involve joint distributions that are invariant under a time shift. The following theorem is a proposed inner bound along these lines.

Theorem 3: For a broadcast channel $q(y, z|x)$, consider two i.i.d. copies (U_1, V_1, W_1) and (U_2, V_2, W_2) and a conditional pmf $r(x|u_1, v_1, w_1, u_2, v_2, w_2)$. Assume that $U_1, V_1, W_1, U_2, V_2, W_2, X_1, X_2, Y_1, Y_2, Z_1, Z_2$ are distributed according to

$$p(u_1, v_1, w_1, u_2, v_2, w_2, x_1, y_1, z_1, x_2, y_2, z_2) =$$
$$r(u_1, v_1, w_1)r(u_2, v_2, w_2)\cdot$$
$$r(x_2|u_1, v_1, w_1, u_2, v_2, w_2)q(y_2, z_2|x_2)\cdot$$
$$\tilde{r}(x_1|u_1, v_1, w_1)q(y_1, z_1|x_1),$$

where $\tilde{r}(x|u, v, w)$ is defined as

$$\sum_{u'\in\mathcal{U}, v'\in\mathcal{V}, w'\in\mathcal{W}} r(x|u', v', w', u, v, w)r(u', v', w').$$

Then a rate triple (R_0, R_1, R_2) is achievable if

$$R_0, R_1, R_2 \geq 0,$$
$$R_0 + R_1 < I(U_2 W_2; Y_1 Y_2 U_1 W_1),$$
$$R_0 + R_2 < I(V_2 W_2; Z_1 Z_2 V_1 W_1),$$
$$R_0 + R_1 + R_2 < I(V_2; Z_1 Z_2 V_1 W_1|W_2) + I(U_2 W_2; Y_1 Y_2 U_1 W_1) - I(U_2; V_2|W_2),$$
$$R_0 + R_1 + R_2 < I(U_2; Y_1 Y_2 U_1 W_1|W_2) + I(V_2 W_2; Z_1 Z_2 V_1 W_1) - I(U_2; V_2|W_2),$$
$$2R_0 + R_1 + R_2 < I(U_2 W_2; Y_1 Y_2 U_1 W_1) + I(V_2 W_2; Z_1 Z_2 V_1 W_1) - I(U_2; V_2|W_2)$$

for some $U_1, V_1, W_1, U_2, V_2, W_2, X_1, X_2$ that satisfy the above conditions.

Remark 2: The above inner bound reduces to Marton's inner bound if the conditional distribution $r(x|u_1, v_1, w_1, u_2, v_2, w_2) = r(x|u_2, v_2, w_2)$, i.e. $U_1 V_1 W_1 \to U_2 V_2 W_2 \to X$ form a Markov chain.

IV. Proofs

Proof of Theorem 1: Let $\mathcal{R}_1$ denote the region given in the statement of Theorem 1. We would like to show that $\mathcal{R}_1 = \mathcal{R}_{NE}$. Our proof resembles and generalizes the one provided by Nair and Zizhou [5] for the case of $R_0 = 0$. We first show that $\mathcal{R}_{NE} = \mathcal{R}_2$, where $\mathcal{R}_2$ consists of the set of non-negative rate triples (R_0, R_1, R_2) satisfying equations (6)-(9) for some random variables $U, V, W, X, Y, Z \sim p(u, v, w, x)q(y, z|x)$. Clearly $\mathcal{R}_{NE} \subset \mathcal{R}_2$, since in $\mathcal{R}_2$, we take the union over all $p(u, v, w, x)$, and $I(X; Z|UW) \geq I(V; Z|UW)$, $I(X; Y|VW) \geq I(U; Y|VW)$. In order to show that $\mathcal{R}_2 \subset \mathcal{R}_{NE}$, take some arbitrary $p(u,v,w,x)$. Without loss of generality assume that $\mathcal{U} = \{0,1,2,...,|\mathcal{U}|-1\}$, $\mathcal{V} = \{0,1,2,...,|\mathcal{V}|-1\}$ and $\mathcal{X} = \{0,1,2,...,|\mathcal{X}|-1\}$.

Let $\widetilde{U}$, $\widetilde{V}$, $\widetilde{X}_1$, $\widetilde{X}_2$, $\widetilde{X}_3$, $\widetilde{X}_4$ be *uniform* random variables on sets $\mathcal{U}$, $\mathcal{V}$, $\mathcal{X}$, $\mathcal{X}$, $\mathcal{X}$, $\mathcal{X}$ respectively. We assume that $\widetilde{U}$, $\widetilde{V}$, $\widetilde{X}_1$, $\widetilde{X}_2$, $\widetilde{X}_3$ and $\widetilde{X}_4$ are *mutually independent* of each other and of U,V,W,X,Y,Z. Let us define random variables $\widehat{U}$, $\widehat{V}$, $\widehat{W}$ and $\widehat{X}$, $\widehat{Y}$ and $\widehat{Z}$ as follows:

- $\widehat{U} = (\widetilde{U} \oplus U, X \oplus \widetilde{X}_1 \oplus \widetilde{X}_4, \widetilde{X}_3)$,
- $\widehat{V} = (\widetilde{V} \oplus V, X \oplus \widetilde{X}_2 \oplus \widetilde{X}_3, \widetilde{X}_4)$,
- $\widehat{W} = (W, \widetilde{U}, \widetilde{V}, \widetilde{X}_1, \widetilde{X}_2)$,
- $\widehat{X} = X$, $\widehat{Y} = Y$, $\widehat{Z} = Z$.

It can be verified that $\widehat{U}$ is independent of $\widehat{V}$. Furthermore

- $I(W;Y) \leq I(\widehat{W};\widehat{Y})$, $I(W;Z) \leq I(\widehat{W};\widehat{Z})$,
- $I(UW;Y) \leq I(\widehat{U}\widehat{W};\widehat{Y})$, $I(VW;Z) \leq I(\widehat{V}\widehat{W};\widehat{Z})$,
- $I(UW;Y)+I(X;Z|UW) \leq I(\widehat{U}\widehat{W};\widehat{Y})+I(\widehat{V};\widehat{Z}|\widehat{U}\widehat{W})$,
- $I(VW;Z)+I(X;Y|VW) \leq I(\widehat{V}\widehat{W};\widehat{Z})+I(\widehat{U};\widehat{Y}|\widehat{V}\widehat{W})$.

Therefore $\mathcal{R}_2 \subset \mathcal{R}_{NE}$, and $\mathcal{R}_2 = \mathcal{R}_{NE}$. It remains to show that $\mathcal{R}_1 = \mathcal{R}_2$. The sketch of the rest of the proof is as follows (the details can be found [11]): we show that in evaluating $\mathcal{R}_2$ it suffices to take the union over all $p(u,v,w,x)$ of the form $p(w,x)p(u|w,x)p(v|w,x)$, and then use the strengthened Carathéodory theorem of Fenchel and Eggleston to establish cardinality bounds on the auxiliary random variables. ■

Proof of Lemma 1: Consider the degraded broadcast channel $p(y,z|x) = p(y|x)p(z|y)$, where the channel from X to Y is a BSC(0.3) and the channel from Y to Z is as follows: $p_{Z|Y}(0|0) = 0.6$, $p_{Z|Y}(1|0) = 0.4$, $p_{Z|Y}(0|1) = 0$, $p_{Z|Y}(1|1) = 1$. We show that the private message capacity region for this channel is strictly larger than Marton's inner bound without W.

In the following we provide a formal proof; see [11] for an intuitive discussion of the proof. The maximum of $R_1 + 2.4R_2$ over pairs (R_1, R_2) in the capacity region, is equal to $\max_{V\rightarrow X\rightarrow YZ} I(X;Y|V) + 2.4I(V;Z)$. Take the joint pmf of $p(v,x)$ to be as follows: $P(V=0,X=0) = 0$, $P(V=0,X=1) = 0.41$, $P(V=1,X=0) = 0.48$, $P(V=1,X=1) = 0.11$. For this choice of $p(v,x)$, $I(X;Y|V) + 2.4I(V;Z) = 0.1229....$ Therefore the maximum of $R_1 + 2.4R_2 \geq 0.1229....$ The maximum of $R_1 + 2.4R_2$ over Marton's inner bound without W is equal to $\sup_{UV\rightarrow X\rightarrow YZ} I(U;V) + 2.4I(V;Z) - I(U;V)$. Using the perturbation method of [9], one can show that the supremum is indeed a minimum, and that the cardinality of U and V can be bounded from above by $|\mathcal{X}|$. Furthermore X can be assumed to be a deterministic function of (U,V). Since X is a binary random variable, we need to search over binary random variables U, V. Numerical simulations show that the maximum is equal to $0.1215... < 0.1229...$ and occurs when $X = V$ and $U = constant$. Therefore Marton's inner bound without W is not tight for this broadcast channel. ■

Proof of Observation 1: This observation was exploited in section 3.1.1 of [12], but no proof for it was given in [12]. We have provided a proof in [11] for completeness. Here we sketch the proof. The observation essentially claims that it is legitimate to exchange the maximum and minimum operators in $\max_{p(u,v,w,x)} \min_{\lambda\in[0,1]} \lambda I(W;Y) + (1-\lambda)I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W)$. The proof begins by showing that the union over all $p(u,v,w,x)$ of real pairs (d_1, d_2) satisfying

$$d_1 \leq I(W;Y) + I(U;Y|W) + I(V;Z|W) - I(U;V|W),$$
$$d_2 \leq I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W).$$

is convex and closed. Furthermore, the point (Marton's Sum Rate, Marton's Sum Rate) is on the boundary of this region. Therefore, one can use the supporting hyperplane theorem to conclude the existence of a supporting hyperplane to the region at this boundary point. We then show that this supporting hyperplane must have the equation $\lambda^* d_1 + (1-\lambda^*)d_2 = T(\lambda^*)$ for some $\lambda^* \in [0,1]$. The next main step is to show that the maximum sum rate for Marton's inner bound equals $T(\lambda^*)$. ■

Proof of Lemma 2: Take a triple (u,v,w) such that $p(u,w)$ and $p(v,w)$ are positive. There must exist some x such that $p(u,w,x) > 0$. Since the transition matrices have positive entries and $p(u,w,y) \geq p(u,w,x)q(y|x)$, $p(u,w,y)$ will be positive for all y. A similar statement could be proved for $p(v,w,z)$. Assume that $p(u,v,w) = 0$. Take some u', v' such that $p(u',v',w) > 0$. Let us reduce $p(u',v',w)$ by ϵ and increase $p(u,v,w)$ by ϵ. Furthermore, have (u,v,w) mapped to the same x that (u',v',w) was mapped to; this ensures that the joint distribution of W and X is preserved. One can then verify that the first derivative of $T(\lambda)$ with respect to ϵ, at $\epsilon = 0$, will be positive. This is a contradiction since $p(u,v,w,x)$ was assumed to maximize $T(\lambda)$. ■

Proof of Theorem 2: A sketch of the proof follows. See [11] for the details. From the set of pmfs $p(u,v,w,x)$ that maximize the expression $\lambda I(W;Y) + (1-\lambda)I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W)$, let $p_0(u,v,w,x)$ be the one that achieves the largest value of $I(W;Y) + I(W;Z)$. The first step is to show that one can find $p(\widehat{u},\widehat{v},\widehat{w},\widehat{x})$ for which the constraints in the first and second bullets are satisfied, and furthermore $I(\widehat{W};\widehat{Y}) + I(\widehat{W};\widehat{Z})$ is equal to $I(W;Y) + I(W;Z)$. Next, we show that the third bullet of Theorem 2 holds for any joint distribution that maximizes the expression $\lambda I(W;Y) + (1-\lambda)I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W)$. The proof for this part considers the optimization problem of $I(U;Y) + I(V;Z) - I(U;V)$ over all $p(u,v,x)$ subject to a fixed marginal distribution on X. Let $p^*(u,v,x)$ be an answer to this optimization problem. Fix some x, and consider a function that takes in a pair (u,v), and outputs the first partial derivative of $I(U;Y) + I(V;Z) - I(U;V)$ with respect to $p(u,v,x)$ at $p^*(u,v,x)$. It is then proved that this function attains its maximum at any pair (u,v) where $p^*(u,v,x) > 0$. This fact implies the third bullet of Theorem 2.

The last step is to show that the fourth bullet of Theorem 2 holds for any joint distribution that maximizes the expression $\lambda I(W;Y) + (1-\lambda)I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W)$, and at the same time has the largest possible value of $I(W;Y) + I(W;Z)$. ■

Proof of Lemma 3: It suffices to show that

$$\max_{(R_0,R_1,R_2)\in\mathcal{C}(q(y,z|x))}(\lambda_0 R_0+\lambda_1 R_1+\lambda_2 R_2)\leq$$
$$\max\{\max_{(R_0,R_2)\in\mathcal{C}_{d_1}(q(y,z|x))}(\lambda_0 R_0+\lambda_2 R_2),$$
$$\max_{(R_0,R_1)\in\mathcal{C}_{d_2}(q(y,z|x))}(\lambda_0 R_0+\lambda_1 R_1)\}.$$

The key step is to show that if (R_0,R_1,R_2) is in the capacity region of a broadcast channel, then $(R_0+\min\{R_1,R_2\},R_1-\min\{R_1,R_2\},R_2-\min\{R_1,R_2\})$ is also in the capacity region. Since $\lambda_0\geq\lambda_1+\lambda_2$, we then have that $\lambda_0(R_0+\min\{R_1,R_2\})+\lambda_1(R_1-\min\{R_1,R_2\})+\lambda_2(R_2-\min\{R_1,R_2\})\geq\lambda_0R_0+\lambda_1R_1+\lambda_2R_2$, so at the maximum we must have $\min(R_1,R_2)=0$. One can prove this property using the result of Willems [10], which shows that the *maximal* probability of error capacity region is equal to the *average* probability of error capacity region. Willems' proof, however, is rather involved. Instead, we provide a simple direct proof. Consider an arbitrary code $(M_0,M_1,M_2,X^n,\epsilon)$. We show that

$$\frac{\lambda_0}{n}H(M_0)+\frac{\lambda_1}{n}H(M_1)+\frac{\lambda_2}{n}H(M_2)-O(\epsilon)\leq$$
$$\max(\max_{(R_0,R_2)\in\mathcal{C}_{d_1}(q(y,z|x))}\lambda_0R_0+\lambda_2R_2,$$
$$\max_{(R_0,R_1)\in\mathcal{C}_{d_2}(q(y,z|x))}\lambda_0R_0+\lambda_1R_1)$$

where $O(\epsilon)$ denotes a constant (depending only on $|\mathcal{X}|$, $|\mathcal{Y}|$, $|\mathcal{Z}|$) times ϵ.

Assume without loss of generality that $H(M_2)\leq H(M_1)$, i.e. $R_2\leq R_1$. Let $\widehat{W}=M_0M_2$, $\widehat{X}=X^n$, $\widehat{Y}=Y^n$, $\widehat{Z}=Z^n$. Note that $q(\widehat{y},\widehat{z}|\widehat{x})$ is the n-fold version of $q(y,z|x)$. Let us look at $\mathcal{C}_{d_1}(q(\widehat{y},\widehat{z}|\widehat{x}))$, evaluated at the joint pmf $p(\widehat{w},\widehat{x})$:

$$\begin{aligned}\widehat{R_0}&\leq I(\widehat{W};\widehat{Z}),\\ \widehat{R_1}&\leq I(\widehat{X};\widehat{Y}|\widehat{W}),\\ \widehat{R_0}+\widehat{R_1}&\leq I(\widehat{X};\widehat{Y}).\end{aligned}$$

Note that, by Fano's inequality,

$$I(\widehat{W};\widehat{Z})=I(M_0M_2;Z^n)=H(M_0)+H(M_2)-O(n\epsilon),$$
$$I(\widehat{X};\widehat{Y}|\widehat{W})=I(X^n;Y^n|M_0M_2)=H(M_1)-O(n\epsilon),$$
$$I(\widehat{X};\widehat{Y})=I(X^n;Y^n)=H(M_0)+H(M_1)-O(n\epsilon).$$

Therefore $\widehat{R_0}=H(M_0)+H(M_2)-O(n\epsilon)=n(R_0+R_2)-O(n\epsilon)$ and $\widehat{R_1}=H(M_1)-H(M_2)=n(R_1-R_2)-O(n\epsilon)$ is in $\mathcal{C}_{d_1}(q(\widehat{y},\widehat{z}|\widehat{x}))$. Since $q(\widehat{y},\widehat{z}|\widehat{x})$ is the n-fold version of $q(y,z|x)$ and $\mathcal{C}_{d_1}(q(\widehat{y},\widehat{z}|\widehat{x}))$ is the degraded message set capacity region for $q(\widehat{y},\widehat{z}|\widehat{x})$, we must have: $\mathcal{C}_{d_1}(q(\widehat{y},\widehat{z}|\widehat{x}))=n\cdot\mathcal{C}_{d_1}(q(y,z|x))$, where the multiplication here is pointwise. Thus, $(\frac{\widehat{R_0}}{n},\frac{\widehat{R_1}}{n})\in\mathcal{C}_{d_1}(q(y,z|x))$. We can complete the proof by letting $\epsilon\to0$, and conclude that $(R_0+R_2,R_1-R_2,0)\in\mathcal{C}_{d_1}(q(y,z|x))$, and thus also in the capacity region. ■

Proof of Theorem 3: Consider a natural number n, and define the super symbols $\tilde{X}=X_1...X_n$, $\tilde{Y}=Y_1...Y_n$, $\tilde{Z}=Z_1...Z_n$ representing n-inputs and n-outputs of the product broadcast channel $q^n(y_1y_2...y_n,z_1z_2...z_n|x_1x_2...x_n)=\prod_{i=1}^n q(y_i,z_i|x_i)$. Since the capacity region of the product channel $q^n(\tilde{y},\tilde{z}|\tilde{x})$ is n times the capacity region of $q(y,z|x)$, one can show that given an *arbitrary* joint pmf $p(u^n,v^n,w^n,x^n)$, the following region is an inner bound to $\mathcal{C}(q(y,z|x))$ for $U^n,V^n,W^n,X^n,Y^n,Z^n\sim p(u^n,v^n,w^n,x^n)q(y^n,z^n|x^n)$:

$$R_0,R_1,R_2\geq0,$$
$$R_0+R_1\leq\frac{1}{n}I(U^nW^n;Y^n),\tag{13}$$
$$R_0+R_2\leq\frac{1}{n}I(V^nW^n;Z^n),\tag{14}$$
$$R_0+R_1+R_2\leq\frac{1}{n}\big[I(U^nW^n;Y^n)+I(V^n;Z^n|W^n)-I(U^n;V^n|W^n)\big],\tag{15}$$
$$R_0+R_1+R_2\leq\frac{1}{n}\big[I(U^n;Y^n|W^n)+I(V^nW^n;Z^n)-I(U^n;V^n|W^n)\big],\tag{16}$$
$$2R_0+R_1+R_2\leq\frac{1}{n}\big[I(U^nW^n;Y^n)+I(V^nW^n;Z^n)-I(U^n;V^n|W^n)\big].\tag{17}$$

Assume that

$$p(u^n,v^n,w^n)=\prod_{i=1}^n r(u_i,v_i,w_i),$$
$$p(x_2^n|u^n,v^n,w^n)=\prod_{i=2}^n r(x_i|u_{i-1},v_{i-1},w_{i-1},u_i,v_i,w_i),$$
$$X_1=constant.$$

We can get the inner bound by evaluating $I(U^nW^n;Y^n)$, $I(V^nW^n;Z^n)$, etc. and plugging this into equations (13)-(17), and then letting $n\to\infty$. The details can be found in [11]. ■

Acknowledgment

The research was partially supported by NSF grants CCF-0500234, CCF-0635372, CNS-0627161 and ARO MURI grant W911NF-08-1-0233 "Tools for the Analysis and Design of Complex Multi-Scale Networks."

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, 1991.
[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Budapest, Hungary: Akadmiai Kiad, 1981.
[3] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. IT*, 25 (3): 306-311 (1979).
[4] S. I. Gelfand and M. S. Pinsker, "Capacity of a broadcast channel with one deterministic component," *Probl. Inf. Transm.*, 16 (1): 17-25 (1980).
[5] C. Nair and V.W. Zizhou, "On the inner and outer bounds for 2-receiver discrete memoryless broadcast channels," *Proceedings of the ITA workshop*, San Diego, 226-229 (2008).
[6] Y. Liang, G. Kramer, and H.V. Poor, "Equivalence of two inner bounds on the capacity region of the broadcast channel," *46th Annual Allerton Conf. on Commun., Control and Comp.*, 1417-1421 (2008).
[7] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. IT*, 23 (1): 60-64 (1977).
[8] C. Nair and A. El Gamal, "An outer bound to the capacity region of the broadcast channel," *IEEE Trans. IT*, 53 (1): 350-355 (2007).
[9] A. A. Gohari and V. Anantharam, "Evaluation of Marton's Inner Bound for the General Broadcast Channel," Submitted to *IEEE Trans. IT*.
[10] F. M. J. Willems, "The maximal-error and average-error capacity region of the broadcast channel are identical," *Problems of Control and Information Theory*, 19 (4): 339-347 (1990).
[11] A. A. Gohari, A. El Gamal and V. Anantharam, "On an Outer bound and an Inner Bound for the General Broadcast Channel," available at http://arxiv.org
[12] V. Jog and C. Nair, "An information inequality for the BSSC channel," available at http://arxiv.org/abs/0901.1492