# On broadcast channels with binary inputs and symmetric outputs

Yanlin Geng, *Member, IEEE,* Chandra Nair, *Member, IEEE,* Shlomo Shamai, *Fellow, IEEE,* and Zizhou Wang

***Abstract*—We establish capacity regions for some classes of broadcast channels with binary inputs and symmetric outputs. We investigate the more capable partial order and establish that the binary erasure channel and the binary symmetric channel form the two extremes for channels having the same capacity. Further, we apply the results to identify a class of broadcast channels for which the best known inner and outer bounds on the capacity region differ.**

## I. Introduction

In [1], Cover introduced the notion of a broadcast channel through which one sender transmits information to two or more receivers. For the purpose of this paper we focus our attention on broadcast channels with precisely two receivers.

A *broadcast channel*, denoted by $[X \to (Y_1, Y_2)]$, consists of an input alphabet $\mathcal{X}$, two output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, and a transition probability function $\mathrm{q}(y_1, y_2|x)$. A $(\lceil 2^{nR_1}\rceil, \lceil 2^{nR_2}\rceil, n)$ code for a broadcast channel consists of an encoder

$$x^n : \{1, 2, \dots, \lceil 2^{nR_1}\rceil\} \times \{1, 2, \dots, \lceil 2^{nR_2}\rceil\} \to \mathcal{X}^n,$$

and two decoders

$$\hat{\mathcal{W}}_1 : \mathcal{Y}_1^n \to \{1, 2, \dots, \lceil 2^{nR_1}\rceil\}$$
$$\hat{\mathcal{W}}_2 : \mathcal{Y}_2^n \to \{1, 2, \dots, \lceil 2^{nR_2}\rceil\}.$$

The probability of error $P_e^{(n)}$ is defined to be the probability of the event that either of the receivers decodes incorrectly, i.e.,

$$P_e^{(n)} = \mathbf{P}\left(\{\hat{\mathcal{W}}_1(Y_1^n) \neq \mathcal{W}_1\} \cup \{\hat{\mathcal{W}}_2(Y_2^n) \neq \mathcal{W}_2\}\right)$$

where the message is assumed to be uniformly distributed over $\{1, 2, \dots, \lceil 2^{nR_1}\rceil\} \times \{1, 2, \dots, \lceil 2^{nR_2}\rceil\}$.

A rate pair $(R_1, R_2)$ is said to be *achievable* for a broadcast channel if there exists a sequence of $(\lceil 2^{nR_1}\rceil, \lceil 2^{nR_2}\rceil, n)$ codes with $P_e^{(n)} \to 0$ as $n \to \infty$. The *capacity region* of a broadcast channel is the closure of the set of achievable rate pairs. *The capacity region of a general two-receiver discrete memoryless broadcast channel is unknown.*

Y. Geng and C. Nair are with The Chinese University of Hong Kong.
S. Shamai is with Technion-Israel Institute of Technology.
Z. Wang is with Altai Technologies Ltd., Hong Kong.

The work of C. Nair was partially supported by the following grants from the University Grants Committee of the Hong Kong Special Administrative Region, China: a) (Project No. AoE/E-02/08), b) GRF Project 415810. He also acknowledges the support from the Institute of Theoretical Computer Science and Communications (ITCSC) at The Chinese University of Hong Kong.

The work of S. Shamai was supported by the Israel Science Foundation (ISF).

The paper was presented, in part, at the International Symposium on Information Theory, 2010.

The capacity region is known [2] for classes of broadcast channels such as degraded, less noisy, more capable, essentially less noisy, and essentially more capable. In each of the classes mentioned above, there is a "dominant receiver" and it has been shown that superposition coding, where the dominant receiver is able to decode the message for the other receiver, is optimal.

For a pair of random variables $(X, Y)$ distributed according to $p(x, y)$, the mutual information is denoted as $I_p(X; Y)$. When the underlying distribution $p(x, y)$ is clear from the context, it is sometimes omitted.

**Definition 1** ([3])**.** A channel $T_1 : [X \to Y_1]$ is *less noisy* than another channel $T_2 : [X \to Y_2]$, if for all $p(u, x)$ such that $U \to X \to (Y_1, Y_2)$ is Markov we have $I(U; Y_1) \geq I(U; Y_2)$.

**Definition 2** ([3])**.** A channel $T_1 : [X \to Y_1]$ is *more capable* than another channel $T_2 : [X \to Y_2]$, denoted by $T_1 \overset{mc}{\geq} T_2$, if for all $p(x)$ we have $I(X; Y_1) \geq I(X; Y_2)$.

**Definition 3** ([4])**.** A class of distributions $\mathcal{P} = \{p(x)\}$ on the input alphabet $\mathcal{X}$ is said to be a *sufficient class* of distributions for a broadcast channel $[X \to (Y_1, Y_2)]$ if the following holds: Given any triple of random variables $(U, V, X)$ such that $(U, V) \to X \to (Y_1, Y_2)$ is Markov, there exists a distribution $q(\tilde{u}, \tilde{v}, x)$ such that $(\tilde{U}, \tilde{V}) \to X \to (Y_1, Y_2)$ is Markov and satisfies

$$\begin{aligned} q(x) &\in \mathcal{P}, \\ I_p(U; Y_2) &\leq I_q(\tilde{U}; Y_2), \\ I_p(V; Y_1) &\leq I_q(\tilde{V}; Y_1), \\ I_p(U; Y_2) + I_p(X; Y_1|U) &\leq I_q(\tilde{U}; Y_2) + I_q(X; Y_1|\tilde{U}), \\ I_p(V; Y_1) + I_p(X; Y_2|V) &\leq I_q(\tilde{V}; Y_1) + I_q(X; Y_2|\tilde{V}). \end{aligned} \tag{1}$$

*Remark* 1. Note that the definition of a sufficient class depends only on the pair of channels $[X \to Y_1]$ and $[X \to Y_2]$ rather than the joint transition probability of the broadcast channel $[X \to (Y_1, Y_2)]$.

**Definition 4** ([4])**.** A channel $T_1 : [X \to Y_1]$ is *essentially less noisy* than another channel $T_2 : [X \to Y_2]$, denoted by $T_1 \overset{eln}{\geq} T_2$, if there exists a sufficient class of distributions $\mathcal{P}$ such that for all $p(u, x)$ with $p(x) \in \mathcal{P}$ and $U \to X \to (Y_1, Y_2)$ is Markov we have $I_p(U; Y_1) \geq I_p(U; Y_2)$.

*Remark* 2. Note that essentially less noisy comparison may not induce a partial order among all channels because the sufficient class of distributions depends on the pair of channels under consideration. (In this regard, see Remark 9.)

*Remark* 3. Sometimes we say that receiver $Y_1$ is less noisy (more capable, essentially less noisy) than $Y_2$, with the un-

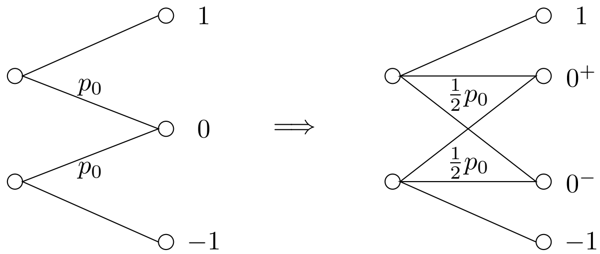


Fig. 1. Splitting of a binary erasure channel.

derstanding that the corresponding channels satisfy the corresponding relationship.

In this paper, we restrict ourselves to a class of discrete memoryless channels with binary inputs and symmetric outputs (BISO) defined below[1].

**Definition 5.** A binary input symmetric output (BISO) discrete memoryless channel $[X \to Y]$ has an input alphabet $\mathcal{X} = \{0, 1\}$, an output alphabet $\mathcal{Y} = \{0, \pm 1, \ldots, \pm l\}$, and transition probabilities that satisfy

$$p_k = \mathrm{P}(Y = k|X = 0) = \mathrm{P}(Y = -k|X = 1), -l \le k \le l.$$

A binary symmetric channel (BSC) and a binary erasure channel (BEC) are examples of BISO channels. It is easy to see that uniform input distribution $u(x)$ maximizes the mutual information (or in other words, $u(x)$ is a capacity achieving distribution), i.e.

$$C = \max_{p(x)} I_p(X;Y) = I_u(X;Y).$$

In the rest of this paper we assume without loss of generality that $\mathcal{Y} = \{\pm 1, \ldots, \pm l\}$. This can be done because one can always split the output $Y = 0$ into two outputs $Y = 0^+$, $Y = 0^-$ such that $\mathrm{P}(Y = 0^+|X = 0) = \mathrm{P}(Y = 0^-|X = 0) = \frac{p_0}{2}$. This new receiver is essentially indistinguishable from the original one as either receiver can "simulate" the other receiver locally and hence the probability of error corresponding to any decoding rule in one receiver can be achieved in the other receiver. We illustrate this splitting for a BEC in Figure 1.

For notation, we use $\{p_k, p_{-k} : k = 1, \ldots, l\}$ to denote the transition probabilities $\mathrm{P}(Y = k|X = 0)$, sometimes shortened to $\{p_k, p_{-k}\}_k$.

**Definition 6.** A binary input symmetric output broadcast channel is a broadcast channel where the channels $[X \to Y_1]$ and $[X \to Y_2]$ are both BISO channels.

*Remark* 4. Our interest is primarily in studying the capacity regions of discrete memoryless broadcast channels without feedback; it is known and easy to see that the capacity region depends only on the marginals. Therefore, we shall treat all broadcast channels with a given pair of marginals to belong to an equivalence class.

In this paper, we study the notions of more capable receivers and essentially less noisy receivers by focusing on the class of binary input symmetric output broadcast channels. We establish several results which are summarized below. These results can be considered as a natural generalization of the results in [4].

### A. *Summary of results*

- Any BISO channel with capacity $C$ is more capable than a BSC with capacity $C$. (Corollary 1).
- A BEC with capacity $C$ is more capable than any BISO channel with capacity $C$. (Corollary 2)
- Any two BISO channels with the same capacity and whose outputs have cardinalities at most 3, are more capable comparable, i.e. one receiver is more capable than the other receiver. (Corollary 3)
- For any two BISO channels with the same capacity, a receiver $Y_1$ is more capable than receiver $Y_2$ *if and only if* receiver $Y_2$ is essentially less noisy than $Y_1$. (They go in reverse directions[2].) (Lemma 3)
- The superposition coding region is the capacity region for a BISO broadcast channel if either of the channels is a BSC or a BEC. (Corollary 4)
- For two BISO channels with the same capacity, superposition coding is optimal if and only if the channels are more capable comparable. (Corollary 5)
- For two BISO channels with the same capacity Marton's inner bound differs from the outer bound [6] unless the channels are more capable comparable. (Theorem 3)
- It suffices to consider $[U \to X]$ to be BSC to compute the boundary of the superposition coding region for BISO broadcast channels. (Lemma 7). channels (Lemma 7). This generalizes a result of Wyner and Ziv [7] for degraded BSC broadcast channel.

### B. *Preliminaries*

Define $h(x) := -x \log_2 x - (1 - x) \log_2(1 - x)$ to be the binary entropy function, and $x = h^{-1}(y)$ be the inverse of $h(x)$, $x \in [0, 0.5]$.

Partition $P$ of an interval $[a, b]$ is a finite sequence of points $\{t_k\}$ such that $a = t_0 < t_1 < t_2 < \ldots < t_N = b$. A partition $P$ is finer than $Q$ if points of partition $P$ contain those of $Q$. A common refinement of two partitions $P$ and $Q$ is a new partition consisting of all the points of $P$ and $Q$.

**Definition 7** (BISO partition and BISO curve)**.** For a BISO channel with transition probabilities $\{p_k, p_{-k}\}_k$, rearrange $h(\frac{p_k}{p_k + p_{-k}})$ in the ascending order and denote the permutation as $\pi$. *BISO partition* is defined as the partition of $[0, 1]$ with points $t_0 = 0$ and $t_k = \sum_{i=1}^{k}(p_{\pi_i} + p_{-\pi_i})$. *BISO curve* is defined as the step function $f(t)$ such that $f(0) = 0$ and $f(t) = h(\frac{p_{\pi_k}}{p_{\pi_k} + p_{-\pi_k}})$ on $(t_{k-1}, t_k]$.

For the channel $BSC(p)$, we have the partition as $t_0 = 0$, $t_1 = 1$ and the curve as $f(t) = h(p)$ on $(0, 1]$. For the channel $BEC(e)$, we have the partition as $t_0 = 0$, $t_1 = 1 - e$, $t_2 = 1$, and the curve as $f(t) = 0$ on $(0, 1 - e]$ and $f(t) = 1$ on $(1 - e, 1]$. These two BISO curves are illustrated in Figure 2.

[1]This class has also been alternately referred to as Binary Input Output Symmetric (BIOS) [5].

[2] Superposition is optimal by taking either of the receivers as the weak receiver, since capacity region matches the time-division multiplexing region.

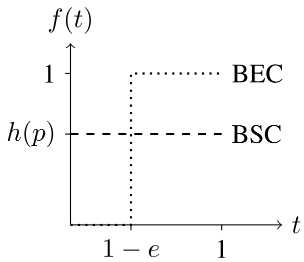


Fig. 2. BISO curves for $BSC(p)$ and $BEC(e)$.

**Definition 8** (Lorenz curve of a BISO channel)**.** For a BISO channel with BISO curve $f(t)$, the Lorenz[3] curve (or the cumulative function) $F(t)$ is defined as $F(t) = \int_0^t f(\tau)\mathrm{d}\tau$.

*Properties of the Lorenz curve:*

Since $0 \leq f(t) \leq 1$ and $f(t)$ is a non-decreasing step function on $[0, 1]$ we have

1) $F(t)$ is non-negative, piecewise linear and convex.
2) The slope of the line segments of $F(t)$ is at most 1.

We illustrate the Lorenz curves for a BSC, a BEC and a generic BISO channel having the same capacity in Figure 3.

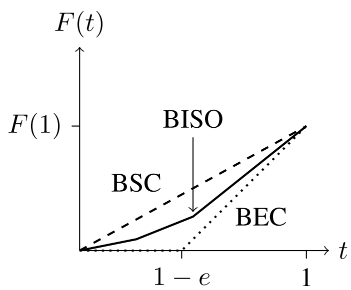


Fig. 3. Lorenz curves for a BSC, a BEC, and a generic BISO channel such that they have the same value at $t = 1$.

Denote $*$ as the binary convolution, that is $a * b := a(1 - b) + (1 - a)b$. Let $x = \mathrm{P}(X = 0)$, elementary calculations yield

$$\begin{aligned} &I(X;Y) \\ &= \sum_{k>0}(p_k + p_{-k})\left(h(x * \frac{p_k}{p_k + p_{-k}}) - h(\frac{p_k}{p_k + p_{-k}})\right) \quad (2) \\ &\stackrel{(a)}{=} \int_0^1 h(x * h^{-1}(f(\tau)))\mathrm{d}\tau - \int_0^1 f(\tau)\mathrm{d}\tau \end{aligned}$$

where $(a)$ follows from the definition of BISO curve. Thus for channels that have the same Lorenz curve, the mutual information and in particular the channel capacities, are the same regardless of the output alphabet sizes. Indeed the capacity, achieved by $x = 0.5$, is $C = 1 - F(1)$.

[3]The authors adopt this name from economics and is sometimes used for cumulative distribution functions obtained after rearrangement of terms from the smallest to the largest.

## II. On partial orders and capacity regions for classes of BISO broadcast channels

### A. On more capable comparability of BISO channels

We will establish a sufficient condition for two BISO channels to be more capable comparable. Before we state our sufficient condition for more capable comparability, we need the following lemmas.

**Lemma 1** (Mrs. Gerber's Lemma: Lemma 2 in [7])**.** *The function $h(x * h^{-1}(y))$ is convex in $y$ for $x \in [0, 1]$.*

**Lemma 2** (Lemma 1 in [8])**.** *Let $x_1, ..., x_l$ and $y_1, ..., y_l$ be nondecreasing sequences of real numbers. Let $\xi_1, ..., \xi_l$ be a sequence of real numbers such that*

$$\sum_{j=k}^{l} \xi_j x_j \geq \sum_{j=k}^{l} \xi_j y_j, \quad 1 \leq k \leq l$$

*with equality for $k = 1$. Then for any convex function $\Lambda$,*

$$\sum_{j=1}^{l} \xi_j \Lambda(x_j) \geq \sum_{j=1}^{l} \xi_j \Lambda(y_j).$$

*Remark* 5. The above inequality is related to majorization and one can trace it back to Result 108, Page 89 in [9].

**Theorem 1** (A sufficient condition)**.** *Consider BISO channels $[X \to Y_1]$ and $[X \to Y_2]$ with Lorenz curves $F(t)$ and $G(t)$, respectively. Further let $F(1) = G(1)$, i.e. channels have the same capacity. If $F \leq G$ then $Y_1$ is more capable than $Y_2$.*

*Proof.* Let the common refinement of these two BISO partitions be $\{t_k : k = 0, \dots, \hat{N}\}$, and $\xi_k := t_k - t_{k-1}$. Let the BISO curves be $f(t)$ and $g(t)$ respectively. Then

$$F(t_i) = \sum_{k=1}^{i} \xi_k f(t_k) \leq \sum_{k=1}^{i} \xi_k g(t_k) = G(t_i), \quad i = 1, \dots, \hat{N}.$$

Since $F(1) = G(1)$ we have equality at $i = \hat{N}$. Using Lemma 2 and by noticing that $f(t_k)$ and $g(t_k)$ are both nondecreasing we have

$$\sum_{j=1}^{\hat{N}} \xi_j \Lambda(f(t_j)) \geq \sum_{j=1}^{\hat{N}} \xi_j \Lambda(g(t_j))$$

for any convex function $\Lambda$. Taking $\Lambda(y) = h(x * h^{-1}(y)) - y$ we obtain that

$$\begin{aligned} &\sum_{j=1}^{\hat{N}} \xi_j h\big(x * h^{-1}(f(t_j))\big) - \sum_{j=1}^{\hat{N}} \xi_j f(t_j) \\ &\geq \sum_{j=1}^{\hat{N}} \xi_j h\big(x * h^{-1}(g(t_j))\big) - \sum_{j=1}^{\hat{N}} \xi_j g(t_j). \end{aligned}$$

From (2) this is equivalent to

$$I(X;Y_1) \geq I(X;Y_2), \forall p(x).$$

Thus the theorem is established. □

*Remark* 6. This is not, however, a necessary condition. Consider $\mathcal{Y}_1 = \mathcal{Y}_2 = \{-2,-1,1,2\}$, $p_k = \mathrm{P}(Y_1 = k|X=0)$, $q_k = \mathrm{P}(Y_2 = k|X=0)$, such that

$$p = [0.05,\ 0.2,\ 0.2,\ 0.55],$$
$$q = [0,\ 0.222434268,\ 0.509103154,\ 0.268462578].$$

One can verify that $I(X;Y_2) \geq I(X;Y_1)$, and the minimum value of $I(X;Y_2)/I(X;Y_1)$, is attained at $\mathrm{P}(X=0)=0.5$, and equals $1+3.5\times10^{-8}$. Now, append a channel $[Y_2 \to \tilde{Y}_2]$ so that $\mathrm{P}(\tilde{Y}_2 = k|X=i) = (1-e)\mathrm{P}(Y_2=k|X=i), i=0,1$ and $\mathrm{P}(\tilde{Y}_2 = E|X=0) = \mathrm{P}(\tilde{Y}_2 = E|X=1) = e$. Then, it is immediate that $I(X;\tilde{Y}_2) = (1-e)I(X;Y_2)$ for all $p(x)$. Choose $e$ to make the ratio of $I(X;\tilde{Y}_2)/I(X;Y_1)$, at $\mathrm{P}(X=0)=0.5$, to be one. Since the ratio $I(X;\tilde{Y}_2)/I(X;Y_1) = (1-e)I(X;Y_2)/I(X;Y_1)$ and hence the minimum of the ratio is still attained at $\mathrm{P}(X=0)=0.5$, and the new pair of channels remain remains more capable comparable. However, the Lorenz curves, shown in Figure 4, don't satisfy the sufficient condition in Theorem 1.

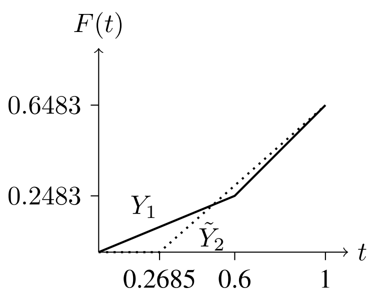


Fig. 4. Lorenz curves for two BISO channels with the same capacity and more capable comparable (see Remark 6).

For reasons that will be apparent later (cf. Lemma 4, Theorem 3, etc.) it is useful to shift our focus to the following subclass of BISO channels.

Let $\mathcal{C}(C)$ be the class of BISO channels with capacity $C$, and $BISO(C)$ denote an arbitrary BISO channel in this class.

For instance when $1-h(p) = C$, $BSC(p)$ belongs to this class. Similarly $BEC(e)$ belongs to this class when $1-e=C$. Using an abuse of notation, we denote $BSC(C)$ and $BEC(C)$ as the binary symmetric channel and the binary erasure channel with capacity $C$, respectively.

**Corollary 1.** $BISO(C) \overset{mc}{\geq} BSC(C)$.

*Proof.* From Theorem 1 it suffices that Lorenz curves satisfy $F_{BISO} \leq F_{BSC}$. Observe that $F_{BISO}(0) = F_{BSC}(0) = 0$, $F_{BISO}(1) = F_{BSC}(1)$, and $F_{BSC}(t)$ is the straight line connecting 0 and $F_{BSC}(1)$. The convexity of $F_{BISO}(t)$ (Property 1) implies that $F_{BISO} \leq F_{BSC}$. □

*Remark* 7. The least capable property of BSC was independently established in Chapter 7 of [10].

**Corollary 2.** $BEC(C) \overset{mc}{\geq} BISO(C)$.

*Proof.* Similar to above it suffices that the Lorenz curves satisfy $F_{BEC} \leq F_{BISO}$. Notice $F_{BEC}(t) = 0,\ t\in[0,1-e]$ hence $F_{BEC}(t) \leq F_{BISO}(t),\ t\in[0,1-e]$. Combining $F_{BEC}(1) = F_{BISO}(1)$ and by comparing slopes $f_{BEC}(t) = 1 \geq f_{BISO}(t),\ t\in(1-e,1]$, we have $F_{BEC} \leq F_{BISO}$. □

*1) Relation to information combining:* Some of the results, more precisely Corollaries 1 and 2, can be seen in the light of results in [11], [12]. For instance, from [12], when $[U \to X] \sim BSC(s)$, for a BISO, a BSC and a BEC that have the same capacity, one has

$$I(X;U,Y_{BSC}) \leq I(X;U,Y_{BISO}) \leq I(X;U,Y_{BEC}),$$

which then yields

$$I(X;Y_{BSC}|U) \leq I(X;Y_{BISO}|U) \leq I(X;Y_{BEC}|U).$$

But conditioning on $U$, where $[U \to X] \sim BSC(s)$ is the same, by symmetry, as taking $X \sim \mathrm{P}(X=0) = 1-s$. One could also obtain the same conclusion by using the results in [4]. However here we have used a different approach, via Theorem 1, to establish the extreme properties of BSC and BEC.

**Corollary 3.** *Let $BISO_1(C)$ and $BISO_2(C)$ be two BISO channels with output alphabet sizes at most 3. Then either $BISO_1(C) \overset{mc}{\geq} BISO_2(C)$ or $BISO_2(C) \overset{mc}{\geq} BISO_1(C)$, i.e. two such channels are always more capable comparable.*

*Proof.* For a BISO channel $[X \to Y]$ with transition probabilities $\{p_{-1}, p_0, p_1\}$, $Y=0$ is split equally into $0^+$ and $0^-$. Thus the Lorenz curve $F(t)$ contains two sloping lines: one with slope $h(\frac{p_{0^+}}{p_{0^+}+p_{0^-}}) = 1$, and the other not bigger than one. Note that for binary output case (i.e. BSC or $p_0 = 0$) the Lorenz curve is a straight line with slope at most one. Given two Lorenz curves of this kind, $F_1(t)$ and $F_2(t)$, with $F_1(1) = F_2(1)$, then either $F_1 \leq F_2$ or $F_1 \geq F_2$ (Figure 5). According to Theorem 1, these two channels are more capable comparable. □

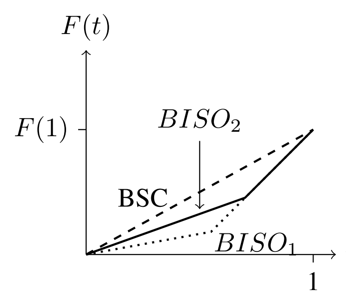


Fig. 5. Lorenz curves for BISO channels with the same capacity and output sizes at most 3.

*Remark* 8. Not all BISO channels with the same capacity are more capable comparable. A counter example is the following: Consider $\mathcal{Y}_1 = \mathcal{Y}_2 = \{-2,-1,1,2\}$, $p_k = \mathrm{P}(Y_1 = k|X=0)$, $q_k = \mathrm{P}(Y_2 = k|X=0)$, such that

$$p = [0.05,\ 0.2,\ 0.2,\ 0.55],$$
$$q = [0,\ 0.177713558,\ 0.732286442,\ 0.09].$$

One can verify that the capacities, attained at $\mathrm{P}(X=0)=0.5$, have a difference $I(X;Y_1) - I(X;Y_2) = 8.2\times10^{-10}$. Now similar to Remark 6, we append an erasure channel to $Y_1$ to get $\tilde{Y}_1$, such that $[X \to \tilde{Y}_1]$ and $[X \to Y_2]$ have the same capacity. Now for $I(X;\tilde{Y}_1) - I(X;Y_2)$, we have the value $0.001630643$ at $\mathrm{P}(X=0) = 0.1932$ and $-0.001678222$ at $\mathrm{P}(X=0) = 0.0177$. So they are not more capable comparable.

### B. On more capable and essentially less noisy orders in BISO channels

We first establish that there is a partial order induced by essentially less noisy comparison within the class of BISO broadcast channels. Further we will establish that, when restricted to $\mathcal{C}(C)$, the more capable and essentially less noisy partial orders are inverse of each other. It is worth noting that more capable and essentially less noisy are two notions of saying that one receiver is superior to another receiver, since superposition coding adapted to the (corresponding) stronger receiver is optimal in both cases.

Note that a BISO broadcast channel is a special case of *c-symmetric broadcast channel* considered in [4]. Thus the following result follows from Lemma 2 in [4].

**Claim 1** (Lemma 2 in [4]). *For a BISO broadcast channel, the uniform input distribution* $\mathrm{P}(X=0) = 0.5$ *forms a sufficient class.*

*Remark* 9. Since the uniform distribution forms a sufficient class for all BISO broadcast channels, it is immediate that the essentially less noisy comparison induces a partial order within the class of BISO channels.

**Lemma 3.** $BISO_1(C) \overset{mc}{\geq} BISO_2(C)$ *if and only if* $BISO_2(C) \overset{eln}{\geq} BISO_1(C)$.

*Proof.* Let the two channels be $[X \to Y_1]$ and $[X \to Y_2]$, respectively. Assume $Y_1 \overset{mc}{\geq} Y_2$. From Claim 1 we know that $\mathrm{P}(X=0) = 0.5$ is a sufficient distribution for these two channels. Therefore, when $\mathrm{P}(X=0) = 0.5$ we have for all $U$ such that $U \to X \to (Y_1, Y_2)$

$$\begin{aligned} I(U;Y_1) &= I(X;Y_1) - I(X;Y_1|U) \\ &= C - I(X;Y_1|U) \\ &= I(X;Y_2) - I(X;Y_1|U) \\ &= I(U;Y_2) + I(X;Y_2|U) - I(X;Y_1|U) \\ &\leq I(U;Y_2), \end{aligned}$$

where the last inequality follows from $Y_1 \overset{mc}{\geq} Y_2$. Thus we obtain that $Y_2 \overset{eln}{\geq} Y_1$.

Assume $Y_2 \overset{eln}{\geq} Y_1$. The proof follows by contradiction. Suppose there is a value $x$ such that when $\mathrm{P}(X=0) = x$, $I(X;Y_2) - I(X;Y_1) = \delta > 0$, then consider a $U$ such that $\mathrm{P}(U=0) = \mathrm{P}(U=1) = 0.5$, $\mathrm{P}(X=0|U=0) = x = \mathrm{P}(X=1|U=1)$. From the symmetry, $I(X;Y_2|U) - I(X;Y_1|U) = \delta > 0$. However since $\mathrm{P}(X=0) = 0.5$, using a similar decomposition we see that

$$\begin{aligned} I(U;Y_1) &= I(U;Y_2) + I(X;Y_2|U) - I(X;Y_1|U) \\ &= I(U;Y_2) + \delta > I(U;Y_2), \end{aligned}$$

contradicting the assumption $Y_2 \overset{eln}{\geq} Y_1$. Thus $Y_1 \overset{mc}{\geq} Y_2$. ☐

Recall our notation: $BSC(C)$ - a binary symmetric channel with capacity $C$; $BEC(C)$ - a binary erasure channel with capacity $C$; and $BISO(C)$ - an arbitrary binary input symmetric output channel with capacity $C$. The following lemma is an immediate consequence of Corollaries 1, 2, and Lemma 3.

**Lemma 4.** *The following relations hold:*

$(i)$ $BEC(C) \overset{mc}{\geq} BISO(C) \overset{mc}{\geq} BSC(C)$,
$(ii)$ $BSC(C) \overset{eln}{\geq} BISO(C) \overset{eln}{\geq} BEC(C)$.

This leads us to the following result.

**Theorem 2.** *For any three numbers* $0 \leq C_1 \leq C_2 \leq C_3$ *the following relations hold:*

$(i)$ $BEC(C_3) \overset{mc}{\geq} BISO(C_2) \overset{mc}{\geq} BSC(C_1)$,
$(ii)$ $BSC(C_3) \overset{eln}{\geq} BISO(C_2) \overset{eln}{\geq} BEC(C_1)$.

*Proof.* If $C_a < C_b$ then $BSC(C_a)$, $BEC(C_a)$ are degraded versions of $BSC(C_b)$, $BEC(C_b)$ respectively. Hence from Lemma 4

$$\begin{aligned} BEC(C_3) &\overset{mc}{\geq} BEC(C_2) \overset{mc}{\geq} BISO(C_2) \\ &\overset{mc}{\geq} BSC(C_2) \overset{mc}{\geq} BSC(C_1), \text{ and} \\ BSC(C_3) &\overset{eln}{\geq} BSC(C_2) \overset{eln}{\geq} BISO(C_2) \\ &\overset{eln}{\geq} BEC(C_2) \overset{eln}{\geq} BEC(C_1). \end{aligned}$$ ☐

*Remark* 10. In [4] the capacity region of a BSC/BEC broadcast channel was established. Corollary 4 generalizes this result to only requiring that one of the BISO channels is a BEC or a BSC.

## III. Comparison of inner and outer bounds for BISO channels

In this section we focus on BISO broadcast channels and consider inner bounds and outer bounds to the capacity region. The main result in this section is that if two BISO channels of the same capacity are not more capable comparable then the best known inner and outer bounds differ for the corresponding BISO broadcast channel.

The following are some commonly used inner bounds (or achievable rate regions) and outer bounds for the capacity region (CR):

- Time-division region (TD): This region is characterized by the set of points $(R_1, R_2)$ satisfying

  $$\begin{aligned} R_1 &\leq \alpha C_1 \\ R_2 &\leq (1-\alpha) C_2 \end{aligned}$$

  for some $\alpha \in [0,1]$, where $C_1$ and $C_2$ are the capacities of the two channels, respectively. The rates are achieved by transmitting at rate $C_1$ to the first receiver for fraction $\alpha$ of the time, and at rate $C_2$ to the second receiver for the remaining fraction.
- Randomized time-divison region (RTD): This corresponds to a time-division strategy except that the slots

for which communication occurs to one receiver is also drawn from a codebook which conveys additional information. This region is characterized by the set of points $(R_1, R_2)$ satisfying

$$\begin{aligned} R_1 &\le I(W;Y_1) + \mathrm{P}(W=0) I(X;Y_1|W=0) \\ R_2 &\le I(W;Y_2) + \mathrm{P}(W=1) I(X;Y_2|W=1) \\ R_1 + R_2 &\le \min\{I(W;Y_1), I(W;Y_2)\} \\ &\quad + \mathrm{P}(W=0) I(X;Y_1|W=0) \\ &\quad + \mathrm{P}(W=1) I(X;Y_2|W=1) \end{aligned}$$

for some random variables $(W, X)$ such that $W$ is binary and $W \to X \to (Y_1, Y_2)$ is Markov. The binary random variable $W$ characterizes the slots which distinguish communication to one receiver over the other.

- Marton's inner bound (MIB): This is the best known achievable rate region. The region is characterized by the set of rate pairs $(R_1, R_2)$ satisfying

$$\begin{aligned} R_1 &\le I(U, W; Y_1) \\ R_2 &\le I(V, W; Y_2) \\ R_1 + R_2 &\le \min\{I(W;Y_1), I(W;Y_2)\} \\ &\quad + I(U;Y_1|W) + I(V;Y_2|W) - I(U;V|W) \end{aligned}$$

for some $(U, V, W, X)$ such that $(U, V, W) \to X \to (Y_1, Y_2)$ is Markov. Observe that setting $U = X$, $V = \emptyset$ when $W = 0$ and $V = X$, $U = \emptyset$ when $W = 1$ reduces MIB to the RTD region.

**Lemma 5** ([13]). *For a binary input broadcast channel, the maximum sum rate implied by Marton's inner bound matches that of randomized time-divison region.*

- Outer bound (OB): The following region [6] represents an outer bound to the capacity region. The region is characterized by the set of rate pairs $(R_1, R_2)$ satisfying

$$\begin{aligned} R_1 &\le I(U;Y_1) \\ R_2 &\le I(V;Y_2) \\ R_1 + R_2 &\le I(U;Y_1) + I(X;Y_2|U) \\ R_1 + R_2 &\le I(V;Y_2) + I(X;Y_1|V) \end{aligned}$$

for some random variables $(U, V, X)$ such that $(U, V) \to X \to (Y_1, Y_2)$ is Markov.

*Remark* 11. For BISO channels since $\mathrm{P}(X = 0) = 0.5$ is a common sufficient distribution, it can be shown that this outer bound matches an earlier outer bound due to Körner and Marton [14].

It is clear that these regions satisfy the following relationship for any broadcast channel:

$$TD \subseteq RTD \subseteq MIB \subseteq CR \subseteq OB.$$

Another achievable region that we deal with in this paper is the superposition coding region [1].

- Superposition coding region (SC): This region is characterized by the set of rate pairs $(R_1, R_2)$ satisfying

$$\begin{aligned} R_1 &\le I(U;Y_1) \\ R_1 + R_2 &\le I(X;Y_2) \\ R_1 + R_2 &\le I(U;Y_1) + I(X;Y_2|U) \end{aligned}$$

for some random variables $(U, X)$ such that $U \to X \to (Y_1, Y_2)$ is Markov.

In the above representation, we treat $Y_2$ as the receiver capable of decoding the message for $Y_1$. One could also interchange the roles of the two receivers and obtain a similar region. It will be usually clear from context as to which of the two representations (or in other words, which of the two receivers plays the role of $Y_2$ above) we employ.

The following corollary to Theorem 2 is immediate.

**Corollary 4.** *Superposition coding region is the capacity region for a BISO broadcast channel if any one of the channels is either a BSC or a BEC.*

*Proof.* Superposition coding region is known to be optimal both for more capable comparable channels [15] and for essentially less noisy comparable channels [4]. From Theorem 2, if any one of the channels is either a BSC or a BEC, then the channels are either more capable comparable or essentially less noisy comparable. □

**Lemma 6.** *Consider a two-receiver broadcast channel where both $[X \to Y_1]$ and $[X \to Y_2]$ are BISO channels. Consider the following convex region formed by taking the union of rate pairs $(R_1, R_2)$ satisfying*

$$\begin{aligned} R_2 &\le I(U;Y_2) \\ R_2 + R_1 &\le I(U;Y_2) + I(X;Y_1|U) \\ R_1 &\le I(X;Y_1) \end{aligned} \tag{3}$$

*over all $(U, X)$ such that $U \to X \to (Y_1, Y_2)$ is Markov. Then any extreme point of this convex region can be realized by restricting to a binary $U$ such that $[U \to X] \sim BSC$ and $\mathrm{P}(X = 0) = 0.5$.*

*Proof.* The proof is presented in Appendix A. □

Consider a Markov chain $(U, V) \to X \to (Y_1, Y_2)$ such that $[U \to X] \sim BSC(s_1)$, $[V \to X] \sim BSC(s_2)$ and $\mathrm{P}(U = 0) = \mathrm{P}(V = 0) = 0.5$. Note this implies $\mathrm{P}(X = 0) = 0.5$.

Using this set of random variables define

$$\begin{aligned} f_1(s_1) &:= I(U;Y_1), \\ f_2(s_2) &:= I(V;Y_2). \end{aligned}$$

It is clear from symmetry that $f_1(s) = f_1(1 - s)$, $f_2(s) = f_2(1 - s)$. When $\mathrm{P}(X = 0) = s$ then note that $I(X;Y_1) = C - f_1(s)$ and $I(X;Y_2) = C - f_2(s)$. To see this, construct $[U \to X] \sim BSC(s)$ with $\mathrm{P}(U = 0) = 0.5$. Then

$$C = I(X;Y_1) = I(U;Y_1) + I(X;Y_1|U) = f_1(s) + I(X;Y_1|U)$$

From Lemma 6 and Remark 11 it follows that OB can be written as the union of rate pairs $(R_1, R_2)$ satisfying

$$\begin{aligned} R_1 &\le f_1(s_1) \\ R_2 &\le f_2(s_2) \\ R_1 + R_2 &\le f_1(s_1) + C - f_2(s_1) \\ R_1 + R_2 &\le f_2(s_2) + C - f_1(s_2) \end{aligned} \tag{4}$$

for some $0 \le s_1, s_2 \le 0.5$.

Let

$$I = \{s \in [0, 0.5] : f_1(s) > f_2(s)\},$$
$$J = \{s \in [0, 0.5] : f_1(s) < f_2(s)\}.$$

The following result about BISO channels with the same capacity, relates the equivalence of various bounds and their relation to whether the channels are more capable comparable.

**Theorem 3.** *For a BISO broadcast channel with channels $BISO_1(C)$ and $BISO_2(C)$, the following statements are equivalent:*

*(a) $BISO_1(C)$ and $BISO_2(C)$ are not more capable comparable*
*(b) $TD \subsetneq OB$*
*(c) There exists $s_1 \in I, s_2 \in J$ such that $f_1(s_1)+f_2(s_2) > C$*
*(d) $TD \subsetneq MIB$*
*(e) $MIB \subsetneq OB$.*

*Proof.* The proof of this equivalence is presented in Appendix B. □

**Corollary 5.** *For a BISO broadcast channel with channels $BISO_1(C)$ and $BISO_2(C)$ superposition coding region is optimal if and only if the channels are more capable comparable.*

*Proof.* Since both channels have the same capacity, the superposition coding region reduces to $\{(R_1, R_2) \in \Re_+^2 : R_1+R_2 \leq C\}$, i.e. the time-division region. Now the corollary is immediate from Theorem 3. □

*Remark* 12. A characterization of when superposition coding is optimal for two-receiver broadcast channels is open in general. It is known that superposition coding is optimal when the channels are either essentially more capable comparable or essentially less noisy comparable. However, it is not known whether such a comparison is necessary for superposition coding to be optimal.

*Remark* 13. From Remark 8 we know that there exists a pair of channels $BISO_1(C)$ and $BISO_2(C)$ which are not more capable comparable. Hence from Theorem 3 we know that the capacity region is strictly larger than TD. However, if we replace $BISO_2(C)$ by $BEC(C)$, a more capable channel, then the capacity of the broadcast channel formed by $BISO_1(C)$ and $BEC(C)$ is the TD region (Corollary 2 and the proof of Corollary 5). Thus replacing by a more capable channel can *strictly* reduce the capacity region.

This observation leads to an operational definition of a better receiver and a partial order as follows.

### A. A new partial order

We now introduce a natural operational partial order among broadcast channels.

**Definition 9.** Receiver $Z_2$ is a *better receiver* than $Y_2$ if the capacity region of broadcast channel $[X \to (Y_1, Z_2)]$ contains that of $[X \to (Y_1, Y_2)]$ for every channel $[X \to Y_1]$. In other words, if we replace receiver $Y_2$ by receiver $Z_2$ then the capacity region will not decrease.

*Remark* 14. Since the capacity region of a broadcast channel only depends on the marginal channels $[X \to Y_1]$ and $[X \to Y_2]$, the above operational partial order is well-defined.

From Remark 13 we know that a more capable receiver is not necessarily a better receiver. However we will show that a less noisy receiver is a better receiver.

**Proposition 1.** *If $Z_2$ is a less noisy receiver than $Y_2$, then $Z_2$ is a better receiver than $Y_2$.*

*Proof.* The capacity region of a discrete memoryless broadcast channel has the following $n$-letter characterization. Consider the region $\mathcal{R}_n$ defined as the closure of the union of rate pairs $(R_1, R_2)$ that satisfy

$$R_1 \leq \frac{1}{n} I(U; Y_{1,1}^n)$$
$$R_2 \leq \frac{1}{n} I(V; Y_{2,1}^n)$$

for some $p(u)p(v)p(x^n|u, v)$. It is known that the capacity region is $\cup_n \mathcal{R}_n$. (It is clear that this is achievable, and a converse follows by setting $U = M_1$ and $V = M_2$ and applying Fano's inequality.) Observe that for $j = n, \ldots, 1$

$$\begin{aligned}
&I(V; Y_{2,1}^j, Z_{2,j+1}^n) \\
&= I(V; Y_{2,1}^{j-1}, Z_{2,j+1}^n) + I(V; Y_{2j}|Y_{2,1}^{j-1}, Z_{2,j+1}^n) \\
&\leq I(V; Y_{2,1}^{j-1}, Z_{2,j+1}^n) + I(V; Z_{2j}|Y_{2,1}^{j-1}, Z_{2,j+1}^n) \\
&= I(V; Y_{2,1}^{j-1}, Z_{2,j}^n).
\end{aligned}$$

By taking the extreme points of this chain we obtain that $I(V; Y_{2,1}^n) \leq I(V; Z_{2,1}^n)$. The proposition follows from the expression of the capacity region stated above. □

## IV. Conclusion

We look at partial orders induced by the more capable and less noisy relations in binary input symmetric output broadcast channels. We establish the capacity regions for a class of them and also show various other results related to the evaluation of various bounds. We also show the optimality of certain auxiliary channels, thus generalizing earlier results.

## V. Acknowledgements

The authors are thankful for valuable suggestions from the anonymous reviewers as well as Young-Han Kim, the associate editor.

## References

[1] T. Cover, "Broadcast channels," *IEEE Trans. Info. Theory*, vol. IT-18, pp. 2–14, January, 1972.
[2] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2012.
[3] J. Körner and K. Marton, "Comparison of two noisy channels," *Topics in Inform. Theory(ed. by I. Csiszar and P.Elias), Keszthely, Hungary*, pp. 411–423, August, 1975.
[4] C. Nair, "Capacity regions of two new classes of two-receiver broadcast channels," *Information Theory, IEEE Transactions on*, vol. 56, no. 9, pp. 4207–4214, sep. 2010.
[5] A. Viterbi and J. Omura, *Principles of digital communication and coding*, ser. McGraw-Hill series in electrical engineering. McGraw-Hill, 1979. [Online]. Available: http://books.google.com.hk/books?id=8fdSAAAAMAAJ

[6] C. Nair and A. El Gamal, "An outer bound to the capacity region of the broadcast channel," *IEEE Trans. Info. Theory*, vol. IT-53, pp. 350–355, January, 2007.
[7] A. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications: Part I," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 6, pp. 769–772, Nov 1973.
[8] B. Hajek and M. Pursley, "Evaluation of an achievable rate region for the broadcast channel," *IEEE Trans. Info. Theory*, vol. IT-25, pp. 36–46, January, 1979.
[9] G. Hardy, J. Littlewood, and G. Pólya, *Inequalities*, ser. Cambridge Mathematical Library. Cambridge University Press, 1952. [Online]. Available: http://books.google.com.hk/books?id=t1RCSP8YKt8C
[10] E. Sasoglu, "Polar Coding Theorems for Discrete Systems," Ph.D. dissertation, IC, Lausanne, 2011.
[11] I. Land, S. Huettinger, P. Hoeher, and J. Huber, "Bounds on information combining," *Information Theory, IEEE Transactions on*, vol. 51, no. 2, pp. 612–619, 2005.
[12] I. Sutskover, S. Shamai, and J. Ziv, "Extremes of information combining," *Information Theory, IEEE Transactions on*, vol. 51, no. 4, pp. 1313–1325, April 2005.
[13] C. Nair, Z. V. Wang, and Y. Geng, "An information inequality and evaluation of Marton's inner bound for binary input broadcast channels," *International Symposium on Information Theory*, 2010.
[14] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Info. Theory*, vol. IT-25, pp. 306–311, May, 1979.
[15] A. El Gamal, "The capacity of a class of broadcast channels," *IEEE Trans. Info. Theory*, vol. IT-25, pp. 166–169, March, 1979.

## Appendix

### A. Proof to Lemma 6

*Proof.* For each $(U, X)$, construct $(\tilde{U}, \tilde{X}) = ((U, Q), \tilde{X})$ such that $Q \sim \text{Bern}(\frac{1}{2})$ and

$$\begin{aligned} \mathrm{P}(\tilde{U} = (u,0)) = \mathrm{P}(\tilde{U} = (u,1)) = \frac{1}{2}\mathrm{P}(U = u) \\ \mathrm{P}(\tilde{X} = 0|\tilde{U} = (u,0)) = \mathrm{P}(X = 0|U = u) \\ \mathrm{P}(\tilde{X} = 0|\tilde{U} = (u,1)) = 1 - \mathrm{P}(X = 0|U = u) \end{aligned}$$

Denote this class of $(\tilde{U}, \tilde{X})$ as $\mathcal{Q}$. Notice that $\mathrm{P}(\tilde{X} = 0) = 0.5$ and by symmetry we have

$$\begin{aligned} I(\tilde{U}; \tilde{Y}_2) \geq I(U; \tilde{Y}_2|Q) = I(U; Y_2), \\ I(\tilde{X}; \tilde{Y}_1|\tilde{U}) = I(X; Y_1|U), \\ I(\tilde{X}; \tilde{Y}_1) \geq I(X; Y_1). \end{aligned}$$

Thus for every $(U, X)$, replacing it with $(\tilde{U}, \tilde{X})$ only enlarges the region given by (3). Thus a uniform distribution on $X$ is sufficient. We proceed to show that taking $[U \to X] \sim BSC$ is sufficient.

From above, it suffices to maximize over all $(U, X) \in \mathcal{Q}$. Since $\mathrm{P}(X = 0) = 0.5$ is fixed, the third inequality remains constant. Therefore, to compute the extreme points, we proceed to compute the distribution $(U, X) \in \mathcal{Q}$ that maximizes $\lambda I(U; Y_2) + \big(I(U; Y_2) + I(X; Y_1|U)\big)$. Rewrite the expression as

$$(\lambda + 1)I(X; Y_2) + I(X; Y_1|U) - (\lambda + 1)I(X; Y_2|U).$$

Let $f(x) = I(X; Y_1) - (\lambda+1)I(X; Y_2)$, where $x = \mathrm{P}(X = 0)$. Notice that $f(x) = f(1-x)$; and let $x_\lambda$ and x$1-x_\lambda$ $1-x_\lambda$ (by symmetry) maximize $f(x)$. Construct $[U \to X] \sim BSC(x_\lambda)$, then $I(X; Y_1|U) - (\lambda + 1)I(X; Y_2|U)$ is maximized; let $U \sim \text{Bern}(\frac{1}{2})$, then $I(X; Y_2)$ is maximized since $\mathrm{P}(X = 0) = 0.5$. Notice this construction falls into class $\mathcal{Q}$, hence the proof is finished. □

The same proof can also be used to establish the following lemma.

**Lemma 7.** *Consider a two-receiver broadcast channel where both $[X \to Y_1]$ and $[X \to Y_2]$ are BISO channels. Consider the following superposition coding region formed by taking the union of rate pairs $(R_1, R_2)$ satisfying*

$$\begin{aligned} R_2 &\leq I(U; Y_2) \\ R_2 + R_1 &\leq I(U; Y_2) + I(X; Y_1|U) \\ R_2 + R_1 &\leq I(X; Y_1) \end{aligned}$$

*over all $(U, X)$ such that $U \to X \to (Y_1, Y_2)$ is Markov. Then the extreme points of the region can be realized by restricting to a binary $U$ such that $[U \to X] \sim BSC(s)$ and $\mathrm{P}(X = 0) = 0.5$.*

*Remark* 15. This generalizes a result by Wyner and Ziv [7] for BSC broadcast channels. In [15] it was shown that superposition coding is optimal when the two channels are more capable comparable. Hence the extreme points of the capacity region for any more capable comparable BISO broadcast channel can be obtained by setting $[U \to X] \sim BSC(s)$ and $U$ to be uniformly distributed.

### B. Proof to Theorem 3

*Proof.* (a) ⇒ (b): As defined earlier, let

$$\begin{aligned} I = \{s \in [0, 0.5] : f_1(s) > f_2(s)\}, \\ J = \{s \in [0, 0.5] : f_1(s) < f_2(s)\}. \end{aligned}$$

Since the channels are not more capable comparable, we know that $I$ and $J$ are both non-empty. Let $s_1 \in I, s_2 \in J$ be two points from these two sets. Construct $(\tilde{U}, X)$, where $\tilde{U} = (U', Q)$ with binary $U'$ and $Q$, and probabilities

$$\begin{aligned} &\mathrm{P}(\tilde{U} = (0,0)) = \frac{1-\varepsilon}{2} && \mathrm{P}(X = 0|\tilde{U} = (0,0)) = 1 \\ &\mathrm{P}(\tilde{U} = (0,1)) = \frac{\varepsilon}{2} && \mathrm{P}(X = 0|\tilde{U} = (0,1)) = s_1 \\ &\mathrm{P}(\tilde{U} = (1,0)) = \frac{1-\varepsilon}{2} && \mathrm{P}(X = 1|\tilde{U} = (1,0)) = 1 \\ &\mathrm{P}(\tilde{U} = (1,1)) = \frac{\varepsilon}{2} && \mathrm{P}(X = 1|\tilde{U} = (1,1)) = s_1. \end{aligned}$$

Thus, $[U' \mapsto X] \sim BSC(0)$ conditioned on the event $Q = 0$, $[U' \mapsto X] \sim BSC(1 - s_1)$ conditioned on $Q = 1$, and further $U'$ is independent of $Q$ with $\mathrm{P}(U' = 0) = 0.5$. We can see that $Q$ is independent of $X$ and hence of $Y_1, Y_2$; thus $I(Q; Y_1) = I(Q; Y_2) = 0$. Now

$$\begin{aligned} I(\tilde{U}; Y_1) &= I(U', Q; Y_1) = I(U'; Y_1|Q) + I(Q; Y_1) \\ &= I(U'; Y_1|Q) \\ &= (1 - \varepsilon)I(X; Y_1) + \varepsilon I(U'; Y_1|Q = 1) \\ &= (1 - \varepsilon)C + \varepsilon f_1(s_1). \end{aligned}$$

Similarly, we obtain

$$I(\tilde{U}; Y_2) = (1 - \varepsilon)C + \varepsilon f_2(s_1).$$

Let $[V \to X] \sim BSC(s_2)$. Thus OB reduces to

$$\begin{aligned} R_1 &\leq (1-\varepsilon)C + \varepsilon f_1(s_1) \\ R_2 &\leq f_2(s_2) \\ R_1 + R_2 &\leq I(\tilde{U};Y_1) + I(X;Y_2|\tilde{U}) \\ &= I(\tilde{U};Y_1) + I(X;Y_2) - I(\tilde{U};Y_2) \\ &= (1-\varepsilon)C + \varepsilon f_1(s_1) + C - (1-\varepsilon)C - \varepsilon f_2(s_1)] \\ &= C + \varepsilon(f_1(s_1) - f_2(s_1)) \qquad (> C) \\ R_1 + R_2 &\leq I(V;Y_2) + I(X;Y_1|V) \\ &= f_2(s_2) + C - f_1(s_2) \qquad (> C). \end{aligned}$$

To show that we can have $(1-\varepsilon)C+\varepsilon f_1(s_1)+f_2(s_2) > C$, we just need to choose small $\varepsilon$ to ensure $f_2(s_2) > \varepsilon(C - f_1(s_1))$. Since this is clearly possible, we have $OB \supsetneq TD$.

(b) $\Rightarrow$ (c): Let $[U \to X] \sim BSC(s_1)$ and $[V \to X] \sim BSC(s_2)$. From Equation (4), we have the following expression of the boundary of the outer bound,

$$\begin{aligned} R_1 &\leq I(U;Y_1) = f_1(s_1) \\ R_2 &\leq I(V;Y_2) = f_2(s_2) \\ R_1 + R_2 &\leq I(U;Y_1) + I(X;Y_2|U) = f_1(s_1) + C - f_2(s_1) \\ R_1 + R_2 &\leq I(V;Y_2) + I(X;Y_1|V) = f_2(s_2) + C - f_1(s_2) \end{aligned}$$

Clearly for every $s_1 \in I, s_2 \in J$ if $f_1(s_1) + f_2(s_2) \leq C$ then from above $OB = TD$. However since $OB \supsetneq TD$, there exists $s_1 \in I, s_2 \in J$ such that $f_1(s_1) + f_2(s_2) > C$.

(c) $\Rightarrow$ (d): In general, $TD \subseteq RTD \subseteq MIB$. So now it suffices to show there exists an example where the sum rate of the RTD region is strictly larger than the TD region.

We now compute the maximum sum rate of the RTD region. From Lemma 5 we know that this matches the maximum sum rate of the MIB region.

Consider an auxiliary channel $W \to X$ such that

$$\begin{aligned} \mathrm{P}(W=0) = a, &\quad \mathrm{P}(W=1) = 1-a \\ \mathrm{P}(X=0|W=0) = s_2, &\quad \mathrm{P}(X=1|W=1) = s_1 \end{aligned}$$

where $as_2 + (1-a)(1-s_1) = 0.5$. Clearly $a \notin \{0,1\}$ since $s_1, s_2 < 0.5$ as at $s = 0.5$ we have $I(X;Y_1) = I(X;Y_2) = C$.

It is straightforward to check the following

$$\begin{aligned} I(X;Y_1|W{=}0) &= C - f_1(s_2), \\ I(X;Y_1|W{=}1) &= C - f_1(s_1), \\ I(X;Y_2|W{=}0) &= C - f_2(s_2), \\ I(X;Y_2|W{=}1) &= C - f_2(s_1), \\ I(X;Y_1) &= I(X;Y_2) = C. \end{aligned}$$

Then observe that

$$\begin{aligned} &I(W;Y_1) + \mathrm{P}(W=0)I(X;Y_1|W=0) \\ &\quad + \mathrm{P}(W=1)I(X;Y_2|W=1) \\ &= I(X;Y_1) + \mathrm{P}(W{=}1)\big(I(X;Y_2|W{=}1) - I(X;Y_1|W{=}1)\big) \\ &= C + (1-a)(f_1(s_1) - f_2(s_1)) \end{aligned}$$

Similarly

$$\begin{aligned} &I(W;Y_2) + \mathrm{P}(W=0)I(X;Y_1|W=0) \\ &\quad + \mathrm{P}(W=1)I(X;Y_2|W=1) \\ &= C + a(f_2(s_2) - f_1(s_2)). \end{aligned}$$

Therefore the sum rate of RTD (equivalently that of MIB) for this choice of $(W,X)$ is given by

$$C + \min\{(1-a)(f_1(s_1) - f_2(s_1)), a(f_2(s_2) - f_1(s_2))\}. \quad (5)$$

Therefore if $(c)$ is satisfied, i.e. there exists $s_1 \in I, s_2 \in J$, then there exists a $(W,X)$ so that equation (5) gives a sum rate strictly larger than $C$.

*Remark* 16. A careful reader will notice that the above argument only requires $s_1 \in I, s_2 \in J$ and does not even require $f_1(s_1) + f_2(s_2) > C$. But the existence of any $s_a \in I, s_b \in J$ will imply that $(a)$ holds and hence $(c)$ holds.

(d) $\Rightarrow$ (e): The maximum sum rate of MIB is achieved using RTD (Lemma 5). Since $TD \subsetneq MIB$, to compute the maximum sum rate of MIB, it suffices to maximize over $s_1 \in I$, $s_2 \in J$, $0 < a < 1$ the term

$$C + \min\{(1-a)(f_1(s_1) - f_2(s_1)), a(f_2(s_2) - f_1(s_2))\}. \quad (6)$$

Consider any triple $s_1 \in I, s_2 \in J, 0 < a < 1$. Pick any $\varepsilon > 0$ small enough (will show later how small we require it).

Define $(U,X) = (Q,U_1,X)$ where $\mathrm{P}(Q=0) = 1-a+\varepsilon, \mathrm{P}(Q=1) = a-\varepsilon$; and $[U_1 \mapsto X] \sim BSC(s_1)$ conditioned on $Q=0$, and $[U_1 \mapsto X] \sim BSC(0)$ conditioned on $Q=1$. Further take $\mathrm{P}(U_1=0|Q=0) = \mathrm{P}(U_1=0|Q=1) = 0.5$. Observe that this induces $\mathrm{P}(X=0) = \mathrm{P}(X=1) = 0.5$.

Similarly define $(V,X) = (Q',V_1,X)$ where $\mathrm{P}(Q'=0) = a+\varepsilon$, $\mathrm{P}(Q'=1) = 1-a-\varepsilon$; and $[V_1 \mapsto X] \sim BSC(s_2)$ conditioned on $Q'=0$, and $[V_1 \mapsto X] \sim BSC(0)$ conditioned on $Q'=1$. Further take $\mathrm{P}(V_1=0|Q'=0) = \mathrm{P}(V_1=0|Q'=1) = 0.5$. Observe that this also induces $\mathrm{P}(X=0) = \mathrm{P}(X=1) = 0.5$.

Since the distribution of $X$ is consistent there exists a triple $(U,V,X)$ with the same pairwise marginals $(U,X)$ and $(V,X)$ as described earlier. With this choice, OB reduces to

$$\begin{aligned} R_1 &\leq I(U;Y_1) = (1-a+\varepsilon)f_1(s_1) + (a-\varepsilon)C \\ R_2 &\leq I(V;Y_2) = (a+\varepsilon)f_2(s_2) + (1-a-\varepsilon)C \\ R_1 + R_2 &\leq I(U;Y_1) + I(X;Y_2|U) \\ &= C + (1-a+\varepsilon)(f_1(s_1) - f_2(s_1)) \\ R_1 + R_2 &\leq I(V;Y_2) + I(X;Y_1|V) \\ &= C + (a+\varepsilon)(f_2(s_2) - f_1(s_2)). \end{aligned}$$

Clearly the maximum sum rate of the above region is minimum of the terms

$$\begin{aligned} &\{C + (1-a+\varepsilon)(f_1(s_1) - f_2(s_1)), \\ &C + (a+\varepsilon)(f_2(s_2) - f_1(s_2)), \\ &(1-2\epsilon)C + (1-a+\varepsilon)f_1(s_1) + (a+\varepsilon)f_2(s_2)\}. \end{aligned} \quad (7)$$

We pick $\varepsilon > 0$ to satisfy

$$\begin{aligned}&(1-2\epsilon)C + (1-a+\varepsilon)f_1(s_1) + (a+\varepsilon)f_2(s_2)\\&\quad > C + (1-a)(f_1(s_1) - f_2(s_1)), \quad \text{or equivalently}\\&(1-a)f_2(s_1) + af_2(s_2) > \varepsilon(2C - f_1(s_1) - f_2(s_2)),\end{aligned}$$

and also satisfy

$$af_1(s_2) + (1-a)f_1(s_1) > \varepsilon(2C - f_1(s_1) - f_2(s_2)).$$

The above two constraints imply that the third term in (7) is strictly larger than both the terms in (6). Comparing the first two terms in (7) to those in (6) it is immediate that the sum rate of the OB expression (7) will be strictly bigger than that of MIB region (6). Since this is possible for every $s_1 \in I, s_2 \in J, 0 < a < 1$, the maximum sum rate of OB is strictly larger than that of MIB. Since $TD \subsetneq MIB$ the maximum of $MIB$ is not achieved when $a \in \{0,1\}$. Therefore $OB \supsetneq MIB$ or $(e)$ holds.

(e) ⇒ (a): $MIB \subsetneq OB$ clearly implies the channels are not more capable comparable. This is because when the channels are more capable comparable we know from [15] that superposition coding is optimal and that $MIB = CR = OB$. □

**Yanlin Geng** Yanlin Geng received his B.Sc. (mathematics) and M.Eng. (signal and information processing) from Peking University, and Ph.D. (information engineering) from The Chinese University of Hong Kong in 2006, 2009, and 2012, respectively. He is currently a postdoctoral researcher in the Information Engineering department at The Chinese University of Hong Kong.

**Shlomo Shamai** Shlomo Shamai (Shitz) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, in 1975, 1981 and 1986 respectively.

During 1975-1985 he was with the Communications Research Labs, in the capacity of a Senior Research Engineer. Since 1986 he is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, where he is now a Technion Distinguished Professor, and holds the William Fondiller Chair of Telecommunications. His research interests encompasses a wide spectrum of topics in information theory and statistical communications.

Dr. Shamai (Shitz) is an IEEE Fellow, a member of the Israeli Academy of Sciences and Humanities and a Foreign Associate of the US National Academy of Engineering. He is the recipient of the 2011 Claude E. Shannon Award. He has been awarded the 1999 van der Pol Gold Medal of the Union Radio Scientifique Internationale (URSI), and is a co-recipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003, and the 2004 joint IT/COM societies paper award, the 2007 IEEE Information Theory Society Paper Award, the 2009 European Commission FP7, Network of Excellence in Wireless COMmunications (NEWCOM++) Best Paper Award, and the 2010 Thomson Reuters Award for International Excellence in Scientific Research. He is also the recipient of 1985 Alon Grant for distinguished young scientists and the 2000 Technion Henry Taub Prize for Excellence in Research. He has served as Associate Editor for the Shannon Theory of the IEEE Transactions on Information Theory, and has also served twice on the Board of Governors of the Information Theory Society. He is a member of the Executive Editorial Board of the IEEE Transactions on Information Theory

**Chandra Nair** Chandra Nair received his Bachelor of Technology (B.Tech) degree in Electrical Engineering from the Indian Institute of Technology (IIT), Madras in 1999. Concurrently, he also completed a four year nurture program in Mathematics at the Institute of Mathematical Sciences (IMSc) under the auspices of the National Board of Higher Mathematics (NBHM). He received a Masters (2002) and PhD (2005) in electrical engineering from Stanford University. Subsequently he was a postdoctoral fellow at the theory group in Microsoft Research (Redmond) for two years. Following this he joined the IE department, CUHK, as an assistant professor in Fall 2007. His research interests are on fundamental problems in various interdisciplinary pursuits involving information theory, combinatorial optimization, statistical physics, and algorithms.

**Zizhou Vincent Wang** Vincent Wang is a system engineer at Altai Technologies located in the Hong Kong Science and Technology Park. He obtained his PhD from the department of Information Engineering at the Chinese University of Hong Kong in 2010. His research work mainly consisted of studying inner and outer bounds for two and three receiver broadcast channels.