Capacity Regions of Two New Classes of Two-Receiver Broadcast Channels

Chandra Nair, *Member, IEEE*

***Abstract*—Motivated by a simple broadcast channel, we generalize the notions of a *less noisy* receiver and a *more capable* receiver to an *essentially less noisy receiver* and an *essentially more capable* receiver, respectively. We establish the capacity regions of these classes by borrowing on existing techniques; however, these new classes contain additional interesting classes of broadcast channels, including the BSC/BEC broadcast channel. We also establish the relationships between the new classes and the existing classes.**

***Index Terms*—Broadcast channel, capacity region, superposition coding.**

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* (BC) consists of a finite input alphabet $\mathcal{X}$, finite output alphabets $\mathcal{Y}_1$, $\mathcal{Y}_2$, and a probability transition function $p(y_1, y_2 \,|\, x)$. A $((2^{nR_1}, 2^{nR_2}), n)$ code for a broadcast channel consists of an encoder

$$X^n : \mathcal{W}_1 \times \mathcal{W}_2 \to \mathcal{X}^n$$

and two decoders

$$\hat{W}_1 : \mathcal{Y}_1^n \to \mathcal{W}_1$$
$$\hat{W}_2 : \mathcal{Y}_2^n \to \mathcal{W}_2$$

where $\mathcal{W}_1 = \{1, 2, \ldots, 2^{nR_1}\}, \mathcal{W}_2 = \{1, 2, \ldots, 2^{nR_2}\}$.

The probability of error $\mathrm{P}_e^{(n)}$ is defined to be the probability that the decoded message is not equal to the transmitted message, i.e.,

$$\mathrm{P}_e^{(n)} = \mathrm{P}\left(\left\{\hat{W}_1\left(Y_1^n\right) \neq W_1\right\} \cup \left\{\hat{W}_2\left(Y_2^n\right) \neq W_2\right\}\right)$$

where the message pair (W_1, W_2) is assumed to be uniformly distributed over $\mathcal{W}_1 \times \mathcal{W}_2$.

A rate pair (R_1, R_2) is said to be *achievable* for the broadcast channel if there exists a sequence of $((2^{nR_1}, 2^{nR_2}), n)$ codes with $\mathrm{P}_e^{(n)} \to 0$. The *capacity region* of the broadcast channel is the closure of the set of achievable rates. *The capacity region of the two user discrete memoryless channel is unknown*.

A. Motivation

This paper is motivated directly by a simple broadcast channel setting, posed to the author by Montanari (see Fig. 1), consisting of a BSC(p) and BEC(e). As we shall see below, the capacity of this simple channel was not known under certain regimes of (e, p). In this paper, we establish the capacity region for all values of parameters, $0 \leq e, p \leq 1$.

B. Notions of Dominance: A Review

One may say that a receiver is dominant compared to the other receiver if superposition coding achieves the capacity region. Various notions of dominance have been considered earlier in the literature and a brief summary of them is presented below.

1) Degraded Receiver: A receiver Y_2 is said to be a degraded version [1] of the receiver Y_1 if there exists a probability transition matrix $p(y_2 \,|\, y_1)$ such that $p(y_2 \,|\, x) = \sum_{y_1} p(y_2 \,|\, y_1) p(y_1 \,|\, x)$.

The capacity region of a degraded broadcast channel is given [2], [3] by the union of rate pairs (R_1, R_2) satisfying

$$R_2 \leq I(U; Y_2)$$
$$R_1 + R_2 \leq I(U; Y_2) + I(X; Y_1 \,|\, U)$$

over all (U, X) such that $U \to X \to (Y_1, Y_2)$ form a Markov chain.

2) Less-Noisy Receiver: A receiver Y_2 is said to be less-noisy [4] than receiver Y_1 if for every $U \to X \to (Y_1, Y_2)$ the inequality $I(U; Y_1) \geq I(U; Y_2)$ holds.

Remark 1: An equivalent definition (Theorem 2 in [5]): A receiver Y_2 is less noisy than a receiver Y_1 if and only if the function $I(X; Y_1) - I(X; Y_2)$ is concave in the input distribution, $p(x)$.

Clearly, from the data-processing inequality, a degraded receiver is also a less-noisy receiver but not necessarily vice-versa [4].

The capacity region of a less-noisy broadcast channel is given [4] by the union of rate pairs (R_1, R_2) satisfying

$$R_2 \leq I(U; Y_2)$$
$$R_1 + R_2 \leq I(U; Y_2) + I(X; Y_1 \,|\, U)$$

over all (U, X) such that $U \to X \to (Y_1, Y_2)$ form a Markov chain.

3) More-Capable Receiver: A receiver Y_2 is said to be more-capable [4] than the receiver Y_1 if for every $p(x)$ the inequality $I(X; Y_1) \geq I(X; Y_2)$ holds.

Clearly less-noisy receiver is also a more-capable receiver but not necessarily vice-versa (see Ahlswede's example in [4]).

The capacity region of a more-capable broadcast channel is given [6] by the union of rate pairs (R_1, R_2) satisfying

$$R_2 \leq I(U; Y_2)$$
$$R_1 + R_2 \leq I(U; Y_2) + I(X; Y_1 \,|\, U)$$
$$R_1 + R_2 \leq I(X; Y_1)$$

Manuscript received April 07, 2009; revised April 14, 2010. Date of current version August 18, 2010. This work was supported in part by the Direct Grant for research at the Chinese University of Hong Kong.

The author is with the Department of Information Engineering, Chinese University of Hong Kong, Sha Tin, N.T. (e-mail: chandra.nair@gmail.com).

Communicated by E. Erkip, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2010.2054310

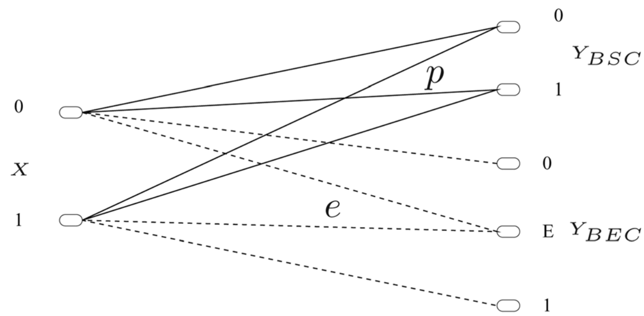

Fig. 1. Broadcast channel consisting of a BSC(p) and BEC(e).

over all (U, X) such that $U \to X \to (Y_1, Y_2)$ form a Markov chain.

C. Inner and Outer Bound

1) Superposition Coding Inner Bound: ([1], [7]) The union of rate pairs (R_1, R_2) satisfying

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

over all (U, X) such that $U \to X \to (Y_1, Y_2)$ form a Markov chain is achievable.

2) U, V Outer Bound: ([6], [8]) The union of rate pairs (R_1, R_2) satisfying

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

over all (U, V, X) such that $(U, V) \to X \to (Y_1, Y_2)$ form a Markov chain constitutes an outer bound to the capacity region.

There is a very similar outer bound that can be found in [7], in which only one of U or V is present. However, that representation is sometimes strictly worse than the above representation. See [8] for an example.

3) An Observation: Suppose we wish to show that the superposition coding inner bound is indeed the capacity region, then one could do so by showing that the U, V outer bound is contained in the inner bound, i.e., the regions coincide.

Remark 2: If one could obtain an additional constraint that $R_1 + R_2 \leq I(X; Y_1)$, then the (U, V) bound will have more constraints than the achievable region, and hence, will be contained in the inner bound. For the degraded and the less-noisy cases observe that $I(U; Y_2) \leq I(U; Y_1)$, and hence, the inequality $R_1 + R_2 \leq I(U; Y_2) + I(X; Y_1 \,|\, U)$ implies that $R_1 + R_2 \leq I(U; Y_1) + I(X; Y_1 \,|\, U) = I(X; Y_1)$ (here we use the fact that $U \to X \to (Y_1, Y_2)$ is Markov). For the more capable case, observe that $I(X; Y_2 \,|\, V) \leq I(X; Y_1 \,|\, V)$, and hence, the inequality $R_1 + R_2 \leq I(V; Y_1) + I(X; Y_2 \,|\, V)$ implies that $R_1 + R_2 \leq I(V; Y_1) + I(X; Y_1 \,|\, V) = I(X; Y_1)$.

In general, to show that the U, V outer bound is contained in the inner bound, it suffices (since the region is convex and closed) to show that the points on the boundary of the U, V outer bound is contained in the inner bound. Suppose the distributions $p(x)$ that achieve the boundary points of the outer bound satisfy ordering properties like $I(U; Y_2) \leq I(U; Y_1)$ or $I(X; Y_2 \,|\, V) \leq I(X; Y_1 \,|\, V)$, it suffices as one could use the same arguments as in the above remark to show the equivalence. This is exactly what is done in this paper. So technically, it is a reasonably simple observation; yet it establishes the capacity regions of additional interesting classes of broadcast channels.

D. Organization of the Paper

In the next section, we present the various definitions and also establish the capacity region of the two new classes of broadcast channels that we introduce. In Section III, we analyze and compute the capacity region of the BSC/BEC broadcast channel for various settings of the parameters. Finally in Section IV, we compare inclusion relationships among the various classes of channels where one receiver is dominant than the other.

II. Main

A. Definitions and Notation

Definition 1: 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 two-receiver broadcast channel if the following holds: Given any triple of random variables (U, V, X) distributed[1] according to $p(u, v, x)$, there exists a distribution $q(\tilde{u}, \tilde{v}, x)$ that satisfies

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

The notation $I(U; Y_1)_p$ denotes the mutual information between U and Y_1 when the input is generated using $p(u, v, x)$.

Remark 3: It follows from the definition of a sufficient class $\mathcal{P}$ that to compute the boundary points of the (U, V)-outer bound it suffices to consider the union over $(U, V) - X - (Y_1, Y_2)$, where $q(x) \in \mathcal{P}$. By setting $V = X$, we also observe that to compute the boundary points of the

[1] In all cases, we assume that the tuple (U, V, X, Y_1, Y_2) satisfies $(U, V) \to X \to (Y_1, Y_2)$ forms a Markov chain. In a discrete memoryless broadcast channel with no feedback, this assumption is "automatically" satisfied. However, it is necessary to state it explicitly to prevent choices like $U = Y_1$ (except when $X \to Y_1$ is deterministic) and other strange choices.

superposition coding inner bound as well, it suffices to consider the union over $U - X - (Y_1, Y_2)$, where $q(x) \in \mathcal{P}$.

Definition 2: A receiver Y_1 is an *essentially less noisy* compared to receiver Y_2 if there exists a sufficient class of distributions $\mathcal{P}$ such that whenever $p(x) \in \mathcal{P}$, for all $U \to X \to (Y_1, Y_2)$, we have

$$I(U;Y_2) \leq I(U;Y_1).$$

Remark 4: Setting $\mathcal{P}$ to be the entire set of distributions $p(x)$ shows that a less noisy receiver is in particular an essentially less noisy receiver.

Definition 3: A receiver Y_1 is *essentially more capable* than receiver Y_2 if there exists a sufficient class of distributions $\mathcal{P}$ such that whenever $p(x) \in \mathcal{P}$, for all $U \to X \to (Y_1, Y_2)$, we have

$$I(X;Y_2 \,|\, U) \leq I(X;Y_1 \,|\, U).$$

Remark 5: Clearly, the above condition holds when Y_1 is a more capable receiver than Y_2, since $I(X;Y_2 \,|\, U = u) \leq I(X;Y_1 \,|\, U = u)$. Thus, by setting $\mathcal{P}$ to be entire set of distributions on $\mathcal{X}$; if Y_1 is a more-capable receiver, then it is also an essentially more capable receiver.

Now we define some classes of symmetric channels that are used in this paper.

Definition 4: A channel with input alphabet $\mathcal{X}$ ($\mathcal{X} = \{0, 1, \dots m-1\}$), output alphabet $\mathcal{Y}$ (of size n) is said to be *c-symmetric*[2] if, for each $j = 0, \dots, m-1$, there is a permutation $\pi_j(\cdot)$ of $\mathcal{Y}$ such that $\mathrm{P}(Y = \pi_j(y) \,|\, X = (i+j)_m) = \mathrm{P}(Y = y \,|\, X = i), \forall i$, where $(i+j)_m = (i+j) \,|\, m$. A broadcast channel with input alphabet $\mathcal{X}$ and output alphabets $\mathcal{Y}_1, \mathcal{Y}_2$ is said to be *c-symmetric* if both the channels $X \to Y_1$ and $X \to Y_2$ are c-symmetric.

Observe that BSC and BEC are examples of c-symmetric channels. To see this, for BSC set $\pi_0(y) = y, \pi_1(y) = 1 - y$ where $y \in \{0, 1\}$. For BEC, set the following: $\pi_0(y) = y, y \in \{0, E, 1\}$ and $\pi_1(0) = 1, \pi_1(E) = E, \pi_1(1) = 0$.

Remark 6: This definition is a natural generalization of binary input symmetric output broadcast channels that have been studied in literature.

Definition 5: In a c-symmetric broadcast channel Y_1 is said to be a *dominantly c-symmetric receiver* if the following condition holds: for every $p(x)$

$$I(X;Y_1)_p - I(X;Y_2)_p \leq I(X;Y_1)_u - I(X;Y_2)_u$$

where $u(x)$ is the uniform distribution.

In other words, uniform distribution also maximizes the difference $I(X;Y_1) - I(X;Y_2)$.

The following well known Lemma (also called Mrs. Gerber's Lemma) shall be employed often.

Lemma 1: [9] The function $f(x) = H(p * H^{-1}(x))$ is strictly convex on x, $0 \leq x \leq 1$

[2]c-symmetric $\equiv$ circularly-symmetric.

Here, logarithms are to the base 2, $H(\cdot)$ is the binary entropy function, and $x * p = x(1-p) + (1-p)x$—the binary convolution.

B. Capacity Regions

Theorem 1: Let $\mathcal{P}$ be any *sufficient class* of distributions that makes the receiver Y_1 essentially less noisy compared to receiver Y_2. Then, the capacity region of a two-receiver broadcast channel $X \to (Y_1, Y_2)$ is given by the union of rate pairs (R_1, R_2) such that

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

for some $U \to X \to (Y_1, Y_2)$ and $p(x) \in \mathcal{P}$.

Proof: The theorem is established following the observation in Section I.C.3, i.e., by showing that the superposition coding regions and the U, V outer bound yields the same region. From the fact that Y_1 essentially less noisy compared to receiver Y_2 and $p(x) \in \mathcal{P}$, we have $I(U;Y_2) \leq I(U;Y_1)$, and hence, $R_1 + R_2 \leq I(U;Y_2) + I(X;Y_1 \,|\, U) \leq I(X;Y_1)$. Thus, any rate pair (R_1, R_2) satisfying the constraints of the Theorem is achievable by superposition coding.

To show the converse, observe that from Remark 3 we can compute the boundary points of the (U, V)-outer bound by restricting ourselves to $\mathcal{P}$. On $\mathcal{P}$, however, $I(U;Y_2) \leq I(U;Y_1)$, and hence, $R_1 + R_2 \leq I(U;Y_2) + I(X;Y_1 \,|\, U) \leq I(X;Y_1)$. This constraint implies that any (R_1, R_2) on the boundary of the (U, V) outer bound is indeed achievable by superposition coding. $\square$

Theorem 2: Let $\mathcal{P}$ be any *sufficient class* of distributions that makes the receiver Y_1 essentially more capable compared to receiver Y_2. Then, the capacity region of a two-receiver broadcast channel $X \to (Y_1, Y_2)$ is given by the union of rate pairs (R_1, R_2) such that

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

for some $U \to X \to (Y_1, Y_2)$ and $p(x) \in \mathcal{P}$.

Proof: Clearly the achievability of any rate pair (R_1, R_2) satisfying the constraints follows using superposition coding (see Section I.C.2).

To show the converse, again from Remark 3 we can compute the boundary points of the (U, V)-outer bound by restricting ourselves to $\mathcal{P}$. On $\mathcal{P}$, however, $I(X;Y_2 \,|\, V) \leq I(X;Y_1 \,|\, V)$, and hence, $R_1 + R_2 \leq I(V;Y_1) + I(X;Y_2 \,|\, V) \leq I(X;Y_1)$. Thus, any (R_1, R_2) on the boundary of the (U, V) outer bound is indeed achievable by superposition coding. $\square$

III. BSC/BEC BROADCAST CHANNEL

In this section, we compute the capacity region of the $\mathrm{BSC}(p)/\mathrm{BEC}(e)$ broadcast channel. As we shall see below, for various choices of (p, e) this channel will belong to the various notions of dominance and superposition coding is optimal in all regimes. W.l.o.g, we assume that $0 \leq p \leq \frac{1}{2}$ and $0 \leq e \leq 1$.

Theorem 3: The broadcast channel consisting of a $\mathrm{BSC}(p)$ and $\mathrm{BEC}(e)$ has the following properties:

1) Y_{BSC} is a degraded version of Y_{BEC} if and only if $0 \le e \le 2p$.
2) Y_{BEC} is a less noisy receiver than Y_{BSC} if and only if $0 \le e \le 4p(1-p)$. Thus, when $2p < e \le 4p(1-p)$, Y_{BEC} is a less noisy version of Y_{BSC} (but the channel is not degraded).
3) Y_{BEC} is a more capable receiver than Y_{BSC} if and only if $0 \le e \le H(p)$. Thus, when $4p(1-p) < e \le H(p)$, Y_{BEC} is more capable than Y_{BSC} (but not less noisy).
4) When $H(p) < e < 1$, the channel does not fall into any of degraded, less noisy or more capable definitions of dominance.
5) Y_{BSC} is an essentially less noisy receiver than Y_{BEC} if and only if $H(p) \le e \le 1$.

Proof: The proof of (1) is known in literature. Indeed, it is straightforward to see that if $0 \le e \le 2p$, then Y_{BSC} is a degraded version of Y_{BEC}. The proof of the other direction can be established using an equivalent characterization of degraded channels with binary inputs and symmetric outputs using $|\mathcal{D}|$-*distributions*. (See Theorem 4.76 in Modern Coding Theory [10], preliminary version October 2007). The proofs of (2)–(5) is new here.

The proof of (2) follows the equivalent characterization of less noisy receiver stated in Remark 1. Let $\mathrm{P}(X=0)=x$; then define

$$\begin{aligned}D(x) &:= I(X;Y_{\mathrm{BEC}}) - I(X;Y_{\mathrm{BSC}})\\ &= (1-e)H(x) - H(x*p) + H(p)\end{aligned}$$

where $H(\cdot)$ is the binary entropy function, and $x*p = x(1-p)+(1-p)x$—the binary convolution. The second derivative of $D(x)$ is given by

$$\frac{\partial^2 D}{\partial x^2} = -\frac{1-e}{x(1-x)\ln 2} + \frac{(1-2p)^2}{(x*p)(1-x*p)\ln 2}.$$

Elementary algebra shows that $\frac{\partial^2 D}{\partial x^2} \le 0$ for all $x \in [0,1]$ if and only if $e \le 4p(1-p)$. Thus, (2) is established.

The proof of (3) follows from Lemma 1. Observe that for all $0 \le x \le 1$

$$\begin{aligned}&(1-H(p))H(x) + H(p)\\ &\quad= (1-H(p))H(x*H^{-1}(0))\\ &\qquad+ H(p)H(x*H^{-1}(1))\\ &\quad\ge H(x*H^{-1}(H(p))) = H(x*p)\end{aligned}$$

where the last inequality follows from Lemma 1. Hence, when $0 \le e \le H(p)$ we have $I(X;Y_{\mathrm{BEC}}) = (1-e)H(x) \ge (1-H(p))H(x) \ge H(x*p) - H(p) = I(X;Y_{\mathrm{BSC}})$. On the other hand if $1 \ge e > H(p)$ then when $\mathrm{P}(X=0)=\frac{1}{2}$ we have $I(X;Y_{\mathrm{BEC}}) = 1-e < 1-H(p) = I(X;Y_{\mathrm{BSC}})$. The proof of (4) is again straightforward. Note that $\frac{\partial D}{\partial x} = (1-e)J(x) - (1-2p)J(x*p)$ where $J(x) = \log_2 \frac{1-x}{x}$. Thus, as $x \to 0+$, we have $\frac{\partial D}{\partial x} \to \infty$ or for small values of x, we have $D(x) > 0$,, i.e., $I(X;Y_{\mathrm{BEC}}) > I(X;Y_{\mathrm{BSC}})$. However, $D(\frac{1}{2}) = H(p) - e < 0$. Hence, in the regime $1 \ge e > H(p)$, neither receiver is more capable than the other receiver.

The proof of (5) is a bit more involved. To prevent repetition of arguments, we will establish the results for a larger class of channels.

Lemma 2: The uniform distribution on $\mathcal{X}$ forms a sufficient class $\mathcal{P}$ for a c-symmetric broadcast channel.

Proof: Let $\mathcal{X} = \{0,1,\ldots,m-1\}$. Given a triple (U,V,X), construct a tuple (W',U',V',X') as follows:

$$\begin{aligned}&\mathrm{P}(W'=j,U'=u,V'=v,X'=i)\\ &\qquad= \frac{1}{m}\mathrm{P}(U=u,V=v,X=(i+j)_m).\end{aligned}$$

Further set $(W',U',V') \to X' \to (Y_1',Y_2')$ to be a Markov chain with the same channel transition probability.

Observe that

$$\begin{aligned}&\mathrm{P}(W'=j,U'=u,Y_1'=y)\\ &\quad= \sum_i \mathrm{P}(W'=j,U'=u,X'=i,Y_1'=y)\\ &\quad= \sum_i \mathrm{P}(W'=j,U'=u,X'=i)\\ &\qquad \times \mathrm{P}(Y_1'=y \,|\, X'=i)\\ &\quad= \frac{1}{m}\sum_i \mathrm{P}(U=u,X=(i+j)_m)\\ &\qquad \times \mathrm{P}(Y_1=\pi_j(y) \,|\, X=(i+j)_m)\\ &\quad= \frac{1}{m}\sum_i \mathrm{P}(U=u,X=(i+j)_m,Y_1=\pi_j(y))\\ &\quad= \frac{1}{m}\mathrm{P}(U=u,Y_1=\pi_j(y)).\end{aligned} \tag{2}$$

Similarly

$$\mathrm{P}(W'=j,U'=u,Y_2'=y) = \frac{1}{m}\mathrm{P}(U=u,Y_2=\sigma_j(y)). \tag{3}$$

It is easy to see that the following holds:

$$\begin{aligned}\mathrm{P}(X'=i) &= \frac{1}{m}\forall i\\ I(X';Y_i' \,|\, W'=j) &= I(X;Y_i)\forall j, i=1,2\\ I(U';Y_i' \,|\, W'=j) &= I(U;Y_i)\forall j, i=1,2\\ I(X';Y_i' \,|\, U',W'=j) &= I(X;Y_i \,|\, U), \forall j, i=1,2\end{aligned}$$

where all equalities (except the first one) follow from (2), (3), and that entropy is unchanged by relabeling. Similar conditions also holds for the pair (W',V').

Therefore, setting $\tilde{U} = (W',U'), \tilde{V} = (W',V')$ and $q(u,v,x)$ to be the distribution induced by $(\tilde{U},\tilde{V},X')$ it is easy to see that the inequalities (1) are satisfied. As $\mathrm{P}(X'=i) = \frac{1}{m}\forall i$ this establishes the sufficiency of the uniform distribution. □

Lemma 3: In a c-symmetric broadcast channel, if Y_1 is a dominantly c-symmetric receiver then Y_1 is also an essentially less noisy receiver.

Proof: Since the uniform distribution on $\mathcal{X}$ forms a sufficient class (Lemma 2); it suffices to show that for all (V,X) such that $p(x)$ is uniform, we have

$$I(V;Y_1) \ge I(V;Y_2). \tag{4}$$

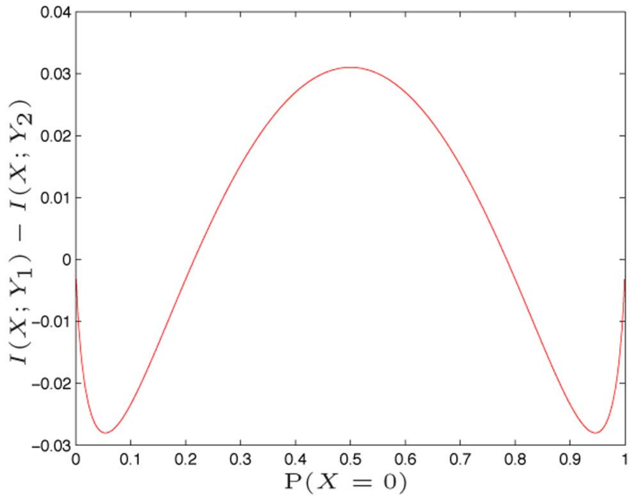

Fig. 2. Function $I(X;Y_1) - I(X;Y_2)$ for BSC(0.1) and BEC(0.5).

Given a pair (V,X) let $p_v(x)$ be the distribution on $\mathcal{X}$ when $V=v$. Y_1 is a dominantly c-symmetric receiver implies

$$I(X;Y_1)_{p_v} - I(X;Y_2)_{p_v} \le I(X;Y_1)_u - I(X;Y_2)_u.$$

Therefore

$$\begin{aligned} I(X;Y_1\,|\,V) &- I(X;Y_2\,|\,V) \\ &= \sum_v \mathrm{P}(V=v)(I(X;Y_1)_{p_v} - I(X;Y_2)_{p_v}) \\ &\le \sum_v \mathrm{P}(V=v)(I(X;Y_1)_u - I(X;Y_2)_u) \\ &= I(X;Y_1)_u - I(X;Y_2)_u. \end{aligned} \tag{5}$$

Since $V \to X \to (Y_1,Y_2)$ is Markov and $p(x)$ is uniform, observe

$$\begin{aligned} I(X;Y_1\,|\,V) &- I(X;Y_2\,|\,V) \\ &= I(X;Y_1)_u - I(V;Y_1) - (I(X;Y_2)_u - I(V;Y_2)) \\ &= I(X;Y_1)_u - I(X;Y_2)_u - (I(V;Y_1) - I(V;Y_2)). \end{aligned} \tag{6}$$

The required inequality (4) follows from (5) and (6), respectively. □

The proof of part 5 of Theorem 3 will be completed by establishing the following lemma.

Claim 1: For the BSC(p), BEC(e) broadcast channel Y_1 is a dominantly c-symmetric receiver when $H(p) \le e \le 1$.

Proof: We wish to show that when $1 \ge e \ge H(p)$ then the difference $I(X;Y_{\mathrm{BSC}}) - I(X;Y_{\mathrm{BEC}})$ is maximized by the uniform distribution. In particular, we wish to show that $I(X;Y_{\mathrm{BSC}}) - I(X;Y_{\mathrm{BEC}}) = H(x*p) - H(p) - (1-e)H(x) \le 1 - H(p) - (1-e) = e - H(p)$. By symmetry it suffices to consider $0 \le x \le \frac{1}{2}$. Observe that $x * H^{-1}(y)$ increases from $x \to \frac{1}{2}$ as y increases from $0 \to 1$. Hence, $H(x * H^{-1}(y))$ is increasing in y. Therefore

$$\begin{aligned} H(x*p) = H(x*H^{-1}(H(p))) &\le H(x*H^{-1}(e)) \\ &\le eH(x*H^{-1}(1)) + (1-e)H(x*H^{-1}(0)) \\ &= e + (1-e)H(x). \end{aligned}$$

Here, the first inequality follows from the fact that $H(x * H^{-1}(y))$ is increasing in y and the second one from its convexity (Lemma 1). Thus, $e - H(p) \ge H(x*p) - H(p) - (1-e)H(x)$, when $1 \ge e \ge H(p)$ as desired. □

Fig. 2 plots a typical $I(X;Y_{\mathrm{BSC}}) - I(X;Y_{\mathrm{BEC}})$ in the regime $1 \ge e \ge H(p)$

This completes the proof of Theorem 3. □

Theorem 4: For the BSC/BEC broadcast channel, the capacity region is given by the following:

1) When $0 \le e \le 4p(1-p)$ the boundary of the capacity region if formed by the set of rate pairs
$$R_2 = 1 - H(s*p), R_1 = (1-e)H(s)$$
for $s \in [0, \frac{1}{2}]$.

2) When $4p(1-p) < e < H(p)$, the boundary of the capacity region is formed by the set of rate pairs
$$R_2 = 1 - H(s*p), R_1 = (1-e)H(s)$$
for $s \in [0, s_0]$, and the line segment connecting the pair $((1-e)H(s_0), 1 - H(s_0*p))$ to $(1-e, 0)$. Here, $s_0 \in (0, \frac{1}{2})$ is the unique point satisfying $1 - H(s_0*p) + (1-e)H(s_0) = 1 - e$.

3) When $H(p) \le e \le 1$ the capacity region is given by time-divison, i.e., the set of rate pairs
$$\begin{aligned} R_2 &= \alpha(1 - H(p)) \\ R_1 &= (1-\alpha)(1-e) \end{aligned}$$
for $0 \le \alpha \le 1$.

Proof: We will first establish some preliminary results that will help in establishing Parts 1 and 2 of the Theorem.

We know that in the regime $0 \le e \le H(p)$, the capacity region is achieved by superposition coding. Thus, the capacity region is the union of rate pairs satisfying

$$\begin{aligned} R_2 &\le I(U;Y_{\rm BSC}) \\ R_1+R_2 &\le I(U;Y_{\rm BSC})+I(X;Y_{\rm BEC}\,|\,U) \\ R_1+R_2 &\le I(X;Y_{\rm BEC}) \end{aligned}$$

over all (U,X) such that $U \to X \to (Y_1,Y_2)$ form a Markov chain.

It is well-known and easy to see (by considering corner points) that the above region is also equivalent to the following region: the union of rate pairs satisfying

$$\begin{aligned} R_2 &\le I(U;Y_{\rm BSC}) \\ R_1 &\le I(X;Y_{\rm BEC}\,|\,U) \\ R_1+R_2 &\le I(X;Y_{\rm BEC}) \end{aligned} \tag{7}$$

over all (U,X) such that $U \to X \to (Y_1,Y_2)$ form a Markov chain. Observe that the equivalence of the regions still hold even if we fix the distribution $\mathrm{P}(x)$.

Further from Remark 3 and Lemma 3 we know that it suffices to consider $\mathrm{P}(X=0)=\frac{1}{2}$ to compute the boundary of the superposition coding inner bound. By the equivalence of the regions mentioned above, it further suffices to compute the boundary of (7). Since $\mathrm{P}(X=0)=\frac{1}{2}$, the third constraint becomes $R_1+R_2 \le (1-e)$. Thus, to compute the boundary of the superposition coding inner bound it suffices to compute the intersection of the following two convex regions: (i) the union of (non-negative) rate pairs satisfying

$$\begin{aligned} R_2 &\le I(U;Y_{\rm BSC}) \\ R_1 &\le I(X;Y_{\rm BEC}\,|\,U) \end{aligned} \tag{8}$$

over all (U,X) such that $U \to X \to (Y_1,Y_2)$ form a Markov chain, and $\mathrm{P}(X=0)=\frac{1}{2}$; (ii) non-negative rate pairs that satisfy $R_1+R_2 \le 1-e$.

Claim 2: When $0 \le e \le H(p)$, the capacity region is given by the convex closure of the set of rate pairs satisfying

$$\begin{aligned} R_2 &\le 1-H(s*p) \\ R_1 &\le (1-e)H(s) \\ R_1+R_2 &\le (1-e) \end{aligned}$$

for $0 \le s \le \frac{1}{2}$, (i.e., choosing $U \to X \sim \mathrm{BSC}(s)$).

Proof: To establish the claim, we use the equivalent representation of the capacity region given by the intersection of the two regions: one given by (8) and the other $R_1+R_2 \le 1-e$. Since the region given by (8) is convex, it suffices to show that there is some $s_\lambda \in [0,1]$ such that $U \to X \sim \mathrm{BSC}(s_\lambda)$ maximizes the expression $\lambda I(U;Y_{\rm BSC})+(1-\lambda)I(X;Y_{\rm BEC}\,|\,U)$ among all $U \to X \to (Y_{\rm BEC},Y_{\rm BSC})$ that forms a Markov chain and for all $\lambda \in [0,1]$.

Consider any $U \to X$ defined by $\mathrm{P}(U=i)=u_i$ and $\mathrm{P}(X=0\,|\,U=i)=s_i$. Since $\mathrm{P}(X=0)=\frac{1}{2}$, we have $\sum_i u_i s_i=\frac{1}{2}$. Observe that

$$\begin{aligned} &\lambda I(U;Y_{\rm BSC})+(1-\lambda)I(X;Y_{\rm BEC}\,|\,U) \\ &= \sum_i u_i(\lambda(1-H(s_i*p))+(1-\lambda)(1-e)H(s_i)) \\ &\le \max_i \lambda(1-H(s_i*p))+(1-\lambda)(1-e)H(s_i). \end{aligned}$$

Thus, we have shown[3] that the maximum is obtained by some $U \to X \sim \mathrm{BSC}(s_\lambda)$. □

Claim 3: The set of points $R_1,R_2 \ge 0$ satisfying

$$\begin{aligned} R_2 &\le 1-H(s*p) \\ R_1+R_2 &\le (1-e)H(s)+1-H(s*p) \end{aligned}$$

for $0 \le s \le \frac{1}{2}$ is a convex set.

Proof: Since this region consists of the lines $R_1=0$, $R_2=0$ and the curve characterized by the points $(1-H(s*p),(1-e)H(s))$. Given and $s_1,s_2 \in [0,\frac{1}{2}], s_1 \ne s_2$ then let $(1-e)H(s_3)=\frac{1}{2}((1-e)H(s_1)+(1-e)H(s_2))$. To show that the region is convex, it suffices to show that the curve is concave (we will indeed establish strict concavity), i.e.,

$$1-H(s_3*p) > \frac{1}{2}(1-H(s_1*p)+1-H(s_2*p))$$

or equivalently $H(s_3*p) < \frac{1}{2}(H(s_1*p)+H(s_2*p))$.

Let $s_1=H^{-1}(x_1), s_2=H^{-1}(x_2), s_3=H^{-1}(x_3)$. Then from $(1-e)H(s_3)=\frac{1}{2}((1-e)H(s_1)+(1-e)H(s_2))$, we have $x_3=\frac{1}{2}(x_1+x_2)$. Hence, from Lemma 1 (using strict convexity) it follows that

$$H(p*H^{-1}(x_3)) < \frac{1}{2}(H(p*H^{-1}(x_1))+H(p*H^{-1}(x_2))).$$

This implies that $H(s_3*p) < \frac{1}{2}(H(s_1*p)+H(s_2*p))$ as desired. Indeed this establishes that the curve defined by the points $(1-H(s*p),(1-e)H(s))$ is strictly concave.

Proof of Part 1: From Theorem 3 (Part 2) we know that when $0 \le e \le 4p(1-p)$ then receiver $Y_{\rm BEC}$ is less noisy than $Y_{\rm BSC}$; hence, $I(U;Y_{\rm BSC})+I(X;Y_{\rm BEC}\,|\,U) \le I(X;Y_{\rm BEC})$. Therefore, the third inequality in superposition coding is irrelevant. Hence, from Claims 2 and 3 the result for Part 1 follows. □

Proof of Part 2: In the proof of the Claim 3, we established that the curve characterized by $(1-H(s*p),(1-e)H(s))$ is strictly concave, and hence, this curve can only intersect the line $R_1+R_2=1-e$ in at most two locations. At $s=\frac{1}{2}$ they coincide; hence, there is at most another distinct point of intersection in the region $s \in (0,\frac{1}{2})$. To show that there is a point of intersection it suffices to show that as $s \to (\frac{1}{2})^-$, the curve lies above the line. (Note that at $s=0$ the curve is strictly below the line when $e<H(p)$, and hence, by the intermediate value theorem, the curve will intersect the line at some $s \in (0,\frac{1}{2})$.)

Therefore, it suffices to show that for $s \in (\frac{1}{2}-\delta,\frac{1}{2})$, and $\delta>0$, we have

$$1-H(s*p) > (1-e)(1-H(s)).$$

Taking Taylor's series expansion about $s=\frac{1}{2}$, this is true if

$$-\left.\frac{\partial^2 H(s*p)}{\partial s^2}\right|_{s=\frac{1}{2}} > -(1-e)\left.\frac{\partial^2 H(s)}{\partial s^2}\right|_{s=\frac{1}{2}}.$$

[3]The author thanks Young-Han Kim and Amin Gohari for suggestions that simpified author's original way of establishing this part.

(We consider the second derivative since the function values as well as the first derivatives match as $s = \frac{1}{2}$.) Clearly this is true when $(1-2p)^2 > 1-e$ or $e > 4p(1-p)$. Thus, when $4p(1-p) < e < H(p)$ we have exactly one point of intersection in the interval $(0, \frac{1}{2})$, say s_0.

From Claim 2 and Claim 3, the capacity region is the intersection of two convex sets: one defined by the set of (non-negative) points bounded by the curve $R_2 = (1 - H(s*p), R_1 = (1-e)H(s))$, $0 \le s \le \frac{1}{2}$, and the other defined by the line $R_1 + R_2 \le 1 - e$. From the concavity of the curve established in Claim 3, it follows that the intersection is given by the curve $R_2 = (1 - H(s*p), R_1 = (1-e)H(s))$, $0 \le s \le s_0$, and line segment connecting the pair $((1-e)H(s_0), 1 - H(s_0 * p))$ to $(1-e, 0)$. This estabishes Part 2 of Theorem 4. □

Proof of Part 3: In this regime, we know (Part 5 of Theorem 3) that Y_{BSC} is an essentially less noisy receiver than Y_{BEC}, and further $\mathrm{P}(X=0) = \frac{1}{2}$ constitutes a sufficient class. Therefore, from Theorem 1 the capacity region is given by the union of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\le I(U; Y_{\mathrm{BEC}}) \\ R_1 + R_2 &\le I(U; Y_{\mathrm{BEC}}) + I(X; Y_{\mathrm{BSC}} \mid U) \end{aligned}$$

for some $U \to X \to (Y_1, Y_2)$ and $\mathrm{P}(X=0) = \frac{1}{2}$.

Let $\mathrm{P}(U=i) = u_i$ and $\mathrm{P}(X=0 \mid U=i) = s_i$, where $\sum_i u_i s_i = \frac{1}{2}$. Then

$$\begin{aligned} I(U; Y_{\mathrm{BEC}}) &= (1-e)\left(1 - \sum_i u_i H(s_i)\right) \\ I(X; Y_{\mathrm{BSC}} \mid U) &= \sum_i u_i (H(s_i * p) - H(p)) \\ &= \sum_i u_i (H(p * H^{-1}(H(s_i))) - H(p)). \end{aligned}$$

From the convexity of $H(p * H^{-1}(u))$ in u (see Lemma 1), it follows that

$$\begin{aligned} H(p * s_i) &= H(p * H^{-1}(H(s_i))) \\ &\le H(s_i) H(p * H^{-1}(1)) \\ &\quad + (1 - H(s_i)) H(p * H^{-1}(0)) \\ &= H(s_i) + (1 - H(s_i)) H(p) \\ &= H(p) + H(s_i)(1 - H(p)). \end{aligned}$$

Substituting this, we obtain that

$$\begin{aligned} I(X; Y_{\mathrm{BSC}} \mid U) &= \sum_i u_i (H(p * s_i) - H(p)) \\ &\le \sum_i u_i H(s_i)(1 - H(p)). \end{aligned}$$

Setting $\alpha = \sum_i u_i H(s_i)$, we see that

$$\begin{aligned} R_1 &\le I(U; Y_{\mathrm{BEC}}) = (1-\alpha)(1-e) \\ R_1 + R_2 &\le I(U; Y_{\mathrm{BEC}}) + I(X; Y_{\mathrm{BSC}} \mid U) \\ &\le (1-\alpha)(1-e) + \alpha(1 - H(p)) \end{aligned}$$

as desired. □

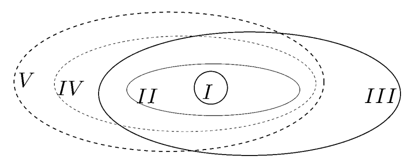

Fig. 3. Classes of broadcast channels with a *superior* receiver. I—degraded, II—less noisy, III—essentially less noisy, IV—more capable, V—essentially more capable.

IV. On Inclusion Relationships Between Classes of Broadcast Channels

In this section, we present the various relationships between the classes of two-receiver broadcast channels that were discussed in the paper.

Claim 4: We claim that the following relationships, as shown in Fig. 3, hold:

(i) Degraded $\subsetneq$ less noisy $\subsetneq$ more capable.
(ii) less noisy $\subsetneq$ essentially less noisy.
(iii) essentially less noisy $\not\subset$ more capable.
(iv) essentially less noisy $\not\subset$ essentially more capable.
(v) more capable $\not\subset$ essentially less noisy.
(vi) more capable $\subsetneq$ essentially more capable.

Proof: Part (i) is known and was established in [4]. The strict inclusions can also be inferred from Theorem 3. The inclusion of Part (ii) follows from Remark 4. The strict inclusion follows from Parts (4), (5) of Theorem 3. Part (iii) again follows from Parts (4), (5) of Theorem 3. See also Fig. 2.

Part (iv): This is again easy to deduce from Fig. 2. Take U to be a binary random variable with $\mathrm{P}(U=0) = \frac{1}{2}$ and take $\mathrm{P}(X=0 \mid U=0) = \epsilon$, $\mathrm{P}(X=0 \mid U=1) = 1-\epsilon$. For sufficiently small ϵ, we have that $I(X; Y_{\mathrm{BEC}} \mid U) > I(X; Y_{\mathrm{BSC}} \mid U)$ (see proof of Part 4 of Theorem 3), and hence, Y_{BSC} is not an *essentially more capable* receiver than Y_{BEC}. Note that $p(x)$ is uniform, and hence, any sufficient class $\mathcal{P}$ must contain the uniform distribution.

Part (v): Let $X \to Y_1$ be BEC(0.5) and $X \to Y_2$ be BSC(0.1101). Observe that $0.5 = 1-e > 1-H(p) \approx 0.4998$, and from part 3 of Theorem 3 we can see that Y_1 is a more capable receiver than Y_2. Let $U \to X$ be BSC(0.05), and set $\mathrm{P}(U=0) = 0.5$. This implies $\mathrm{P}(X=0) = 0.5 \in \mathcal{P}$ and it is easy to see that $0.3568 \approx I(U; Y_1) < I(U; Y_2) \approx 0.3924$ and thus Y_1, is not an essentially less noisy receiver than Y_2. (Note that this also implies that *essentially more capable* $\not\subset$ *essentially less noisy*.)

Part (vi): The inclusion follows from Remark 5. Hence, it suffices to prove that *essentially more capable* $\not\subset$ *more capable*. To this end, consider the following channel. The alphabets are given by $\mathcal{X} = \{0,1,2,3\}, \mathcal{Y}_1 = \mathcal{Y}_2 = \{0,1\}$. The channel $X \to Y_1$ is a perfectly clean channel when $\mathcal{X} \in \{0,1\}$, and is the completely noisy BSC(0.5) when $\mathcal{X} \in \{2,3\}$. The channel $X \to Y_2$ is a BSC(0.1) when $\mathcal{X} \in \{0,1\}$ and BSC(0.4) when $\mathcal{X} \in \{2,3\}$. When $p(x)$ is uniform on $\mathcal{X} = \{2,3\}$ we have $I(X; Y_2) > I(X; Y_1)$; implying Y_2 is not a more capable receiver than Y_1. However, it is easy to show that $p(x)$, that is uniform on $\mathcal{X} = \{0,1\}$, forms a sufficient class, and clearly on this sufficient class Y_1 is a more capable receiver than Y_2. This example shows that there are essentially more capable receivers that need not be more capable. □

Acknowledgment

The author would like to thank A. Montanari for bringing up the motivating question as well as some interesting discussions. The author also wishes to thank A. El Gamal, Y.-H. Kim, A. Gohari, S. Shamai (Shitz), and the anonymous referees for comments that improved the presentation.

References

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

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

[3] P. F. Bergmans, "Coding theorem for broadcast channels with degraded components," *IEEE Trans. Inf. Theory*, vol. IT-15, pp. 197–207, Mar. 1973.

[4] J. Körner and K. Marton, "Comparison of two noisy channels," Topics in Inf. Theory (ed. by I. Csiszar and P. Elias), Keszthely, Hungary, pp. 411–423, Aug. 1975.

[5] M. Van Dijk, "On a special class of broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 712–714, Mar. 1997.

[6] A. El Gamal, "The capacity of a class of broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-25, pp. 166–169, Mar. 1979.

[7] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-25, pp. 306–311, May 1979.

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

[9] A. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications: Part I," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 6, pp. 769–772, Nov. 1973.

[10] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2008 [Online]. Available: http://www.cambridge.org/catalogue/catalogue.asp?isbn=9780521852296

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

He is an Assistant Professor with the Information Engineering Department, Chinese University of Hong Kong. Previously, he was a postdoctoral researcher with the Theory group at Microsoft Research, Redmond. His research interests are in probabilistic and combinatorial problems arising in Electrical Engineering and Computer Science: especially those in discrete optimization and information theory. He 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.