Capacity regions of two new classes of 2-receiver broadcast channels

Chandra Nair
Department of Information Engineering
Chinese University of Hong Kong
Sha Tin, N.T., Hong Kong
Email: chandra@ie.cuhk.edu.hk

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 to obtain the characterization of the capacity region for certain new and interesting classes of broadcast channels. We also establish the relationships between the new classes and the existing classes.

1 Introduction

This paper is motivated directly by a simple broadcast channel setting, posed by Andrea Montanari(see Figure 1), consisting of a BSC(p) and BEC(e). Clearly if $e \leq 2p$, then the channel is *degraded*[1] and the capacity[2, 3] is given by the union of rate pairs (R_1, R_2) satisfying

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

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

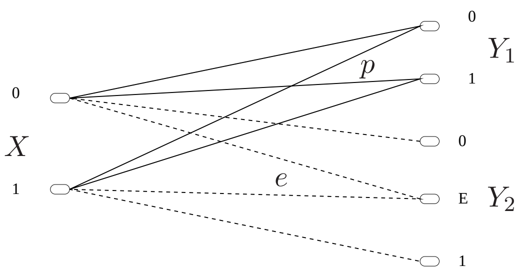

Figure 1: A broadcast channel consisting of a BSC(p) and BEC(e)

If $1 - H(p) \leq 1 - e$ then Y_2 would be a *more capable*[4] receiver than Y_1 (see Parts 1,2 of Claim 3 in the Appendix) and in this case capacity[5] is given by the union of rate pairs satisfying

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

over all (U, X) such that $U \to X \to (Y_1, Y_2)$ form a Markov chain. Therefore, the interesting case for determining the capacity region occurs when $1 - H(p) > 1 - e$.

Hence we restrict ourselves to the case when $1 - H(p) > 1 - e$, i.e. when the channel $X \to Y_1$ has a higher capacity than the channel $X \to Y_2$. The Figure 2 plots $I(X; Y_1) - I(X; Y_2)$ for the case $p = 0.1, e = 0.5$. It is clear that neither is more capable than the other. In particular this setting does not fall into any class of broadcast channels for which the capacity region has been characterized. We address this regime and establish the capacity region.

In fact we establish the capacity region of a whole new class of broadcast channels (motivated of course by this example) that contains this broadcast channel, under the regime $1-H(p) > 1-e$, as a special case.

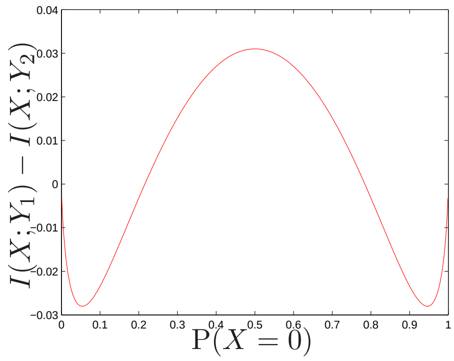

Figure 2: The function $I(X; Y_1) - I(X; Y_2)$ for BSC(0.1) and BEC(0.5)

1.1 Observation

The common theme between degraded, less noisy, and more capable channels is the existence of a dominant receiver who manages to decode the private messages for both the users. Let us recall the definition of the *less noisy*[4] receiver. One requires $I(U; Y_2) \leq I(U; Y_1)$ to hold true for *every* $p(u, x)$ to classify receiver Y_1 as a less noisy receiver than Y_2. In this paper we remove the need for $I(U; Y_2) \leq I(U; Y_1)$ holding true for *every* $p(u, x)$ and replace it by $I(U; Y_2) \leq I(U; Y_1)$ holding true for *a sufficiently large* class of distributions $p(u, x)$. We then show that for this relaxed definition of a less noisy receiver as well, the capacity regions can be obtained.

A similar development has been done for a more capable receiver as well.

Remark 1. The key contribution of this paper is the identification of capacity regions for interesting classes of broadcast channels by proving that these channels belong to a slightly tweaked definition of a less noisy or a more capable receiver. Secondly, the tweaking of the definitions is done in such a way that the existing techniques are sufficient to establish the capacity regions.

The organization of the paper is as follows. In section 2 we make the formal definitions and set up the required notation. In Section 3 we establish the capacity region of a class of two-receiver broadcast channels that has one receiver who is essentially less noisy when compared to the other. In Section 4 we identify certain interesting classes of channels that are neither less noisy nor more capable but are essentially less noisy. In Section 5 we establish the capacity region of a class of two-receiver broadcast channels that has one receiver who is essentially more capable than the other. Finally in Section 6 we establish the various inclusions among these classes.

2 Definitions and Notation

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 an input alphabet $\mathcal{X}$ and output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ 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 $\mathcal{W}_1 = \{1, 2, ..., 2^{nR_1}\}, \mathcal{W}_2 = \{1, 2, ..., 2^{nR_2}\}$.

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.,

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

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 with is the closure of the set of achievable rates. *The capacity region of the two user discrete memoryless channel is unknown.*

The capacity region is known for lots of special cases such as degraded, less noisy, more capable, deterministic, semi-deterministic, etc. - see [6] and the references therein. In this paper we establish the capacity region for two more classes of broadcast channels, one where one receiver is *essentially less noisy* compared to the other receiver; and the other where one receiver is *essentially more capable* than the other receiver.

A channel $X \to Y_2$ is said to be a *degraded* version of the channel $X \to Y_1$ if $X \to Y_1 \to Y_2'$ is a Markov chain and the pair (X, Y_2') is identically distributed as the pair (X, Y_2). A receiver Y_1 is said to be *less noisy*[4] compared to Y_2 if

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

for all $p(u, x)$ such that $U \to X \to (Y_1, Y_2)$ forms a Markov chain. Finally, a receiver Y_1 is said to be *more capable*[4] compared to Y_2 if

$$I(X; Y_2) \leq I(X; Y_1)$$

for all $p(x)$.

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 2-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 $q(u, v, x)$

[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.

that satisfies

$$\begin{aligned}
&q(x) \in \mathcal{P}, \\
&I(U;Y_i)_p \leq I(U;Y_i)_q,\ i = 1,2, \\
&I(V;Y_i)_p \leq I(V;Y_i)_q,\ i = 1,2, \\
&I(X;Y_i|U)_p \leq I(X;Y_i|U)_q,\ i = 1,2, \\
&I(X;Y_i|V)_p \leq I(X;Y_i|V)_q,\ i = 1,2, \\
&I(X;Y_i)_p \leq I(X;Y_i)_q,\ i = 1,2,
\end{aligned} \tag{1}$$

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

Definition 2. A receiver Y_1 is *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 2. 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. However, in Section 4 we will show that there are essentially less noisy receivers that are not less noisy.

Definition 3. A receiver Y_1 is *essentially more capable* 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

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

Remark 3. Clearly the above condition holds when Y_1 is a more capable receiver than Y_1, since it holds under each conditioning of U. Thus by setting $\mathcal{P}$ to be entire set of distributions on $\mathcal{X}$; if Y_1 is also a more-capable receiver than it is also an essentially more capable receiver.

3 The capacity region of a broadcast channel with an essentially less noisy receiver

Theorem 1. *The capacity region of a two-receiver broadcast channel where Y_1 is essentially less noisy compared to 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 $p(x) \in \mathcal{P}$. Here $\mathcal{P}$ denotes any sufficient class *of distributions that makes the receiver Y_1 essentially less noisy compared to receiver Y_2.*

Proof. The theorem follows in a straightforward manner from the known achievability regions and outer bounds for the two receiver broadcast channels, as shown below.

The direct part

It is well-known [7] that the set of all rate pairs (R_1, R_2) satisfying

$$\begin{aligned} R_2 &\leq I(U;Y_2) \\ R_1 + R_2 &\leq I(U;Y_2) + I(X;Y_1|U) \\ R_1 + R_2 &\leq I(X;Y_1) \end{aligned} \tag{2}$$

for any $U \to X \to (Y_1, Y_2)$ is achievable via superposition coding. Further if $p(x) \in \mathcal{P}$, we have $I(U;Y_2) \leq I(U;Y_1)$ and thus (2) reduces to the region in Theorem 1 and completes the proof of the achievability.

The converse part

It is well-known [5, 7] that the set of all rate pairs (R_1, R_2) satisfying

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

over all $U \to X \to (Y_1, Y_2)$ forms an outer bound to the capacity region of the broadcast channel. Clearly from the definition of the sufficient class $\mathcal{P}$ it is clear that one can restrict the union to be over $p(x) \in \mathcal{P}$. Further if $p(x) \in \mathcal{P}$, we have $I(U;Y_2) \leq I(U;Y_1)$ and thus

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

This implies that (3) reduces to the region in Theorem 1 and completes the proof of the converse to the capacity region.

4 A class of symmetric broadcast channels with an essentially less noisy receiver

In this section we prove that the class with an essentially less noisy receiver of channels is strictly larger than the class where where one receiver is *less noisy* [4] compared to the other receiver. In particular this class contains the channel that served as the motivation behind this paper - the broadcast channel when one of the channels is BSC(p) and the other is BEC(e); where the pair (p, e) satisfies $1 - H(p) \geq 1 - e$.

Definition 4. A channel with input alphabet $\mathcal{X}$ ($\mathcal{X} = \{0, 1, ...m-1\}$), output alphabet $\mathcal{Y}$ (of size n) is said to be *c-symmetric* if, for each $j = 0, .., m-1$, 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 $(i+j)_m = (i+j) \mod m$.

Observe that BSC and BEC are examples of c-symmetric channels.

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.

Lemma 1. *The uniform distribution on $\mathcal{X}$ forms a sufficient class $\mathcal{P}$ for a c-symmetric broadcast channel.*

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

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

Further set $(W', U', V') \to X' \to (Y_1', Y_2')$ to be a Markov chain with $p(y_1', y_2'|x') \equiv p(y_1, y_2|x)$ (i.e. the channel transition probability remains the same).

Observe that

$$\begin{aligned} &\mathrm{P}(W' = j, U' = u, Y_1' = y) \\ &\quad = \sum_i \mathrm{P}(W' = j, U' = u, X' = i, Y_1' = y) \\ &= \frac{1}{m}\sum_i \mathrm{P}(U = u, X = (i+j)_m, Y_1 = \pi_j(y)) \\ &= \frac{1}{m}\mathrm{P}(U = u, Y_1 = \pi_j(y)). \end{aligned} \tag{4}$$

Similarly

$$\mathrm{P}(W' = j, U' = u, Y_2' = y) = \frac{1}{m}\mathrm{P}(U = u, Y_2 = \sigma_j(y)). \tag{5}$$

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 equations (4), (5), 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. □

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 $p(x)$

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

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

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

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

The proof follows from part 3 of Claim 3 in the appendix; also see Figure 2.

Lemma 2. *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; by Lemma 1 it suffices to show that for all (V, X) such that $p(x)$ is uniform we have

$$I(V; Y_1) \geq I(V; Y_2). \tag{6}$$

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

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

Therefore

$$\begin{aligned} &I(X; Y_1|V) - I(X; Y_2|V) \\ &\quad= \sum_v \mathrm{P}(V = v) \left(I(X; Y_1)_{p_v} - I(X; Y_2)_{p_v}\right) \\ &\quad\leq \sum_v \mathrm{P}(V = v) \left(I(X; Y_1)_u - I(X; Y_2)_u\right) \\ &\quad= I(X; Y_1)_u - I(X; Y_2)_u. \end{aligned} \tag{7}$$

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

$$\begin{aligned} &I(X; Y_1|V) - I(X; Y_2|V) \\ &\quad= I(X; Y_1)_u - I(V; Y_1) - (I(X; Y_2)_u - I(V; Y_2)) \\ &\quad= I(X; Y_1)_u - I(X; Y_2)_u - (I(V; Y_1) - I(V; Y_2)). \end{aligned} \tag{8}$$

The required inequality (6) follows from (7) and (8) respectively.

5 The capacity region of a broadcast channel with an essentially more capable receiver

Theorem 2. *The capacity region of a two-receiver broadcast channel where Y_1 is essentially more capable compared to 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 $p(x) \in \mathcal{P}$. Here $\mathcal{P}$ denotes any sufficient class *of distributions that makes the receiver Y_1 essentially more capable compared to receiver Y_2.*

The direct part

It is well-known [7] that the set of all rate pairs (R_1, R_2) satisfying

$$\begin{aligned} R_2 &\leq I(U; Y_2) \\ R_1 + R_2 &\leq I(U; Y_2) + I(X; Y_1|U) \\ R_1 + R_2 &\leq I(X; Y_1) \end{aligned} \tag{9}$$

for any $U \to X \to (Y_1, Y_2)$ is achievable via superposition coding. Restricting ourselves to $p(x) \in \mathcal{P}$, we have the region in Theorem 1 and completes the proof of the achievability.

The converse part

It is well-known [5, 8] that the set of all rate pairs (R_1, R_2) satisfying

$$\begin{aligned} R_2 &\leq I(U;Y_2) \\ R_1 + R_2 &\leq I(U;Y_2) + I(X;Y_1|U) \\ R_1 &\leq I(V;Y_1) \\ R_1 + R_2 &\leq I(V;Y_1) + I(X;Y_2|V) \end{aligned} \tag{10}$$

over all $(U,V) \to X \to (Y_1, Y_2)$ forms an outer bound to the capacity region of the broadcast channel. Clearly from the definition of the sufficient class $\mathcal{P}$ it is clear that one can restrict the union to be over $p(x) \in \mathcal{P}$. Further if $p(x) \in \mathcal{P}$, since Y_1 is an essentially more capable receiver we have $I(X;Y_2|V) \leq I(X;Y_1|V)$ and thus

$$I(V;Y_1) + I(X;Y_2|V) \leq I(X;Y_1).$$

This implies that (10) is contained inside the (achievable) region in Theorem 1 and completes the proof of the converse to the capacity region. (Indeed it is easy to see that setting $V = X$ is optimal and thus reduces the region in (10) to the region in Theorem 2.) $\square$

6 On inclusion relationships between classes of broadcast channels

In this section, we present the various relationships between the classes of 2-receiver broadcast channels that were discussed in the paper.

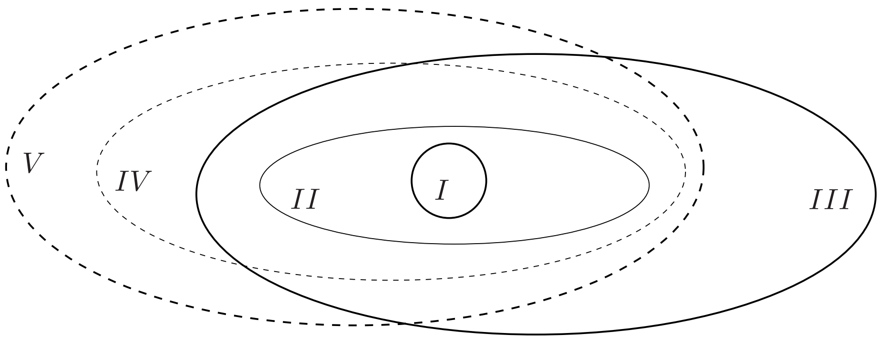

Figure 3: The classes of broadcast channels with a *superior* receiver. I - degraded, II - less noisy, III- essentially less noisy, IV - more capable, V - essentially more capable

Claim 2. *We claim that the following relationships, as shown in Figure 3, hold*

(i) Degraded $\subset$ less noisy $\subset$ more capable,

(ii) less noisy $\subset$ essentially less noisy,

(iii) essentially less noisy $\not\Rightarrow$ more capable,

(iv) essentially less noisy $\not\Rightarrow$ essentially more capable,

(v) more capable $\not\Rightarrow$ essentially less noisy.

(*vi*) more capable $\subset$ essentially more capable,

Proof. Part (i) was established in [4]. Part (ii) follows from Remark 2 and Section 4. Part (iii) follows from Figure 2 and Section 4.

Part (iv): This is again easy to deduce from Figure 2. Take U to be a binary random variable with $\mathrm{P}(U=0)=\frac{1}{2}$ and take $\mathrm{P}(X=0|U=0)=\epsilon$, $\mathrm{P}(X=0|U=1)=1-\epsilon$. For sufficiently small ϵ we have that $I(X;Y_2|U)>I(X;Y_1|U)$ and hence Y_1 is not an *essentially more capable* receiver than Y_2. 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 2 of Claim 3 in the Appendix 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* $\not\Rightarrow$ *essentially less noisy*.)

Part (vi): From Remark 3 it is clear that *more capable* $\subseteq$ *essentially more capable*. Hence it suffices to prove that *essentially more capable* $\not\Rightarrow$ *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_1 is not a more capable receiver than Y_2. However it is easy to show that $p(x)$ 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. $\square$

Acknowledgement

The author wishes to that Andrea Montanari for bringing up the motivating question as well as some interesting discussions. The author also wishes to thank Abbas El Gamal and the anonymous referees for comments that improved the presentation.

References

[1] T. Cover, "Broadcast channels," *IEEE Trans. Info. Theory*, vol. IT-18, pp. 2–14, January, 1972.

[2] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Probl. Peredac. Inform.*, vol. 10(3), pp. 3–14, 1974.

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

[4] J. Körner and K. Marton, "A source network problem involving the comparison of two channels ii," *Trans. Colloquim Inform. Theory, Keszthely, Hungary*, August, 1975.

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

[6] T. Cover, "Comments on broadcast channels," *IEEE Trans. Info. Theory*, vol. IT-44, pp. 2524–2530, October, 1998.

[7] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Info. 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. Info. Theory*, vol. IT-53, pp. 350–355, January, 2007.

Consider a broadcast channel with two receivers. Let $X \to Y_1$ be BSC(p), $0 \le p \le \frac{1}{2}$ and $X \to Y_2$ be BEC(e). Let

$$\begin{aligned} D(x) &\stackrel{\triangle}{=} I(X;Y_1) - I(X;Y_2) \\ &= H(x * p) - (1-e)H(x) - H(p) \end{aligned}$$

be the difference $I(X;Y_1) - I(X;Y_2)$ conditioned on $\mathrm{P}(X=0) = x$. Observe that the function is symmetric about $x = \frac{1}{2}$, i.e. $D(x) = D(1-x)$.

Claim 3. *The function $D(x)$ has the following properties:*

1. *When $e \le 2p$, $D(x)$ monotonically decreases in the interval $[0, \frac{1}{2}]$.*
2. *When $2p < e \le H(p)$, $D(x)$ monotonically decreases in the interval $[0, r]$, and monotonically increases in the interval $[r, \frac{1}{2}]$ for some $r \in (0, \frac{1}{2}]$. The maximum occurs at $x = 0$, i.e. $D(x) \le 0, \forall x \in [0,1]$.*
3. *When $H(p) < e \le 1$, $D(x)$ monotonically decreases in the interval $[0, r]$, and monotonically increases in the interval $[r, \frac{1}{2}]$ for some $r \in (0, \frac{1}{2})$. The maximum occurs at $x = \frac{1}{2}$, i.e. $D(x) \le D(\frac{1}{2}), \forall x \in [0,1]$.*

Proof. Let $J(x) = \log_2 \frac{1-x}{x}$. Observe that

$$\frac{d}{dx} D(x) = (1-2p)J(x*p) - (1-e)J(x). \tag{11}$$

For $x \in [0, \frac{1}{2}]$, when $e \le 2p$ we have $\frac{d}{dx}D(x) \le 0$, since $0 < J(x*p) < J(x)$ and establishes Part 1.

From (11) any x such that $\frac{d}{dx}D(x) = 0$ must satisfy

$$\left((\frac{1-x}{x})^c + 1 \right)^{-1} = x(1-p) + p(1-x), \tag{12}$$

where $c = \frac{1-e}{1-2p}$. Define

$$L(x) = \left((\frac{1-x}{x})^c + 1 \right)^{-1}.$$

When $e \ge 2p$, we have $0 < c < 1$. Then it is easy to see that $L(x)$ is concave in $x \in [0, \frac{1}{2}]$. Observe that

$$\frac{d}{dx}L(x) = \frac{c}{\big(x(1-x)\big)^{1-c}\big[(1-x)^c + x^c\big]^2 x^{1-c}}$$

and since the functions: $\big(x(1-x)\big)^c, (1-x)^c + x^c, x^{1-c}$ increase in $x \in [0, \frac{1}{2}]$, we have $\frac{d^2}{dx^2}L(x) \le 0$. This implies that $L(x)$ can intersect the line $R(x) = x(1-p) + p(1-x)$ at possibly no more than two points on $x \in [0, \frac{1}{2}]$. Since $L(\frac{1}{2}) = \frac{1}{2} = R(\frac{1}{2})$, there is at most one other solution $r \in (0, \frac{1}{2})$ to (12) when $x \in (0, \frac{1}{2})$.

Since $\frac{d}{dx}D(x) \to -\infty$ as $x \to 0^+$, it is clear that $D(x)$ decreases in $[0, r]$ and increases in $[r, \frac{1}{2}]$. The maximum can therefore be obtained by comparing $D(0) = 0$ and $D(\frac{1}{2}) = e - H(p)$. This establishes Parts 2,3. □

Claim 4. *Consider a broadcast channel with two receivers. Let $X \to Y_1$ be BSC(p), $0 \leq p \leq \frac{1}{2}$ and $X \to Y_2$ be BEC(e). Then the following holds:*

1. $0 \leq e \leq 2p$*: Y_1 is a degraded version of Y_2.*
2. $2p \leq e \leq 4p(1-p)$*: Y_2 is less noisy than Y_1, but is not a degraded version.*
3. $4p(1-p) \leq e \leq H(p)$*: Y_2 is more capable (but not less noisy) than Y_2.*
4. $H(p) \leq e \leq 1$*: Y_1 is essentially less noisy than Y_2.*

Proof. Part 1 is well-known and easy to establish. Part 4 follows from Claim 3 and Section 4. We also know from Parts 1 and 2 of Claim 3 that when $0 \leq e \leq H(p)$ $D(x) \leq 0$, i.e. Y_2 is a more capable receiver than Y_1. Therefore to complete the proof of the claim it suffices to show that Y_2 is a less noisy receiver than Y_1 *if and only if* $0 \leq e \leq 4p(1-p)$.
The following statements are equivalent (the proof is immediate and omitted):
(i) Y_2 is a less noisy receiver than Y_1;
(ii) $\forall U \to X \to (Y_1, Y_2)$, $I(X; Y_1|U) - I(X; Y_2|U) \geq I(X; Y_1) - I(X; Y_2)$;
(iii) $I(X; Y_1)_{p(x)} - I(X; Y_2)_{p(x)}$ is a convex function of $p(x)$.

Therefore Y_2 is a less noisy receiver than Y_1 *if and only if* $D(x)$ is convex for $x \in [0, 1]$. It is straight forward to see that $D(x)$ is convex *if and only if* $0 \leq e \leq 4p(1-p)$. □