Superposition Coding is Almost Always Optimal for the Poisson Broadcast Channel

Hyeji Kim
Department of Electrical Engineering
Stanford University
Stanford, CA 94305, USA
Email: hyejikim@stanford.edu

Benjamin Nachman
Department of Physics
Stanford University
Stanford, CA 94305, USA
Email: nachman@stanford.edu

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

***Abstract*—This paper shows that the capacity region of the continuous-time Poisson broadcast channel is achieved via superposition coding for most channel parameter values. Interestingly, the channel in some subset of these parameter values does not belong to any of the existing classes of broadcast channels for which superposition coding is optimal (e.g., degraded, less noisy, more capable). For the rest of the channel parameter values, we show that there is a gap between Marton's inner bound and the UV outer bound.**

I. Introduction

The continuous-time Poisson channel is a canonical model for optical communication. Wyner [1] established the capacity region of this channel using an elementary method in which the capacity is shown to be the limit of the capacity of a discrete memoryless channel. In [2], Lapidoth, Telatar, and Urbanke extended Wyner's approach to study the Poisson broadcast channel. They considered a 2-receiver continuous-time Poisson broadcast channel (PBC) with input $X(t) \in [0,1]$ and outputs $Y_j(t)$, $j = 1,2$, such that $Y_j(t)|\{X(t) = x(t)\}$ is a Poisson process with an instantaneous rate $A_j(x(t) + s_j)$ for $t \geq 0$. Note that in this model the peak power of the transmitted signal $X(t)$ is constrained by 1.

Following Wyner's approach, Lapidoth et al. [2] approximated the PBC by the $(1/\Delta)$-extension of the binary-input binary-output broadcast channel (BPBC) depicted in Figure 1, where $\Delta > 0$. They showed that the capacity region of the PBC is the same as that of the $(1/\Delta)$-extension of the BPBC as $\Delta \to 0$. Hence, determining the capacity region of the PBC reduces to finding the capacity region of the BPBC.

Lapidoth et al. [2] showed that for the PBC with $s_1 \leq s_2$, Y_2 is a degraded version of Y_1 if $\alpha = A_1/A_2 \geq 1$; hence the capacity region in this parameter range is achieved via superposition coding.

In this paper, we show that superposition coding is optimal for a much larger set of channel parameter values. Our result is illustrated via the graphs in Figure 2 which show examples of the range of PBC parameters for which superposition coding is optimal. We further show that for a certain range of channel parameter values, superposition coding is optimal but the channel does not belong to any of the known classes for which superposition coding is optimal. Finally we show that in general there is a gap between Marton's inner bound [3] and the UV outer bound [4].

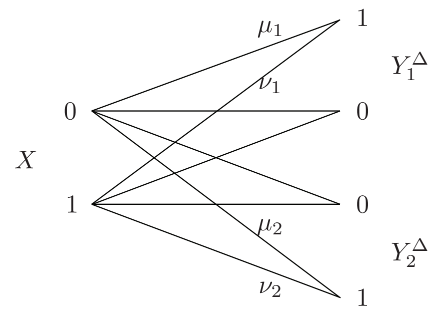

Fig. 1. BPBC: $\mu_1 = A_1 s_1 \Delta + O(\Delta^2)$, $\mu_2 = A_2 s_2 \Delta + O(\Delta^2)$, $\nu_1 = A_1(1+s_1)\Delta + O(\Delta^2)$, $\nu_2 = A_2(1+s_2)\Delta + O(\Delta^2)$.

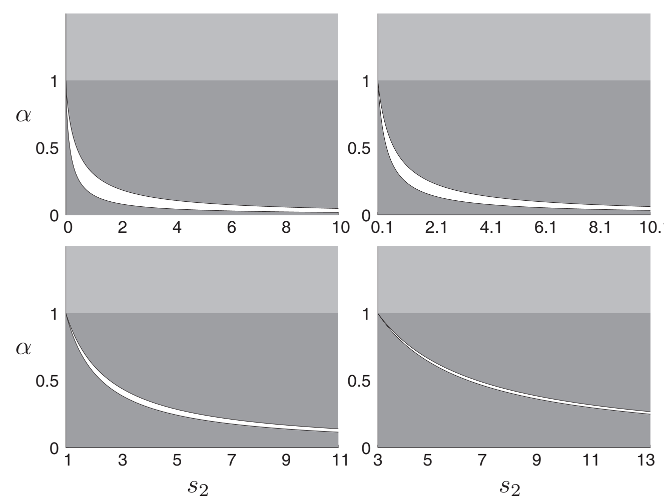

Fig. 2. Plots of α versus s_2 for $s_1 = 0$ (top left), $s_1 = 0.1$ (top right), $s_1 = 1$ (bottom left) and $s_1 = 3$ (bottom right). The shaded areas in each plot are where superposition coding is optimal and the light shaded area ($\alpha \geq 1$) is where Y_2 is a degraded version of Y_1.

In the following section we introduce the class of *effectively* less noisy class of broadcast channels for which superposition coding is optimal. Our main result is presented in Section III. In Section IV, we show that in the parameter range for which superposition coding is not optimal, in general, there is a small gap between Marton's inner bound and the UV outer bound.

Hyeji Kim is supported by the Alma M. Collins Stanford Graduate Fellowship. Benjamin Nachman is supported by the NSF Graduate Research Fellowship under Grant No. DGE-4747 and by the Stanford Graduate Fellowship.

Hence, the capacity region of the PBC is still not known in general.

II. Effectively less noisy broadcast channels

Consider a 2-receiver discrete memoryless broadcast channel (DM-BC) $p(y_1, y_2|x)$. It is well known that superposition coding is optimal for the following classes of DM-BC:

Degraded [5]: Receiver Y_1 is said to be a degraded version of Y_2 if there exists a random variable Y_2' such that $X \to Y_2' \to Y_1$ form a Markov chain.

Less noisy [6]: Receiver Y_2 is said to be less noisy than Y_1 if $I(U;Y_2) \geq I(U;Y_1)$ for all $p(u,x)$.

More capable [6]: Receiver Y_2 is said to be more capable than receiver Y_1 if $I(X;Y_2) \geq I(X;Y_1)$ for all $p(x)$.

It is also well known that degraded implies less noisy which implies more capable [6], and that the capacity region for the more capable class [7] is the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\leq I(U;Y_1), \\ R_2 &\leq I(X;Y_2|U), \\ R_1 + R_2 &\leq I(X;Y_2) \end{aligned} \tag{1}$$

for some $p(u,x)$ with $|\mathcal{U}| \leq \min\{|\mathcal{X}|, |\mathcal{Y}_1| \cdot |\mathcal{Y}_2|\} + 1$.

Note that when the channel is less noisy or degraded the sum bound in (1) is always inactive.

We now introduce the following new class of broadcast channels. Let $\mathcal{P}$ be the set of pmfs $p(x)$ such that for every $0 \leq \lambda \leq 1$,

$$\begin{aligned} &\max_{p(x)} \max_{p(u|x)} \{I(U;Y_1) + \lambda I(X;Y_2|U)\} \\ &\quad = \max_{p(x) \in \mathcal{P}} \max_{p(u|x)} \{I(U;Y_1) + \lambda I(X;Y_2|U)\}. \end{aligned} \tag{2}$$

Effectively less noisy: Receiver Y_2 is said to be *effectively less noisy* than Y_1 if $I(U;Y_2) \geq I(U;Y_1)$ for every $p(x) \in \mathcal{P}$ and $p(u|x)$, equivalently, by [8], if

$$\mathfrak{C}[I(X;Y_2) - I(X;Y_1)] = I(X;Y_2) - I(X;Y_1) \tag{3}$$

for every $p(x) \in \mathcal{P}$, where $\mathfrak{C}[f(x)]$ denotes the upper concave envelope of $f(x)$.

Remark 1. Clearly if the channel is less noisy, then it is effectively less noisy. Moreover, if the channel is essentially less noisy as defined in [8], it is also effectively less noisy. To show this note that the sufficient class $\mathcal{P}_o$ in [8] satisfies

$$\begin{aligned} &\max_{p(x)} \max_{p(u|x)} \{I(U;Y_1) + \lambda I(X;Y_2|U)\} \\ &\quad \leq \max_{p(x) \in \mathcal{P}_o} \max_{p(u|x)} \{I(U;Y_1) + \lambda I(X;Y_2|U)\}. \end{aligned}$$

Hence $\mathcal{P} \subseteq \mathcal{P}_o$, and if $I(U;Y_1) \leq I(U;Y_2)$ for $p(x) \in \mathcal{P}_o$ and every $p(u|x)$, then $I(U;Y_1) \leq I(U;Y_2)$ for $p(x) \in \mathcal{P}$ and every $p(u|x)$. We do not know if the two notions are identical, however. As we will see in the next section, effectively less noisy neither implies nor is implied by more capable in general.

We now establish the capacity region of effectively less noisy broadcast channels.

Proposition 1. The capacity region of a broadcast channel in which Y_2 is effectively less noisy than Y_1 is the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\leq I(U;Y_1), \\ R_2 &\leq I(X;Y_2|U) \end{aligned} \tag{4}$$

for some $p(x) \in \mathcal{P}$ and $p(u|x)$ with $|\mathcal{U}| \leq |\mathcal{X}| + 1$.

In Appendix A, we show that

$$\max_{(R_1,R_2) \in \bar{\mathcal{R}}} R_1 + \lambda R_2 = \max_{(R_1,R_2) \in \mathcal{R}} R_1 + \lambda R_2,$$

where $\mathcal{R}$ and $\bar{\mathcal{R}}$ denote the region in (4) and an outer bound on the UV outer bound [4], respectively. To complete the proof of Proposition 1 we use the following.

Lemma 1. [9] Let $\mathcal{R} \in \mathbb{R}^d$ be convex and $\mathcal{R}_1 \subseteq \mathcal{R}_2$ be two bounded convex subsets of $\mathcal{R}$, closed relative to $\mathcal{R}$. If every supporting hyperplane of $\mathcal{R}_2$ intersects $\mathcal{R}_1$, then $\mathcal{R}_1 = \mathcal{R}_2$.

III. Optimality conditions of superposition coding

Consider the 2-receiver Poisson broadcast channel defined in Section I. We show that superposition coding is optimal for most channel parameter values. Assume throughout and without loss of generality that $s_1 \leq s_2$ and define

$$I_i(q) = \lim_{\Delta \to 0} (1/\Delta) I(X; Y_i^\Delta), \; i = 1, 2,$$

for $X \sim \mathrm{Bern}(q)$, $q \in [0,1]$ where X and Y_i^Δ are the input and outputs of BPBC as depicted in Figure 1. In [1], Wyner showed that

$$\begin{aligned} I_i(q) = A_i\{&-(q+s_i)\log(q+s_i) \\ &+ q(1+s_i)\log(1+s_i) + (1-q)s_i\log(s_i)\}, \end{aligned}$$

and is maximized at

$$q_i = \frac{(1+s_i)^{1+s_i}}{e s_i^{s_i}} - s_i.$$

For a 2-receiver PBC, we say that receiver Y_2 is more capable than Y_1 if $I_2(q) \geq I_1(q)$ for every $q \in [0,1]$. Less noisy and effectively less noisy PBC classes can also be defined in a similar manner using the conditions for the DM–BC cases in Section II.

To establish the parameter ranges for which superposition coding is optimal for the PBC, define the following breakpoints of α:

$$\begin{aligned} \alpha_1 &= \frac{s_2 \log(1 + q_1/s_2) - q_1}{s_1 \log(1 + q_1/s_1) - q_1}, \\ \alpha_2 &= \frac{s_2 \log(1 + 1/s_2) - 1}{s_1 \log(1 + 1/s_1) - 1}, \\ \alpha_3 &= \frac{(1+s_2)\log\big((1+s_2)/(q_2+s_2)\big) - 1 + q_2}{(1+s_1)\log\big((1+s_1)/(q_2+s_1)\big) - 1 + q_2}. \end{aligned}$$

It can be shown that $q_1 \leq q_2$ and $\alpha_1 \leq \alpha_2 \leq \alpha_3$ for $s_1 \leq s_2$.

We now present our main result.

Theorem 1. For the PBC with $s_1 \leq s_2$:

1. If $0 \le \alpha \le \alpha_1$, Y_2 is effectively less noisy than Y_1 and the capacity region is the set of rate pairs (R_1, R_2) such that

$$R_1 \le I_1(\beta p_0 + \bar{\beta} p_1) - \beta I_1(p_0) - \bar{\beta} I_1(p_1),$$
$$R_2 \le \beta I_2(p_0) + \bar{\beta} I_2(p_1)$$

for some $0 \le \beta, p_0, p_1 \le 1$.

2. If $\alpha \ge \alpha_3$, Y_1 is effectively less noisy than Y_2 and the capacity region is the set of rate pairs (R_1, R_2) such that

$$R_1 \le \beta I_1(p_0) + \bar{\beta} I_1(p_1),$$
$$R_2 \le I_2(\beta p_0 + \bar{\beta} p_1) - \beta I_2(p_0) - \bar{\beta} I_2(p_1)$$

for some $0 \le \beta, p_0, p_1 \le 1$.

3. If $\alpha_2 \le \alpha < \alpha_3$, Y_1 is more capable than Y_2 and the capacity region is the set of rate pairs (R_1, R_2) such that

$$R_1 \le \beta_0 I_1(p_0) + \beta_1 I_1(p_1) + \beta_2 I_1(p_2),$$
$$R_2 \le I_2(p) - \beta_0 I_2(p_0) - \beta_1 I_2(p_1) - \beta_2 I_2(p_2),$$
$$R_1 + R_2 \le I_1(p)$$

for some $0 \le \beta_0, \beta_1, \beta_2, p_0, p_1, p_2 \le 1$, where $\beta_0 + \beta_1 + \beta_2 = 1$ and $p = \beta_0 p_0 + \beta_1 p_1 + \beta_2 p_2$.

To prove this theorem, we need the following lemma which characterizes the upper concave envelope of $(I_2(q) - I_1(q))$ and $(I_1(q) - I_2(q))$.

Define the following additional breakpoints of α:

$$\alpha_{01} = s_1/s_2,$$
$$\alpha_{02} = \frac{(1+s_2)\log(1+1/s_2) - 1}{(1+s_1)\log(1+1/s_1) - 1},$$
$$\alpha_{31} = (1+s_1)/(1+s_2).$$

It can be shown that $\alpha_{01} \le \alpha_{02} \le \alpha_1 \le \alpha_2 \le \alpha_3 \le \alpha_{31}$.

Lemma 2. Consider a PBC.

1. Let

$$r = \begin{cases} 0 & \text{if } 0 \le \alpha \le \alpha_{01}, \\ g_1^{-1}(\alpha) & \text{if } \alpha_{01} < \alpha < \alpha_2, \\ 1 & \text{if } \alpha \ge \alpha_2, \end{cases}$$

where

$$g_1(q) = \frac{s_2 \log(1 + q/s_2) - q}{s_1 \log(1 + q/s_1) - q}.$$

Note that $g_1^{-1}(\alpha_{01}) = 0$ and $g_1^{-1}(\alpha_2) = 1$. Then the upper concave envelope of $(I_2(q) - I_1(q))$ is

$$\mathfrak{C}[I_2(q) - I_1(q)] = \begin{cases} (q/r)(I_2(r) - I_1(r)) & \text{for } 0 \le q < r, \\ I_2(q) - I_1(q) & \text{for } q \ge r. \end{cases}$$

Indeed $\mathfrak{C}[I_2(q) - I_1(q)] > I_2(q) - I_1(q)$ for $0 < q < r$.

2. Let

$$t = \begin{cases} 0 & \text{if } 0 \le \alpha \le \alpha_{02}, \\ g_2^{-1}(\alpha) & \text{if } \alpha_{02} < \alpha < \alpha_{31}, \\ 1 & \text{if } \alpha \ge \alpha_{31}, \end{cases}$$

where

$$g_2(q) = \frac{(1+s_2)\log\big((1+s_2)/(q+s_2)\big) - 1 + q}{(1+s_1)\log\big((1+s_1)/(q+s_1)\big) - 1 + q}.$$

Note that $g_2^{-1}(\alpha_{02}) = 0$ and $g_2^{-1}(\alpha_{31}) = 1$. Then the upper concave envelope of $(I_1(q) - I_2(q))$ is

$$\mathfrak{C}[I_1(q) - I_2(q)] = \begin{cases} I_1(q) - I_2(q) & \text{for } 0 \le q \le t, \\ (1-q)(I_1(t) - I_2(t))/(1-t) & \text{for } q > t. \end{cases}$$

Indeed $\mathfrak{C}[I_1(q) - I_2(q)] > I_1(q) - I_2(q)$ for $q > t$.

In Figure 3, the area where $q \ge r$ and $q \le t$ are shaded on the left and on the right plot, respectively for $s_1 = 0.1, s_2 = 1$. Note that each shaded region denotes where $\mathfrak{C}[I_2(q) - I_1(q)] = I_2(q) - I_1(q)$ and $\mathfrak{C}[I_1(q) - I_2(q)] = I_1(q) - I_2(q)$ respectively.

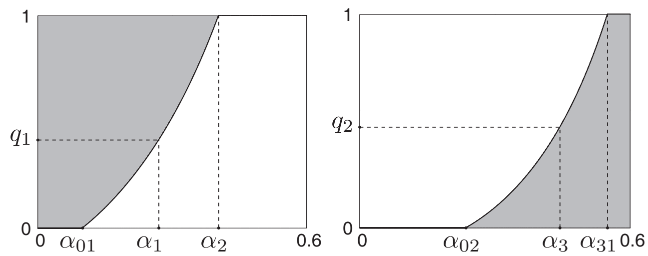

Fig. 3. Plots of q vs α for $s_1 = 0.1, s_2 = 1$. The shaded area are where $\mathfrak{C}[I_2(q) - I_1(q)] = I_2(q) - I_1(q)$ (left) and $\mathfrak{C}[I_1(q) - I_2(q)] = I_1(q) - I_2(q)$ (right).

To prove Lemma 2, we first prove the following:

Lemma 3. The function $(I_2(q) - I_1(q))$ can change concavity at most once in the range $q \in [0, 1]$. In particular, the function $I_2(q) - I_1(q)$ is (i) concave in $[0, 1]$ if $\alpha \le \alpha_{01}$, (ii) convex in $[0, 1]$ if $\alpha \ge \alpha_{31}$, or (iii) convex in $[0, \kappa]$ and concave in $[\kappa, 1]$ otherwise, where

$$\kappa = \frac{\alpha s_2 - s_1}{1 - \alpha} \in [0, 1].$$

Proof: The proof of this lemma follows immediately from the fact that the second derivative of $I_2(q) - I_1(q)$,

$$I_2''(q) - I_1''(q) = \frac{(\alpha - 1)q + \alpha s_2 - s_1}{(q + s_1)(q + s_2)}, \tag{5}$$

has a unique zero. ■

The proof of Lemma 2 is in Appendix B.

We are now ready to prove Theorem 1.

Proof:

1. For $0 \le \alpha \le \alpha_1$, we show that Y_2 is effectively less noisy than Y_1.

We first show that

$$\max_{q \in [0,1]} \{I_1(q) + \mathfrak{C}[\lambda I_2(q) - I_1(q)]\} = \max_{q \in [q_1, 1]} \{I_1(q) + \mathfrak{C}[\lambda I_2(q) - I_1(q)]\}.$$

For $\alpha' = \alpha/\lambda$ and $r' = r(\alpha')$, by Lemma 2, the derivative of $\mathfrak{C}[I_2(q) - I_1(q)/\lambda]$ is $I_2'(q) - I_1'(q)/\lambda \geq 0$ if $q_1 \geq r'$ and $(I_2(r') - I_1(r'))r'^{-1} \geq 0$, otherwise. Since $\mathfrak{C}[\lambda I_2(q) - I_1(q)]$ and $I_1(q)$ are concave and increasing at q_1, they are increasing in $[0, q_1]$.

We now show that the condition (3) holds with $\mathcal{P} = \{p(x)\colon p_X(1) \geq q_1\}$. Note that $0 < q_1 < 1$. By Lemma 2,

$$\mathfrak{C}[I_2(q) - I_1(q)] = I_2(q) - I_1(q) \text{ for } q \geq q_1$$

if $0 \leq \alpha \leq g_1(q_1) = \alpha_1$. As an example, see Figure 3.

The condition $0 \leq \alpha \leq \alpha_1$ is also necessary for Y_2 to be effectively less noisy than Y_1. This is because $\mathcal{P}$ must include $\mathrm{Bern}(q_1)$ and by Lemma 2, if $\mathfrak{C}[I_2(q_1) - I_1(q_1)] = I_2(q_1) - I_1(q_1)$, then α has to be in $[0, \alpha_1]$.

Hence for $0 \leq \alpha \leq \alpha_1$, the capacity region is given by (4). To complete the proof, we evaluate the mutual information terms in (4) for $p_U(0) = \beta, p_U(1) = 1 - \beta, p_{X|U}(1|0) = p_0, p_{X|U}(1|1) = p_1$.

2. For $\alpha \geq \alpha_3$, we can show that Y_1 is effectively less noisy than Y_2 in a similar fashion. We can show that $\mathcal{P} = \{p(x)\colon p_X(1) \leq q_2\}$ satisfies the condition (3) with $I(X;Y_1)$ and $I(X;Y_2)$ interchanged, and if (and only if) $\alpha \geq \alpha_3$,

$$\mathfrak{C}[I_1(q) - I_2(q)] = I_1(q) - I_2(q) \text{ for } q \leq q_2.$$

As an example, see Figure 3.

Hence the capacity region is given by (4) with Y_1 replacing Y_2 and R_1 replacing R_2. To complete the proof, we evaluate the resulting mutual information terms for $p_U(0) = \beta, p_U(1) = 1 - \beta, p_{X|U}(1|0) = p_0, p_{X|U}(1|1) = p_1$.

3. For $\alpha_2 \leq \alpha < \alpha_3$, we show that Y_1 is more capable than Y_2. By Lemma 2, $\mathfrak{C}[I_2(q) - I_1(q)] = 0$ for $q \in [0, 1]$ if and only if $\alpha \geq \alpha_2$. Since $I_2(q) - I_1(q) \leq 0$ if and only if $\mathfrak{C}[I_2(q) - I_1(q)] = 0$, we conclude that Y_1 is more capable than Y_2 if and only if $\alpha \geq \alpha_2$.

Hence the capacity region is given by (1). To complete the proof, we evaluate the mutual information terms for $p_U(0) = \beta_0, p_U(1) = \beta_1, p_U(2) = \beta_2, p_{X|U}(1|0) = p_0, p_{X|U}(1|1) = p_1, p_{X|U}(1|2) = p_2$. ■

Theorem 1 established the range of parameters for which superposition coding is optimal for the PBC. We now provide a more detailed breakdown of the ranges in which the channel is degraded, less noisy, more capable, and effectively less noisy. This shows that for some range of parameters, the PBC is effectively less noisy but is neither less noisy nor more capable.

Proposition 2. For the PBC, the following holds.

1. If $0 \leq \alpha \leq \alpha_{01}$, Y_2 is less noisy than Y_1 but the channel is not degraded.
2. If $\alpha_{01} < \alpha \leq \alpha_{02}$, Y_2 is more capable and effectively less noisy but not less noisy than Y_1.
3. If $\alpha_{02} < \alpha \leq \alpha_1$, Y_2 is effectively less noisy than Y_1 but not more capable.
4. If $\alpha_2 \leq \alpha < \alpha_3$, Y_1 is more capable than Y_2 but not effectively less noisy.
5. If $\alpha_3 \leq \alpha < \alpha_{31}$, Y_1 is more capable and effectively less noisy than Y_2 but not less noisy.
6. If $\alpha_{31} \leq \alpha < 1$, Y_1 is less noisy than Y_2 but the channel is not degraded.
7. For $\alpha \geq 1$, Y_2 is a degraded version of Y_1.

Proof: In Theorem 1, we proved the necessary and sufficient condition for effectively less noisy channels and Y_1 being more capable than Y_2. The necessary and sufficient condition for Y_2 being more capable than Y_1 can be shown in a similar manner. The condition for less noisy channels can be established using Lemma 3, which shows that the channel is less noisy if and only if $I_2(q) - I_1(q)$ is concave or convex. It can be also shown that the channel is degraded if and only if $\alpha \geq 1$. ■

Figure 4 illustrates the parameter ranges in Proposition 2. Note that in the range $\alpha_1 < \alpha < \alpha_2$, the channel does not fall into any of the classes for with superposition coding is optimal.

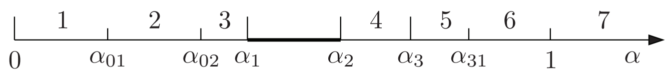

Fig. 4. Illustration of Proposition 2.

Remark 2. It can be shown that the fraction of the channel parameter space for which superposition coding is optimal approaches one; hence superposition coding is in a sense almost always optimal for the PBC. In comparison, the fraction of the parameter space for which the PBC is degraded is always bounded away from 1.

Remark 3. For every binary input BC, the essentially more capable condition in [8] coincides with the more capable condition.

IV. Gap between Marton and UV

We showed that the capacity region of the PBC is achieved using superposition coding in the ranges $\alpha \leq \alpha_1$ and $\alpha \geq \alpha_2$. We now consider the remaining range.

Proposition 3. For the PBC with $s_1 \leq s_2$, if $\alpha_1 < \alpha < \alpha_2$, neither Y_1 nor Y_2 is more capable or effectively less noisy than the other receiver.

This proposition follows immediately from the if and only if conditions established in Proposition 2.

We evaluated Marton's lower bound on the sum rate for $\alpha_1 \leq \alpha \leq \alpha_2$ using Theorem 3 in [10], which shows that for any binary input broadcast channel, randomized time-division strategy achieves the Marton sum rate. For the upper bound, we evaluated the UV upper bound [4] with $|\mathcal{U}| = |\mathcal{V}| = 3$ (In [11], it is shown $|\mathcal{U}|, |\mathcal{V}| \leq |\mathcal{X}| + 1$ is sufficient). Figure 5 plots the lower and upper bounds on the sum rate for $\alpha_1 \leq \alpha \leq \alpha_2$ when $s_1 = 0.1$ and $s_2 = 1$ ($\alpha_1 = 0.27, \alpha_2 = 0.4$). Note that for $0.27 \leq \alpha \leq 0.286$, the bounds on the sum rate coincide. Indeed it can be shown that superposition coding is optimal in this range. However, for the rest of the range there is a small gap between Marton inner bound and UV outer bound. In particular for $\alpha = 0.34$, the gap is approximately 0.0039.

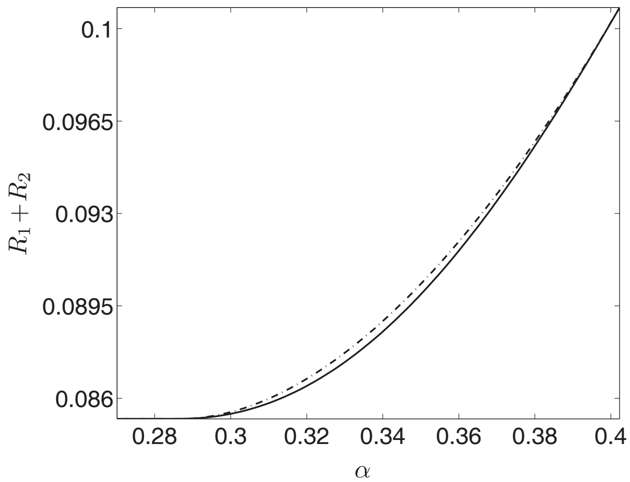

Fig. 5. Plots of Marton sum-rate (solid) and UV sum-rate (dash) for $s_1 = 0.1$, $s_2 = 1$, and $\alpha \in [\alpha_1, \alpha_2] = [0.27, 0.4]$.

REFERENCES

[1] A. Wyner, "Capacity and error exponent for the direct detection photon channel. i,ii," *IEEE Trans. Inform. Theory*, vol. 34, no. 6, Nov 1988.

[2] A. Lapidoth, I. Telatar, and R. Urbanke, "On wide-band broadcast channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 12, Dec 2003.

[3] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, 1979.

[4] C. Nair and A. El Gamal, "An outer bound to the capacity region of the broadcast channel," *IEEE Trans. Inform. Theory*, vol. 53, no. 1, Jan 2007.

[5] T. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. 18, no. 1, 1972.

[6] J. Körner and K. Marton, "Comparison of two noisy channels," *Colloquia Mathematica Societatis, Janos Bolyai*, vol. 16, pp. 411–424, 1977.

[7] A. Gamal, "The capacity of a class of broadcast channels," *IEEE Trans. Inform. Theory*, 1979.

[8] C. Nair, "Capacity regions of two new classes of two-receiver broadcast channels," *IEEE Trans. Inform. Theory*, vol. 56, no. 9, Sept 2010.

[9] H. G. Egglestone, *Convexity*. Cambridge University Press, Cambridge, 1958.

[10] Y. Geng, V. Jog, C. Nair, and Z. Wang, "An information inequality and evaluation of Marton's inner bound for binary input broadcast channels," *IEEE Trans. Inform. Theory*, vol. 59, no. 7, July 2013.

[11] C. Nair and V. Wang Zizhou, "On the inner and outer bounds for 2-receiver discrete memoryless broadcast channels," in *Inform. Theory and Applications Workshop*, Jan 2008.

APPENDIX A
PROOF OF PROPOSITION 1

Let $\bar{\mathcal{R}}$ be the set of rate pairs (R_1, R_2) such that

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

for some $p(u,x)$. Note that $\bar{\mathcal{R}}$ is in general larger than the UV outer bound; hence it is an outer bound on the capacity region of any broadcast channel.

We now show that every supporting hyperplane of $\bar{\mathcal{R}}$ intersects $\mathcal{R}$, i.e., for every $\lambda \ge 0$, there exists a rate pair $(R_1, R_2) \in \mathcal{R}$ such that $R_1 + \lambda R_2 = \max_{(r_1,r_2)\in\bar{\mathcal{R}}} r_1 + \lambda r_2$. Consider

$$\begin{aligned}
&\max_{(r_1,r_2)\in\bar{\mathcal{R}}} r_1 + \lambda r_2 \\
&= \begin{cases} \max_{p(u,x)} \{I(U;Y_1) + \lambda I(X;Y_2|U)\} & \text{if } 0 \le \lambda \le 1, \\ \max_{p(u,x)} \lambda\{I(U;Y_1) + I(X;Y_2|U)\} & \text{if } \lambda > 1 \end{cases} \\
&= \begin{cases} \max_{p(u,x):\, p(x)\in\mathcal{P}} \{I(U;Y_1) + \lambda I(X;Y_2|U)\} & \text{if } 0 \le \lambda \le 1, \\ \max_{p(u,x):\, p(x)\in\mathcal{P}} \lambda\{I(U;Y_1) + I(X;Y_2|U)\} & \text{if } \lambda > 1 \end{cases} \\
&= \max_{(R_1,R_2)\in\mathcal{R}} R_1 + \lambda R_2.
\end{aligned}$$

APPENDIX B
PROOF OF LEMMA 2

We prove part 1 of Lemma 2. The proof of part 2 follows similarly. Let $f_1(q) = (I_2(q) - I_1(q))/q$. We first show that $r = \arg\max_{q\in[0,1]} f_1(q)$. Consider

$$f_1'(q) = (\alpha - g_1(q))(q - s_1 \log(1 + q/s_1))q^{-2},$$

where $g_1(q)$ is defined in (2). Clearly, $f_1'(q)$ and $(\alpha - g_1(q))$ have the same sign for $0 < q < 1$. Also it can be shown that $g_1(q)$ is increasing for $q > 0$, so $\alpha - g_1(1) \le \alpha - g_1(0)$.

(i) If $\alpha \le g_1(0) = \alpha_{01}$, then $f_1'(q) \le 0$ in $[0,1]$.

(ii) If $\alpha \ge g_1(1) = \alpha_2$, then $f_1'(q) \ge 0$ in $[0,1]$.

(iii) Otherwise, $f_1'(q) \ge 0$ in $[0, g_1^{-1}(\alpha)]$ and $f_1'(q) \le 0$ in $[g_1^{-1}(\alpha), 1]$.

Let

$$L_1(q) = \begin{cases} (q/r)(I_2(r) - I_1(r)) & \text{if } 0 \le q < r, \\ I_2(q) - I_1(q) & \text{if } q \ge r. \end{cases}$$

Then $I_2(q) - I_1(q) \le L_1(q)$ because $f_1(r) \ge f_1(q)$ for any q.

On the other hand, $\mathfrak{C}[I_2(q) - I_1(q)] \ge L_1(q)$ because for $0 \le q \le r$, $L_1(q)$ is a convex combination of $I_2(r) - I_1(r)$ and $I_2(0) - I_1(0)$. To complete the proof we now show that $\mathfrak{C}[L_1(q)] = L_1(q)$. By Lemma 3, $I_2(q) - I_1(q)$ is concave in $[\kappa, 1]$ for some κ. To show $I_2(q) - I_1(q)$ is concave in $[r, 1]$, we show $\kappa \le r$, i.e. $f_1'(\kappa) \ge 0$. Consider

$$\begin{aligned}
f_1'(q) &= \frac{I_2'(q) - I_1'(q)}{q} - \frac{I_2(q) - I_1(q)}{q^2} \\
&= (I_2'(q) - I_1'(q))/q - (I_2'(u) - I_1'(u))/q
\end{aligned}$$

for some $u \le q$. By Lemma 3, $I_2''(q) - I_1''(q) > 0$ in $[0, \kappa)$. Thus $f'(\kappa) \ge 0$. We conclude that $L_1(q)$ is concave because it is concave in $[0, r]$ and $[r, 1]$ and differentiable at r.