Superposition Coding Is Almost Always Optimal for the Poisson Broadcast Channel

Hyeji Kim, *Student Member, IEEE*, Benjamin Nachman, and Abbas El Gamal, *Fellow, IEEE*

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, and more capable). In particular, we introduce the notion of effectively less noisy broadcast channel and show that it implies less noisy but is not in general implied by 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.

Index Terms—Broadcast channel, superposition coding, poisson channel, generalized binary skew channel.

I. INTRODUCTION

THE continuous-time Poisson channel is a canonical model of the point to point optical communication channel in the low power regime [1]–[3]. The capacity of this channel was established using different approaches by Kabanov [4], Davis [5], and Wyner [6], [7]. In particular, Wyner [6], [7] established the capacity using an elementary method in which the capacity is shown to be the the limit of the capacity of a certain memoryless binary channel. Wyner's approach spurred several generalizations to multiple user Poisson channels. In [8] Lapidoth and Shamai established the capacity region of the Poisson multiple-access channel. In [9], Lai, Liang, and Shamai studied the Poisson interference channel. In [10], Bross, Lapidoth, and Shamai studied the Poisson channel with side information at the transmitter. In [11], Lapidoth, Telatar, and Erbanke studied the Poisson broadcast channel and established the condition under which the channel is degraded; hence the capacity region is achieved using superposition coding [12].

In this paper, which is an expanded and a more complete version of [13], we show that for the Poisson broadcast channel, superposition coding is optimal much beyond
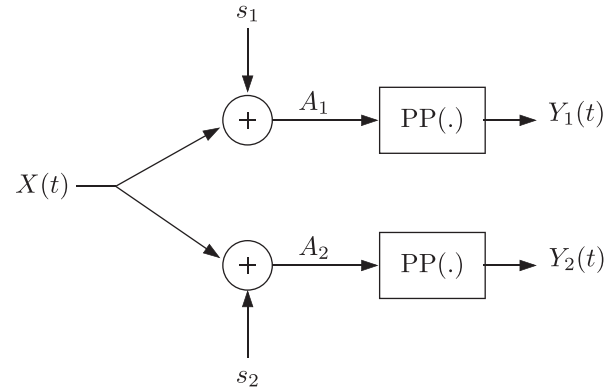
Fig. 1.   Two receiver continuous time Poisson broadcast channel.

the parameter ranges for which the channel is degraded. We consider the 2-receiver continuous-time Poisson broadcast channel (P-BC) depicted in Figure 1. The channel input signal $X(t) \in [0, 1]$, $t \geq 0$, that is, we assume a *peak power* constraint of 1 on $X(t)$. Given $X(t) = x(t)$, the output $Y_i(t)$ is a Poisson process (PP) with instantaneous rate $A_i(x(t) + s_i)$ for $i = 1, 2$, i.e., for $0 \leq w \leq w + \tau \leq T$,

$$P\{Y_i(w + \tau) - Y_i(w) = k | X(t) = x(t), \ t \in [0, T]\}$$
$$= \frac{\Gamma^k(w, \tau)}{k!} e^{-\Gamma(w, \tau)}, \quad k \in \mathbb{N},$$

where

$$\Gamma(w, \tau) = \int_w^{w+\tau} A_i(x(t) + s_i) \, dt.$$

The parameter A_i is the channel gain for receiver $i = 1, 2$. The parameter $s_i \geq 0$ is the rate of the input referred dark noise for receiver i.

We consider the private message setting in which the sender X wishes to communicate a message M_1 to receiver Y_1 at rate R_1 and a message M_2 to receiver Y_2 at rate R_2, where M_1 and M_2 are independent and uniformly distributed over $[1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$. The results we establish on the optimality of superposition coding can be readily extended to the case with common message [14]. We define a $(2^{nR_1}, 2^{nR_2}, n)$ code, achievability, and the capacity region for this setting in the standard way [14].

As in [11], we use Wyner's approach for the point-to-point Poisson channel to study the capacity region of the Poisson broadcast channel. As depicted in Figure 2-(a), time is quantized into intervals of length $\Delta > 0$ and the continuous time P-BC is approximated in each interval by the binary
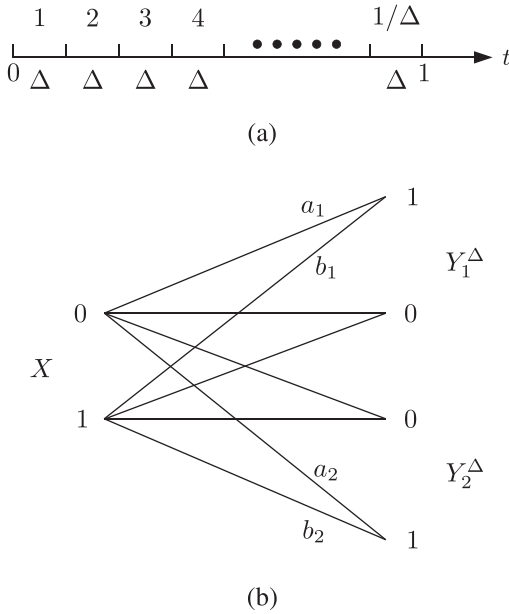
Fig. 2. (a) Time quantization. (b) Binary P-BC.

memoryless broadcast channel (binary P-BC) depicted in Figure 2-(b), with transition probabilities:

$$a_1 = A_1 s_1 \Delta + O(\Delta^2), \quad a_2 = A_2 s_2 \Delta + O(\Delta^2),$$
$$b_1 = A_1(1 + s_1)\Delta + O(\Delta^2), \quad b_2 = A_2(1 + s_2)\Delta + O(\Delta^2). \tag{1}$$

Following Wyner's arguments, Lapidoth, Telatar, and Urbanke [11] showed that the capacity region of the P-BC is equal to the capacity region of the $1/\Delta$-extension of this binary channel as Δ tends to zero.

In this paper, we show the surprising fact that superposition coding is optimal for almost all channel parameter values. We find explicit analytical expressions for the ranges of parameter values in which the channel is less noisy and more capable [15]. We then introduce the new class of *effectively less noisy* broadcast channels for which superposition coding is optimal and show that it includes the less noisy class but is not strictly included in the more capable class. The key idea is that the less noisy condition needs to hold only for channel input distributions that attain the maximum weighted sum rate. We find explicit analytical expressions for the ranges of parameter values for which the P-BC is effectively less noisy. By further strengthening the effectively less noisy condition, we show numerically that superposition coding can be optimal even when the channel is not more capable or effectively less noisy. Finally we show that for the remaining set of parameter values. there is a gap between Marton's inner bound [16] and the UV outer bound [17].

The rest of the paper is organized as follows. In Section II we review superposition coding and the optimal superposition coding inner bound for binary input broadcast channels, and summarize known classes of broadcast channels for which superposition coding inner bound is tight. In Section III, we establish the parameter ranges for which a P-BC is less noisy and more capable. In Section IV, we introduce the new class of effectively less noisy broadcast channels for which

superposition coding is also optimal, and obtain the parameter range for which a P-BC is effectively less noisy. In Section V, we show that for certain range of parameter values, there is a gap between Marton's inner bound and the UV outer bound. Hence, the capacity region of the P-BC is still not known in general. In Section VI, we extend our results to the average power constraint case. Finally in Section VII, we remark on the optimality of superposition coding for general binary input broadcast channels. We demonstrate via an example that our intuition about when superposition coding is optimal for the broadcast channel can be quite misleading.

II. SUPERPOSITION CODING INNER BOUND

Consider a 2-receiver discrete memoryless broadcast channel $p(y_1, y_2|x)$. The superposition coding scheme [12] is motivated by broadcast channels for which one receiver is "stronger" than the other. This suggests a layered coding approach in which the weaker receiver (say Y_2) recovers only its own message M_2 carried by the auxiliary random variable U, while the stronger receiver (Y_1) recovers both messages (M_1, M_2) carried by X. This coding scheme leads to the inner bound on the capacity region of the general discrete memoryless broadcast channel (DM-BC) $p(y_1, y_2|x)$ that consists of all rate pairs (R_1, R_2) such that

$$R_1 < I(X; Y_1|U),$$
$$R_2 < I(U; Y_2),$$
$$R_1 + R_2 < I(X; Y_1) \tag{2}$$

for some pmf $p(u, x)$, and $|\mathcal{U}| \leq |\mathcal{X}| + 1$ [18].

Let $\mathcal{R}_1$ denote the region (2). Note that a second superposition coding inner bound can be readily obtained by exchanging Y_1 and Y_2 and R_1 and R_2 in (2). Denoting the second region by $\mathcal{R}_2$, it can be shown that if the capacity of the channel $p(y_1|x)$, C_1, is larger than the capacity of the channel $p(y_2|x)$, C_2, then $\mathcal{R}_2 \subseteq \{(R_1, R_2): R_1/C_1 + R_2/C_2 \leq 1\}$.

Remark 1: In [19], Wang, Sasoglu, Bandemer, and Kim compared two superposition encoding schemes for the broadcast channel. The first is the UX scheme above, and the second is the (U, V) scheme in which M_1 is carried by V, M_2 is carried by U, and X is a function of (U, V). The weaker receiver Y_2 again recovers M_2 carried by U and the stronger receiver Y_1 recovers M_1 carried by V. They showed that the optimal inner bound achieved by the UV scheme can be strictly larger than that achieved by the UX scheme. It turns out, however, that if the broadcast channel has binary input, the optimal inner bound (assuming $C_1 \geq C_2$) is the convex hull of $(\mathcal{R}_1 \cup \{(0, C_2)\})$ [20]. Since in this paper we are concerned with the binary P-BC, which has binary input and outputs, we focus only on conditions under which either $\mathcal{R}_1$ or $\mathcal{R}_2$ is optimal.

It is well known that region $\mathcal{R}_1$ is tight for the following classes of DM-BC.

Definition 1 (Degraded Broadcast Channel [12]): For a DM-BC $p(y_1, y_2|x)$ Y_2 is said to be a degraded version of Y_1 if there exists a random variable Y_1' such that $Y_1'|\{X = x\} \sim p_{Y_1|X}(y_1'|x)$, i.e., Y_1' has the same conditional pmf as Y_1 (given X) and $X \rightarrow Y_1' \rightarrow Y_2$ form a Markov chain.

Definition 2 (Less Noisy Channel [15]): For a DM-BC $p(y_1, y_2|x)$ receiver Y_1 is said to be *less noisy* than receiver Y_2 if $I(U; Y_1) \geq I(U; Y_2)$ for all $p(u, x)$.

Van-Dijk [21] showed that receiver Y_1 is less noisy than receiver Y_2 if $I(X; Y_1) - I(X; Y_2)$ is concave in $p(x)$, or equivalently, $I(X; Y_1) - I(X; Y_2)$ is equal to its upper concave envelope $\mathfrak{C}[I(X; Y_1) - I(X; Y_2)]$ (the smallest concave function that is greater than or equal to $I(X; Y_1) - I(X; Y_2)$). As we will see, this alternative condition of less noisy is significantly simpler to compute than the original condition.

Definition 3 (More Capable Channel [15]): For a DM-BC $p(y_1, y_2|x)$ receiver Y_1 is said to be *more capable* than receiver Y_2 if $I(X; Y_1) \geq I(X; Y_2)$ for all $p(x)$.

The more capable condition can also be recast in terms of the concave envelope: Receiver Y_1 is more capable than Y_2 if $\mathfrak{C}[I(X; Y_2) - I(X; Y_1)] = 0$ for every $p(x)$.

It is also well known that degraded implies less noisy which implies more capable [15], but the converses do not always hold. In [22], Nair generalized the notions of less noisy and more capable. Let $\mathcal{P}_o$ be a class of pmfs $p(u, v, x)$ such that for any triple of random variables $(U, V, X) \sim p(u, v, x)$, there exists a pmf $q(\tilde{u}, \tilde{v}, x)$ such that

$$q(x) \in \mathcal{P}_o,$$
$$I(V; Y_1)_p \leq I(\tilde{V}; Y_1)_q,$$
$$I(U; Y_2)_p \leq I(\tilde{U}; Y_2)_q,$$
$$I(V; Y_1)_p + I(X; Y_2|V)_p \leq I(\tilde{V}; Y_1)_q + I(X; Y_2|\tilde{V})_q,$$
$$I(U; Y_2)_p + I(X; Y_1|U)_p \leq I(\tilde{U}; Y_2)_q + I(X; Y_1|\tilde{U})_q.$$
$$\tag{3}$$

The notation $I(V; Y_1)_p$ refers to the mutual information between V and Y_1 when the input is generated according to $p(u, v, x)$.

Definition 4 (Essentially Less Noisy [22]): For a DM-BC $p(y_1, y_2|x)$, Y_1 is said to be *essentially less noisy* than Y_2 if there exists a sufficient class of pmfs $\mathcal{P}_o$ as defined in (3) such that $I(U; Y_1) \geq I(U; Y_2)$ for every $p(x) \in \mathcal{P}_o$ and all $U \rightarrow X \rightarrow (Y_1, Y_2)$.

Definition 5 (Essentially More Capable [22]): For a DM-BC $p(y_1, y_2|x)$, Y_1 is said to be *essentially more capable* than Y_2 if there exists a sufficient class of pmfs $\mathcal{P}_o$ as defined in (3) such that $I(X; Y_1|U) \geq I(X; Y_2|U)$ for every $p(x) \in \mathcal{P}_o$ and all $U \rightarrow X \rightarrow (Y_1, Y_2)$.

It can be easily seen from the definitions that less noisy implies essentially less noisy and more capable implies essentially more capable. However the converses do not always hold. Also it is shown in [22] that essentially less noisy neither implies nor is implied by essentially more capable. The capacity region for the essentially less noisy and the essentially more capable classes [23] is region $\mathcal{R}_1$ as defined in (2). Also note that when the channel is essentially less noisy, the sum bound in (2) is always inactive.

III. LESS NOISY AND MORE CAPABLE P-BC

We evaluate the conditions for the less noisy and more capable stated in the previous section for the P-BC and show that the P-BC is less noisy for almost all channel parameter values. We assume without loss of generality that $s_1 \leq s_2$. The results for $s_1 > s_2$ can be similarly established. Also assume that $A_1 = \alpha$ and $A_2 = 1$. The result for $A_2 \neq 1$ can be established using the fact that the capacity for $(A_1, A_2) = (a_1, a_2)$ is equal to a_2 times the capacity region for $(A_1, A_2) = (a_1/a_2, 1)$.

In [11], Lapidoth et al. showed that Y_2 is a degraded version of Y_1 if $\alpha \geq 1$. This can be seen either by directly inspecting the P-BC channel model or its binary P-BC counterpart.

We now determine the conditions under which the P-BC is less noisy and more capable.

We use the binary P-BC with the parameters given in (1) to extend the definitions of less noisy and more capable to the P-BC. Let

$$I_i(q) = \lim_{\Delta \to 0} \frac{1}{\Delta} I(X; Y_i^{\Delta}), \quad X \sim \text{Bern}(q), \quad i = 1, 2,$$

where X and Y_i^{Δ} are the input and outputs of binary P-BC as depicted in Figure 2-(b). Using Wyner's results [6], [7], it immediately follows that

$$I_1(q) = \alpha \big(-(q + s_1) \log(q + s_1) + q(1 + s_1) \log(1 + s_1) + (1 - q)s_1 \log(s_1) \big),$$
$$I_2(q) = -(q + s_2) \log(q + s_2) + q(1 + s_2) \log(1 + s_2) + (1 - q)s_2 \log(s_2), \tag{4}$$

and that $I_i(q)$ is maximized at

$$q_i = \frac{(1 + s_i)^{1 + s_i}}{e s_i^{s_i}} - s_i \tag{5}$$

for $i = 1, 2$.

We define the less noisy condition for P-BC as follows.

Definition 6: For the 2-receiver P-BC, receiver Y_1 is less noisy than Y_2 if $I_1(q) - I_2(q)$ is concave in $q \in [0, 1]$, i.e., if $\mathfrak{C}[I_1(q) - I_2(q)] = I_1(q) - I_2(q)$.

Similarly we define the more capable condition for P-BC as follows.

Definition 7: For the 2-receiver P-BC, receiver Y_1 is more capable than Y_2 if $I_1(q) - I_2(q) \geq 0$ for every $q \in [0, 1]$, i.e., if $\mathfrak{C}[I_2(q) - I_1(q)] = 0$.

To establish the parameter ranges for less noisy and more capable P-BC, we define the following breakpoints of α:

$$\alpha_1 = \frac{1 + s_1}{1 + s_2},$$
$$\alpha_2 = \frac{s_2 \log(1 + 1/s_2) - 1}{s_1 \log(1 + 1/s_1) - 1},$$
$$\alpha_3 = \frac{(1 + s_2) \log(1 + 1/s_2) - 1}{(1 + s_1) \log(1 + 1/s_1) - 1},$$
$$\alpha_4 = \frac{s_1}{s_2}.$$

We also need the following lemma which characterizes the upper concave envelope of $(I_1(q) - I_2(q))$ and $(I_2(q) - I_1(q))$.

Lemma 1: Consider a P-BC.

1. Let

$$t = \begin{cases} 0 & \text{if } 0 \leq \alpha \leq \alpha_3, \\ g_1^{-1}(\alpha) & \text{if } \alpha_3 < \alpha < \alpha_1, \\ 1 & \text{if } \alpha \geq \alpha_1, \end{cases} \tag{6}$$
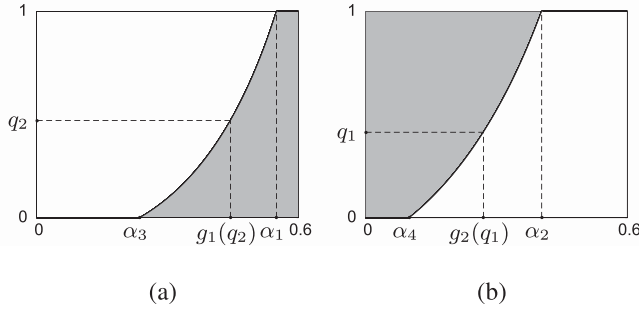
(a)　　　　　　　　(b)

Fig. 3.　(a) Plot of t vs α for $s_1 = 0.1, s_2 = 1$. The shaded area is where $\mathfrak{C}[I_1(q) - I_2(q)] = I_1(q) - I_2(q)$, i.e., $0 \leq q \leq t$. (b) Plot of r vs α for $s_1 = 0.1, s_2 = 1$. The shaded area is where $\mathfrak{C}[I_2(q) - I_1(q)] = I_2(q) - I_1(q)$, i.e., $r \leq q \leq 1$.

where

$$g_1(x) = \frac{(1 + s_2) \log \left((1 + s_2)/(x + s_2)\right) - 1 + x}{(1 + s_1) \log \left((1 + s_1)/(x + s_1)\right) - 1 + x}. \quad (7)$$

Then, the upper concave envelope of $(I_1(q) - I_2(q))$ is

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

2. Let

$$r = \begin{cases} 0 & \text{if } 0 \leq \alpha \leq \alpha_4, \\ g_2^{-1}(\alpha) & \text{if } \alpha_4 < \alpha < \alpha_2, \\ 1 & \text{if } \alpha \geq \alpha_2, \end{cases} \quad (8)$$

where

$$g_2(x) = \frac{s_2 \log(1 + x/s_2) - x}{s_1 \log(1 + x/s_1) - x}. \quad (9)$$

Then, the upper concave envelope of $(I_2(q) - I_1(q))$ is

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

The proof of Lemma 1 is in Appendix A. Figure 3-(a) plots t vs α. The shaded area is where $\mathfrak{C}[I_1(q) - I_2(q)] = I_1(q) - I_2(q)$. Figure 3-(b) plots r vs α. The shaded region is where $\mathfrak{C}[I_2(q) - I_1(q)] = I_2(q) - I_1(q)$. Figure 4 plots $I_1(q) - I_2(q)$ and $\mathfrak{C}[I_1(q) - I_2(q)]$ for $\alpha = 0.3\alpha_3 + 0.7g_1(q_2)$ and $\alpha = g_1(q_2)$ for $(s_1, s_2) = (0.1, 1)$. Figure 5 plots $I_2(q) - I_1(q)$ and $\mathfrak{C}[I_2(q) - I_1(q)]$ for $\alpha = 0.3\alpha_4 + 0.7g_2(q_1)$ and $\alpha = g_2(q_1)$ for $(s_1, s_2) = (0.1, 1)$.

We are now ready to state the conditions for less noisy and more capable.

Theorem 1: For $s_1 \leq s_2$:

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

$$R_1 \leq \beta I_1(p_0),$$
$$R_2 \leq I_2(\beta p_0 + \bar{\beta}) - \beta I_2(p_0) \quad (10)$$

for some $0 \leq \beta, p_0 \leq 1$ and $\bar{\beta} = 1 - \beta$.
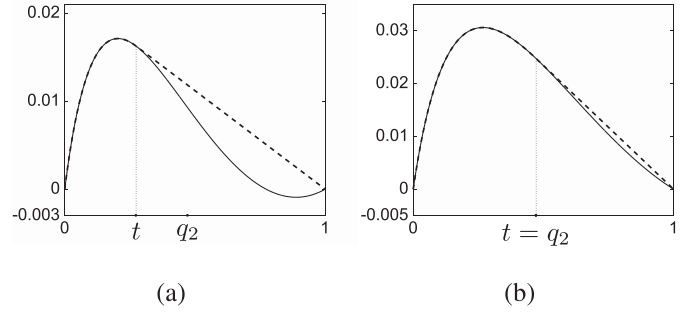

(a)　　　　　　　　(b)

Fig. 4.　Plots of $I_1(q) - I_2(q)$ (solid) and $\mathfrak{C}[I_1(q) - I_2(q)]$ (dashed line) vs q for (a) $\alpha = 0.3\alpha_3 + 0.7g_1(q_2)$ and (b) $\alpha = g_1(q_2)$ $(s_1 = 0.1, s_2 = 1)$.
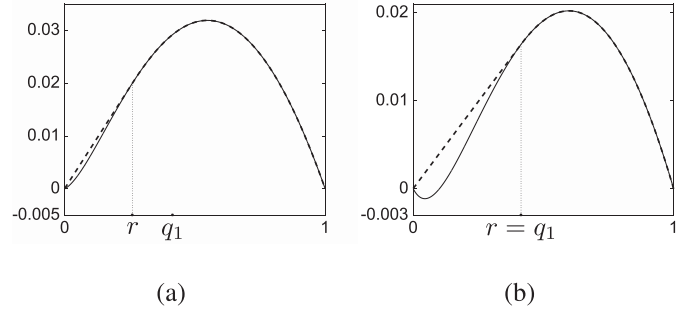

(a)　　　　　　　　(b)

Fig. 5.　Plots of $I_2(q) - I_1(q)$ (solid) and $\mathfrak{C}[I_2(q) - I_1(q)]$ (dashed line) vs q for (a) $\alpha = 0.3\alpha_4 + 0.7g_2(q_1)$ and (b) $\alpha = g_2(q_1)$ $(s_1 = 0.1, s_2 = 1)$.

2. If $\alpha_2 \leq \alpha \leq \alpha_1$, Y_1 is more capable than Y_2 and the capacity region is the set of rate pairs (R_1, R_2) such that

$$R_1 \leq \sum_{i \in \{0,1,2\}} \beta_i I_1(p_i),$$
$$R_2 \leq I_2(p) - \sum_{i \in \{0,1,2\}} \beta_i I_2(p_i),$$
$$R_1 + R_2 \leq I_1(p) \quad (11)$$

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

3. If $0 \leq \alpha \leq \alpha_4$, Y_2 is less noisy than Y_1 and the capacity region is the set of rate pairs (R_1, R_2) such that

$$R_1 \leq I_1(\beta p_0 + \bar{\beta}) - \beta I_1(p_0),$$
$$R_2 \leq \beta I_2(p_0) \quad (12)$$

for some $0 \leq \beta, p_0 \leq 1$.

4. If $\alpha_4 \leq \alpha \leq \alpha_3$, Y_2 is more capable than Y_1 and the capacity region is the set of rate pairs (R_1, R_2) such that

$$R_1 \leq I_1(p) - \sum_{i \in \{0,1,2\}} \beta_i I_1(p_i),$$
$$R_2 \leq \sum_{i \in \{0,1,2\}} \beta_i I_2(p_i),$$
$$R_1 + R_2 \leq I_2(p)$$

for some $0 \leq \beta_0, \beta_1, \beta_2, p_0, p_1, p_2 \leq 1$, where $\beta_0 + \beta_1 + \beta_2 = 1$ and $p = \beta_0 p_0 + \beta_1 p_1 + \beta_2 p_2$.
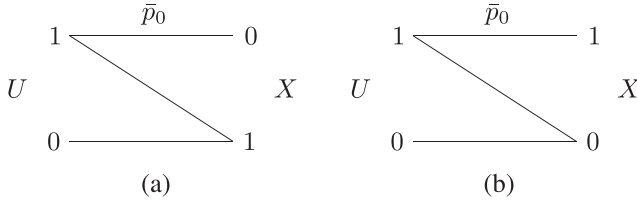
Fig. 6. (a) $p(x|u)$ for Y_1 less noisy than Y_2. (b) $p(x|u)$ for Y_2 less noisy than Y_1.

The capacity region for Y_1 less noisy than Y_2 is the superposition rate region $\mathcal{R}_1$ without the sum rate bound for $U \sim \text{Bern}(\beta)$ and $p(x|u)$ is the Z-channel shown in Figure 6-(a) where $\beta, p_0 \in [0, 1]$; and the capacity region for Y_2 less noisy than Y_1 is $\mathcal{R}_2$ for $U \sim \text{Bern}(\beta)$ and $p(x|u)$ is the Z-channel shown in Figure 6-(b) where $\beta, p_0 \in [0, 1]$.

The capacity region for Y_1 more capable than Y_2 is the superposition rate region $\mathcal{R}_1$ with $U \in \{0, 1, 2\}$, $p_U(j) = \beta_j$, and $X|\{U = j\} \sim \text{Bern}(p_j)$, $j = 0, 1, 2$; and the capacity region for Y_2 more capable than Y_1 is the superposition rate region $\mathcal{R}_2$ with $U \in \{0, 1, 2\}$, $p_U(j) = \beta_j$, and $X|\{U = j\} \sim \text{Bern}(p_j)$ for $j = 0, 1, 2$.

Figure 7 illustrates the ranges of α for which the P-BC is degraded, less noisy and more capable. From Theorem 1 the P-BC is less noisy if $\alpha \leq \alpha_4$ or $\alpha \geq \alpha_1$ and more capable if $\alpha \leq \alpha_3$ or $\alpha \geq \alpha_2$.

Proof of Theorem 1: We first find the condition on α for one receiver being less noisy or more capable than the other receiver.

1. By part 1 of Lemma 1, $\mathfrak{C}[I_1(q) - I_2(q)] = I_1(q) - I_2(q)$ if and only if $\alpha \geq \alpha_1$. Thus Y_1 is less noisy than Y_2 iff $\alpha \geq \alpha_1$.
2. By part 2 of Lemma 1, $\mathfrak{C}[I_2(q) - I_1(q)] = 0$ if and only if $\alpha \geq \alpha_2$. Thus Y_1 is more capable than Y_2 iff $\alpha \geq \alpha_2$.
3. By part 2 of Lemma 1, $\mathfrak{C}[I_2(q) - I_1(q)] = I_2(q) - I_1(q)$ if and only if $0 \leq \alpha \leq \alpha_4$. Thus Y_2 is less noisy than Y_1 iff $0 \leq \alpha \leq \alpha_4$.
4. By part 1 of Lemma 1, $\mathfrak{C}[I_1(q) - I_2(q)] = 0$ if and only if $0 \leq \alpha \leq \alpha_3$. Thus Y_2 is more capable than Y_1 iff $0 \leq \alpha \leq \alpha_3$.

We now obtain the capacity region for P-BC such that Y_1 is more capable than Y_2 (part 2) and for P-BC such that Y_1 is less noisy than Y_2 (part 1). By exchanging Y_1 and Y_2 and $I_1(\cdot)$ and $I_2(\cdot)$, the capacity expression for part 4 and part 3 of Theorem 1 can be obtained similarly.

For the $1/\Delta$-extension binary P-BC shown in Figure 2-(b), superposition coding inner bound is the set of rate pairs (R_1, R_2) such that

$$
\begin{aligned}
R_1 &< (1/\Delta)I(X; Y_1^\Delta|U), \\
R_2 &< (1/\Delta)I(U; Y_2^\Delta), \\
R_1 + R_2 &< (1/\Delta)I(X; Y_1^\Delta)
\end{aligned}
\tag{13}
$$

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

Let $U \in \{0, 1, 2\}$ where $p_U(i) = \beta_i$ and $p_{X|U}(1|i) = p_i \in [0, 1]$. As $\Delta \to 0$, the region in (13) is equivalent to the region in (11) in Theorem 1. If a P-BC is more capable, the region in (11) is indeed the capacity region. This is because the

UV outer bound for $1/\Delta$-extension binary P-BC as $\Delta \to 0$ reduces to the inner bound in (11) under the condition for more capable P-BC.

For part 1 of Theorem 1, note that the capacity region is the region in (11) with an inactive sum bound, i.e., the set of rate pairs (R_1, R_2) that satisfy

$$
R_1 \leq \sum_{i \in \{0,1,2\}} \beta_i I_1(p_i),
$$

$$
R_2 \leq I_2\left(\sum_{i \in \{0,1,2\}} \beta_i p_i\right) - \sum_{i \in \{0,1,2\}} \beta_i I_2(p_i)
\tag{14}
$$

for some $\beta_i, p_i \in [0, 1]$ such that $\sum_{i \in \{0,1,2\}} \beta_i = 1$. Let $\mathcal{R}'$ and $\mathcal{R}''$ denote the region in (10) and (14) respectively. Note that $\mathcal{R}'$ is the set of rate pairs (R_1, R_2) that satisfy inequalities in (14) for $(\beta_0, \beta_1, \beta_2) = (\beta, \bar{\beta}, 0)$ for some $\beta \in [0, 1]$ and $p_0 \in [0, 1]$ and $p_1 = 1$. Thus $\mathcal{R}' \subseteq \mathcal{R}''$. We now show that every supporting hyperplane of $\mathcal{R}''$ intersects $\mathcal{R}'$, i.e., $\max_{(R_1, R_2) \in \mathcal{R}''}(\lambda R_1 + R_2) \leq \max_{(r_1, r_2) \in \mathcal{R}'}(\lambda r_1 + r_2)$. Consider

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

$$
= \max_{\beta_i, p_i \in [0,1]} \left(I_2\left(\sum_{i \in \{0,1,2\}} \beta_i p_i\right) + \sum_{i \in \{0,1,2\}} \beta_i(\lambda I_1(p_i) - I_2(p_i)) \right)
$$

$$
= \max_{p \in [0,1]} \left(I_2(p) + \max_{\substack{\beta_i, p_i: \\ \sum \beta_i p_i = p}} \sum_{i \in \{0,1,2\}} \beta_i(\lambda I_1(p_i) - I_2(p_i)) \right)
$$

$$
= \max_{p \in [0,1]} \left(\mathfrak{C}[\lambda I_1(p) - I_2(p)] + I_2(p) \right)
$$

$$
\overset{(a)}{=} \max\left(\max_{p \in [0, t']} (\lambda I_1(p) - I_2(p) + I_2(p)), \right.
$$

$$
\left. \max_{p \in [t', 1]} \left(\frac{(1 - p)(\lambda I_1(t') - I_2(t'))}{1 - t'} + I_2(p) \right) \right)
\tag{15}
$$

$$
\overset{(b)}{\leq} \max_{\beta, p_0 \in [0,1]} \left(\beta(\lambda I_1(p_0) - I_2(p_0)) + I_2(\beta p_0 + \bar{\beta}) \right)
\tag{16}
$$

$$
= \max_{(r_1, r_2) \in \mathcal{R}'} (\lambda r_1 + r_2).
$$

Step (a) holds by Lemma 1. To prove step (b), note that the two terms in (15) are obtained by letting $(\beta, p_0) = (1, p)$ and $(\beta, p_0) = ((1 - p)/(1 - t'), t')$ in (16).

We now show that $\mathcal{R}'$ is convex. Suppose $\mathcal{R}'$ is not convex, i.e., $\text{co}\{\mathcal{R}'\} \neq \mathcal{R}'$ where $\text{co}\{\mathcal{R}'\}$ denotes the convex hull of $\mathcal{R}'$. Then there exists a rate pair $(r_1, r_2) \in \text{co}\{\mathcal{R}'\}$ on the boundary of $\text{co}\{\mathcal{R}'\}$, i.e., $\mu r_1 + r_2 = \max_{(R_1, R_2) \in \text{co}\{\mathcal{R}'\}}(\mu R_1 + R_2)$ for some $\mu \geq 0$ such that $(r_1, r_2) \notin \mathcal{R}'$. Note that $(r_1, r_2) = \eta(r_{10}, r_{20}) + (1 - \eta)(r_{11}, r_{21})$ for some $0 < \eta < 1$ and $(r_{10}, r_{20}), (r_{11}, r_{21}) \in \mathcal{R}'$. Since $\max_{(R_1, R_2) \in \text{co}\{\mathcal{R}'\}}(\mu R_1 + R_2) = \max_{(R_1, R_2) \in \mathcal{R}'}(\mu R_1 + R_2)$, the two rate pairs (r_{10}, r_{20}) and (r_{11}, r_{21}) satisfy

$$
\mu r_{10} + r_{20} = \mu r_{11} + r_{21} = \max_{(R_1, R_2) \in \mathcal{R}'} (\mu R_1 + R_2).
\tag{17}
$$

We now show that the equality (17) cannot hold for $(r_{10}, r_{20}) \neq (r_{11}, r_{21})$, i.e., there exists a unique rate pair (R_1, R_2) such that $\mu R_1 + R_2 = \max_{(r_1, r_2) \in \mathcal{R}'}(\mu r_1 + r_2)$.
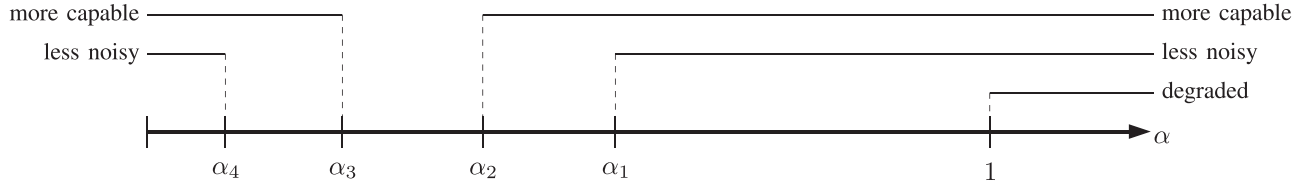
Fig. 7. Illustration of the ranges of α for which the channel is degraded, less noisy, and more capable as stated in Theorem 1.

Consider

$$\max_{(r_1, r_2) \in \mathcal{R}'} \left(\mu r_1 + r_2 \right)$$
$$= \max_{\beta, p_0 \in [0,1]} \left(I_2(\beta p_0 + \bar{\beta}) + \mu \beta I_1(p_0) - \beta I_2(p_0) \right).$$

We show that (β, p_0) that achieves the maximum is unique. Note that

$$\max_{\beta, p_0} \left(I_2(\beta p_0 + \bar{\beta}) + \mu \beta I_1(p_0) - \beta I_2(p_0) \right)$$
$$= \max_p \left(I_2(p) + \mathfrak{C}[\mu I_1(p) - I_2(p)] \right).$$

Since $I_2(p)$ is strictly concave, $I_2(p) + \mathfrak{C}[\mu I_1(p) - I_2(p)]$ is strictly concave. Let p^* be the unique solution that maximizes $I_2(p) + \mathfrak{C}[\mu I_1(p) - I_2(p)]$. Then

$$\max_{\beta, p_0} \left(I_2(\beta p_0 + \bar{\beta}) + \mu \beta I_1(p_0) - \beta I_2(p_0) \right)$$
$$= I_2(p^*) + \mathfrak{C}[\mu I_1(p^*) - I_2(p^*)]$$
$$= \max_{\beta, p_0 : \beta p_0 + \bar{\beta} = p^*} \left(I_2(p^*) + \mu \beta I_1(p_0) - \beta I_2(p_0) \right).$$

Finally by Lemma 1, there exists a unique (β, p_0) such that $\beta p_0 + \bar{\beta} = p^*$ and

$$\mathfrak{C}[\mu I_1(p^*) - I_2(p^*)] = \mu \beta I_1(p_0) - \beta I_2(p_0).$$

Thus $\mathcal{R}' = \mathrm{co}\{\mathcal{R}'\}$. To complete the proof for $\mathcal{R}' = \mathcal{R}''$, we use Lemma 2 below. $\qquad\square$

Lemma 2 [24]: Let $\mathcal{R} \subseteq \mathbb{R}^d$ be convex and $\mathcal{R}' \subseteq \mathcal{R}''$ be two bounded convex subsets of $\mathcal{R}$, closed relative to $\mathcal{R}$. If every supporting hyperplane of $\mathcal{R}''$ intersects $\mathcal{R}'$, then $\mathcal{R}' = \mathcal{R}''$.

As mentioned in the first part of this section, superposition coding is also optimal for the essentially less noisy and essentially more capable classes. Can we extend the range of parameter α for which superposition coding is optimal by evaluating the conditions for these two classes?

To answer this question, first note that for binary input broadcast channels, the essentially more capable condition in [22] reduces to that for the more capable class; hence essentially more capable does not extend the range of α for which superposition coding is optimal beyond more capable. To see this, consider a binary input broadcast channel which is not more capable. Then there exists $p, q \in (0, 1)$ such that $I(X; Y_2)_p - I(X; Y_1)_p > 0$ and $I(X; Y_1)_q - I(X; Y_2)_q > 0$. Then

$$\mathfrak{C}[I(X; Y_2)_r - I(X; Y_1)_r] > 0 \quad \text{for every } r \in (0, 1),$$
$$\mathfrak{C}[I(X; Y_1)_r - I(X; Y_2)_r] > 0 \quad \text{for every } r \in (0, 1).$$

Thus, for every $X \sim \mathrm{Bern}(r)$ for $r \in (0, 1)$, there exists U_1 and U_2 such that $I(X; Y_2|U_1) - I(X; Y_1|U_1) > 0$ and

$I(X; Y_1|U_2) - I(X; Y_2|U_2) > 0$. Hence if a binary input broadcast channel is not more capable it is also not essentially more capable.

The answer for essentially less noisy is less clear. It appears to be quite difficult to evaluate the set of pmfs $\mathcal{P}_o$ that satisfy the condition in (3). In the following section, we define a new class of broadcast channels for which the condition can be easily evaluated and which includes the essentially less noisy class.

IV. Effectively Less Noisy P-BC

Consider outer bound $\bar{\mathcal{R}}_1$ on the capacity region of the DM-BC $p(y_1, y_2|x)$ which consists of all rate pairs (R_1, R_2) such that

$$R_2 \le I(U; Y_2),$$
$$R_1 + R_2 \le I(U; Y_2) + I(X; Y_1|U) \tag{18}$$

for some $p(u, x)$. To see that this is indeed an outer bound, note that it is simply the Körner–Marton [16] outer bound without the sum rate bound.

The outer bound $\bar{\mathcal{R}}_1$ can be alternatively represented in terms of its supporting hyperplanes as:

$$\max_{(r_1, r_2) \in \bar{\mathcal{R}}_1} (\lambda r_1 + r_2)$$
$$= \begin{cases} \max_{p(u,x)} (\lambda I(X; Y_1|U) + I(U; Y_2)) & \text{if } \lambda \in [0, 1], \\ \max_{p(u,x)} \lambda (I(U; Y_2) + I(X; Y_1|U)) & \text{if } \lambda > 1. \end{cases} \tag{19}$$

Now consider the supporting hyperplane representation of the superposition rate region $\mathcal{R}_1$ in (2):

$$\max_{(R_1, R_2) \in \mathcal{R}_1} (\lambda R_1 + R_2)$$
$$= \begin{cases} \max_{p(u,x)} \left(\lambda \min\{I(X; Y_1) - I(U; Y_2), \ I(X; Y_1|U)\} \right. \\ \qquad\qquad \left. + I(U; Y_2) \right) & \text{if } \lambda \in [0, 1], \\ \max_{p(u,x)} \lambda \left(I(U; Y_1) + I(X; Y_1|U) \right) & \text{if } \lambda > 1. \end{cases}$$

Note that $\bar{\mathcal{R}}_1$ and $\mathcal{R}_1$ differ only in the first term when $0 \le \lambda \le 1$: for $\bar{\mathcal{R}}_1$, the term is $\lambda I(X; Y_1|U)$, while for $\mathcal{R}_1$ the term is $\lambda \min\{I(X; Y_1|U), I(X; Y_1) - I(U; Y_2)\}$. When $\lambda > 1$, for $\bar{\mathcal{R}}_1$, the term is $\lambda I(U; Y_2)$, while for $\mathcal{R}_1$ the term is $\lambda I(U; Y_1)$.

Now it is easy to see that if the DM-BC is less noisy, i.e., $I(U; Y_1) \ge I(U; Y_2)$ for all $p(u, x)$, then these two bounds coincide. The key observation that leads to a more general class than less noisy is that the inequality $I(U; Y_1) \ge I(U; Y_2)$ does not need to hold for every $p(u, x)$. For example,
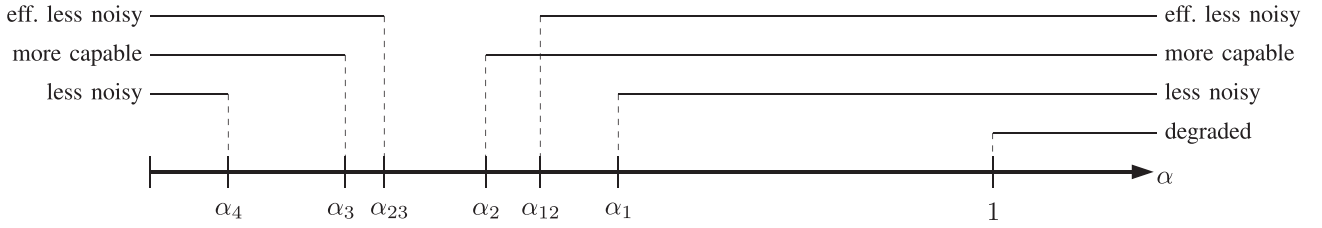
Fig. 8.   Illustration of the ranges of α for which the channel is degraded, less noisy, more capable, and effectively less noisy as stated in Theorems 1 and 2.

if $I(U; Y_1) \geq I(U; Y_2)$ for every $p(x) \in \mathcal{P}$ and every $p(u|x)$ such that

$$\max_{p(x)} \max_{p(u|x)} (\lambda I(X; Y_1|U) + I(U; Y_2))$$
$$= \max_{p(x) \in \mathcal{P}} \max_{p(u|x)} (\lambda I(X; Y_1|U) + I(U; Y_2)) \quad (20)$$

for every $0 \leq \lambda \leq 1$, then $\mathcal{R}_1$ and $\bar{\mathcal{R}}_1$ coincide.

This particular example is quite interesting because both the condition for the set $\mathcal{P}$ and the inequality $I(U; Y_1) \geq I(U; Y_2)$ can be expressed in terms of the upper concave envelope of $I(X; Y_1) - I(X; Y_2)$, which makes their evaluation quite straightforward especially for binary broadcast channels.

We are now ready to introduce a new class of broadcast channels for which superposition coding is optimal.

Definition 8 (Effectively Less Noisy Broadcast Channels): For a DM-BC, $p(y_1, y_2|x)$, let $\mathcal{P}$ be the set of pmfs $p(x)$ such that for every $0 \leq \lambda \leq 1$,

$$\max_{p(x)} \left(I(X; Y_2) + \mathfrak{C}[\lambda I(X; Y_1) - I(X; Y_2)] \right)$$
$$= \max_{p(x) \in \mathcal{P}} \left(I(X; Y_2) + \mathfrak{C}[\lambda I(X; Y_1) - I(X; Y_2)] \right). \quad (21)$$

Receiver Y_1 is said to be *effectively less noisy* than receiver Y_2 if $I(X; Y_1) - I(X; Y_2) = \mathfrak{C}[I(X; Y_1) - I(X; Y_2)]$ for every $p(x) \in \mathcal{P}$.

Clearly if the DM-BC is less noisy, then it is effectively less noisy. We can further show that if the channel is essentially less noisy as defined in [22], it is also effectively less noisy. To show this note that the sufficient class $\mathcal{P}_o$ in (3) must satisfy

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

for every $0 \leq \lambda \leq 1$. Hence $\mathcal{P}_o \supseteq \mathcal{P}$. If $I(U; Y_1) \geq I(U; Y_2)$ for $p(x) \in \mathcal{P}_o$ and every $p(u|x)$, then $I(U; Y_1) \geq I(U; Y_2)$ for $p(x) \in \mathcal{P}$ and every $p(u|x)$. We do not know if the condition for effectively less noisy is strictly weaker than that for essentially less noisy, however. As we will see in the next section, effectively less noisy neither implies nor is implied by more capable in general.

The definition of effectively less noisy can be readily extended to the P-BC in the same manner as the less noisy and more capable we presented in the previous section.

Definition 9 (Effectively Less Noisy P-BC): For the 2-receiver P-BC, let $\mathcal{Q} \subseteq [0, 1]$ be such that for

every $0 \leq \lambda \leq 1$,

$$\max_{q \in [0,1]} (I_2(q) + \mathfrak{C}[\lambda I_1(q) - I_2(q)])$$
$$= \max_{q \in \mathcal{Q}} (I_2(q) + \mathfrak{C}[\lambda I_1(q) - I_2(q)]). \quad (22)$$

Receiver Y_1 is said to be effectively less noisy than Y_2 if $I_1(q) - I_2(q) = \mathfrak{C}[I_1(q) - I_2(q)]$ for every $q \in \mathcal{Q}$.

To establish the parameter ranges for effectively less noisy P-BC, we need the following additional breakpoints of α:

$$\alpha_{12} = g_1(q_2),$$
$$\alpha_{23} = g_2(q_1)$$

where the functions $g_1(\cdot)$ and $g_2(\cdot)$ are defined in (7) and (9), respectively.

Theorem 2: Consider a 2-receiver P-BC and assume that $s_1 \leq s_2$.

1. If $\alpha \geq \alpha_{12}$, Y_1 is effectively less noisy than Y_2, and the capacity region is the set of rate pairs that satisfy (10).
2. If $\alpha \leq \alpha_{23}$, Y_2 is effectively less noisy than Y_1, and the capacity region is the set of rate pairs that satisfy (12).

Note that as for the less noisy case, the capacity region for effectively less noisy channels is also attained using binary U and a Z-channel from U to X. Figure 8 illustrates the parameter ranges for which superposition coding is optimal.

It can be shown that $\alpha_4 \leq \alpha_3 \leq \alpha_{23} \leq \alpha_2 \leq \alpha_{12} \leq \alpha_1$ (see Appendix B). The fact that $\alpha_2 \leq \alpha_{12}$ implies that if Y_1 is effectively less noisy than Y_2, it is also more capable than Y_2 (see Figure 4-(b) for an example). Thus effectively less noisy does not offer any new range of parameters for which superposition coding is optimal. On the other hand, the fact that $\alpha_3 \leq \alpha_{23}$ implies that for $\alpha \in (\alpha_3, \alpha_{23})$, Y_2 is effectively less noisy but is not more capable than Y_1 (see Figure 5-(b) for an example). Thus effectively less noisy offers a new range of parameters for which superposition coding is optimal.

Proof of Theorem 2:

1. We first show that $\mathcal{Q} = [0, q_2]$ satisfies (22). Specifically we show that $I_2(q) + \mathfrak{C}[\lambda I_1(q) - I_2(q)]$ is decreasing in $[q_2, 1]$. Since $I_2(q) + \mathfrak{C}[\lambda I_1(q) - I_2(q)]$ is concave, it suffices to show that the derivative of $I_2(q) + \mathfrak{C}[\lambda I_1(q) - I_2(q)]$ at $q = q_2$ is nonpositive. Consider

$$\left. \frac{d(I_2(q) + \mathfrak{C}[\lambda I_1(q) - I_2(q)])}{dq} \right|_{q=q_2}$$
$$\stackrel{(a)}{=} \begin{cases} \lambda I_1'(q_2) & \text{if } q_2 \geq t', \\ I_2'(q_2) - (\lambda I_1(t') - I_2(t'))/(1-t') & \text{otherwise.} \end{cases}$$
$$\stackrel{(b)}{\leq} 0$$

for $t' = t(\alpha')$ where $\alpha' = \lambda\alpha$. Step (a) follows by rewriting $\lambda I_1(q)$ for $A_1 = \alpha$ as $I_1(q)$ for $A_1 = \lambda\alpha$ (see (4)) and then applying Lemma 1. Step (b) holds because $\lambda I_1'(q_2) \leq 0$ and $-(\lambda I_1(t') - I_2(t'))/(1-t') \leq 0$. The first inequality holds since $s_1 \leq s_2$ implies $q_1 \leq q_2$ and $\lambda I_1'(q_1) = 0$. The second inequality holds by (30). By Lemma 1, $\mathfrak{C}[I_1(q) - I_2(q)] = I_1(q) - I_2(q)$ for $q \in [0, q_2]$ if and only if $\alpha \geq g_1(q_2) = a_{12}$ (also see Figure 3 (a)).

2. We first show that $\mathcal{Q} = [q_1, 1]$ satisfies (22) with $I_1(\cdot)$ and $I_2(\cdot)$ interchanged. Specifically we show that $I_1(q) + \mathfrak{C}[\lambda I_2(q) - I_1(q)]$ is increasing in $[0, q_1]$. Since $I_1(q) + \mathfrak{C}[\lambda I_2(q) - I_1(q)]$ is concave, it suffices to show that the derivative of $I_1(q) + \mathfrak{C}[\lambda I_2(q) - I_1(q)]$ at $q = q_1$ is nonnegative. Consider

$$\frac{d\left(I_1(q) + \mathfrak{C}[\lambda I_2(q) - I_1(q)]\right)}{dq}\bigg|_{q=q_1}$$

$$= \frac{d\left(I_1(q) + \lambda\mathfrak{C}[I_2(q) - I_1(q)/\lambda]\right)}{dq}\bigg|_{q=q_1}$$

$$\overset{(a)}{=} \begin{cases} \lambda I_2'(q_1) & \text{if } q_1 \geq r', \\ I_1'(q_1) + (\lambda I_2(r') - I_1(r'))/r' & \text{otherwise}. \end{cases}$$

$$\overset{(b)}{\geq} 0$$

for $r' = r(\alpha')$ where $\alpha' = \alpha/\lambda$. Step (a) follows by rewriting $I_1(q)/\lambda$ for $A_1 = \alpha$ as $I_1(q)$ for $A_1 = \alpha/\lambda$ (see (4)) and then applying Lemma 1. Step (b) holds because $\lambda I_2'(q_1) \geq 0$ and $(\lambda I_2(r') - I_1(r'))/r' \geq 0$. The first inequality holds since $s_1 \leq s_2$ implies $q_1 \leq q_2$ and $\lambda I_2'(q_2) = 0$. The second inequality holds by (31). By Lemma 1, $\mathfrak{C}[I_2(q) - I_1(q)] = I_2(q) - I_1(q)$ for $q \in [q_1, 1]$ if and only if $\alpha \leq g_2(q_1) = a_{23}$ (also see Figure 3 (b)). $\qquad \square$

Figure 9 illustrates the parameter ranges for which superposition coding is optimal. As can be seen the area in the α-s_2 plane where superposition coding is not optimal becomes smaller as s_1 increases. In Appendix C, we show that the fraction of the channel parameter space for which superposition coding is optimal approaches one as s_1 goes to infinity; hence superposition coding is in a sense almost always optimal for the P-BC. In comparison, the fraction of the parameter space for which the P-BC is degraded is always bounded away from 1.

As illustrated in Figure 8, the capacity region of the P-BC is achieved using superposition coding in the ranges $\alpha \leq \alpha_{23}$ and $\alpha \geq \alpha_2$. Note that for the remaining range of α the channel is not more capable or effectively less noisy, which follows immediately from the if and only if conditions established in Theorems 1 and 2. In the following section we explore bounds on the capacity region in the range for $\alpha \in (\alpha_{23}, \alpha_2)$.

V. GAP BETWEEN MARTON AND UV

The best known inner and outer bounds on the capacity region of the DM-BC are the Marton inner bound [16] and the UV outer bound [17], respectively. We show that for $\alpha \in (\alpha_{23}, \alpha_2)$ there can be a gap between these two bounds.
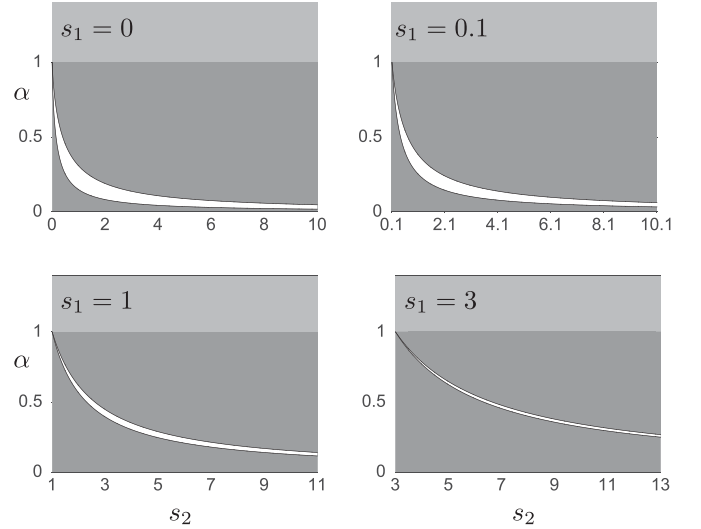

Fig. 9. Plots of α versus s_2 for $s_1 = 0$ (top left), $s_1 = 0.1$ (top right), $s_1 = 1$ (bottom left) and $s_1 = 3$ (bottom right). The shaded areas in each plot are where superposition coding is optimal and the light shaded area ($\alpha \geq 1$) is where Y_2 is a degraded version of Y_1.

In [25], Anantharam, Gohari, and Nair showed that for binary input broadcast channels, Marton's inner bound reduces to the set of rate pairs (R_1, R_2) such that

$$R_1 < I(W; Y_1) + \sum_{j=1}^{k} \beta_j I(X; Y_1 | W = j),$$

$$R_2 < I(W; Y_2) + \sum_{j=k+1}^{5} \beta_j I(X; Y_2 | W = j),$$

$$R_1 + R_2 < \min\{I(W; Y_1), I(W; Y_2)\}$$
$$+ \sum_{j=1}^{k} \beta_j I(X; Y_1 | W = j) + \sum_{j=k+1}^{5} \beta_j I(X; Y_2 | W = j) \tag{23}$$

for some $p_W(j) = \beta_j$, $j \in [1:5]$, $k \in [1:5]$, and $p(x|w)$. This region is achieved using *randomized time-division* [26]. This ingenious insight helps simplify the computation of Marton's inner bound for the binary P-BC; hence for the P-BC itself.

The UV outer bound on the capacity region of the DM-BC is the set of rate pairs (R_1, R_2) such that

$$R_1 \leq I(V; Y_1),$$
$$R_2 \leq I(U; Y_2),$$
$$R_1 + R_2 \leq I(V; Y_1) + I(U; Y_2 | V),$$
$$R_1 + R_2 \leq I(V; Y_1 | U) + I(U; Y_2) \tag{24}$$

for some $p(u, v)$ and function $x(u, v)$, $|\mathcal{U}|, |\mathcal{V}| \leq |\mathcal{X}| + 1$. Computing this bound even for binary input broadcast channels is quite difficult. Hence, we compute the maximum sum rates for the Marton and the UV bounds instead of the complete bounds. Figure 10 plots the maximum sum rates for $\alpha_{23} \leq \alpha \leq \alpha_2$ when $s_1 = 0.1$ and $s_2 = 1$ ($\alpha_{23} = 0.27$, $\alpha_2 = 0.4$). Note that for $0.27 \leq \alpha \leq 0.286$, the sum
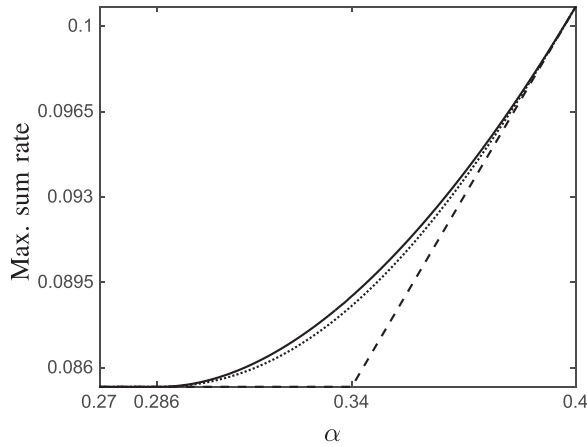
Fig. 10. Plots of the maximum Marton sum rate (dot), superposition sum rate (dash), and UV sum rate for $s_1 = 0.1$, $s_2 = 1$, and $\alpha \in [\alpha_{23}, \alpha_2]$.

rates coincide. For the rest of the range there is a small gap between the Marton and the UV bound sum rates. In particular for $\alpha = 0.34$, the gap is approximately 0.00039.

It turns out that superposition coding is optimal in the range where Marton's sum rate and the UV outer bound sum rate in Figure 10 coincide, i.e., for $\alpha \in [0.27, 0.286]$. To show this, note that the condition for effectively less noisy in (20) can be tightened further. It is clearly sufficient that $I(U; Y_1) \geq I(U; Y_2)$ for every $p^*(u, x) \in \Pr(U, X)$ and every $0 \leq \lambda \leq 1$ such that

$$\max_{p(u,x)} \left(\lambda I(X; Y_1|U) + I(U; Y_2) \right)$$
$$= \max_{p^*(u,x) \in \Pr(U,X)} \left(\lambda I(X; Y_1|U) + I(U; Y_2) \right). \quad (25)$$

This is the condition satisfied for the P-BC in the range $\alpha \in [0.27, 0.286]$. However, unlike the looser condition (20), where we are able to express the class of pmfs in terms of the concave envelope of $I(X; Y_1) - I(X; Y_2)$ and determine analytically the range in which the P-BC is effectively less noisy, we can only numerically evaluate the above condition.

VI. AVERAGE POWER CONSTRAINT

The results on superposition coding in sections II- IV can be readily extended to the case when there is also an average power constraint. Suppose that in addition to the maximum power constraint $X(t) \leq 1$ (which is needed for the capacity to be finite) there is an average power constraint , i.e.,

$$\frac{1}{T} \int_0^T X(t)dt \leq \sigma. \quad (26)$$

The capacity region of P-BC for the average power constraint setting is equal to the capacity region of the corresponding binary P-BC for $\mathsf{E}[X] \leq \sigma$. This is a simple extension for Wyner's argument that the capacity of point-to-point Poisson channel is equal to the capacity of the corresponding binary channel for $\mathsf{E}[X] \leq \sigma$ [6], [7].

Note that if a broadcast channel is less noisy or more capable, the UV outer bound in (24) with an input constraint coincides with the superposition inner bound with the same

input constraint. Hence, if a P-BC is less noisy, it is also less noisy under the average power constraint, and similarly if a P-BC is more capable, it is also more capable under the average power constraint.

In contrast, if a P-BC is effectively less noisy, it is not necessarily effectively less noisy under the average power constraint. The results of effectively less noisy, however, can be easily extended to the average power constraint setting.

Theorem 3: For P-BC with an average power constraint in (26),
1. Receiver Y_1 is effectively less noisy than Y_2 if $\alpha \geq g_1(\min\{\sigma, q_2\})$,
2. Receiver Y_2 is effectively less noisy than Y_1 if $\alpha \leq g_2(\min\{\sigma, q_1\})$,

where the functions $g_1(\cdot)$ and $g_2(\cdot)$ are defined in (7) and (9), respectively.

Proof: Note that $X \sim \text{Bern}(q)$ for $q \in [0, \sigma]$ satisfies the average power constraint. The condition in Definition 9 can be modified to the average power constraint setting. Let $\mathcal{Q} \subseteq [0, \sigma]$ be such that for every $0 \leq \lambda \leq 1$,

$$\max_{q \in [0,\sigma]} (I_2(q) + \mathfrak{C}[\lambda I_1(q) - I_2(q)])$$
$$= \max_{q \in \mathcal{Q}} (I_2(q) + \mathfrak{C}[\lambda I_1(q) - I_2(q)]). \quad (27)$$

Receiver Y_1 is effectively less noisy than Y_2 under the average power constraint if $I_1(q) - I_2(q) = \mathfrak{C}[I_1(q) - I_2(q)]$ for every $q \in \mathcal{Q}$.

1. Recall that for P-BC without an average power constraint, Y_1 is effectively less noisy than Y_2 if

   $$I_1(q) - I_2(q) = \mathfrak{C}[I_1(q) - I_2(q)] \text{ for } q \in [0, q_2].$$

   For P-BC with an average power constraint $q \leq \sigma$, Y_1 is effectively less noisy than Y_2 if

   $$I_1(q) - I_2(q) = \mathfrak{C}[I_1(q) - I_2(q)] \text{ for } q \in [0, \min\{\sigma, q_2\}].$$

   If $\alpha \geq g_1(\min\{\sigma, q_2\})$, Y_1 is effectively less noisy than Y_2 under the average power constraint.

2. Recall that for P-BC without an average power constraint, Y_2 is effectively less noisy than Y_1 if

   $$I_2(q) - I_1(q) = \mathfrak{C}[I_2(q) - I_1(q)] \text{ for } q \in [q_1, 1].$$

   If there is an average power constraint $q \leq \sigma$ for some $\sigma \geq q_1$, then Y_2 is effectively less noisy than Y_1 under the average power constraint if

   $$I_2(q) - I_1(q) = \mathfrak{C}[I_2(q) - I_1(q)] \text{ for } q \in [q_1, \sigma].$$

   Thus for $\alpha \leq g_2(q_1)$, Y_2 is effectively less noisy than Y_1 under the average power constraint.
   If there is an average power constraint $q \leq \sigma$ for some $\sigma < q_1$, then Y_2 is effectively less noisy under the average power constraint than Y_1 if

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

   Thus for $\alpha \leq g_2(\sigma)$, Y_2 is effectively less noisy under the average power constraint than Y_1. $\square$
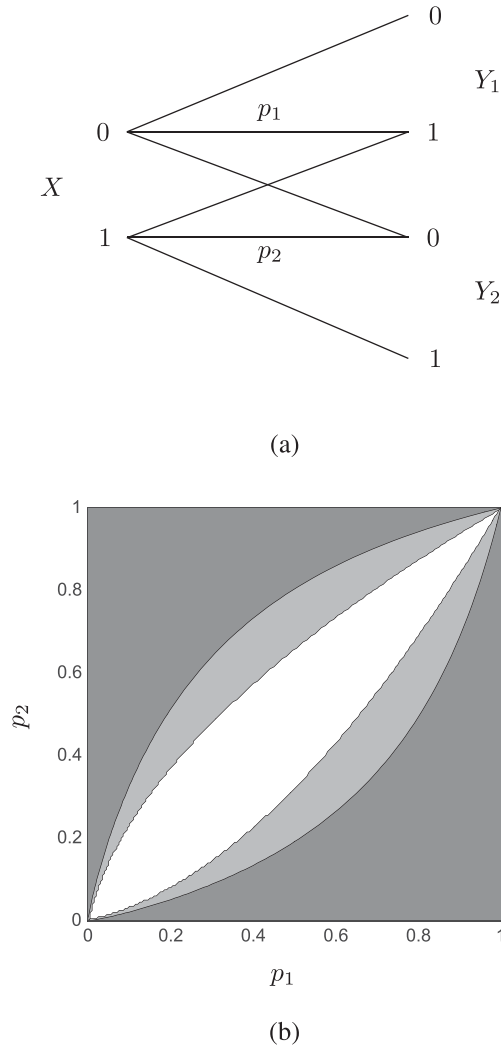
(a)

(b)

Fig. 11.   (a) Skewed binary broadcast channel (b) Plot of p_1 vs p_2. The lightly shaded area is where superposition coding is optimal and the dark shaded area is where one receiver is effectively less noisy than the other receiver.

VII. FINAL REMARKS

We showed that superposition coding is optimal for almost all Poisson broadcast channels, and that when superposition is not optimal, there is a gap between Marton's inner bound and the UV outer bound. Hence the capacity region for the P-BC is still not known in general. We note that in [27], Geng, Nair, Shamai, and Wang similarly showed that for the class of binary input symmetric output broadcast channels (which do not include the binary P-BC) either superposition coding is optimal or there is a gap between Marton's inner bound and the UV outer bound.

We introduced the effectively less noisy broadcast channels for which superposition coding is optimal. This condition for effectively less noisy can be further extended, but we can only verify it numerically.

Why is this the case and does it hold for general binary input broadcast channels?

The intuitive reason superposition coding is almost always optimal for the P-BC is that the binary P-BC is more capable (and even less noisy) for most parameter ranges as established

in Appendix C. Hence one channel is almost always stronger than the other.

For other classes of binary input broadcast channels, the more capable condition is much less likely to be satisfied. As an extreme case, consider the skewed binary broadcast channel in Figure 11-(a), which is a generalization of the skew symmetric BC in [26]. We can show that the channel is not more capable (hence also not less noisy or degraded) for every $p_1, p_2 \in (0, 1)$. It turns out, however, that for (p_1, p_2) in the dark shaded area in Figure 11-(b), the channel is effectively less noisy, and for (p_1, p_2) in the lightly shaded area in Figure 11-(b), the BC is not effectively less noisy but superposition coding is still optimal (which is shown by verifying the stronger condition in (25)). These shaded areas constitute 76% of the parameter space area! This clearly demonstrates that our intuition about when superposition coding is optimal does not always hold. The unshaded area in Figure 11-(b) is where Marton rate region is strictly larger than the superposition rate region, which implies that superposition coding is not optimal.

To perform the analytical and computational evaluations, we relied heavily on the concave envelope method that has been used in other applications, including [28] which establishes the optimality of dirty paper coding and superposition coding for MIMO broadcast channels with common message, and [29] which shows that UV outer bound is not tight.

Finally, it would be interesting to find a similar extension of the more capable to the notion of effectively less noisy. The difficulty is finding an outer bound similar to the one we used for effectively less noisy.

APPENDIX A
PROOF OF LEMMA 1

We first prove that 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 $\alpha \le \alpha_4$, (ii) convex in $[0, 1]$ if $\alpha \ge \alpha_1$, or (iii) convex in $[0, \kappa]$ and concave in $[\kappa, 1]$ otherwise, where

$$\kappa = \frac{\alpha s_2 - s_1}{1 - \alpha} \in [0, 1]. \tag{28}$$

To show the above statement, note that the second derivative of $I_2(q) - I_1(q)$,

$$I_2''(q) - I_1''(q) = \frac{(\alpha - 1)q + \alpha s_2 - s_1}{(q + s_1)(q + s_2)}, \tag{29}$$

has a unique zero, is nonnegative if $\alpha \ge \alpha_1$, and is nonpositive if $\alpha \le \alpha_4$. Now we are ready to prove Lemma 1.

1. Let $f_1(q) = (I_1(q) - I_2(q))/(1 - q)$. We first show that t defined in (6) satisfies

$$t = \arg\max_{q \in [0,1]} f_1(q). \tag{30}$$

It can be shown that

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

where $g_1(q)$ is defined in (7). It can be easily seen that $f_1'(q)$ and $(\alpha - g_1(q))$ have the same sign for $0 < q < 1$ and $g_1(q)$ is increasing for $q > 0$. Hence

   a) if $\alpha \le g_1(0) = \alpha_3$, then $f_1'(q) \le 0$ in $[0, 1]$ and $f_1(q)$ is maximized at t.

   b) if $\alpha \ge \lim_{q \to 1} g_1(q) = \alpha_1$, then $f_1'(q) \ge 0$ in $[0, 1]$ and $f_1(q)$ is maximized at t.

   c) otherwise, $f_1'(q) \ge 0$ in $[0, g_1^{-1}(\alpha)]$ and $f_1'(q) \le 0$ in $[g_1^{-1}(\alpha), 1]$ and $f_1(q)$ is maximized at t.

Let

$$L_1(q) = \begin{cases} I_1(q) - I_2(q) & \text{if } q \in [0, t], \\ (1-q)(I_1(t) - I_2(t))/(1-t) & \text{if } q > t. \end{cases}$$

Then $L_1(q) \ge I_1(q) - I_2(q)$ because $f_1(t) \ge f_1(q)$ for any $q \in [0, 1]$.

On the other hand, note that for $q > t$, $L_1(q)$ can be expressed as

$$L_1(q) = \big((1 - q)/(1 - t)\big)(I_1(t) - I_2(t)) \\ + \big((q - t)/(1 - t)\big)(I_1(1) - I_2(1)).$$

Since $q = \big((1 - q)/(1 - t)\big) \cdot t + \big((q - t)/(1 - t)\big) \cdot 1$, it follows that for $q > t$, $L_1(q)$ is a convex combination of $I_1(t) - I_2(t)$ and $I_1(1) - I_2(1)$. Thus $\mathfrak{C}[I_1(q) - I_2(q)] \ge L_1(q)$.

Finally to argue that $L_1(q) = \mathfrak{C}[I_1(q) - I_2(q)]$, we are left to show that $L_1(q)$ is indeed concave. We first show that $L_1(q)$ is concave in $[0, t]$. Recall that $I_1(q) - I_2(q)$ is concave in $[0, \kappa]$ for κ in (28). Thus it suffices to show that $t \le \kappa$. Equivalently we show that $f_1'(\kappa) \le 0$. Consider

$$f_1'(q) \\ = \frac{I_1'(q) - I_2'(q)}{1 - q} - \frac{(I_1(1) - I_2(1)) - (I_1(q) - I_2(q))}{(1 - q)^2} \\ = (I_1'(q) - I_2'(q))/(1 - q) - (I_1'(u) - I_2'(u))/(1 - q)$$

for some $u \in [q, 1]$. The inequality $f_1'(\kappa) \le 0$ holds since $I_1''(q) - I_2''(q) \ge 0$ for $q \ge \kappa$ as can be seen from (29). We conclude that $L_1(q)$ is concave because it is concave in $[0, t]$ and $[t, 1]$ and is differentiable at t.

2. Let $f_2(q) = (I_2(q) - I_1(q))/q$. We show that r defined in (8) satisfies

$$r = \arg\max_{q \in [0, 1]} f_2(q). \tag{31}$$

It can be shown that

$$f_2'(q) = q^{-2}(\alpha - g_2(q))(q - s_1 \log(1 + q/s_1)),$$

where $g_2(q)$ is defined in (9). It can be easily seen that $f_2'(q)$ and $(\alpha - g_2(q))$ have the same sign for $0 < q < 1$. Also it can be shown that $g_2(q)$ is increasing for $q > 0$. Hence

   a) if $\alpha \le \lim_{q \to 0} g_2(q) = \alpha_4$, then $f_2'(q) \le 0$ in $[0, 1]$ and $f_2(q)$ is maximized at r.

   b) if $\alpha \ge g_2(1) = \alpha_2$, then $f_2'(q) \ge 0$ in $[0, 1]$ and $f_2(q)$ is maximized at r.

   c) otherwise, $f_2'(q) \ge 0$ in $[0, g_2^{-1}(\alpha)]$ and $f_2'(q) \le 0$ in $[g_2^{-1}(\alpha), 1]$ and $f_2(q)$ is maximized at r.

Let

$$L_2(q) = \begin{cases} (q/r)(I_2(r) - I_1(r)) & \text{if } 0 \le q < r, \\ I_2(q) - I_1(q) & \text{if } q \ge r. \end{cases}$$

Then $I_2(q) - I_1(q) \le L_2(q)$ because $f_2(r) \ge f_2(q)$ for any $q \in [0, 1]$.

On the other hand, $\mathfrak{C}[I_2(q) - I_1(q)] \ge L_2(q)$ because for $0 \le q \le r$, $L_2(q)$ 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 $L_2(q)$ is concave. Recall that $I_2(q) - I_1(q)$ is concave in $[\kappa, 1]$ for κ in (28). To show that $L_2(q)$ is concave in $[r, 1]$, we show $\kappa \le r$, i.e. $f_2'(\kappa) \ge 0$. Consider

$$f_2'(q) = \frac{I_2'(q) - I_1'(q)}{q} - \frac{I_2(q) - I_1(q)}{q^2} \\ = (I_2'(q) - I_1'(q))/q - (I_2'(u) - I_1'(u))/q$$

for some $u \in [0, q]$. Since $I_2''(q) - I_1''(q) \ge 0$ in $[0, \kappa]$, it follows that $f_2'(\kappa) \ge 0$. We conclude that $L_2(q)$ is concave because it is concave in $[0, r]$ and $[r, 1]$ and is differentiable at r.

APPENDIX B
ORDER OF BREAKPOINTS

Since less noisy channels are also more capable it follows that $\alpha_4 \le \alpha_3$ and $\alpha_2 \le \alpha_1$. Also Y_1 and Y_2 cannot be more capable simultaneously unless the two channels are identical. Thus, $0 \le \alpha_4 \le \alpha_3 \le \alpha_2 \le \alpha_1$. To completely characterize the order, we need to show the following:

$$\alpha_2 \overset{(a)}{\le} \alpha_{12} \overset{(b)}{\le} \alpha_1 \quad \text{and} \quad \alpha_3 \overset{(c)}{\le} \alpha_{23} \overset{(d)}{\le} \alpha_2.$$

Note that inequalities (b) and (d) follow by Lemma 1. The remaining inequalities were shown through simulation. To show (a), note that

$$\alpha_2 - \alpha_{12} = \frac{u_1(s_2)u_2(s_1) - u_1(s_1)u_2(s_2)}{u_1(s_1)u_2(s_1)},$$

where $u_1(x) = x \log(1 + 1/x) - 1$ and $u_2(x) = (1 + x) \log((1 + x)/(q_2 + x)) - 1 + q_2$, where q_2 is defined in (5). It can be easily shown that the denominator is always nonpositive, hence the numerator must be nonnegative for every $s_1 \le s_2$ in order for the inequality (a) to hold. Using the MINIMIZE function of Mathematica and Maple, we verified that the numerator is indeed nonnegative for every $s_1 \le s_2$. To show (c), note that

$$\alpha_3 - \alpha_{23} = \frac{u_3(s_2)u_4(s_1) - u_3(s_1)u_4(s_2)}{u_3(s_1)u_4(s_1)},$$

where $u_3(x) = (1 + x) \log(1 + 1/x) - 1$ and $u_4(x) = x \log(1 + q_1/x) - q_1$, where q_1 is defined in (5). It can be easily shown that the denominator is always nonpositive, hence the numerator must be nonnegative for every $s_1 \le s_2$ in order for the inequality (c) to hold. Using the MINIMIZE function of Mathematica and Maple, we verified that the numerator is indeed nonnegative for every $s_1 \le s_2$. The difficulty in

proving (a) and (c) analytically comes from the fact that $\alpha_2 - \alpha_{12}$ and $\alpha_3 - \alpha_{23}$ depend on q_i, $i = 1, 2$, which are in turn functions of s_i, $i = 1, 2$.

APPENDIX C
VOLUME

Consider the set of channel parameters: $(\alpha, s_1, s_2) \in [0, 1] \times [0, b] \times [0, kb]$ for $b, k \geq 0$. Let $b_0 = \min\{b, kb\}$. The fraction of the set of P-BC parameters for which the channel is degraded is

$$\frac{1}{kb^2} \iint_{s_1 \geq s_2} ds_2 \, ds_1 = \begin{cases} 0.5k^{-1} & \text{if } k \geq 1, \\ 1 - 0.5k & \text{if } k < 1. \end{cases}$$

The fraction of the set of P-BC parameters for which superposition coding is optimal is lower bounded by the fraction for which the P-BC is less noisy. Consider

$$1 - \frac{1}{kb^2} \iint_{s_1 \leq s_2} \int_{s_1/s_2}^{(1+s_1)/(1+s_2)} d\alpha \, ds_2 \, ds_1$$

$$= 1 - \frac{1}{kb^2} \int_0^{b_0} \int_{s_1}^{kb} \int_{s_1/s_2}^{(1+s_1)/(1+s_2)} d\alpha \, ds_2 \, ds_1$$

$$= 1 - \frac{1}{kb^2} \int_0^{b_0} \int_{s_1}^{kb} \left(\frac{1+s_1}{1+s_2} - \frac{s_1}{s_2} \right) ds_2 \, ds_1$$

$$= 1 - \frac{1}{kb^2} \int_0^{b_0} \left((1+s_1) \log \frac{1+kb}{1+s_1} - s_1 \log \frac{kb}{s_1} \right) ds_1$$

$$= 1 - \frac{1}{kb^2} \left(\frac{(1+b_0)^2 - 1}{2} \log(1+kb) - \frac{b_0^2}{2} \log kb \right.$$
$$\left. - \frac{1+b_0^2}{2} \log(1+b_0) + \frac{b_0^2 \log b_0}{2} + \frac{b_0}{2} \right)$$

$$= 1 - \frac{1}{kb^2} \left(\frac{b_0^2}{2} \left(\log \frac{1+kb}{kb} - \log \frac{1+b_0}{b_0} \right) - \frac{\log(1+b_0)}{2} \right.$$
$$\left. + b_0 \left(\log(1+kb) - \log(1+b_0) + \frac{1}{2} \right) \right),$$

which approaches to 1 as $b \to \infty$.

REFERENCES

[1] M. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, pp. 710–715, Nov. 1980.

[2] D. L. Snyder and I. B. Rhodes, "Some implications of the cutoff-rate criterion for coded direct-detection optical communication systems," *IEEE Trans. Inf. Theory*, vol. 26, no. 3, pp. 327–338, May 1980.

[3] I. Bar-David and G. Kaplan, "Information rates of photon-limited overlapping pulse position modulation channels," *IEEE Trans. Inf. Theory*, vol. 30, no. 3, pp. 455–464, May 1984.

[4] Y. M. Kabanov, "The capacity of a channel of the Poisson type," *Theory Probab. Appl.*, vol. 23, no. 1, pp. 143–147, 1978.

[5] M. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, pp. 710–715, Nov. 1980.

[6] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel. I," *IEEE Trans. Inf. Theory*, vol. 34, no. 6, pp. 1449–1461, Nov. 1988.

[7] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel. II," *IEEE Trans. Inf. Theory*, vol. 34, no. 6, pp. 1462–1471, Nov. 1988.

[8] A. Lapidoth and S. Shamai (Shitz), "The Poisson multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 488–501, Mar. 1998.

[9] L. Lai, Y. Liang, and S. Shamai (Shitz), "On the capacity bounds for Poisson interference channels," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 223–238, Jan. 2015.

[10] S. Bross, A. Lapidoth, and L. Wang, "The Poisson channel with side information," in *Proc. 47th Annu. Allerton Conf. Commun., Control, Comput.*, Sep./Oct. 2009, pp. 574–578.

[11] A. Lapidoth, I. E. Telatar, and R. Urbanke, "On wide-band broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3250–3258, Dec. 2003.

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

[13] H. Kim, B. Nachman, and A. El Gamal, "Superposition coding is almost always optimal for the Poisson broadcast channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2015, pp. 834–838.

[14] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge: Cambridge Univ. Press, 2011.

[15] J. Körner and K. Marton, "Comparison of two noisy channels," in *Topics in Information Theory* (Colloquia Mathematica Societatis Janos Bolyai), I. Csiszár and P. Elias, Eds. Amsterdam, The Netherlands: North Holland, 1977, pp. 411–423.

[16] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 3, pp. 306–311, May 1979.

[17] 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.

[18] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inf. Theory*, vol. 19, no. 2, pp. 197–207, Mar. 1973.

[19] L. Wang, E. Şaşoğlu, B. Bandemer, and Y.-H. Kim, "A comparison of superposition coding schemes," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 2970–2974.

[20] C. Nair, H. Kim, and A. El Gamal, "Simplified characterization of the superposition coding inner bound for the broadcast channel," to be published.

[21] 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.

[22] C. Nair, "Capacity regions of two new classes of two-receiver broadcast channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4207–4214, Sep. 2010.

[23] A. El Gamal, "The capacity of a class of broadcast channels," *IEEE Trans. Inf. Theory*, vol. 25, no. 2, pp. 166–169, Mar. 1979.

[24] H. G. Eggleston, *Convexity*. Cambridge, U.K.: Cambridge Univ. Press, 1958.

[25] V. Anantharam, A. Gohari, and C. Nair, "Improved cardinality bounds on the auxiliary random variables in Marton's inner bound," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 1272–1276.

[26] B. Hajek and M. B. Pursley, "Evaluation of an achievable rate region for the broadcast channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 1, pp. 36–46, Jan. 1979.

[27] Y. Geng, C. Nair, S. Shamai (Shitz), and Z. V. Wang, "On broadcast channels with binary inputs and symmetric outputs," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 6980–6989, Nov. 2013.

[28] Y. Geng and C. Nair, "The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2087–2104, Apr. 2014.

[29] Y. Geng, A. Gohari, C. Nair, and Y. Yu, "On Marton's inner bound and its optimality for classes of product broadcast channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 22–41, Jan. 2014.

Hyeji Kim (S'14) is a Ph.D. candidate in the Department of Electrical Engineering at Stanford University. She received the B.S. degree with honors in Electrical Engineering from the Korea Advanced Institute of Science and Technology (KAIST) in 2011 and the M.S. degree in Electrical Engineering from Stanford University in 2013. She is a recipient of the Stanford Graduate Fellowship (SGF). Her research interest include information theory, statistical learning, and stochastic control.

Benjamin Nachman is a Ph.D. candidate in the Department of Physics at Stanford University. He received the B.A. degree in Physics, Mathematics, and Economics from Cornell University in 2012 and the MASt in Applied Mathematics and Theoretical Physics from Cambridge University in 2013. He is a recipient of the Stanford Graduate Fellowship (SGF) and the National Science Foundation Graduate Research Fellowship (GRFP). His research interest is applying machine learning techniques to high energy physics.

Abbas El Gamal (S'71–M'73–SM'83–F'00) is the Hitachi America Professor in the School of Engineering and the Chair of the Department of Electrical Engineering at Stanford University. He received his B.Sc. Honors degree from Cairo University in 1972, and his M.S. in Statistics and Ph.D. in Electrical Engineering both from Stanford University in 1977 and 1978, respectively. From 1978 to 1980, he was an Assistant Professor of Electrical Engineering at USC. From 2004 to 2009, he was the Director of the Information Systems Laboratory at Stanford University. His research contributions have been in network information theory, FPGAs, and digital imaging devices and systems. He has authored or coauthored over 230 papers and holds 35 patents in these areas. He is coauthor of the book *Network Information Theory* (Cambridge Press 2011). He received several honors and awards for his research contributions, including the 2016 IEEE Richard Hamming Medal, the 2014 Viterbi Lecture, the 2013 Shannon Memorial Lecture, the 2012 Claude E. Shannon Award, the inaugural Padovani Lecture, and the 2004 INFOCOM Paper Award. He is a member of the US National Academy of Engineering and a Fellow of the IEEE. He has been serving on the Board of Governors of the Information Theory Society since 2009 and is currently Senior Past President. He has played key roles in several semiconductor, EDA, and biotechnology startup companies.