The Stochastic-Calculus Approach to Multi-Receiver Poisson Channels

Nirmal V. Shende
School of Electrical and Computer Engineering
Cornell University
Ithaca, NY 14853
Email: nvs25@cornell.edu

Aaron B. Wagner
School of Electrical and Computer Engineering
Cornell University
Ithaca, NY 14853
Email: wagner@cornell.edu

Abstract—We study two-receiver Poisson channels using tools derived from stochastic calculus. We compute necessary and sufficient conditions under which the continuous-time, continuous-space Poisson channel is less noisy and more capable, which turn out to be distinct from the conditions under which the “sampled” channel is less noisy and more capable. We also determine the capacity region of the more capable Poisson broadcast channel with independent message sets, the more capable Poisson wiretap channel, and the general two-decoder Poisson broadcast channel with degraded message sets.

I. INTRODUCTION

The Poisson channel models a direct-detection optical communication system, where the input to the channel $X_0^T$ represents the strength of the optical input signal, and the output of the channel is a Poisson process with rate $aX_0^T + \lambda$, where $a$ accounts for attenuation and $\lambda$ represents the rate of the dark current. Capacity studies of this channel have been ongoing since it was introduced as a viable model in [1], [2].

Broadly speaking, the channel has been studied using two mathematical approaches. Early work calculated mutual information and related quantities for the channel using stochastic calculus and, in particular, the theory of point process martingales [3], [4]. Most later work followed the approach of Wyner [5] who argued that the encoder and decoder could be restricted to use the channel so that it behaved like a discrete-time, memoryless, binary channel, with no essential loss of performance. One then applies standard techniques for such channels [6]–[9].

We espouse the former approach in this paper, both on the general principle that, when the existing tools are insufficient for a new problem, it is preferable to extend the tools rather than reduce the problem, and for certain pragmatic reasons. The reduction to a discrete-time binary channel is somewhat involved, and at least formally must be reproved for new variations. Once the appropriate stochastic-calculus-based tools have been developed, on the other hand, they can be directly applied to new problems. Moreover, it is unclear how to extend Wyner’s [5] reduction to some setups such as the wiretap version of the channel considered herein.

Of course, the stochastic calculus approach also has its disadvantages: it requires more sophisticated mathematics, and one cannot apply results from the extensive literature on discrete memoryless channels. One cannot even presume that the capacity is governed by the maximal mutual information, for instance, an oversight in the early work that used this approach. On the other hand, once the necessary tools are developed, coding theorems follow expeditiously.

The goal of this paper is to develop those tools that are necessary for various multi-decoder extensions of the Poisson channel. The two-decoder Poisson channel consists of a single transmitter (which inputs process $X_0^T$) and two receivers with output processes $Y_0^T$ and $Z_0^T$, where $Y_0^T$ and $Z_0^T$ are Poisson process with rates $a_0X_0^T + \lambda_y$ and $a_2X_0^T + \lambda_z$, respectively. We assume throughout that $\lambda_y, \lambda_z > 0$. We shall consider both the broadcast channel (either with independent or degraded message sets) and the wiretap channel (where one of the receivers is an eavesdropper).

We derive a general formula for the conditional mutual information over the Poisson channel, which generalizes an existing formula [3], [4] by allowing the use of auxiliary random variables and conditioning. We also obtain a continuous-time Csiszár-sum-like identity for Poisson channels. Using these tools, we obtain necessary and sufficient conditions for which the broadcast channel is less noisy and more capable, and show that these orderings are in fact equivalent. These conditions turn out not to be equivalent, however, to the analogous conditions for the discrete-time binary channel obtained as a reduction of the Poisson channel [10], indicating that some care is required when interpreting results obtained via this reduction. We also rederive the capacity of the more capable broadcast channel with independent message sets (found earlier using the reduction method [10]), extend the secrecy capacity results of the degraded wiretap channel to the more capable wiretap channel, and obtain the capacity of the broadcast channel with degraded message sets. In this paper, we overview the results while providing only proof sketches. Complete proofs are available elsewhere [11].

II. PRELIMINARIES

We will construct a probability space $(\Omega, \mathcal{F}, P)$ on which all stochastic processes considered here are defined. Stochastic
processes are denoted as \( X^T_t = (X_t, 0 \leq t \leq T) \). Let \( N^T_t \) denote the set of point process realizations on \([0, T]\), i.e., if \( N^T_t \in N^T_t \), then for \( t \in [0, T] \), \( N_t \) is integer valued, right continuous, has unit jumps with \( N_0 = 0 \). Fix \( 0 \leq t_1 < t_2 \leq T \), and define a point process \( N^T_{t_2} \) as having the same arrival instants as process \( N^T_t \) for \( t \in [t_1, t_2] \) and no arrivals before \( t_1 \) and after \( t_2 \). For given mappings \( A \) and \( B \) on the probability space, \( P^{A,B} \) denotes their joint distribution under \( P \), \( \text{log}(-) \) denote the natural logarithm.

**Definition 1:** Let \( X^T_0 \) be a non-negative process. A counting process \( N^T_0 \) is called a doubly stochastic Poisson process with rate process \( X^T_0 \) under measure \( P \) if conditioned on \( X^T_0 \), the increments of \( N^T_t \) are independent and the increment over \([s, t]\) is Poisson distributed with mean \( \int_s^t X_\tau \, d\tau \).

### III. CHANNEL MODEL

A code for the two receiver Poisson channel consists of an encoder \( \hat{g}^T_0 \) and two decoders \( g^T_0 \) and \( g^T_1 \). Let \( X^T_0 \) denote the set of all waveforms over \([0, T]\), which are non-negative, right-continuous with left limits, and peak power limited by unity. This is the set of inputs to the channel. The received signal at the first receiver \( Y^T_0 \) is a doubly stochastic Poisson process with rate \( a_y X^T_0 + \lambda_y \). Similarly the received signal at the second receiver is \( Z^T_0 \), where \( Z^T_0 \) is a doubly stochastic Poisson process with rate \( a_z X^T_0 + \lambda_z \).

Let \((X^T_0, \mathcal{F}^T_0, \mathbb{P})\) denote the input space, where \( \mathbb{X}^T \) is the state on \( X^T_0 \), generated by the open sets of \( X^T_0 \) when endowed with the Skorohod topology [12, Section 12]. Similarly, let \((N^T_0, \mathcal{N}^T, \mathbb{P})\) be the first and second receiver’s output space, respectively, where \( \mathbb{N}^T \) and \( \mathbb{Z}^T \) are the state spaces generated by the open sets of \( N^T_0 \) when endowed with the Skorohod topology. Let \( P^0_0 \) (resp. \( P^0_1 \)) be the probability measure on the first receiver’s (resp. second receiver’s) output space such that point process \( Y^T_0 \) (resp. \( Z^T_0 \)) is a unit rate Poisson process. Then we will take the output space of the channel to be the product space \((\mathcal{N}^T \times N^T_0, \mathcal{N}^T \otimes N^T_0, \mathbb{P})\) and our reference measure \( P_0 \) will be the product measure \( P_0 = P^0_1 \times P^0_0 \). Fix \( x^T_0 \in X^T_0 \), and let \( \Xi_{x^T_0} \) denote the transition probability function from the input space \((X^T_0, \mathbb{X}^T)\) to the output space \((\mathcal{N}^T \times N^T_0, \mathcal{N}^T \otimes N^T_0, \mathbb{P})\). Then the channel is modeled through following Radon-Nikodým derivative:

\[
\frac{d\Xi_{x^T_0}}{dP_0}(y^T_0, z^T_0) = \prod_{u=y,z} p_u(x^T_0, u^T_0) \prod_{a=x,z} p_a(x^T_0, u^T_0), \quad \text{where}
\]

\[
p_a(x^T_0, u^T_0) = \exp \left( \int_0^T \log(a_u x_t + \lambda_a) \, du_t + 1 - (a_u x_t + \lambda_a) \, dt \right).
\]

Then due to Girsanov’s theorems [13, Chapter VI, Theorems T2-T4], the Poisson process \( U^T_t \) has rate \( a_u x_t + \lambda_u \) under probability measure \( \Xi_{x^T_0} \) for \((u, U) \in \{(y, Y), (z, Z)\} \).

Let \( M \) be a random variable with a measurable space \((\mathcal{M}, \mathcal{M}^M)\), and let \( \mu_m(dx^M_0) \) denote the transition probability function from \((\mathcal{M}, \mathcal{M}^M)\) to the input space \((X^T_0, \mathcal{N}^T)\). Let \( \nu(dm) \) be a probability measure on \((\mathcal{M}, \mathcal{M}^M)\). Then these measures induce a joint measure \( P \) on \((\Omega, \mathcal{F})\), where

\[
\begin{align*}
\Omega &= \mathcal{M} \times \mathcal{X}^T_0 \times N^T_0 \times N^T_0 \\
\mathcal{F} &= \mathcal{M}^M \otimes \mathcal{N} \otimes N^T \otimes N^T \\
P &= \nu(dm) \mu_m(dx^M_0) P^0_1 (dy^T_0) P^0_0 (dz^T_0) \prod_{u=y,z} p_u(x^T_0, u^T_0). 
\end{align*}
\]

Thus under \( P \), we have \( M \Rightarrow X^T_0 \Rightarrow (Y^T_0, Z^T_0) \) and \( Y^T_0 \Rightarrow X^T_0 \Rightarrow Z^T_0 \). This Markov chain structure will play a dual role in upcoming analysis. First, if \( A \Rightarrow X^T_0 \Rightarrow U^T_t \) is a chain, then it implies the finiteness of mutual information (and hence absolute continuity of measures) of the form \( I(A; U^T_t) \) for \( U \in \{Y, Z\} \) due to the data processing inequality. Also it allows us to compute the likelihood ratio martingales through the intensity of point process \( U^T_t \).

**Definition 2:** For \( u \in \{y, z\} \), \( \phi_u(x) = (a_u x + \lambda_u) \log(a_u x + \lambda_u) \), with convention that \( 0 \log(0) = 0 \). Note that \( \phi_u(x) \) is convex and continuous for \( x \in [0, 1] \).

We now provide expressions for the above Radon-Nikodým derivatives, generalizing earlier expressions [3, 4] by allowing conditioning and auxiliary random variables. The proof uses the generalization of the converse Radon-Nikodým theorem [13, Theorem T12, Chapter VI].

**Theorem I:** Fix \( 0 \leq t_1 < t_2 \leq T \), and let \((u, U) \in \{(y, Y), (z, Z)\} \) with \( \lambda_u > 0 \).

1) Let \( A \Rightarrow X^T_0 \Rightarrow U^T_t \) be a Markov chain, and denote by

\[
\hat{P}^A U^T_t = P^A \times P^0_0.
\]

Then

\[
\log \left( \frac{dP^A U^T_t}{d\hat{P}^A U^T_t} \right) = \int_{t_1}^{t_2} \log(a_u \Pi_t + \lambda_u) \, du_t + 1 - (a_u \Pi_t + \lambda_u) \, dt,
\]

where the above equality is \( P^A U^T_t \)-a.s., and for \( t \in [t_1, t_2] \)

\[
\Pi_t = \mathbb{E}[X_t | A, U^T_t] \quad \text{\( P^A U^T_t \)-a.s.}
\]

2) Suppose that Markov chain \((A, B) \Rightarrow X^T_0 \Rightarrow U^T_t \) holds. Then

\[
I(A; U^T_t | B) = \int_{t_1}^{t_2} \mathbb{E}[\phi_u(E[X_t | U^T_t, A, B])] dt - \mathbb{E}[\phi_u(E[X_t | U^T_t, B])] dt - \mathbb{E}[\phi_u(E[X_t | U^T_t, B])] dt.
\]

We now discuss some properties of \( I(A; U^T_t | B) \). Lemma 1 and Lemma 2 can be proved using Theorem 1, the right-continuity of \( \sigma \)-fields, and boundedness and “right continuous with left limits” property of the rate processes.

**Lemma 1:** If \((A, B) \Rightarrow X^T_0 \Rightarrow U^T_t \) is a Markov chain, then

\[
\lim_{\delta \rightarrow 0^+} \frac{1}{\delta} I(A; U^T_{t+\delta} | B) = \mathbb{E}[\phi_u(E[X_t | U^T_t, A, B])] - \mathbb{E}[\phi_u(E[X_t | U^T_t, B])],
\]

361

2016 IEEE International Symposium on Information Theory
\[
\lim_{\delta \to 0^+} \frac{1}{\delta} I (X_{t-\delta} ; U_T^T, B) = \mathbb{E}[\phi_u (\mathbb{E}[X_t ; U^T, A, B])] \\
- \mathbb{E}[\phi_u (\mathbb{E}[X_t ; U^T, B])].
\]

**Lemma 2:** If \((A, B) \equiv X_0^T \equiv U^T_0\) is a Markov chain, then both \(\frac{1}{\delta} I (A; U_{s+\delta}^T | U_0^T, B)\) and \(\frac{1}{\delta} I (A; U_{s-\delta}^T | U_0^T, B)\) are bounded uniformly over \(s\) and \(\delta > 0\).

Using Lemma 1 and 2, we obtain the chain rule of mutual information in continuous time.

**Lemma 3** (Chain Rule in Continuous Time): If \((A, B) \equiv X_0^T \equiv U^T_0\) is a Markov chain, then
\[
I (A; U_0^T | B) = \lim_{\delta \to 0^+} \frac{1}{\delta} \int_0^T I (A; U_{s+\delta}^T | U_0^T, B) \, ds, \\
I (A; U_T^T | B) = \lim_{\delta \to 0^+} \frac{1}{\delta} \int_0^T I (A; U_{s-\delta}^T | U_0^T, B) \, ds.
\]
The following identity parallels the Csizár-sum identity \([14]\) for the discrete memoryless channels.

**Lemma 4:** With the channel model in (2)
\[
\lim_{\epsilon \to 0^+} \frac{1}{\epsilon} \int_0^T \frac{1}{\epsilon} I (Z_{t-\epsilon}; Y_0^T | T^T, M) \, dt = \lim_{\epsilon \to 0^+} \frac{1}{\epsilon} \int_0^T \frac{1}{\epsilon} I (Y^{t+\epsilon}; Z_{t}^T | Y_0^T, M) \, dt.
\]
This gives
\[
\int_0^T [\mathbb{E}[\phi_y (\mathbb{E}[X_t | Y_0^T, M])] - \mathbb{E}[\phi_z (\mathbb{E}[X_t | Z_{t}^T, M])] \] dt = \int_0^T [\mathbb{E}[\phi_y (\mathbb{E}[X_t | Y_0^T, Z_{t}^T, M])] - \mathbb{E}[\phi_z (\mathbb{E}[X_t | Z_{t}^T, M])] \] dt.

**Proof Sketch:** Using [15, Lemma 3.3] we get
\[
\int_0^T I (Z_{t-\epsilon}; Y_0^T | Z_{t}^T, M) \, dt = \int_0^T I (Z_{t-\epsilon}; Z_{t}^T; Y_0^T | M) \, dt \\
- \int_0^T I (Z_{t}^T; Y_0^T | M) \, dt = \int_0^T I (Z_{t}^T; Y_0^T | M) \, dt - \int_0^T I (Z_{t}^T; Y_0^T | M) \, dt.
\]
Similarly
\[
\int_0^T I (Y^{t+\epsilon}; Z_{t}^T | Y_0^T, M) \, dt = \int_0^T I (Y^{t+\epsilon}; Y_0^T; Z_{t}^T | M) \, dt \\
- \int_0^T I (Y_0^T; Z_{t}^T | M) \, dt
\]
Now dividing the integrands by \(\epsilon\), then taking \(\epsilon \to 0\), and using the properties of the mutual information we can show that the first part of the lemma holds. Then using Lemma 1 and 2 we can prove the second part.

**IV. LESS NOISY AND MORE CAPABLE TWO-RECEIVER POISSON CHANNEL**

Motivated by the corresponding definition for discrete memoryless channels \([8]\), we define less noisy and more capable receivers for the Poisson channel as follows.

**Definition 3** (Less Noisy Receiver): Receiver 1 is said to be less noisy than receiver 2 if \(I (M; Y_0^T) \geq I (M; Z_0^T)\) for all possible \(M\) in (2), where \(M \equiv X_0^T \equiv (Y_0^T, Z_0^T)\) is a Markov chain.

**Definition 4** (More Capable Receiver): Receiver 1 is said to be more capable than receiver 2 if \(I (X_0^T; Y_0^T) \geq I (X_0^T; Z_0^T)\) for all probability measures on the input space \((X_0^T, \mathcal{X})\).

We call a two-receiver Poisson channel less noisy if it has a less noisy receiver similarly for more capable.

**Theorem 2:** In a two user Poisson channel the following conditions are equivalent:

(I) \(\Phi (x) = \phi_y (x) - \phi_z (x)\) is a convex function over \([0,1]\).

(II) Receiver 1 is less noisy than receiver 2.

(III) Receiver 1 is more capable than receiver 2.

\(\Phi (x)\) is a convex function if and only if
\[ a_y \geq a_z \text{ and } a_y^2 \lambda_z \geq a_z^2 \lambda_y; \text{ or } 0 < a_y < a_z \text{ and } a_y^2 (a_z + \lambda_z) \geq a_z^2 (a_y + \lambda_y). \]

**Proof Sketch:** To prove (I) implies (II), Theorem 1 yields
\[
I (M; Y_0^T) - I (M; Z_0^T) = \int_0^T [\mathbb{E}[\phi_y (\mathbb{E}[X_t | Y_0^T, M])] - \mathbb{E}[\phi_y (\mathbb{E}[X_t | Y_0^T])] \] dt \\
- \int_0^T [\mathbb{E}[\phi_z (\mathbb{E}[X_t | Z_0^T, M])] - \mathbb{E}[\phi_z (\mathbb{E}[X_t | Z_0^T])] \] dt \equiv \int_0^T \mathbb{E}[\Phi (\mathbb{E}[X_t | Y_0^T, Z_0^T, M])] - \mathbb{E}[\Phi (\mathbb{E}[X_t | Y_0^T, Z_0^T])] \] dt
\]
where (a) is due to Lemma 4. Since \(\Phi (x)\) is a convex function, 
Jensen’s inequality gives \(I (M; Y_0^T) - I (M; Z_0^T) > 0\).

Note that (II) implies (III) trivially.

We now prove that (III) implies (I). There exists a sequence of input distribution (indexed by \(n\), such that \(X_0^T\) is binary, stationary with following limit \([3], [4]\)
\[
\lim_{n \to \infty} \mathbb{E} [\phi_u (\mathbb{E}[X_t | U^n])] = \phi_u (\mathbb{E}[X_t]).
\]
Thus choosing \(X_t\) such that \(P (X_t = p) = 1 - P (X_t = q) = \alpha, 0 \leq \alpha \leq 1\) and taking the limit, \(I (X_0^T; Y_0^T) \geq I (X_0^T; Z_0^T)\) gives
\[
\alpha \Phi (p) + (1 - \alpha) \Phi (q) \geq \Phi (\alpha p + (1 - \alpha) q).
\]
Hence \(\Phi (x)\) is a convex function.

For the discrete-time, binary, memoryless broadcast channel obtained from the Poisson broadcast channel via Wyner’s method, Kim et al. [10] determine the range of parameter values for which it is less noisy and more capable, and these conditions do not match those in Theorem 2. For \(a_y = 0.4, \lambda_y = 0.01, a_z = \lambda_z = 1\), for instance, the \(Y\) receiver in the discretized channel is more capable [10, Theorem 1].
whereas for the continuous-time, continuous-space channel it is not (Theorem 2). The reason is as follows: for these channel parameters there exists a process $X_T^T$ such that $I(X_T^T; Z_T^T) > I(X_T^T; Y_T^T)$. But $X_T^T$ takes two values that are both close to one, whereas Kim et al. effectively restrict $X_T^T$ to take values in $\{0, 1\}$, which is sufficient for communication purposes. Thus the Kim et al. conditions guarantee that the continuous-time channel is actually something more akin to “essentially” less noisy or more capable [16].

Armed with the above stochastic calculus tools, coding theorems for various multi-decoder Poisson channel track those for discrete memoryless channels.

V. MORE CAPABLE POISSON BROADCAST CHANNEL

An $(L_y, L_z, T)$ code for the Poisson broadcast channel consists of a source (equipped with an encoder $\mathcal{E}_y^T$), two users each with a decoder ($\mathcal{D}_y^T$ and $\mathcal{D}_z^T$). The source has two independent messages $M_y$ and $M_z$ for the first and second user respectively, where $M_y$ and $M_z$ are uniformly distributed on sets $\mathcal{M}_y = \{1, \ldots, L_y\}$ and $\mathcal{M}_z = \{1, \ldots, L_z\}$ respectively. Given messages $M_y$ and $M_z$ the encoder selects a waveform in $X_0^T$. The average probability of error for this code is denoted by $P_e$. A rate pair $(R_y, R_z)$ is said to be achievable if for all $\epsilon > 0$ and sufficiently large $T$, there exists an $(L_y, L_z, T)$ code such that

$$\frac{\log(L_y)}{T} \geq R_y - \epsilon, \quad \frac{\log(L_x)}{T} \geq R_z - \epsilon, \quad P_e \leq \epsilon. \quad (5)$$

The capacity region $(C_y, C_z)$ is the closure of achievable rate pairs.

**Theorem 3:** The capacity of the more capable Poisson broadcast channel when receiver 1 is more capable than receiver 2 is given by union over all $0 \leq \alpha \leq \frac{1}{2}$ and $0 \leq p, q \leq 1$ of rate pairs satisfying

$$R_y \leq C_y = \alpha(p\phi_y(1) + (1 - p)\phi_y(0) - \phi_y(p)) + (1 - \alpha)(q\phi_y(1) + (1 - q)\phi_y(0) - \phi_y(q)) \quad (8)$$

$$R_z \leq C_z = \alpha\phi_z(p) + (1 - \alpha)\phi_z(q) - \phi_z(\alpha p + (1 - \alpha)q). \quad (9)$$

Although the proof of above theorem can be found in [10], we provide alternate proof using the tools derived from the stochastic calculus theory without resorting to discretization. Note that Kim et al. [10] obtain the capacity of the Poisson broadcast channel with independent message sets over wider range of channel parameters than those considered here. Also, currently we only consider the case when both of the users have positive dark current rates, while Kim et al. do not have that restriction.

**A. Achievability**

The achievability argument uses superposition coding. Let $T_n = n\tau$ for $\tau > 0$. Construct an auxiliary process $V_0^T$ to be a piecewise-constant, binary-valued process. We divide the interval $[0, T_n]$ into $n$ subintervals each of equal length $\tau$. The process will be constant on each of these sub-interval with value given by

$$V_t = \tilde{V}_{i} \quad \text{for} \quad (i - 1)\tau \leq t < i\tau, \quad i = 1, 2, \ldots, n \quad (6)$$

where the $\tilde{V}_{i}$’s are independent and identically Bernoulli random variables with $P(V_i = 1) = \alpha$. The input waveform $X_0^T$ is binary, piecewise constant with

$$X_t = \tilde{X}_i \quad \text{for} \quad (i - 1)\tau \leq t < i\tau, \quad i = 1, 2, \ldots, n \quad (7)$$

where

$$P(X_t = 1|V_i = 1) = 1 - P(X_t = 0|V_i = 1) = p$$

$$P(X_t = 1|V_i = 0) = 1 - P(X_t = 0|V_i = 0) = q. \quad (8)$$

Let $Q^{Y_{T_0}, X_{T_0}}_{V_0} = P_{Y_{T_0}}^{V_{T_0}} \times P_{X_{T_0}}^{V_{T_0}}$, where for fixed $V_0^{T_0} \in V_0^{T_0}$, $P_{X_{T_0}}^{V_{T_0}}$ denotes the probability measure on the input space (resp. output space of the first receiver). Then following mutual information densities can be shown to exist:

$$i(X_0^T; Y_0^T) = \log \left( \frac{dP_{X_0^T,Y_0^T}}{dP_{X_0^T,Y_0^T}^{V_0}} \right) \quad (10)$$

$$i(Y_0^T; Z_0^T) = \log \left( \frac{dP_{Y_0^T,Z_0^T}}{dP_{Y_0^T,Z_0^T}^{V_0}} \right) \quad (11)$$

Let $\tilde{C}_y = \alpha\phi_y(p) + (1 - \alpha)\phi_y(q) - \phi_y(\alpha p + (1 - \alpha)q)$. Then using Theorem 1, Lemma 1, and law of large numbers we can show that with the above input to the channel for all $\epsilon > 0$ there exists $\tau$ and $N$ such that if $n \geq N$ and $\tau \leq \tau$ then

$$P \left( \frac{1}{T_n}i(X_0^T; Y_0^T) - C_y + \tilde{C}_y \right) > \epsilon \leq \epsilon \quad (12)$$

$$P \left( \frac{1}{T_n}i(Y_0^T; Z_0^T) - C_z \right) > \epsilon \leq \epsilon \quad (13)$$

With these results the argument for implementing the superposition coding using the typical set decoding can be completed.

**B. Converse**

For the given sequence of $(L_y, L_z, T)$ code, using Wyner’s lemma, Fano’s inequality, and standard manipulations we get

$$R_y \leq \frac{1}{T}I(X_0^T; Y_0^T|M_z) + \varpi_\epsilon, \quad R_z \leq \frac{1}{T}I(M_z; Z_0^T) + \varpi_\epsilon \quad (14)$$

where $\varpi_\epsilon \to 0$ as $\epsilon \to 0$. Then using Theorem 1, Lemma 4, and convexity of $\Phi(x)$ we can show that

$$R_y \leq E[\phi_y(X_S)] - E[\phi_y(E[X_S|Y_0^S, M_z])] + \varpi_\epsilon$$

$$R_z \leq E[\phi_z(E[X_S|Y_0^S, M_z])] - \phi_z(E[X_S]) + \varpi_\epsilon \quad (15)$$

where $S$ is a random variable uniformly distributed on $[0, T]$ and independent of all other variables. Now using Fenchel-Eggleston-Carathéodory’s theorem and taking $\epsilon \to 0$, we get

$$R_y \leq C_y, \quad R_z \leq C_z. \quad (16)$$
VI. MORE CAPABLE POISSON WIRETAP CHANNEL

Here we will consider the first receiver to be the legitimate user and the second receiver to be an eavesdropper. The transmitter (equipped with a stochastic encoder) wishes to communicate a message $M$ which is uniformly distributed on $M = \{1, \ldots, L\}$ to the legitimate user. To transmit message $M = m$, the encoder chooses the input waveform $X_0^T \in X_0^T$. Upon observing $Y_0^T$, the legitimate decoder chooses a symbol $\hat{M} \in M$. We will call such an arrangement an $(L, T)$ code. The capacity region is the closure of the achievable rate pairs. The average probability of error at the legitimate receiver is denoted by $P_e$. The metric to measure the secrecy will be $\frac{1}{T} I(M; Z_0^T)$. A secrecy rate $R_0$ is said to be achievable for the Poisson wiretap channel if for all $\epsilon > 0$ and for all sufficiently large $T$, there exists an $(L, T)$ code such that

$$\frac{\log(L)}{T} \geq R_0 - \epsilon, P_e \leq \epsilon, \frac{1}{T} I(M; Z_0^T) \leq \epsilon. \quad (13)$$

The secrecy capacity is defined to be the supremum of achievable secrecy rates. Theorem 4: The secrecy capacity of the more capable Poisson wiretap channel is

$$C_s = \max_{0 \leq \alpha \leq 1} \alpha \Phi(1) + (1 - \alpha) \Phi(0) - \Phi(\alpha),$$

where we recall $\Phi(x) = \phi_y(x) - \phi_z(x)$ and $\Phi(x)$ is a convex function since the channel is more capable. The achievability argument is identical to that for the degraded Poisson wiretap channel in [17], hence we will only discuss the converse here. Using Wyner’s lemma and Fano’s inequality, we get

$$R_s \leq \frac{1}{T} \left( I(M; Y_0^T) - I(M; Z_0^T) \right) + \varepsilon \epsilon, \quad (14)$$

where $\varepsilon \rightarrow 0$ as $\epsilon \rightarrow 0$. Then using Theorem 1, Lemma 4, and convexity of $\Phi(x)$ we can show that

$$R_s \leq C_s + \varepsilon \epsilon.$$ 

Taking $\epsilon \rightarrow 0$ yields the converse.

VII. GENERAL POISSON BROADCAST CHANNEL WITH DEGRADED MESSAGE SETS

In this section the setting is that a common message $M_o \in \mathcal{M}_o = \{1, \ldots, L_0\}$ for both of the users and a private message $M_y \in \mathcal{M}_y = \{1, \ldots, L_y\}$ for the first user. Messages $M_0$ and $M_y$ are assumed to be independent and uniformly distributed. The transmitter uses an encoder $\Theta_x^T$ which maps these messages into an input $X_0^T$. We will call above setup an $(L_0, L_y, T)$ code. The rate pair $(R_0, R_y)$ is said to be achievable if for any $\epsilon > 0$ and for any sufficiently large $T$, there exists an $(L_0, L_y, T)$ code such that

$$\frac{\log(L_o)}{T} \geq R_0 - \epsilon, \frac{\log(L_y)}{T} \geq R_y - \epsilon, P_e \leq \epsilon, \quad (15)$$

where $P_e$ is average probability of the error for above code. The capacity region is the closure of the achievable rate pairs. Theorem 5: The capacity region of the general Poisson broadcast channel with degraded message sets is given by the union over all $0 \leq \alpha_i, p_i \leq 1, i = 1, 2, 3$ with $\sum_{i=1}^{3} \alpha_i = 1$ of rate pairs satisfying:

$$R_0 \leq C_0, \quad R_0 + R_y \leq \min(\tilde{C}_y + \tilde{C}_z + \tilde{C}_y),$$

where

$$\tilde{C}_z = \sum_{i=1}^{3} \alpha_i \phi_z(p_i) - \phi_z \left( \sum_{i=1}^{3} \alpha_i p_i \right),$$

$$\tilde{C}_y = \sum_{i=1}^{3} \alpha_i (p_i \phi_y(1) + (1 - p_i) \phi_y(0) - \phi_y(p_i)),$$

$$\tilde{C}_y = \sum_{i=1}^{3} \alpha_i \phi_y(p_i) - \phi_y \left( \sum_{i=1}^{3} \alpha_i p_i \right).$$

The achievability argument uses superposition coding and is similar to the superposition coding argument for the broadcast channel with independent message sets. The converse argument can be completed using Theorem 1, Lemma 4, and Fenchel-Eggleston-Carathéodory’s theorem.

REFERENCES

[1] S. Personick, “Receiver design for digital fiber optic communication systems, II,” The Bell System Technical Journal, vol. 52, no. 6, pp. 875–886, July 1973.
[2] J. Mazo and J. Salz, “On optical data communication via direct detection of light pulses,” The Bell System Technical Journal, vol. 55, no. 3, pp. 347–369, Mar 1976.
[3] Y. Kabanov, “The capacity of a channel of the Poisson type,” Theory of Probability and Applications, vol. 23, pp. 143–147, 1978.
[4] M. Davis, “Capacity and cutoff rate for Poisson-type channels,” IEEE Transactions on Information Theory, vol. 26, no. 6, pp. 710–715, Nov 1980.
[5] A. Wyner, “Capacity and error exponent for the direct detection photon channel- Part I and II,” IEEE Transactions on Information Theory, vol. 34, no. 6, pp. 1449–1471, Nov 1988.
[6] C. Shannon, “A mathematical theory of communication,” The Bell System Technical Journal, vol. 27, no. 3, pp. 379–423, July 1948.
[7] P. Bergmans, “Random coding theorem for broadcast channels with degraded components,” IEEE Transactions on Information Theory, vol. 19, no. 2, pp. 197–207, Mar 1973.
[8] J. Körner and K. Marton, “Comparison of two noisy channels,” Topics in Information Theory, no. 16, 1977.
[9] P. Bergmans, “The capacity of a class of broadcast channels,” IEEE Transactions on Information Theory, vol. 25, no. 2, pp. 166–169, Mar 1979.
[10] H. Kim, B. Nachman, and A. E. Gamal, “Superposition coding is almost always optimal for the Poisson broadcast channel,” 2015. [Online]. Available: http://arxiv.org/abs/1508.04228
[11] N. V. Shende and A. B. Wagner, “The stochastic-calculus approach to multi-receiver Poisson channels,” http://arxiv.org, to appear.
[12] P. Billingsley, Convergence of Probability Measures, 2nd ed. Wiley Series in Probability and Statistics, 1999.
[13] P. Brémaud, Point Processes and Queues: Martingale Dynamics. Springer-Verlag, 1981.
[14] I. Csiszár and J. Körner, “Broadcast channels with confidential messages,” IEEE Transactions on Information Theory, vol. 24, no. 3, pp. 339–348, May 1978.
[15] A. Wyner, “A definition of conditional mutual information for arbitrary ensembles,” Information and Control, vol. 38, no. 1, pp. 51 – 59, 1978.
[16] C. Nair, “Capacity regions of two new classes of two-receiver broadcast channels,” IEEE Transactions on Information Theory, vol. 56, no. 9, pp. 4207–4214, Sept 2010.
[17] A. Lourrime and A. B. Wagner, “The degraded Poisson wiretap channel,” IEEE Transactions on Information Theory, vol. 58, no. 12, pp. 7073–7085, Dec 2012.