Properties of the Support of the Capacity-Achieving Distribution of the Amplitude-Constrained Poisson Noise Channel

Alex Dytso, *Member, IEEE*, Luca Barletta, *Member, IEEE*, and Shlomo Shamai (Shitz), *Life Fellow, IEEE*

Abstract—This work considers a Poisson noise channel with an amplitude constraint. It is well-known that the capacity-achieving input distribution for this channel is discrete with finitely many points. We sharpen this result by introducing upper and lower bounds on the number of mass points. Concretely, an upper bound of order $\mathsf{A} \log^2(\mathsf{A})$ and a lower bound of order $\sqrt{\mathsf{A}}$ are established where A is the constraint on the input amplitude. In addition, along the way, we show several other properties of the capacity and capacity-achieving distribution. For example, it is shown that the capacity is equal to $-\log P_{Y^\star}(0)$ where $P_{Y^\star}$ is the optimal output distribution. Moreover, an upper bound on the values of the probability masses of the capacity-achieving distribution and a lower bound on the probability of the largest mass point are established. Furthermore, on the per-symbol basis, a nonvanishing lower bound on the probability of error for detecting the capacity-achieving distribution is established under the maximum a posteriori rule.

Index Terms—Amplitude constraint, Poisson noise channel, optical communications, capacity, discrete distributions, strong data-processing inequality.

I. INTRODUCTION

WE CONSIDER a discrete-time memoryless Poisson channel. The output Y of this channel takes value on the set of nonnegative integers $\mathbb{N}_0$ and the input X takes value on the set of nonnegative real numbers. The conditional probability mass function (pmf) of the output random variable Y given the input X that specifies the channel is given by

$$P_{Y|X}(y|x) = \frac{1}{y!} x^y e^{-x}, \quad x \geq 0, \ y \in \mathbb{N}_0. \quad (1)$$

In (1), we use the standard convention that $0^0 = 1$ and $0! = 1$.

The capacity of this channel where the input X is subject to the amplitude constraint $0 \leq X \leq \mathsf{A}$ is given by

$$C(\mathsf{A}) = \max_{X : 0 \leq X \leq \mathsf{A}} I(X; Y), \quad \mathsf{A} > 0. \quad (2)$$

Finding the capacity of this channel remains to be an elusive task. The goal of this work is to make progress on this problem by studying the properties of the capacity-achieving distribution denoted by $P_{X^\star}$.

A. Contributions and Paper Outline

The outline and the contribution of the paper are as follows. The remaining part of Section I will survey the relevant literature, present our notation, and will go over the key tools needed in our analysis, such as the oscillation theorem and the strong data-processing inequality. Section II presents our main results, which include the following: a new compact representation of the capacity; an upper bound on the values of probability masses of the optimal input distribution $P_{X^\star}$; a lower bound on the probability of the largest mass points (*i.e.*, $P_{X^\star}(\mathsf{A})$); a lower bound on the size of the support of $P_{X^\star}$; an upper bound on the size of the support of $P_{X^\star}$; a nonvanishing lower bound, on the per-symbol basis, on the probability of error for detecting the capacity-achieving distribution. Section III is dedicated to the proofs. Section IV concludes the paper with some final remarks and possible future directions.

B. Prior Work

The now classical approach developed by Witsenhausen in [1] says that if the output alphabet has cardinality n, then the support of the optimal input distribution cannot be more than n, irrespective of the size of the input alphabet. However, since the output alphabet of the Poisson noise channel has a countably infinite cardinality, the Witsenhausen approach does not apply. Instead, the approach that has been applied to the Poisson noise channel largely follows the analyticity idea introduced by Smith in [2] in the context of amplitude-constrained Gaussian noise channel. The interested reader is referred to [3] for a summary of these techniques. In this work, we also follow the latter technique. However, we considerably generalize and improve this approach. In what follows, we summarize the known results on the Poisson noise channel and highlight the elements of the new technique.

The discrete-time Poisson noise channel is suited to model low-intensity, direct-detection optical communication channels [4]; the interested reader is also referred to a survey on free-space optical communications in [5]. The Poisson channel can be seen as a limiting case of the Binomial channel [6], which can be used to model the number of particles released by a sender unit in molecular communications [7]. A key difference in the mathematical formulation between the Poisson and the Binomial channel models is the infinite/finite nature of the output alphabet. In this work, we are concerned with the discrete-time channel; however, there also exists a large literature on continuous-time channels, and the interested reader is referred to a survey in [8].

The first major study of the capacity-achieving distribution for the Poisson channel was undertaken in [9], where the authors considered the capacity with and without an additional average-power constraint on the input,[1] that is

$$\mathbb{E}[X] \leq \mathsf{P}. \tag{3}$$

The authors of [9] derived the Karush-Kuhn-Tucker (KKT) conditions that are necessary to study the structure of the capacity-achieving distribution. These KKT conditions were then used to show that the support of an optimal input distribution for any A can contain at most one mass point in the open interval $(0, 1)$. Moreover, for any $\mathsf{A} \leq 1$, it was shown that the optimal input distribution consists of two mass points at 0 and A and is given by

$$P_{X^\star}(\mathsf{A}) = \frac{1}{e^{\frac{\mathsf{A}}{e^\mathsf{A}-1}} - e^{-\mathsf{A}} + 1}, \quad P_{X^\star}(0) = 1 - P_{X^\star}(\mathsf{A}), \tag{4}$$

and the capacity is given by

$$C(\mathsf{A}) = -\log\left(P_{X^\star}(0) + e^{-\mathsf{A}} P_{X^\star}(\mathsf{A})\right). \tag{5}$$

The KKT conditions proposed in [9] were rigorously derived in [11, Corollary 1] and extended to a more general case that includes the possibility of a nonzero dark current[2] parameter. Moreover, using the analyticity idea of [2], in [11] for $\mathsf{A} < \infty$ and any $\mathsf{P} > 0$ it was shown that the optimizing input distribution is unique and discrete with finitely many mass points. Moreover, for the case of $\mathsf{P} \geq \mathsf{A}$ (*i.e.*, the average-power constraint is not active) and dark current is zero, it was shown that the distribution in (4) continues to be capacity-achieving if and only if $\mathsf{A} \leq \bar{\mathsf{A}}$ where $\bar{\mathsf{A}} \approx 3.3679$.

Further studies of the conditions under which the capacity-achieving distribution is binary have been undertaken by the authors of [13] and [14]. For example, in [13], it was

shown that with both the amplitude and the average-power constraint, the optimal input distribution always contains a mass point at 0. Moreover, in the case of only an amplitude constraint, the optimal input distribution contains mass points at both 0 and A. In [14], it was shown that if $\mathsf{P} < \frac{\mathsf{A}}{2}$ and the dark current is large enough, the following binary distribution is optimal:

$$P_{X^\star}(0) = 1 - \frac{\mathsf{P}}{\mathsf{A}}, \quad P_{X^\star}(\mathsf{A}) = \frac{\mathsf{P}}{\mathsf{A}}. \tag{6}$$

The capacity-achieving distribution with only an average-power constraint was considered in [15] and was shown to be discrete with infinitely many mass points.

The low-average-power and the low-amplitude asymptotics of the capacity have been studied in [16]–[20]. A number of papers have also focused on upper and lower bounds on the capacity. The first upper and lower bounds on the capacity have been derived in [9] for two situations: the case of the average-power constraint only, and the case of both the average-power and the amplitude constraint with $\mathsf{A} \leq 1$. The authors of [12] derived upper and lower bounds, in the case of the average-power constraint only, by focusing on the regime where both P and the dark current tend to infinity with a fixed ratio. Firm upper and lower bounds on the capacity in the case of only the average-power constraint and no dark current have been derived in [16] and [17]. Bounds in [16] and [17] have been further improved in [15] and [21]. The most general bounds on the capacity that consider both the amplitude and the average-power constraints on the input and hold for an arbitrary value of the dark current have been derived in [22]. The bounds in [22] have been shown to be tight in the regime where both the average-power and the amplitude constraint approach infinity with a fixed ratio $\frac{\mathsf{P}}{\mathsf{A}}$. Finally, the authors of [23] sharpened the results of [22] for small values of P and A.

C. Notation

Throughout the paper, the deterministic scalar quantities are denoted by lower-case letters and random variables are denoted by uppercase letters. The set of all real numbers, nonnegative integers and positive integers are denoted by $\mathbb{R}$, $\mathbb{N}_0$, and $\mathbb{N}$, respectively. The function $\mathbb{1}_\mathcal{A}(x)$ denotes the indicator function over the set $\mathcal{A}$ where $\mathbb{1}_\mathcal{A} = 1$ if $x \in \mathcal{A}$ and $\mathbb{1}_\mathcal{A} = 0$ otherwise.

We denote the distribution of a random variable X by P_X. The support set of P_X is denoted and defined as

$$\mathsf{supp}(P_X) = \{x : \text{ for every open set } \mathcal{D} \ni x$$
$$\text{we have that } P_X(\mathcal{D}) > 0\}. \tag{7}$$

The relative entropy between distributions P and Q will be denoted by $\mathsf{D}(P\|Q)$. A Poisson pmf with expected value x is denoted by $\mathcal{P}(x)$.

The number of zeros of a function $f \colon \mathbb{R} \to \mathbb{R}$ on the interval $\mathcal{I}$ is denoted by $\mathrm{N}(\mathcal{I}, f)$. If the set of zeros of f on $\mathcal{I}$ has either countably infinite or uncountable cardinality, we set $\mathrm{N}(\mathcal{I}, f) = \infty$. Let g_1 and g_2 be nonnegative functions, then

---

[1] In the Poisson noise channel the average-power is measured in terms of the first moment of the input X. The term power is appropriate if we consider laser communications where $\mathbb{E}[X] = \mathsf{P}$ represents average photon count. In this case, $\mathsf{P} = \frac{\eta P_t}{h f_0}$ where P_t is the transmit power, f_0 is the optical carrier frequency, h is Plank's constant and η is the proportionality constant [4], [10].

[2] A more general Poisson noise model also incorporates a nonnegative parameter λ known as the dark current parameter, and the channel is given by $P_{Y|X}(y|x) = \frac{1}{y!}(x + \lambda)^y e^{-(x+\lambda)}$, $x \geq 0, y \in \mathbb{N}_0$. The dark current parameter λ represents the intensity of an additional source of noise or interference, which produces on average extra λ photons at a particle counter [4], [10]–[12].

- $g_1(x) = O(g_2(x))$ means that there exists a constant $c > 0$ and x_0 such that $\frac{g_1(x)}{g_2(x)} \leq c$ for all $x > x_0$;
- $g_1(x) = \Omega(g_2(x))$ means that $g_2(x) = O(g_1(x))$; and
- $g_1(x) = o(g_2(x))$ means $\lim_{x\to\infty} \frac{g_1(x)}{g_2(x)} = 0$.

Finally, the Lambert W-function is denoted by W_k where k indicates the branch [24]. Since we are dealing with real numbers, we only use the principle branch W_0 and the lower branch W_{-1}. Recall that, for the real numbers x and y, the Lambert W-function provides a solution to the equation $ye^y = x$, which can be solved for y only if $x \geq e^{-1}$. Moreover, in the regime $x \geq 0$, the solution is unique and is given by $y = W_0(x)$, and in the regime $e^{-1} \leq x < 0$, there are two solutions given by $y = W_0(x)$ and $y = W_{-1}(x)$.

D. Overview of the Key Tools

In this section, we overview the main tools needed in our analysis.

1) Strong Data-Processing Inequality: In order to find an upper bound on the values of probability masses and a lower bound on the number of mass points, we will rely on the strong data-processing inequality for the relative entropy. The study of the strong data-processing inequalities has recently received some attention, and the interested reader is referred to [25]–[27] and references therein.

Fix some channel $Q_{Y|X} : \mathcal{X} \to \mathcal{Y}$ where $\mathcal{X}$ and $\mathcal{Y}$ are the input and output alphabets, respectively. Let Q_Y denote the distribution on $\mathcal{Y}$ induced by the push-forward of the distribution Q_X on $\mathcal{X}$ through $Q_{Y|X}$. We denote this operation by $Q_X \to Q_{Y|X} \to Q_Y$.

The classical data-processing inequality for the relative entropy states that for $Q_X \to Q_{Y|X} \to Q_Y$ and $P_X \to Q_{Y|X} \to P_Y$, we have

$$D(P_Y \| Q_Y) \leq D(P_X \| Q_X). \qquad (8)$$

The strong data-processing inequality is the sharpening of (8) where one seeks to find a coefficient $0 < \eta \leq 1$ such that

$$D(P_Y \| Q_Y) \leq \eta\, D(P_X \| Q_X). \qquad (9)$$

It is not difficult to see that

$$\eta \leq \eta_{\mathsf{KL}}(\mathcal{X}, Q_{Y|X}) := \sup_{\substack{Q_X, P_X: \\ \mathsf{supp}(Q_X) \subseteq \mathcal{X},\, \mathsf{supp}(P_X) \subseteq \mathcal{X} \\ 0 < D(Q_X \| P_X) < \infty \\ Q_X \to Q_{Y|X} \to Q_Y \\ P_X \to Q_{Y|X} \to P_Y}} \frac{D(P_Y \| Q_Y)}{D(P_X \| Q_X)}.$$

$$(10)$$

The quantity $0 < \eta_{\mathsf{KL}}(\mathcal{X}, Q_{Y|X}) \leq 1$ is known as the contraction coefficient. In Section III-B, for the case of Poisson channel, we will provide a nontrivial upper bound on $\eta_{\mathsf{KL}}(\mathcal{X}, Q_{Y|X})$.

2) Oscillation Theorem: To find an upper bound on the number of points in the support of $P_{X^\star}$, we will follow the proof technique developed in [28] for the Gaussian noise channel. The key step that we borrow from [28] is the use of the variation-diminishing property, which is captured by the oscillation theorem of Karlin [29]. To state the oscillation theorem we need the following definition.

Definition 1 (Sign Changes of a Function): The number of sign changes of a function $\xi : \mathcal{X} \to \mathbb{R}$ is given by

$$\mathscr{S}(\xi) = \sup_{m \in \mathbb{N}} \left\{ \sup_{y_1 < \cdots < y_m \subseteq \mathcal{X}} \mathcal{N}\{\xi(y_i)\}_{i=1}^m \right\}, \qquad (11)$$

where $\mathcal{N}\{\xi(y_i)\}_{i=1}^m$ is the number of sign changes of the sequence $\{\xi(y_i)\}_{i=1}^m$ (where a zero is treated as a positive number).

The following theorem, shown in [29, Thm. 3] (see also [30, Thm. 3.1, p. 21]), will be a key step in the proof of the upper bound on the number of mass points.

Theorem 1 (Oscillation Theorem): Given domains $\mathbb{I}_1$ and $\mathbb{I}_2$, let $p \colon \mathbb{I}_1 \times \mathbb{I}_2 \to \mathbb{R}$ be a strictly totally positive kernel.[3] For an arbitrary y, suppose $p(\cdot, y) \colon \mathbb{I}_1 \to \mathbb{R}$ is an n-times differentiable function. Assume that μ is a regular σ-finite measure on $\mathbb{I}_2$, and let $\xi \colon \mathbb{I}_2 \to \mathbb{R}$ be a function with $\mathscr{S}(\xi) = n$. For $x \in \mathbb{I}_1$, define

$$\Xi(x) = \int \xi(y)\, p(x, y)\, \mathrm{d}\mu(y). \qquad (12)$$

If $\Xi \colon \mathbb{I}_1 \to \mathbb{R}$ is an n-times differentiable function, then either $\mathrm{N}(\mathbb{I}_1, \Xi) \leq n$, or $\Xi \equiv 0$.

The above theorem says that the number of zeros of a function $\Xi(x)$, which is the output of integral transformation, is less than the number of sign changes of the function $\xi(y)$, which is the input to the integral transformation. In the rest of this paper, we will take the kernel $p(x, y)$ to be the Poisson transition probability $P_{Y|X}$ in (1), μ to be the counting measure, and the domains will be set to $\mathbb{I}_1 = \{x : x \geq 0\}$ and $\mathbb{I}_2 = \mathbb{N}_0$. The fact that the Poisson transition probability is a strictly totally positive kernel was shown in [29]. The interested reader is also referred to [30, p. 19], where it is shown that members of the exponential family (*e.g.*, Poisson, Gaussian, chi-squared) are positive definite kernels. Figure 1 shows an example of Theorem 1 for the Poisson kernel.

The oscillation theorem allows to upper bound the number of points in the support of $P_{X^\star}$ with the number of sign changes of a function that is related to the output distribution $P_{Y^\star}$. In the case of the Gaussian noise channel, in order to count the number of sign changes, one needs to resort to complex analytic techniques. In contrast, in the Poisson case, due to the discrete nature of the channel, we no longer need to rely on the complex analytic techniques, which simplifies this part of the analysis.

However, the analysis for the Poisson channel is not necessarily simpler than that in the Gaussian case. One crucial step in both proofs relies on finding a lower bound of the output pdf in the Gaussian case and the output pmf in the Poisson case. In the Gaussian case, this can be done by using Jensen's inequality, and the resulting bound is universal and is independent of $P_{X^\star}$. In the Poisson noise case, however, a distribution-independent bound cannot be obtained, and the lower bound on the tail depends on $P_{X^\star}$. Specifically,

---

[3] A function $f : \mathbb{I}_1 \times \mathbb{I}_2 \to \mathbb{R}$ is said to be a strictly totally positive kernel of order n if $\det\left([f(x_i, y_j)]_{i,j=1}^m\right) > 0$ for all $1 \leq m \leq n$, and for all $x_1 < \cdots < x_m \in \mathbb{I}_1$, and $y_1 < \cdots < y_m \in \mathbb{I}_2$. If f is a strictly totally positive kernel of order n for all $n \in \mathbb{N}$, then f is a strictly totally positive kernel.
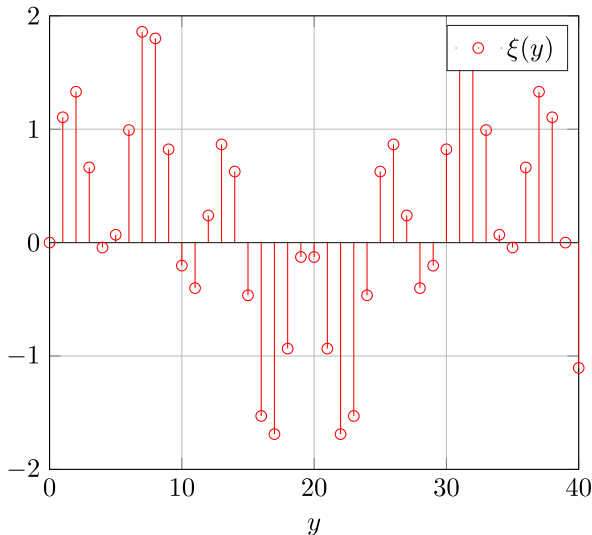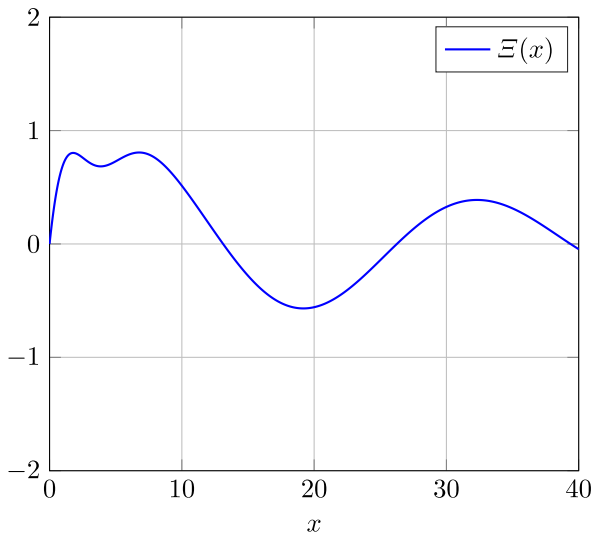
(a) Input function $\xi(y)$.

(b) Output function $\Xi(y)$.

Fig. 1. An example of the oscillation theorem with the Poisson kernel where we set $\xi(y) = \sin\left(\frac{\pi}{3}y\right) + \sin\left(\frac{\pi}{13}y\right)$, $y \in \mathbb{N}_0$.

the lower bound depends on the value of $P_{X^\star}(\mathsf{A})$. Therefore, to complete the proof, we need to also find a lower bound on $P_{X^\star}(\mathsf{A})$.

3) Connection Between Estimation and Information Measures: Some of the key intermediate steps in our proofs will rely on identities that connect information measures and estimation measure. In particular, we will rely on the expression for the conditional expectation derived in [31]. For further connections between estimation and information theoretic measure the interested reader is referred to [32] and [33] and references therein.

II. MAIN RESULTS

The main results of this paper are summarized in the following theorem.

Theorem 2: The capacity and the capacity-achieving distribution $P_{X^\star}$ of an amplitude-constrained Poisson noise channel satisfy the following properties:
- *A New Capacity Expression:* For every $\mathsf{A} \geq 0$, the capacity is given by

$$C(\mathsf{A}) = \log \frac{1}{P_{Y^\star}(0)}, \tag{13}$$

where $P_{Y^\star}$ is the capacity-achieving output distribution induced by $P_{X^\star}$.
- *An Upper Bound on the Probabilities:* For every $\mathsf{A} \geq 0$,

  Universal Bound: For $x \in \mathsf{supp}(P_{X^\star})$

$$P_{X^\star}(x) \leq \mathrm{e}^{-\frac{C(\mathsf{A})}{1-\mathrm{e}^{-\mathsf{A}}}}, \tag{14}$$

  Location Dependent Bound: For $x \in \mathsf{supp}(P_{X^\star}) \setminus \{0\}$

$$P_{X^\star}(x) \leq \mathrm{e}^{-C(\mathsf{A})-x\frac{\mathrm{e}^{-x}}{1-\mathrm{e}^{-x}}}. \tag{15}$$

In addition, if $|\mathsf{supp}(P_{X^\star})| = 2$, then the bound in (15) becomes equality.
- *A Lower Bound on the Probability of the Largest Point:* For all A such that $\mathrm{e}^{\frac{C(\mathsf{A})}{1-\mathrm{e}^{-\mathsf{A}}}} \geq 4$, we have that

$$P_{X^\star}(\mathsf{A}) \geq \frac{1 - 3\,\mathrm{e}^{-\frac{C(\mathsf{A})}{1-\mathrm{e}^{-\mathsf{A}}}}}{2\mathsf{A}^{2\mathsf{Ae}\log(\mathsf{A})+2}\,\mathrm{e}^{-2\mathsf{A}+1}}. \tag{16}$$

- *On the Location of Support Points:* Suppose that $|\mathsf{supp}(P_{X^\star})| \geq 3$ and let $\mathsf{supp}(P_{X^\star}) \setminus \{0, \mathsf{A}\}$. Then,

$$\mathrm{e}^{-\sqrt{2(\log(\mathsf{A})-1)}} \leq \mathsf{A}\mathrm{e}^{W_{-1}\left(-\frac{1}{\mathsf{A}}\right)} \leq x^\star \leq \mathsf{A}\mathrm{e}^{W_0\left(-\frac{1}{\mathsf{A}}\right)} \leq \mathsf{A} - 1. \tag{17}$$

- *A Lower Bound on the Size of the Support:* For every $\mathsf{A} \geq 0$,

$$|\mathsf{supp}(P_{X^\star})| \geq \mathrm{e}^{\frac{C(\mathsf{A})}{1-\mathrm{e}^{-\mathsf{A}}}}. \tag{18}$$

- *An Upper Bound on the Size of the Support:* For every $\mathsf{A} \geq 0$,

$$|\mathsf{supp}(P_{X^\star})| \leq \lceil \mathsf{A} - \log\left(P_{X^\star}(\mathsf{A})\right) - C(\mathsf{A})\rceil + 2. \tag{19}$$

In addition, for all A such that $\mathrm{e}^{\frac{C(\mathsf{A})}{1-\mathrm{e}^{-\mathsf{A}}}} \geq 4$, we have that

$$|\mathsf{supp}(P_{X^\star})| \leq 2\mathrm{e}\,\mathsf{A}\log^2(\mathsf{A}) + 2\log(\mathsf{A}) - \mathsf{A}$$
$$- \log\left(\frac{1 - 3\,\mathrm{e}^{-\frac{C(\mathsf{A})}{1-\mathrm{e}^{-\mathsf{A}}}}}{2}\right) - C(\mathsf{A}) + 4. \tag{20}$$

The proof of Theorem 2 is given in Section III. A few comments are now in order.

A. Numerical Simulations

In order to aid our discussion, we have also numerically computed the optimal input distributions for values of A up to 15. Fig. 2 depicts the output of this simulation.

We note that there are several numerical recipes for generating an optimal input distribution [34]–[36]. However, most of these approaches ultimately optimize over the space of distributions, which is an infinite-dimensional space. As was

already alluded to in [2] and [28], a firm upper bound on the number of mass points, such as the one in Theorem 2, allows us to move the optimization from the space of probability distributions to the space $\mathbb{R}^{2n}$ where n is the number of points. Working in $\mathbb{R}^{2n}$ allows us to employ methods such as the projected gradient ascent [37], which was used to generate the plots in Fig. 2. A quick sketch of how to implement the projected gradient ascent is given in Appendix A. Finally, we would like to point out that the projected gradient ascent becomes too slow for large values of A. Therefore, other methods such as the combination of gradient ascent and Blahut-Arimoto algorithm are necessary; the interested reader is referred to [38] for such an implementation and more detailed numerical analysis of the structure of $P_{X^\star}$.

B. On the Order of the Bounds

First, note that almost all of our bounds depend on the value of $C(\mathsf{A})$, which is currently unknown for $\mathsf{A} \geq \bar{\mathsf{A}} \approx 3.4$. However, this is not a limitation of our result, as we do have access to upper and lower bounds on $C(\mathsf{A})$ that are tight for large A such as those in [22], which suggest that

$$C(\mathsf{A}) = \frac{1}{2}\log(\mathsf{A}) - \frac{1}{2}\log\left(\frac{\pi \mathrm{e}}{2}\right) + o_{\mathsf{A}}(1). \quad (21)$$

Moreover, some of the bounds in Theorem 2, such as those in (16) and (20), while are firm, are meant to be used for large values of A. Therefore, combing the bounds in Theorem 2 with the bound in (21), we arrive at the following:

$$P_{X^\star}(x) \leq O\left(\frac{1}{\sqrt{\mathsf{A}}}\right), \ x \in \mathsf{supp}(P_{X^\star}), \quad (22)$$

$$\Omega(\sqrt{\mathsf{A}}) \leq |\mathsf{supp}(P_{X^\star})| \leq O\left(\mathsf{A}\log^2(\mathsf{A})\right). \quad (23)$$

Thus, the order of the lower bound on the number of points is $\sqrt{\mathsf{A}}$, and the order of the upper bound on the number of points is $\mathsf{A}\log^2(\mathsf{A})$. It is interesting to speculate as to the reason why the bounds do not match and have different orders.

First, note that Theorem 2 presents two upper bounds on the number of points. The first and implicit bound in (19) depends on the value of $P_{X^\star}(\mathsf{A})$. The second bound in (20) is an explicit bound in terms of A and is derived by plugging in the lower bound on $P_{X^\star}(\mathsf{A})$ in (16) into the first bound in (19). We suspect that one of the reason why the bounds do not match is due to the lower bound on $P_{X^\star}(\mathsf{A})$ in (16), which we think is not tight. Hence, one interesting future direction is to improve the lower bound on the value of $P_{X^\star}(\mathsf{A})$ which, in view of (19), would lead to a better upper bound on $|\mathsf{supp}(P_{X^\star})|$. However, to the best of our knowledge, there are no other methods for finding lower or upper bounds on the probabilities of the optimal input distribution. Indeed, one of the contributions of this work is the introduction of two such methods, one for finding an upper bound on the values of probabilities and one for finding a lower bound.

C. On the Upper Bounds on the Probabilities in (14) and (15)

It is also interesting to ask how tight the upper bounds on the probabilities in (14) and (15) are.

Note that the bound in (14) is universal and does not depend on the positions of the probability masses, while the bound in (15) depends on the position of the points. The advantage of the bound in (15) is that it can be tighter than the universal bound in (14). For example, in the regime where $\mathsf{A} \leq \bar{\mathsf{A}}$ where we only have two points in the support, the bound is achieved with equality. The clear disadvantage of the bound in (15) is that we do not know the location of the points (except 0 and A). However, such a bound might become useful once better estimates for the locations of the mass points are found. Some preliminary estimates of the locations are provided in (17).

Fig. 2c plots the upper bounds in (14) and (15) and compares them to the values of $P_{X^\star}$. To create the plot, in the regime $\mathsf{A} \leq \bar{\mathsf{A}}$, we have used the exact expressions for $P_{X^\star}(\mathsf{A})$ and $P_{X^\star}(0)$ in (4). To create the plot in the regime $\mathsf{A} > \bar{\mathsf{A}}$, first note that the upper bounds in (14) can be loosened to

$$P_{X^\star}(x) \leq \mathrm{e}^{-\frac{I(\widetilde{X};\widetilde{Y})}{1-\mathrm{e}^{-\mathsf{A}}}}, \quad x \in \mathsf{supp}(P_{X^\star}), \quad (24)$$

$$P_{X^\star}(x) \leq \mathrm{e}^{-I(\widetilde{X};\widetilde{Y})-x\frac{\mathrm{e}^{-x}}{1-\mathrm{e}^{-x}}}, \quad x \in \mathsf{supp}(P_{X^\star}) \setminus \{0\}, \quad (25)$$

where we have used the fact that $C(\mathsf{A}) \geq I(\widetilde{X};\widetilde{Y})$ for any random variable $\widetilde{X} \in [0, \mathsf{A}]$ and where $\widetilde{Y}$ is induced by $\widetilde{X}$. Therefore, since we can choose any $\widetilde{X} \in [0, \mathsf{A}]$, we selected it to be the one that is the output of the numerical simulation. The bound in (25) is only computed for $x = \mathsf{A}$ as we do not know the locations of other points and only have estimates for these. From the simulations in Fig. 2c, the bounds in (14) and (15) appear to be relatively tight.

The bound in (14) relies on the strong data-processing inequality. Specifically, the factor $\frac{1}{1-\mathrm{e}^{-\mathsf{A}}}$ in the exponent comes from using the strong data-processing inequality. The dotted black curve in Fig. 2c plots the loosened version of the bound in (14) that ignores the contribution of the strong data-processing inequality, that is we plot

$$P_{X^\star}(x) \leq \mathrm{e}^{-I(\widetilde{X};\widetilde{Y})}, \quad x \in \mathsf{supp}(P_{X^\star}). \quad (26)$$

From the comparison Fig. 2c, we see that the contribution of the strong data-processing inequality is nontrivial, especially for small and medium values of A.

D. On the Bound in (17)

In addition to finding bounds on the number of points and the values of the probabilities, we have also provided additional information about the location of the points. Specifically, (17) provides information about the location of support points other than 0 and A.

From the bound in (17), we see that the second-largest point can never be too close to A. Specifically, according to the bound in (17), the gap between A (i.e., the largest point) and the second-largest point is at least one. In fact, the numerical simulations shown in Fig. 2a suggest that this gap is much larger. In particular, the simulations suggest that the gap is not constant but is an increasing function of A. Therefore, one interesting future direction would be to verify this behavior and produce a better bound in (17) than $\mathsf{A} - 1$.

Similarly, from the lower bound in (17), we see that the second smallest point cannot be too close to the zero
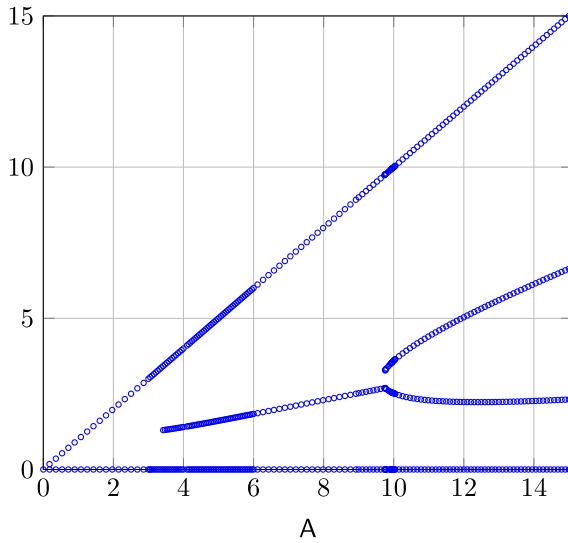
(a) Plot of the locations of the support points of $P_{X^\star}$ vs. A.
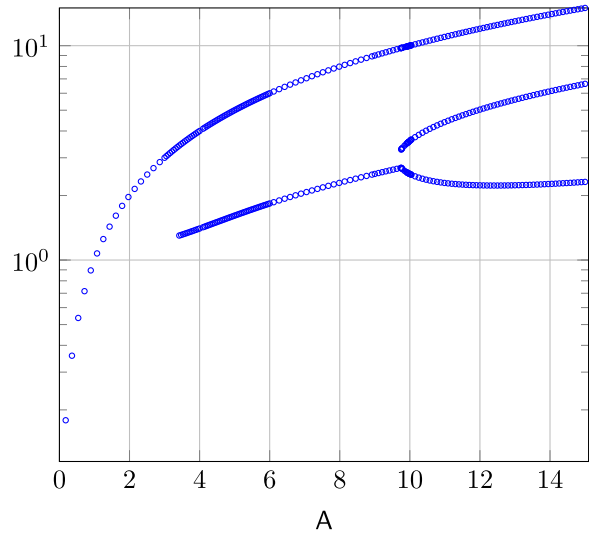
(b) Plot of the locations of the support points of $P_{X^\star}$ vs. A where the log scale is used for the vertical axis.

(c) Plot of the probability values of the support points vs. A where points of the support satisfy $0 \leq x_1 \leq x_2 \leq$ A. The curves show: $P_{X^\star}(0)$ (solid blue line); $P_{X^\star}(A)$ (solid red line); $P_{X^\star}(x_1)$ (solid cyan line); $P_{X^\star}(x_2)$ (solid green line); upper bound in (14) (dashed black line); upper bound on $P_{X^\star}(A)$ in (15) (dashed red line); and upper bound in (26) with no strong data-processing term (dotted black line).

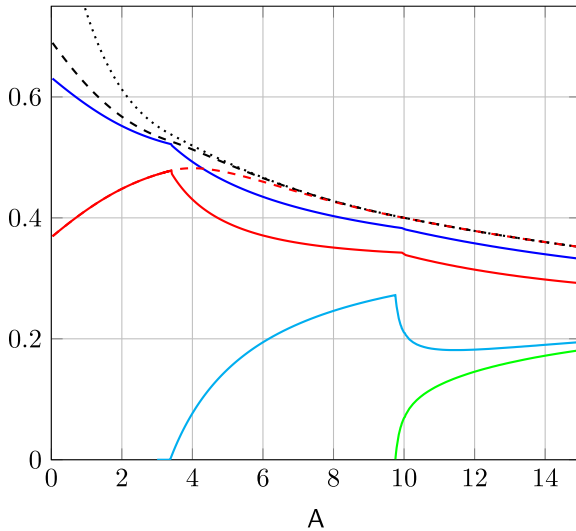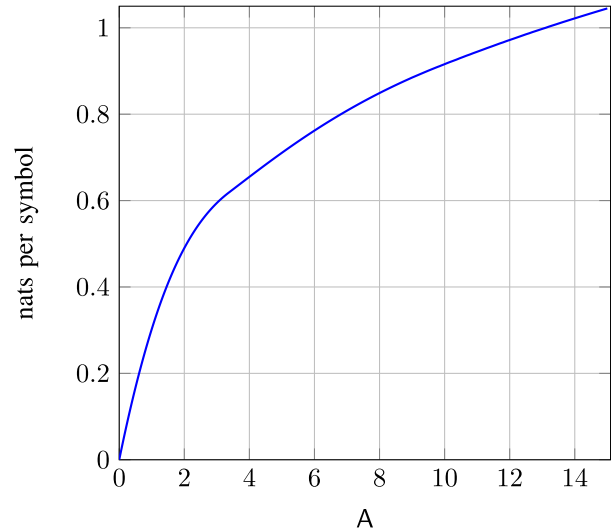(d) Plot of $C(A)$ vs. A.

Fig. 2. Examples of the optimal input distributions and capacity vs. A. The probabilities and the capacity for $A > \bar{A} \approx 3.4$ were computed numerically, and the probabilities and the capacity for $A \leq \bar{A}$ are given in (4) and (5), respectively.

point. However, as A increases, the distance is allowed to get smaller. Note that our limited simulation results suggest a better lower bound, namely $x^\star \geq 1$. Therefore, one interesting future direction would be to either demonstrate the existence of a mass point in the range $(0, 1)$ or show that there is no such mass point. Note that the work of [9] already showed that there is at most one point in the range $(0, 1)$.

Beyond theoretical interest, the existence of the estimates for the mass points' location might also be of interest from the practical point of view. As the existence of such estimates

can also impact the design of practical constellations for the Poisson noise channel.

E. On the Equivocation, Symbol Error Probability, and Entropy

It is well-known that the capacity-achieving distribution should be 'difficult' to detect or estimate on the per-symbol basis. To make this statement explicit, we consider the equivocation $H(X^\star|Y^\star)$ and the probability of error under the

maximum a posteriori (MAP) rule (*i.e.*, $P_e = \mathbb{P}[X^\star \neq \hat{X}(Y^\star)]$ where $\hat{X}(Y^\star)$ is the MAP decoder). The plots in Fig. 3a and Fig. 3b show that the equivocation and P_e for the capacity-achieving input have relatively high values. With Theorem 2 at our disposal, we can now show the following result regarding the asymptotic behavior of the error probability.

Proposition 1: Let $P_e = \mathbb{P}[X^\star \neq \hat{X}(Y^\star)]$ where $\hat{X}(y) = \arg\max_{x \in \mathsf{supp}(P_{X^\star})} P_{X^\star|Y^\star}(x|y)$. Then,

$$\liminf_{\mathsf{A} \to \infty} P_e \geq 1 - \sqrt{\frac{2}{\pi}}, \tag{27}$$

$$\liminf_{\mathsf{A} \to \infty} H(X^\star|Y^\star) \geq 2\left(1 - \sqrt{\frac{2}{\pi}}\right). \tag{28}$$

Proof: The probability of error for the MAP rule can be written as

$$P_e = 1 - \mathbb{E}\left[\max_{x \in \mathsf{supp}(P_{X^\star})} P_{X^\star|Y^\star}(x|Y^\star)\right] \tag{29}$$

$$= 1 - \mathbb{E}\left[\max_{x \in \mathsf{supp}(P_{X^\star})} \frac{P_{Y|X}(Y^\star|x)P_{X^\star}(x)}{P_{Y^\star}(Y^\star)}\right] \tag{30}$$

$$\geq 1 - e^{-\frac{C(\mathsf{A})}{1-e^{-\mathsf{A}}}} \mathbb{E}\left[\max_{x \in \mathsf{supp}(P_{X^\star})} \frac{P_{Y|X}(Y^\star|x)}{P_{Y^\star}(Y^\star)}\right] \tag{31}$$

$$= 1 - e^{-\frac{C(\mathsf{A})}{1-e^{-\mathsf{A}}}} \sum_{y=0}^{\infty} \max_{x \in \mathsf{supp}(P_{X^\star})} P_{Y|X}(y|x) \tag{32}$$

$$= 1 - e^{-\frac{C(\mathsf{A})}{1-e^{-\mathsf{A}}}} \left(\sum_{y<\mathsf{A}} \max_{x \in \mathsf{supp}(P_{X^\star})} P_{Y|X}(y|x)\right.$$

$$\left. + \sum_{y \geq \mathsf{A}} \max_{x \in \mathsf{supp}(P_{X^\star})} P_{Y|X}(y|x)\right) \tag{33}$$

$$\geq 1 - e^{-\frac{C(\mathsf{A})}{1-e^{-\mathsf{A}}}} \left(\sum_{y<\mathsf{A}} P_{Y|X}(y|y) + \sum_{y \geq \mathsf{A}} P_{Y|X}(y|\mathsf{A})\right) \tag{34}$$

$$= 1 - e^{-\frac{C(\mathsf{A})}{1-e^{-\mathsf{A}}}} \left(1 + \sum_{1 \leq y < \mathsf{A}} \frac{y^y e^{-y}}{y!} + \mathbb{P}[Y \geq \mathsf{A}|X = \mathsf{A}]\right) \tag{35}$$

$$\geq 1 - e^{-\frac{C(\mathsf{A})}{1-e^{-\mathsf{A}}}} \left(1 + \sum_{1 \leq y < \mathsf{A}} \frac{1}{\sqrt{2\pi y}} + \mathbb{P}[Y \geq \mathsf{A}|X = \mathsf{A}]\right) \tag{36}$$

$$\geq 1 - e^{-\frac{C(\mathsf{A})}{1-e^{-\mathsf{A}}}} \left(1 + \sqrt{\frac{2\mathsf{A}}{\pi}} + \frac{1}{\sqrt{2\pi\mathsf{A}}}\right.$$

$$\left. + c + \mathbb{P}[Y \geq \mathsf{A}|X = \mathsf{A}]\right), \tag{37}$$

where the expression for the probability of error in (29) comes from [39]; (31) follows by using the bound in (14); (34) follows by using that $\max_{x \in [0,\mathsf{A}]} P_{Y|X}(y|x) \leq P_{Y|X}(y|y)$ for $y < \mathsf{A}$ and $\max_{x \in [0,\mathsf{A}]} P_{Y|X}(y|x) \leq P_{Y|X}(y|\mathsf{A})$ for $y \geq \mathsf{A}$; (36) follows by using Stirling's bound $\frac{y^y e^{-y}}{y!} \leq \frac{1}{\sqrt{2\pi y}}$, $y \geq 1$; and (37) follows by using the bound $\sum_{1 \leq y < \mathsf{A}} \frac{1}{\sqrt{y}} \leq 2\sqrt{y} + y^{-\frac{1}{2}} + c$ where $c > 0$ is some universal constant [40].

To conclude the proof of (27), we use the bound in (37) together with the asymptotic expression for $C(\mathsf{A})$ in (21). The proof of (28) follows by using the following bound [41]:

$$H(X|Y) \geq 2P_e. \tag{38}$$

■

Remark 1: The bound in (28) can be improved, albeit at the expense of a more complicated expression. In particular, the bound relies on the inequality $H(X|Y) \geq 2P_e$. There are several stronger versions of this inequality; the interested reader is referred to [39] for a summary of these inequalities.

The entropy of the optimal input distribution vs. A is plotted in Fig. 3c. We observe a particular behavior of the entropy in the simulated range of A: the rate of increase has finite jumps approximately at the levels $\log(k)$ of entropy, where the cardinality of the optimal input distribution increases from k to $k + 1$ points. These levels correspond to an approximate uniform input distribution on the k amplitude levels: this behavior is also confirmed by the mass probabilities plotted in Fig. 2c. When the rate of increase of the entropy is not compensated by a sufficiently large rate of decrease of the equivocation (see Fig. 3a), the rate of increase of capacity must be sustained by boosting the input entropy: This is done by increasing the cardinality of the input distribution. It is interesting to understand how the rate of increase of capacity is split between entropy and equivocation: If one could prove that the equivocation is upper-bounded by a constant, then this would show that the equivocation does not provide degrees of freedom to channel capacity for large A, and thus the whole rate should be sustained by the input entropy by increasing the cardinality of $\mathsf{supp}(P_{X^\star})$. This hypothesis would imply $|\mathsf{supp}(P_{X^\star})| \approx \sqrt{\mathsf{A}}$ for large A.

III. PROOF OF THE MAIN RESULT

The starting point for most of our proofs are the following KKT conditions shown in [11].

Lemma 1: The capacity-achieving distribution $P_{X^\star}$ and induced capacity-achieving output distribution $P_Y^\star$ satisfy the following:

$$i(x; P_{X^\star}) \leq C(\mathsf{A}), \quad x \in [0, \mathsf{A}] \tag{39a}$$

$$i(x; P_{X^\star}) = C(\mathsf{A}), \quad x \in \mathsf{supp}(P_{X^\star}), \tag{39b}$$

where

$$i(x; P_{X^\star}) = \mathsf{D}(P_{Y|X}(\cdot|x)\|P_{Y^\star}). \tag{40}$$

Lemma 1 asserts that the points of the support of $P_{X^\star}$ are zeros of the function $x \mapsto i(x; P_{X^\star}) - C(\mathsf{A})$. Therefore, we have the following inclusion:

$$\mathsf{supp}(P_{X^\star}) \subseteq \{x \in [0, \mathsf{A}] : i(x; P_{X^\star}) - C(\mathsf{A}) = 0\}. \tag{41}$$

This important observation will be a key step in the proof of the upper bound on the cardinality of $\mathsf{supp}(P_{X^\star})$.
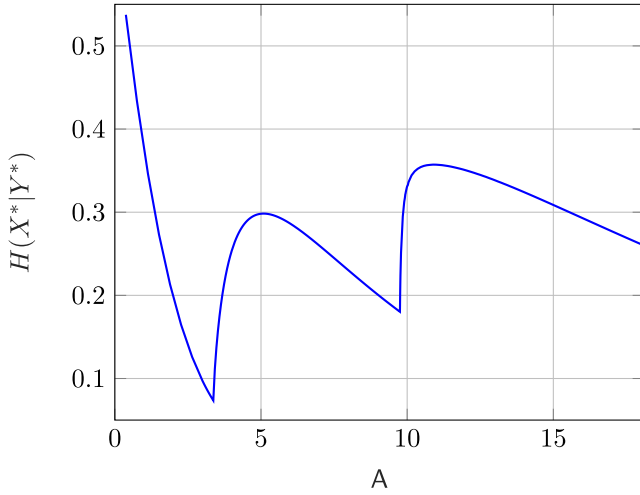
(a) Equivocation $H(X^\star|Y^\star)$ vs. A.
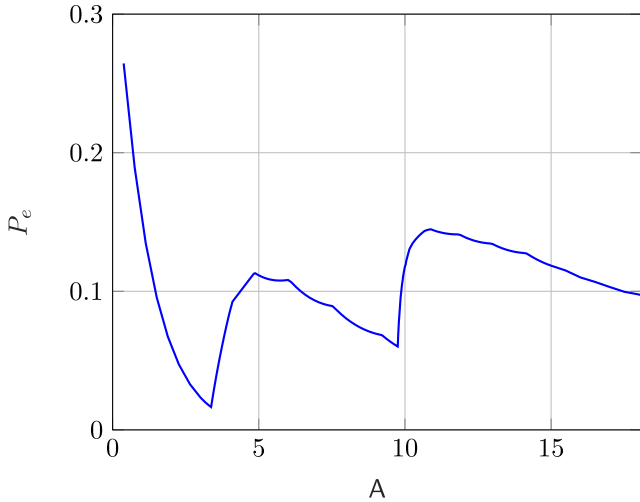
(b) Plot of P_e vs. A.
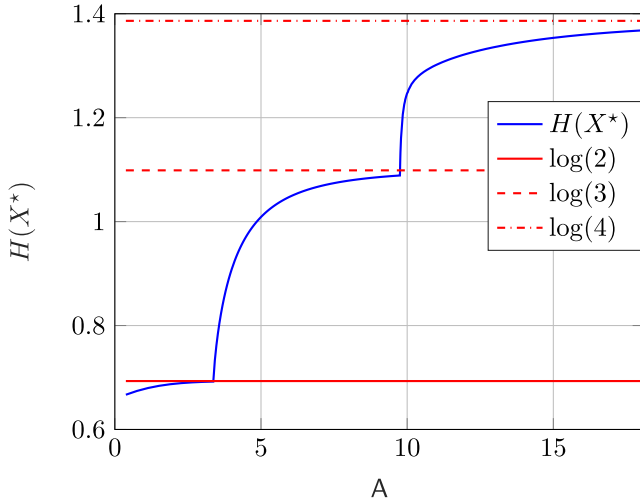
(c) Entropy vs. A.

Fig. 3. Equivocation (i.e., $H(X^\star|Y^\star)$), the probability of error detection under MAP rule of the optimal input distribution, and entropy (i.e., $H(X^\star)$).

A. Proof of the New Capacity Expression in (13)

As was shown in [13], $0 \in \mathsf{supp}(P_{X^\star})$. Therefore, by using (39b), we have that

$$C(\mathsf{A}) = \mathsf{D}(P_{Y|X}(\cdot|0)\|P_{Y^\star}) \tag{42}$$

$$= \log \frac{1}{P_{Y^\star}(0)}. \tag{43}$$

This concludes the proof. Despite the simplicity of the above proof, to the best of our knowledge, the above expression has not been observed in the past.

B. Proof of the Upper Bound in (14) and of the Lower Bound in (18)

In this section, we develop upper bounds on the probability masses and a lower bound on the cardinality of the support. Specifically, we will show that the probabilities are bounded by a term of the form $\mathrm{e}^{-C(\mathsf{A})}$, and the cardinality of the support points are lower bounded by a term of the form $\mathrm{e}^{C(\mathsf{A})}$.

We first show an upper bound on the number of points. We show two methods for finding bounds on the probabilities. The first method relies on the strong data-processing inequality and the second method relies on the exact expression for the values of the probability distribution. An interesting feature of both methods is that they work for all channels for which a capacity-achieving distribution is discrete. In this section, with some abuse of notation, $P_{X^\star}$ and $P_{Y^\star}$ will denote the capacity-achieving input and output distribution, respectively, not only for the Poisson noise channel but for a generic channel $P_{Y|X}$.

Theorem 3: Fix some channel $P_{Y|X}$ and consider the optimization problem

$$C = \max_{X \in \mathcal{X}} I(X;Y). \tag{44}$$

Moreover, suppose that a maximizing distribution $P_{X^\star}$ is discrete. Then,

$$P_{X^\star}(x) \le \exp\left(-\frac{C}{\eta_{\mathsf{KL}}\left(\mathcal{X}; P_{Y|X}\right)}\right), \quad x \in \mathsf{supp}(P_{X^\star}) \tag{45}$$

where $0 < \eta_{\mathsf{KL}}\left(\mathcal{X}; P_{Y|X}\right) \le 1$ is the contraction coefficient defined in (10).

Proof: Let Y_x be the output of the channel $P_{Y|X}$ when the input is $P_X = \delta_x$, where δ_x is the Dirac delta function centered in x. Next, suppose that $x \in \mathsf{supp}(P_{X^\star})$. Then, by using (39), we have that

$$C = \mathsf{D}(P_{Y|X}(\cdot|x)\|P_{Y^\star}) \tag{46}$$

$$= \mathsf{D}(P_{Y_x}\|P_{Y^\star}) \tag{47}$$

$$\le \eta_{\mathsf{KL}}\left(\mathcal{X}; P_{Y|X}\right) \mathsf{D}(\delta_x\|P_{X^\star}) \tag{48}$$

$$= \eta_{\mathsf{KL}}\left(\mathcal{X}; P_{Y|X}\right) \log \frac{1}{P_{X^\star}(x)}, \tag{49}$$

where in (48) we have noted that $\delta_x \to P_{Y|X} \to P_{Y_x}$ and $P_{X^\star} \to P_{Y|X} \to P_{Y^\star}$ and used the strong data-processing inequality in (9). This concludes the proof. ∎

The next result provides an upper bound on the contraction coefficient for the Poisson channel.

Lemma 2: Let $P_{Y|X}$ be a Poisson channel as in (1). Then, for all $\mathsf{A} \ge 0$

$$\eta_{\mathsf{KL}}([0, \mathsf{A}]; P_{Y|X}) \le 1 - \mathrm{e}^{-\mathsf{A}}. \tag{50}$$

Proof: As was shown in [42, Proposition II.4.10], the contraction coefficient is upper bounded by

$$
\begin{aligned}
&\eta_{\mathsf{KL}}([0,\mathsf{A}];P_{Y|X}) \\
&\leq \sup_{x,x':\, 0\leq x\leq \mathsf{A},\, 0\leq x'\leq \mathsf{A}} \mathsf{TV}(P_{Y|X=x}\|P_{Y|X=x'}) \qquad (51) \\
&\leq \sup_{x,x':\, 0\leq x\leq \mathsf{A},\, 0\leq x'\leq \mathsf{A}} (1 - \mathrm{e}^{-|x-x'|}) \qquad (52) \\
&= 1 - \mathrm{e}^{-\mathsf{A}}, \qquad (53)
\end{aligned}
$$

where in (51) $\mathsf{TV}(P_{Y|X=x}\|P_{Y|X=x'})$ is the total variation distance; and (52) follows from the bound in [43, Corollary 3.1]. ∎

Combining (45) and (50) leads to the bound in (14). Next, we show the proof of the second bound in (15). The following result holds for all channels that have a discrete capacity-achieving distribution.

Theorem 4: Fix some channel $P_{Y|X}$ and consider the optimization problem

$$
C = \max_{X\in\mathcal{X}} I(X;Y). \qquad (54)
$$

Moreover, suppose that a maximizing distribution $P_{X^\star}$ is discrete. Then, for $x \in \mathsf{supp}(P_{X^\star})$

$$
P_{X^\star}(x) = \mathrm{e}^{-C}\mathrm{e}^{\mathbb{E}\left[\log\left(P_{X^\star|Y^\star}(x|Y)\right)|X=x\right]}. \qquad (55)
$$

Proof: By using (39), for $x \in \mathsf{supp}(P_{X^\star})$ we have that

$$
\begin{aligned}
C &= \mathsf{D}(P_{Y|X}(\cdot|x)\|P_{Y^\star}) \qquad (56) \\
&= \mathbb{E}\left[\log\left(\frac{P_{Y|X}(Y|x)}{P_{Y^\star}(Y)}\right)\Big| X=x\right] \qquad (57) \\
&= \log\frac{1}{P_{X^\star}(x)} + \mathbb{E}\left[\log\left(P_{X^\star|Y^\star}(x|Y)\right)\big| X=x\right]. \qquad (58)
\end{aligned}
$$

This concludes the proof. ∎

Remark 2: Upon using the bound $P_{X^\star|Y^\star}(x|y) \leq 1$ in (55), we arrive at

$$
P_{X^\star}(x) \leq \mathrm{e}^{-C}, \quad x \in \mathsf{supp}(P_{X^\star}). \qquad (59)
$$

This is clearly a weaker version of the bound in (45). However, as shown next, we can improve this by using a better bound on the term $\mathbb{E}\left[\log\left(P_{X^\star|Y^\star}(x|Y)\right)|X=x\right]$.

Lemma 3: Let $P_{Y|X}$ be a Poisson channel as in (1). Then, for all $\mathsf{A} \geq 0$ and $x \in \mathsf{supp}(P_{X^\star}) \setminus \{0\}$

$$
P_{X^\star}(x) \leq \mathrm{e}^{-C(\mathsf{A})-x\frac{\mathrm{e}^{-x}}{1-\mathrm{e}^{-x}}}. \qquad (60)
$$

In addition, the bound in (60) becomes equality if $|\mathsf{supp}(P_{X^\star})| = 2$.

Proof: An upper bound to the expectation in (55) is as follows: for $x \in \mathsf{supp}(P_{X^\star})$

$$
\begin{aligned}
&\mathbb{E}\left[\log\left(P_{X^\star|Y^\star}(x|Y)\right)\big| X=x\right] \\
&= \sum_{k=0}^{\infty} P_{Y|X}(k|x)\log\frac{P_{Y|X}(k|x)P_{X^\star}(x)}{P_{Y^\star}(k)} \qquad (61) \\
&= P_{Y|X}(0|x)\log\frac{P_{Y|X}(0|x)P_{X^\star}(x)}{P_{Y^\star}(0)} \\
&\quad + \sum_{k=1}^{\infty} P_{Y|X}(k|x)\log P_{X^\star|Y^\star}(x|k) \qquad (62) \\
&= \mathrm{e}^{-x}\log\frac{\mathrm{e}^{-x}P_{X^\star}(x)}{\mathrm{e}^{-C(\mathsf{A})}} + \sum_{k=1}^{\infty} P_{Y|X}(k|x)\log P_{X^\star|Y^\star}(x|k) \\
&\qquad (63) \\
&\leq \mathrm{e}^{-x}\log\frac{\mathrm{e}^{-x}P_{X^\star}(x)}{\mathrm{e}^{-C(\mathsf{A})}} \qquad (64) \\
&= \mathrm{e}^{-x}\left(C(\mathsf{A}) - x + \log P_{X^\star}(x)\right), \qquad (65)
\end{aligned}
$$

where in (63) we have used the expression for the capacity in (13); and in (64) we have used the bound $P_{X^\star|Y^\star}(x|k) \leq 1$.

Next, for $x \in \mathsf{supp}(P_{X^\star}) \setminus \{0\}$, combining (55) and (65), we arrive at

$$
P_{X^\star}(x) \leq \mathrm{e}^{-C(\mathsf{A})+\mathrm{e}^{-x}(C(\mathsf{A})-x+\log P_{X^\star}(x))}. \qquad (66)
$$

Now, solving (66) for $P_{X^\star}(x)$ yields (60). Finally, if $P_{X^\star}$ is binary, then by using Bayes' rule and the fact that $P_{Y|X}(0|0) = 1$, it is not difficult to check that for $k \geq 1$

$$
P_{X^\star|Y^\star}(\mathsf{A}|k) = 1, \qquad (67)
$$

and, therefore, the bound in (64) is tight. This concludes the proof. ∎

This concludes the proof of the upper bounds on the values of the probabilities. The lower bound on the number of points in (20) is now a consequence of the upper bound on the values of $P_{X^\star}$:

$$
1 = \sum_{x\in\mathsf{supp}(P_{X^\star})} P_{X^\star}(x) \leq |\mathsf{supp}(P_{X^\star})|\mathrm{e}^{-\frac{1}{1-\mathrm{e}^{-\mathsf{A}}}C(\mathsf{A})}, \quad (68)
$$

which simplifies to

$$
\mathrm{e}^{\frac{1}{1-\mathrm{e}^{-\mathsf{A}}}C(\mathsf{A})} \leq |\mathsf{supp}(P_{X^\star})|. \qquad (69)
$$

Remark 3: An alternative way of finding a lower bound on $|\mathsf{supp}(P_{X^\star})|$ is by using the following sequence of elementary inequalities:

$$
\begin{aligned}
C(\mathsf{A}) &= I(X^\star;Y^\star) \qquad (70) \\
&\leq H(X^\star) \qquad (71) \\
&\leq \log\left(|\mathsf{supp}(P_{X^\star})|\right). \qquad (72)
\end{aligned}
$$

However, this upper bound is weaker than that in (69).

C. Proof of the Lower Bound on $P_{X^\star}(\mathsf{A})$ in (16) and of the Bounds in (17)

In the remaining of the paper, $P_{X^\star}$ shall again denote the optimal input distribution for the Poisson noise channel.

In this section, we establish bounds on $P_{X^\star}(\mathsf{A})$ and bounds the location of the support points. Let

$$i'(x; P_{X^\star}) = \frac{\mathrm{d}}{\mathrm{d}x}i(x; P_{X^\star}), \qquad (73)$$

$$i''(x; P_{X^\star}) = \frac{\mathrm{d}^2}{\mathrm{d}x^2}i(x; P_{X^\star}). \qquad (74)$$

The starting place for both bounds is the fact that if $x^\star \in \mathsf{supp}(P_{X^\star})$ is a point that is neither equal to 0 nor A, then by using the KKT conditions in (39), we have that

$$i'(x^\star; P_{X^\star}) = 0, \qquad (75)$$

$$i''(x^\star; P_{X^\star}) \leq 0. \qquad (76)$$

Then, by letting $\overline{Y} \sim \mathcal{P}(x^\star)$, we have that

$$0 \geq i''(x^\star; P_{X^\star}) \qquad (77)$$

$$= -\mathbb{E}\left[\log\left(\mathbb{E}[X^\star|Y^\star = \overline{Y}+1]\right)\right] - i'(x^\star; P_{X^\star})$$

$$+ \log(x^\star) + \frac{1}{x^\star} \qquad (78)$$

$$= -\mathbb{E}\left[\log\left(\mathbb{E}[X^\star|Y^\star = \overline{Y}+1]\right)\right] + \log(x^\star) + \frac{1}{x^\star}, \quad (79)$$

where (78) follows by using the identity for the second derivative of $i(x^\star; P_{X^\star})$ in Lemma 13 in Appendix B; and (79) follows by using that $i'(x^\star; P_{X^\star}) = 0$.

The inequality in (79) will be the key to both proofs. We start by showing the bound in (17). We also use a result of [9] to establish a lower bound on the location of the second largest point.

Lemma 4: Suppose that $|\mathsf{supp}(P_{X^\star})| \geq 3$. Then, for $x^\star \in \mathsf{supp}(P_{X^\star}) \setminus \{0, \mathsf{A}\}$

$$\mathrm{e}^{-\sqrt{2(\log(\mathsf{A})-1)}} \leq \mathsf{A}\,\mathrm{e}^{W_{-1}\left(-\frac{1}{\mathsf{A}}\right)} \leq x^\star \leq \mathsf{A}\,\mathrm{e}^{W_0\left(-\frac{1}{\mathsf{A}}\right)} \leq \mathsf{A}-1. \qquad (80)$$

In addition, let $x_0 = \max\{\mathsf{supp}(P_{X^\star}) \setminus \{\mathsf{A}\}\}$ (*i.e.*, the second largest point in the support). Then, for $|\mathsf{supp}(P_{X^\star})| \geq 4$,

$$x_0 \geq 1. \qquad (81)$$

Proof: By using (79), we have that

$$0 \geq \log\left(\frac{1}{\mathsf{A}}\right) + \log(x^\star) + \frac{1}{x^\star} \qquad (82)$$

$$= \log\left(\frac{x^\star}{\mathsf{A}}\right) + \frac{1}{x^\star}, \qquad (83)$$

where the inequality follows by using the bound $\mathbb{E}[X^\star|Y^\star] \leq \mathsf{A}$. We now show that (83) implies (80).

The function $f(x) = \log\left(\frac{x}{\mathsf{A}}\right) + \frac{1}{x}$ is decreasing for $x < 1$ and increasing for $x > 1$, and has two zeros for $\mathsf{A} > \mathrm{e}$. Note that under the assumption $|\mathsf{supp}(P_{X^\star})| \geq 3$, we have that $\mathsf{A} > \bar{\mathsf{A}} \approx 3.4$. Therefore, we operate in the regime where we have two zeros.

The exact solution to (83) is given in terms of branches of the Lambert W-function:

$$\mathsf{A}\mathrm{e}^{W_{-1}\left(-\frac{1}{\mathsf{A}}\right)} \leq x^\star \leq \mathsf{A}\mathrm{e}^{W_0\left(-\frac{1}{\mathsf{A}}\right)}. \qquad (84)$$

To make the above bounds more useful, we further loosen them to involve simpler functions.

To find the upper bound on the largest zero, we use the bound $\log(t) \geq 1 - \frac{1}{t}$, which leads to

$$0 \geq \log\left(\frac{x^\star}{\mathsf{A}}\right) + \frac{1}{x^\star} \geq 1 - \frac{\mathsf{A}}{x^\star} + \frac{1}{x^\star}$$

$$\implies x^\star \leq \mathsf{A} - 1. \qquad (85)$$

We now find a lower bound on the smallest zero. By substituting $x = \mathrm{e}^t$, we can equivalently study the smallest zero of $f(t) = t - \log(\mathsf{A}) + \mathrm{e}^{-t}$, which is the solution $t_{\min} < 0$. Since $f(t)$ is decreasing for $t < 0$, a lower bound to $f(t)$ will provide a lower bound to $t_{\min}$. By using $\mathrm{e}^{-t} \geq 1 - t + \frac{t^2}{2}$, for $t < 0$, we get to study the smallest zero of

$$\underline{f}(t) = t - \log(\mathsf{A}) + 1 - t + \frac{t^2}{2} \qquad (86)$$

$$= -\log(\mathsf{A}) + 1 + \frac{t^2}{2}, \qquad (87)$$

which yields

$$t_{\min} \geq -\sqrt{2(\log(\mathsf{A}) - 1)}, \qquad (88)$$

and finally

$$x^\star \geq \mathrm{e}^{-\sqrt{2(\log(\mathsf{A})-1)}}, \qquad \mathsf{A} > \mathrm{e}. \qquad (89)$$

Finally, under the additional assumption that $|\mathsf{supp}(P_{X^\star})| \geq 4$ and using the fact that there is at most one point on the open interval $(0, 1)$ [9], it follows that the second largest point must satisfy $x_0 \geq 1$. This concludes the proof. ∎

Remark 4: We note that from the statement of Lemma 4, the lower bound in (80) is only useful to bound the location of the second smallest point since by (81) the other larger points are lower-bounded by one.

We now show the lower bound on $P_{X^\star}(\mathsf{A})$. For ease of presentation and to emphasize key steps, the proof is split among three lemmas.

Lemma 5: Suppose that $|\mathsf{supp}(P_{X^\star})| \geq 3$. Let $x_0 = \max\{\mathsf{supp}(P_{X^\star}) \setminus \{\mathsf{A}\}\}$ and $\overline{Y} \sim \mathcal{P}(x_0)$. Then, for every $c \geq 0$ such that

$$\mathbb{P}\left[\overline{Y} > c\right] \leq \frac{1}{\mathsf{A}}, \qquad (90)$$

we have that

$$P_{X^\star}(\mathsf{A}) \geq \frac{-\frac{1}{\mathsf{A}}\log\left(\frac{\mathsf{A}}{x_0}\right) + \frac{1}{x_0}}{\left(\frac{\mathsf{A}}{x_0} - 1\right)\mathbb{E}\left[\frac{P_{Y|X}(\overline{Y}+1|\mathsf{A})}{P_{Y^\star}(\overline{Y}+1)}\mathbb{1}_{\overline{Y}<c}\right]} \geq 0. \qquad (91)$$

Proof: Let $p_{\mathsf{A}} = P_{X^\star}(\mathsf{A})$. Then, starting with (79), we have that

$$0 \geq -\mathbb{E}\left[\log\left(\mathbb{E}\left[\frac{X^\star}{x_0}\middle|Y^\star = \overline{Y}+1\right]\right)\right] + \frac{1}{x_0} \qquad (92)$$

$$= -\mathbb{E}\left[\left(\mathbb{1}_{\overline{Y}\geq c} + \mathbb{1}_{\overline{Y}<c}\right)\log\left(\mathbb{E}\left[\frac{X^\star}{x_0}\middle|Y^\star = \overline{Y}+1\right]\right)\right] + \frac{1}{x_0} \qquad (93)$$

$$\geq -\mathbb{E}\left[\mathbb{1}_{\overline{Y}<c}\log\left(\mathbb{E}\left[\frac{X^\star}{x_0}\middle|Y^\star = \overline{Y}+1\right]\right)\right]$$

$$- \mathbb{P}\left[\overline{Y} \geq c\right]\log\left(\frac{\mathsf{A}}{x_0}\right) + \frac{1}{x_0} \qquad (94)$$

$$\geq -\mathbb{E}\left[\mathbb{1}_{\overline{Y}<c}\log\left(\frac{\left(\frac{\mathsf{A}}{x_0}-1\right)p_\mathsf{A}P_{Y|X}(\overline{Y}+1|\mathsf{A})}{P_{Y^\star}(\overline{Y}+1)}+1\right)\right]$$
$$-\mathbb{P}\left[\overline{Y}\geq c\right]\log\left(\frac{\mathsf{A}}{x_0}\right)+\frac{1}{x_0} \qquad (95)$$

$$\geq -\mathbb{P}\left[\overline{Y}<c\right]$$
$$\cdot\log\left(\left(\frac{\mathsf{A}}{x_0}-1\right)p_\mathsf{A}\mathbb{E}\left[\frac{P_{Y|X}(\overline{Y}+1|\mathsf{A})}{P_{Y^\star}(\overline{Y}+1)}\middle|\overline{Y}<c\right]+1\right)$$
$$-\mathbb{P}\left[\overline{Y}\geq c\right]\log\left(\frac{\mathsf{A}}{x_0}\right)+\frac{1}{x_0}, \qquad (96)$$

where in (94) we have used the bound $\mathbb{E}\left[\frac{X^\star}{x_0}\middle|Y^\star\right]\leq\frac{\mathsf{A}}{x_0}$; and (95) follows by first using that

$$\mathbb{E}\left[\frac{X^\star}{x_0}\middle|Y^\star\right]=\mathbb{E}\left[\frac{X^\star}{x_0}\mathbb{1}_{X^\star\leq x_0}\middle|Y^\star\right]+\mathbb{E}\left[\frac{X^\star}{x_0}\mathbb{1}_{X^\star=\mathsf{A}}\middle|Y^\star\right] \qquad (97)$$

$$=\mathbb{E}\left[\frac{X^\star}{x_0}\mathbb{1}_{X^\star\leq x_0}\middle|Y^\star\right]+\frac{\mathsf{A}}{x_0}p_\mathsf{A}\frac{P_{Y|X}(Y^\star|\mathsf{A})}{P_{Y^\star}(Y^\star)}, \qquad (98)$$

and then bounding the term $\mathbb{E}\left[\frac{X^\star}{x_0}\mathbb{1}_{X^\star\leq x_0}\middle|Y^\star\right]$ as follows:

$$\mathbb{E}\left[\frac{X^\star}{x_0}\mathbb{1}_{X^\star\leq x_0}\middle|Y^\star\right]\leq\mathbb{E}\left[\mathbb{1}_{X^\star\leq x_0}|Y^\star\right] \qquad (99)$$

$$=\mathbb{P}\left[X^\star\leq x_0|Y^\star\right] \qquad (100)$$

$$=1-P_{X^\star|Y^\star}(\mathsf{A}|Y^\star) \qquad (101)$$

$$=1-p_\mathsf{A}\frac{P_{Y|X}(Y^\star|\mathsf{A})}{P_{Y^\star}(Y^\star)}; \qquad (102)$$

and (96) follows by using Jensen's inequality.

Solving (96) for p_A, we obtain

$$p_\mathsf{A}\geq\frac{\exp\left(\left(-\mathbb{P}\left[\overline{Y}\geq c\right]\log\left(\frac{\mathsf{A}}{x_0}\right)+\frac{1}{x_0}\right)\frac{1}{\mathbb{P}[\overline{Y}<c]}\right)-1}{\left(\frac{\mathsf{A}}{x_0}-1\right)\mathbb{E}\left[\frac{P_{Y|X}(\overline{Y}+1|\mathsf{A})}{P_{Y^\star}(\overline{Y}+1)}\middle|\overline{Y}<c\right]}. \qquad (103)$$

Now, the assumption $\mathbb{P}\left[\overline{Y}>c\right]\leq\frac{1}{\mathsf{A}}$ guarantees that the argument of the exponential in (103) is positive, which leads to

$$p_\mathsf{A}$$
$$\geq\frac{\mathbb{P}\left[\overline{Y}<c\right]\exp\left(\left(-\mathbb{P}\left[\overline{Y}\geq c\right]\log\left(\frac{\mathsf{A}}{x_0}\right)+\frac{1}{x_0}\right)\frac{1}{\mathbb{P}[\overline{Y}<c]}\right)-1}{\left(\frac{\mathsf{A}}{x_0}-1\right)\mathbb{E}\left[\frac{P_{Y|X}(\overline{Y}+1|\mathsf{A})}{P_{Y^\star}(\overline{Y}+1)}\mathbb{1}_{\overline{Y}<c}\right]}$$
$$\qquad (104)$$

$$\geq\frac{-\mathbb{P}\left[\overline{Y}\geq c\right]\log\left(\frac{\mathsf{A}}{x_0}\right)+\frac{1}{x_0}}{\left(\frac{\mathsf{A}}{x_0}-1\right)\mathbb{E}\left[\frac{P_{Y|X}(\overline{Y}+1|\mathsf{A})}{P_{Y^\star}(\overline{Y}+1)}\mathbb{1}_{\overline{Y}<c}\right]} \qquad (105)$$

$$\geq\frac{-\frac{1}{\mathsf{A}}\log\left(\frac{\mathsf{A}}{x_0}\right)+\frac{1}{x_0}}{\left(\frac{\mathsf{A}}{x_0}-1\right)\mathbb{E}\left[\frac{P_{Y|X}(\overline{Y}+1|\mathsf{A})}{P_{Y^\star}(\overline{Y}+1)}\mathbb{1}_{\overline{Y}<c}\right]}, \qquad (106)$$

where in (105) we have used the inequality $\exp(x)\geq x+1$; and in (106) we have used that $\mathbb{P}\left[\overline{Y}\geq c\right]\leq\frac{1}{\mathsf{A}}$. This concludes the proof. ∎

To find an explicit lower bound on $P_{X^*}(\mathsf{A})$, we need to find an upper bound on $\mathbb{E}\left[\frac{P_{Y|X}(\overline{Y}+1|\mathsf{A})}{P_{Y^\star}(\overline{Y}+1)}\mathbb{1}_{\overline{Y}<c}\right]$. The next lemma provides such an upper bound.

Lemma 6: Let $x_0=\max\{\mathsf{supp}(P_{X^\star})\setminus\{\mathsf{A}\}\}$, and suppose that the following conditions hold:

$$|\mathsf{supp}(P_{X^\star})|\geq 3, \qquad (107)$$
$$1-3\,\mathrm{e}^{-\frac{C(\mathsf{A})}{1-\mathrm{e}^{-\mathsf{A}}}}>0, \qquad (108)$$
$$c<x_0\mathsf{A}+1. \qquad (109)$$

Then,

$$\mathbb{E}\left[\frac{P_{Y|X}(\overline{Y}+1|\mathsf{A})}{P_{Y^\star}(\overline{Y}+1)}\mathbb{1}_{\overline{Y}<c}\right]\leq\frac{1}{1-3\,\mathrm{e}^{-\frac{C(\mathsf{A})}{1-\mathrm{e}^{-\mathsf{A}}}}}$$
$$\cdot\begin{cases}x_0\mathsf{A}^{x_0-1}, & c\leq x_0\\ x_0\mathsf{A}^{x_0-1}+\frac{\mathsf{A}\mathrm{e}^{-\mathsf{A}+1-x_0+c}}{\sqrt{2\pi c}(x_0\mathsf{A}-c+1)}\left(\frac{x_0\mathsf{A}}{c}\right)^c, & c>x_0,\end{cases}$$
$$\qquad (110)$$

where $\overline{Y}\sim\mathcal{P}(x_0)$.

Proof: See Appendix C. ∎

The final result that we require is the following bound on the tail of the Poisson random variable.

Lemma 7: For $\mathsf{A}\geq\mathrm{e}$ and $x_0\geq 1$

$$\mathbb{P}\left[\overline{Y}\geq\mathrm{e}\,x_0\log(\mathsf{A})\right]\leq\left(\frac{1}{\log(\mathsf{A})}\right)^{\mathrm{e}\log(\mathsf{A})}\mathrm{e}^{-1}\leq\frac{1}{\mathsf{A}}, \quad (111)$$

where $\overline{Y}\sim\mathcal{P}(x_0)$.

Proof: See Appendix D. ∎

With Lemma 5, Lemma 6, and Lemma 7 at our disposal, we are now ready to provide an explicit bound on $P_{X^*}(\mathsf{A})$. Next, we make a choice of c and verify that it satisfies the conditions of the aforementioned lemmas.

To that end, choose $c=\mathrm{e}x_0\log(\mathsf{A})$. Moreover, assume that $|\mathsf{supp}(P_{X^*})|\geq 4$, which by Lemma 4 guarantees that $x_0\geq 1$. This choice of c satisfies the condition of Lemma 5 since $|\mathsf{supp}(P_{X^*})|\geq 4$ implies that $\mathsf{A}>\bar{\mathsf{A}}\approx 3.4$. Furthermore, we assume that $1-3\,\mathrm{e}^{-\frac{C(\mathsf{A})}{1-\mathrm{e}^{-\mathsf{A}}}}>0$, which is needed in Lemma 6. Now using the bound in (18), both $1-3\mathrm{e}^{-\frac{C(\mathsf{A})}{1-\mathrm{e}^{-\mathsf{A}}}}>0$ and $|\mathsf{supp}(P_{X^*})|\geq 4$ hold provided that $\mathrm{e}^{\frac{C(\mathsf{A})}{1-\mathrm{e}^{-\mathsf{A}}}}\geq 4$. To summarize, in order to proceed with the proof, we require that

$$\mathrm{e}^{\frac{C(\mathsf{A})}{1-\mathrm{e}^{-\mathsf{A}}}}\geq 4. \qquad (112)$$

Having verified the conditions of Lemma 5 and Lemma 6, the bound on $P_{X^*}(\mathsf{A})$ now proceeds as follows:

$$P_{X^*}(\mathsf{A})$$
$$\geq\frac{-\frac{1}{\mathsf{A}}\log\left(\frac{\mathsf{A}}{x_0}\right)+\frac{1}{x_0}}{\left(\frac{\mathsf{A}}{x_0}-1\right)\mathbb{E}\left[\frac{P_{Y|X}(\overline{Y}+1|\mathsf{A})}{P_{Y^\star}(\overline{Y}+1)}\mathbb{1}_{\overline{Y}<c}\right]} \qquad (113)$$

$$\geq\frac{\mathsf{A}^{-2}}{\mathbb{E}\left[\frac{P_{Y|X}(\overline{Y}+1|\mathsf{A})}{P_{Y^\star}(\overline{Y}+1)}\mathbb{1}_{\overline{Y}<c}\right]} \qquad (114)$$

$$\geq \frac{\mathsf{A}^{-2}\left(1 - 3\,\mathrm{e}^{-\frac{C(\mathsf{A})}{1 - \mathrm{e}^{-\mathsf{A}}}}\right)}{x_0\mathsf{A}^{x_0 - 1} + \mathsf{A}\mathrm{e}^{-\mathsf{A}+1-x_0+c}\frac{1}{\sqrt{2\pi c}(x_0\mathsf{A} - c + 1)}\left(\frac{x_0\mathsf{A}}{c}\right)^c} \quad (115)$$

$$\geq \frac{\mathsf{A}^{-2}\left(1 - 3\,\mathrm{e}^{-\frac{C(\mathsf{A})}{1 - \mathrm{e}^{-\mathsf{A}}}}\right)}{x_0\mathsf{A}^{x_0 - 1} + \mathsf{A}\mathrm{e}^{-\mathsf{A}+1-x_0+x_0\mathrm{e}\log(\mathsf{A})}\left(\frac{x_0\mathsf{A}}{x_0\mathrm{e}\log(\mathsf{A})}\right)^{x_0\mathrm{e}\log(\mathsf{A})}} \quad (116)$$

$$\geq \frac{1 - 3\,\mathrm{e}^{-\frac{C(\mathsf{A})}{1 - \mathrm{e}^{-\mathsf{A}}}}}{2\mathsf{A}^{2\mathsf{A}\mathrm{e}\log(\mathsf{A})+2}\,\mathrm{e}^{-2\mathsf{A}+1}}, \quad (117)$$

where (113) is due to the bound in Lemma 5; (114) follows from the upper bound $\frac{\mathsf{A}}{x_0} - 1 \leq \mathsf{A}$ because $x_0 \geq 1$, and the fact $x \mapsto -\frac{1}{\mathsf{A}}\log\left(\frac{\mathsf{A}}{x}\right) + \frac{1}{x}$ is minimized at $x = \mathsf{A}$; (115) follows by using the bound in (110); (116) follows from the choice of $c = x_0\mathrm{e}\log(\mathsf{A})$ and the bound $\frac{1}{\sqrt{2\pi c}(x_0\mathsf{A} - c + 1)} \leq 1$; and (117) follows by using $1 \leq x_0 \leq \mathsf{A}$, which implies that

$$x_0\mathsf{A}^{x_0 - 1} + \mathsf{A}\mathrm{e}^{-\mathsf{A}+1-x_0+x_0\mathrm{e}\log(\mathsf{A})}\left(\frac{x_0\mathsf{A}}{x_0\mathrm{e}\log(\mathsf{A})}\right)^{x_0\mathrm{e}\log(\mathsf{A})}$$
$$\leq 2\mathsf{A}^{2\mathsf{A}\mathrm{e}\log(\mathsf{A})}\,\mathrm{e}^{-2\mathsf{A}+1}. \quad (118)$$

This concludes the proof of the lower bound on $P_{X^*}(\mathsf{A})$.

Remark 5: To arrive at the bound in (117), we have used $x_0 \geq 1$, which requires an assumption that $\mathrm{e}^{\frac{C(\mathsf{A})}{1 - \mathrm{e}^{-\mathsf{A}}}} \geq 4$. Alternatively, we could have used the bound $x_0 \geq \mathrm{e}^{-\sqrt{2(\log(\mathsf{A})-1)}}$ in (80), which only requires an assumption that $\mathrm{e}^{\frac{C(\mathsf{A})}{1 - \mathrm{e}^{-\mathsf{A}}}} > 3$, and would result in

$$P_{X^*}(\mathsf{A}) \geq \frac{1 - 3\,\mathrm{e}^{-\frac{C(\mathsf{A})}{1 - \mathrm{e}^{-\mathsf{A}}}}}{2\mathsf{A}^{2\mathsf{A}\mathrm{e}\log(\mathsf{A})\exp\left(-\sqrt{2(\log(\mathsf{A})-1)}\right)+2}\exp(-2\mathsf{A}+1)}. \quad (119)$$

However, asymptotically this bound is weaker than that in (117).

D. Proof of the Upper Bound in (20)

In this section, we establish an upper bound on the number of the support points. Specifically, we will establish a bound of the order $\mathsf{A}\log^2(\mathsf{A})$. To show the bound on the number of points, we need to first present a number of ancillary results. The first result that we will need allows us to bound the number of zeros of f with the number of zeros of f'.

Lemma 8: Suppose that f is continuous on $[-R, R]$ and differentiable on $(-R, R)$. If $\mathrm{N}([-R, R], f) < \infty$, then

$$\mathrm{N}([-R, R], f) \leq \mathrm{N}([-R, R], f') + 1, \quad (120)$$

where f' denotes the derivative of f.

Proof: Let $x_1 < \cdots < x_n$ denote the zeros of f. By Rolle's theorem, each of the intervals (x_i, x_{i+1}) for $i = 1, \ldots, n-1$ contains at least one extreme point. ∎

The final ancillary result that we need is the following lemma.

Lemma 9: Let $f : \mathbb{N}_0 \to \mathbb{R}$. Suppose that there exists $k^\star$ such that one of the following conditions holds:

1) $f(k) \leq 0$ for all $k \geq k^\star$; or
2) $f(k) \geq 0$ for all $k \geq k^\star$.

Then,

$$\mathscr{S}(f) \leq k^\star. \quad (121)$$

Proof: To prove the bound, note that for all $k \geq k^\star$ the function $f(k)$ no longer can change sign. Therefore, there can be at most $k^\star$ sign changes. ∎

The next result provides an upper bound on the number of support points in terms of the number of sign changes of the function related to $P_{Y^\star}$.

Lemma 10: For every $\mathsf{A} > 0$

$$|\mathsf{supp}(P_{X^\star})| \leq \mathscr{S}(\psi(\cdot; P_{X^\star})) + 2, \quad (122)$$

where

$$\psi(k; P_{X^\star}) = k\log\left(\frac{1}{\mathbb{E}[X^\star | Y^\star = k - 1]}\right) + \log(k! P_{Y^\star}(k)) + C(\mathsf{A}) + k, \quad k \in \mathbb{N}_0, \quad (123)$$

and where we define $k\log\left(\frac{1}{\mathbb{E}[X^\star | Y^\star = k - 1]}\right)\Big|_{k=0} = 0$.

Proof: First, using Lemma 13 in Appendix B, we have that

$$i(x; P_{X^\star}) - C(\mathsf{A}) = G(x) - x - C(\mathsf{A}) + x\log(x), \quad (124)$$

where

$$G(x) = \sum_{k=0}^{\infty} P_{Y|X}(k|x)\log\left(\frac{1}{k! P_{Y^\star}(k)}\right) \quad (125)$$

$$= \sum_{k=0}^{\infty} P_{Y|X}(k|x)g(k), \quad (126)$$

and where we let $g(k) = \log\left(\frac{1}{k! P_{Y^\star}(k)}\right)$.

Now, the upper bound can be established as follows:

$$|\mathsf{supp}(P_{X^\star})|$$
$$\leq \mathrm{N}([0, \mathsf{A}], i(x; P_{X^\star}) - C(\mathsf{A})) \quad (127)$$
$$= \mathrm{N}((0, \mathsf{A}], i(x; P_{X^\star}) - C(\mathsf{A})) + 1 \quad (128)$$
$$= \mathrm{N}((0, \mathsf{A}], G(x) - x - C(\mathsf{A}) + x\log(x)) + 1 \quad (129)$$
$$= \mathrm{N}\left((0, \mathsf{A}], \frac{G(x)}{x} - 1 - \frac{C(\mathsf{A})}{x} + \log(x)\right) + 1 \quad (130)$$
$$\leq \mathrm{N}\left((0, \mathsf{A}], \frac{G'(x)}{x} - \frac{G(x)}{x^2} + \frac{C(\mathsf{A})}{x^2} + \frac{1}{x}\right) + 2 \quad (131)$$
$$= \mathrm{N}((0, \mathsf{A}], xG'(x) - G(x) + C(\mathsf{A}) + x) + 2 \quad (132)$$
$$= \mathrm{N}((0, \mathsf{A}], \mathbb{E}[Y(g(Y) - g(Y - 1)) - g(Y) + C(\mathsf{A})$$
$$+ Y | X = x]) + 2 \quad (133)$$
$$= \mathrm{N}((0, \mathsf{A}], \mathbb{E}[\psi(Y; P_{X^\star}) | X = x]) + 2 \quad (134)$$
$$\leq \mathscr{S}(\psi(\cdot; P_{X^\star})) + 2, \quad (135)$$

where (127) follows from the containment in (41); (128) follows by using the fact that $0 \in \mathsf{supp}(P_{X^\star})$; (129) follows by using the definition of $G(x)$ in (124); (130) follows by using the fact that $G(x) - x - C(\mathsf{A}) + x\log(x)$ and $\frac{G(x)}{x} - 1 - \frac{C(\mathsf{A})}{x} + \log(x)$ have the same zeros on $(0, \mathsf{A}]$; (131) follows by using the bound in Lemma 8; (132) follows by using the fact that $\frac{G'(x)}{x} - \frac{G(x)}{x^2} + \frac{C(\mathsf{A})}{x^2} + \frac{1}{x}$ and $xG'(x) - G(x) + C(\mathsf{A}) + x$ have

the same number of zeros on $(0, \mathsf{A}]$; (133) follows by using Lemma 12 in Appendix B which leads to

$$xG'(x) = x\mathbb{E}\left[g(Y+1) - g(Y)|X = x\right]$$
$$= \mathbb{E}\left[Y(g(Y) - g(Y-1))|X = x\right], \tag{136}$$

$$-G(x) + C(\mathsf{A}) + x = \mathbb{E}\left[-g(Y) + C(\mathsf{A}) + Y|X = x\right], \tag{137}$$

where in (136) we assume that $kg(k-1)|_{k=0} = 0$; (134) follows by defining

$$\psi(k; P_{X^\star})$$
$$= k(g(k) - g(k-1)) - g(k) + C(\mathsf{A}) + k \tag{138}$$
$$= k\log\left(\frac{(k-1)!P_{Y^\star}(k-1)}{k!P_{Y^\star}(k)}\right) + \log\left(k!P_{Y^\star}(k)\right) + C(\mathsf{A}) + k \tag{139}$$
$$= k\log\left(\frac{1}{\mathbb{E}[X^\star|Y^\star = k-1]}\right) + \log\left(k!P_{Y^\star}(k)\right) + C(\mathsf{A}) + k, \tag{140}$$

and (135) follows from the oscillation theorem. This concludes the proof. ∎

Remark 6: Note that the goal of steps (127) through (134) is to transform the function $x \mapsto i(x; P_{X^\star}) - C(\mathsf{A})$ to a function of the form $\mathbb{E}\left[\psi(Y; P_{X^\star})|X = x\right]$. The latter formulation is significant as it enables the application of the oscillation theorem. The reader might wonder why such a representation was not possible right after (127)? To see why this cannot be done, note that

$$i(x; P_{X^\star}) - C(\mathsf{A})$$
$$= G(x) - x - C(\mathsf{A}) + x\log(x) \tag{141}$$
$$= \mathbb{E}\left[g(Y) - Y - C(\mathsf{A})|X = x\right] + x\log(x), \tag{142}$$

where in the last step we have used the expression for $G(x)$ in (126) and the fact that $\mathbb{E}[Y|X = x] = x$. Therefore, it remains to answer whether we can write $\mathbb{E}[h(Y)|X = x] = x\log(x)$ for some function h. In other words, can we show that $x\log(x)$ is a statistic of a Poisson distribution with mean x? To answer this question, towards a contradiction, assume that $x\log(x)$ is a statistic of a Poisson distribution and there exists a function $h : \mathbb{N}_0 \to \mathbb{R}$ such that

$$x\log(x) = \mathbb{E}\left[h(Y)|X = x\right] = \sum_{k=0}^{\infty} \frac{h(k)}{k!} x^k e^{-x}. \tag{143}$$

Multiplying the above by e^x, we arrive at

$$e^x x\log(x) = \sum_{k=0}^{\infty} \frac{h(k)}{k!} x^k, \tag{144}$$

which asserts that the function $e^x x\log(x)$ has a power series representation at $x = 0$. This is, of course, not possible and leads to a contradiction. Therefore, we cannot write $\mathbb{E}[h(Y)|X = x] = x\log(x)$. The purpose of steps (127) through (134) was to bypass this issue and eliminate the term $x\log(x)$ while keeping the number of zeros relatively unchanged. Finally, we note that the analyticity argument

leading to the contradiction in (144) has been used before in [15, Thm. 15] to show discreteness of the capacity-achieving distribution for the Poisson noise channel with both peak-power and average-power constraints.

The next result provides an upper bound on the number of sign changes $\mathscr{S}\left(\psi(\cdot; P_{X^\star})\right)$.

Lemma 11: For $\mathsf{A} > 0$

$$\mathscr{S}\left(\psi(\cdot; P_{X^\star})\right) \leq k^\star \tag{145}$$

where

$$k^\star = \lceil \mathsf{A} - \log\left(P_{X^\star}(\mathsf{A})\right) - C(\mathsf{A}) \rceil \geq 0. \tag{146}$$

Proof: We begin by lower bounding $\psi(\cdot; P_{X^\star})$

$$\psi(k; P_{X^\star})$$
$$= k\log\left(\frac{1}{\mathbb{E}[X^\star|Y^\star = k-1]}\right) + \log\left(k!P_{Y^\star}(k)\right) + C(\mathsf{A}) + k \tag{147}$$
$$\geq k\log\left(\frac{1}{\mathsf{A}}\right) + \log\left(k!P_{Y^\star}(k)\right) + C(\mathsf{A}) + k \tag{148}$$
$$\geq k\log\left(\frac{1}{\mathsf{A}}\right) + \log\left(P_{X^\star}(\mathsf{A})\mathsf{A}^k e^{-\mathsf{A}}\right) + C(\mathsf{A}) + k \tag{149}$$
$$= -\mathsf{A} + \log\left(P_{X^\star}(\mathsf{A})\right) + C(\mathsf{A}) + k, \tag{150}$$

where (148) follows by using the bound $\mathbb{E}[X^\star|Y^\star] \leq \mathsf{A}$; and (149) follows by using the fact that there is a mass point at A which leads to the bound

$$P_{Y^\star}(k) = \frac{1}{k!}\mathbb{E}\left[X^{\star k} e^{-X^\star}\right] \geq \frac{1}{k!}P_{X^\star}(\mathsf{A})\mathsf{A}^k e^{-\mathsf{A}}. \tag{151}$$

Therefore, there exists $k^\star = \lceil \mathsf{A} - \log\left(P_{X^\star}(\mathsf{A})\right) - C(\mathsf{A}) \rceil$ such that

$$\psi(k; P_{X^\star}) \geq 0, \qquad k \geq k^\star. \tag{152}$$

It remains to show that $k^\star$ is larger than zero. This follows by using the bound in (45)

$$\mathsf{A} - \log\left(P_{X^\star}(\mathsf{A})\right) - C(\mathsf{A}) \geq \mathsf{A} - \log\left(e^{-C(\mathsf{A})}\right) - C(\mathsf{A}) = \mathsf{A}. \tag{153}$$

The proof is concluded by using Lemma 9. ∎

Remark 7: Note that the value of $P_{X^\star}(\mathsf{A})$ was introduced in the step (149). Since a significant part of this work was dedicated to find a lower bound on $P_{X^\star}(\mathsf{A})$, it is interesting to discuss how tight the bound in (149) actually is. Observe that

$$\lim_{k \to \infty} \frac{k!P_{Y^\star}(k)}{P_{X^\star}(\mathsf{A})\mathsf{A}^k e^{-\mathsf{A}}}$$
$$= \lim_{k \to \infty} \sum_{x \in \text{supp}(P_{X^\star})\setminus\{\mathsf{A}\}} \frac{P_{X^\star}(x)}{P_{X^\star}(\mathsf{A})}\left(\frac{x}{\mathsf{A}}\right)^k e^{-x+\mathsf{A}} + 1 = 1. \tag{154}$$

Therefore, asymptotically the bound used to produce the inequality in (149) is tight.

Now, combining the bounds in (122), (146) and (16), we arrive at

$$|\mathrm{supp}(P_{X^\star})|$$
$$\leq \lceil \mathsf{A} - \log\left(P_{X^\star}(\mathsf{A})\right) - C(\mathsf{A}) \rceil + 2 \tag{155}$$
$$\leq \mathsf{A} - \log\left(P_{X^\star}(\mathsf{A})\right) - C(\mathsf{A}) + 3 \tag{156}$$
$$\leq \mathsf{A} - \log\left(\frac{1 - 3\,\mathrm{e}^{-\frac{C(\mathsf{A})}{1 - \mathrm{e}^{-\mathsf{A}}}}}{2\mathsf{A}^{2\mathsf{A}\mathrm{e}\log(\mathsf{A}) + 2}\,\mathrm{e}^{-2\mathsf{A}+1}}\right) - C(\mathsf{A}) + 3 \tag{157}$$
$$= 2\mathrm{e}\mathsf{A}\log^2(\mathsf{A}) + 2\log(\mathsf{A}) - \mathsf{A} - \log\left(\frac{1 - 3\,\mathrm{e}^{-\frac{C(\mathsf{A})}{1 - \mathrm{e}^{-\mathsf{A}}}}}{2}\right)$$
$$- C(\mathsf{A}) + 4, \tag{158}$$

where the first inequality is due to the bound $\lceil x \rceil \leq x + 1$. This concludes the proof.

IV. CONCLUSION

This work has focused on studying properties of the capacity-achieving distribution for the Poisson noise channel with an amplitude constraint. It was previously known that the capacity-achieving distribution for this channel is discrete with finitely many points. In this work, we sharpened this result in several ways.

First, by using a strong data-processing inequality, an upper bound on the values of the mass points has been shown. This upper bound on the probability values has been shown to lead to the lower bound on the number of support point of the optimal input distribution. Specifically, a lower bound of order $\sqrt{\mathsf{A}}$ has been established on the number of support points where A is the constraint on the amplitude.

Second, by using the variation-diminishing property of the Poisson kernel, the work has also established an upper bound on the number of the support points of the optimal input distribution. Specifically, an order $\mathsf{A}\log^2(\mathsf{A})$ bound has been established.

Finally, along the way, several other results have been shown. For example, a new compact expression for the capacity has been shown. In addition, a lower bound on the probability of the largest points of the optimal input distribution has been established. Furthermore, an estimate on the locations of the support other than 0 and A has been established.

There are few interesting future directions and open questions:

- Some of the ideas in this work, in particular the oscillation theorem, were borrowed from [28] where a Gaussian noise channel was considered. It is interesting to note that in [28], it was shown that the oscillation theorem leads to an order tight bound on the cardinality of support of the capacity-achieving distribution. Currently, we do not have such a claim for the Poisson noise channel. Therefore, an interesting future direction will be to assess whether the bound due to the oscillation theorem is also order tight for the Poisson noise channel.
- The location of the second-largest point plays an important role in our analysis. In particular, it will be interesting to investigate the gap between the largest and the second-largest points. Currently, this gap is lower-bounded by a constant. In other words, the second-largest point never approaches A. In fact, the numerical simulation suggests that this gap increases as a function of A. One interesting future direction is to investigate this gap and determine whether it increases with A.
- This work has developed a method for finding lower bounds on the value of $P_{X^\star}(\mathsf{A})$. This lower bound was instrumental for finding an explicit upper bound on the cardinality of support. However, the current bound on $P_{X^\star}(\mathsf{A})$ appears to be loose. An important future direction is to find ways of improving this bound.
- The general upper bounds on the values of the probabilities can be applied to other channels, and it would be interesting to see how tight these bounds are. Another interesting future direction is to extend the results of this paper to a more general Poisson model that includes the dark current parameter.

APPENDIX A
SKETCH OF THE IMPLEMENTATION OF THE GRADIENT ASCENT

First, to apply the projected gradient ascent [37], we first parametrize the input distribution as a vector in $\mathbb{R}^{2n}$

$$P_X \Longrightarrow \mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2] = [\underbrace{x_1, x_2, \ldots, x_n}_{\text{points of the support}},\ \underbrace{p_1, p_2, \ldots, p_n}_{\text{probability values}}]. \tag{159}$$

From Theorem 2, for a given A, we know that n can be at most equal to (20). Therefore, we set n to be equal to (20). Second, we explicitly write the mutual information as a function of $\mathbf{x}$ and let

$$g(\mathbf{x}) = I(X; Y). \tag{160}$$

In view of Theorem 2, we have that

$$\max_{P_X: 0 \leq X \leq \mathsf{A}} I(X; Y) = \max_{\mathbf{x}:\, \mathbf{x}_1 \in [0, \mathsf{A}]^n,\, \mathbf{x}_2 \in \mathcal{P}} g(\mathbf{x}), \tag{161}$$

where $\mathcal{P}$ is the probability simplex.

The update equation for the projected gradient ascent are now given by

$$\mathbf{x}^{(t+1)} = \mathsf{proj}\left(\mathbf{x}^{(t)} + \lambda \nabla g(\mathbf{x}^{(t)})\right), \ t \in \mathbb{N} \tag{162}$$

where $\lambda > 0$ is some step size (to generate Fig. 2, we have used $\lambda = 0.01$), $\nabla g(\mathbf{x})$ is the gradient of g, and the projection operation proj maps a vector $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2]$ to the set

$$\{\mathbf{x} : \mathbf{x}_1 \in [0, \mathsf{A}]^n, \mathbf{x}_2 \in \mathcal{P}\}. \tag{163}$$

The projection of the first element $\mathbf{x}_1$ onto a cube $[0, \mathsf{A}]^n$ can be done by individually going through each coordinate of $\mathbf{x}_1$. An efficient implementation of the projection of $\mathbf{x}_2$ onto the probability simplex can be found in [44].

Finally, the initial condition $\mathbf{x}^{(1)}$ for given A, is chosen to be the distribution that was a final output of the gradient algorithm for the previous epoch in which we considered $\mathsf{A}' < \mathsf{A}$. For the best performance, the difference between A' and A should be made as small as possible.

APPENDIX B
DERIVATIVES OF $i(\cdot; P_X)$

To find the derivatives of the term $i(\cdot; P_X)$, we will need the following auxiliary results, the proof of which can be found in [31].

Lemma 12: Let $f : \mathbb{N}_0 \to \mathbb{R}$ and assume $P_{Y|X}$ is given in (1). Then, for any P_X, the following identities hold:

$$\frac{\mathrm{d}}{\mathrm{d}x}\mathbb{E}[f(Y)|X = x] = \mathbb{E}[f(Y+1) - f(Y)|X = x], \ x > 0, \tag{164}$$

$$x\mathbb{E}[f(Y)|X = x] = \mathbb{E}[Yf(Y-1)|X = x], \quad x \geq 0 \tag{165}$$

where $kf(k-1)|_{k=0} = 0$ and

$$\mathbb{E}[X|Y = y] = \frac{(y+1)P_Y(y+1)}{P_Y(y)}, \quad y \in \mathbb{N}_0, \tag{166}$$

$$\mathbb{E}[X^k|Y = y] = \prod_{i=1}^{k-1}\mathbb{E}[X|Y = y+i], \quad k \in \mathbb{N}, \ y \in \mathbb{N}_0. \tag{167}$$

The derivatives of $i(\cdot; P_X)$ are given next.
Lemma 13: We have

$$i(x; P_X) = G(x) + x\log(x) - x, \quad x \geq 0, \tag{168}$$

where

$$G(x) = \sum_{k=0}^{\infty} P_{Y|X}(k|x)\log\left(\frac{1}{k!P_Y(k)}\right). \tag{169}$$

Moreover,

- The first and second derivative of $G(x)$ are given by

$$G'(x) = \sum_{k=0}^{\infty} P_{Y|X}(k|x)\log\left(\frac{1}{\mathbb{E}[X|Y = k]}\right), \tag{170}$$

$$G''(x) = \sum_{k=0}^{\infty} P_{Y|X}(k|x)\log\left(\frac{\mathbb{E}[X|Y = k]}{\mathbb{E}[X|Y = k+1]}\right) \tag{171}$$

$$= \sum_{k=0}^{\infty} P_{Y|X}(k|x)\log\left(\frac{\mathbb{E}^2[X|Y = k]}{\mathbb{E}[X^2|Y = k]}\right) \tag{172}$$

$$= \sum_{k=0}^{\infty} P_{Y|X}(k|x)\log\left(\frac{1}{\mathbb{E}[X|Y = k+1]}\right) - G'(x). \tag{173}$$

- The second derivative of $i(x; P_X)$ is given by

$$i''(x; P_X) = \sum_{k=0}^{\infty} P_{Y|X}(k|x)\log\left(\frac{1}{\mathbb{E}[X|Y = k+1]}\right) - i'(x; P_X) + \log(x) + \frac{1}{x}. \tag{174}$$

Proof: We first show the decomposition in (168)

$$i(x; P_X)$$
$$= \sum_{k=0}^{\infty} P_{Y|X}(k|x)\log\left(\frac{P_{Y|X}(k|x)}{P_Y(k)}\right) \tag{175}$$

$$= \sum_{k=0}^{\infty} P_{Y|X}(k|x)\left(\log\left(\frac{1}{k!P_Y(k)}\right) + k\log(x) - x\right) \tag{176}$$

$$= \sum_{k=0}^{\infty} P_{Y|X}(k|x)\log\left(\frac{1}{k!P_Y(k)}\right) + x\log(x) - x \tag{177}$$

$$= G(x) - x + x\log(x), \tag{178}$$

where (176) follows by using that

$$\log\left(P_{Y|X}(k|x)\right) = \log\left(\frac{x^k e^{-x}}{k!}\right) = \log\left(\frac{1}{k!}\right) + k\log(x) - x; \tag{179}$$

and where (177) follows by using that $\mathbb{E}[Y|X = x] = x$.

Next, to find the derivative of $G(x)$, we use the properties in (164) and (166) which lead to

$$G'(x) = \sum_{k=0}^{\infty} P_{Y|X}(k|x)\log\left(\frac{P_Y(k)}{(k+1)P_Y(k+1)}\right) \tag{180}$$

$$= \sum_{k=0}^{\infty} P_{Y|X}(k|x)\log\left(\frac{1}{\mathbb{E}[X|Y = k]}\right). \tag{181}$$

Similarly, by using (164) and (167) we arrive at

$$G''(x) = \sum_{k=0}^{\infty} P_{Y|X}(k|x)\log\left(\frac{\mathbb{E}[X|Y = k]}{\mathbb{E}[X|Y = k+1]}\right) \tag{182}$$

$$= \sum_{k=0}^{\infty} P_{Y|X}(k|x)\log\left(\frac{\mathbb{E}^2[X|Y = k]}{\mathbb{E}[X^2|Y = k]}\right). \tag{183}$$

To show the expression for the second derivative observe that

$$i''(x; P_X)$$
$$= G''(x) + \frac{1}{x} \tag{184}$$

$$= \sum_{k=0}^{\infty} P_{Y|X}(k|x)\log\left(\frac{1}{\mathbb{E}[X|Y = k+1]}\right) - G'(x) + \frac{1}{x} \tag{185}$$

$$= \sum_{k=0}^{\infty} P_{Y|X}(k|x)\log\left(\frac{1}{\mathbb{E}[X|Y = k+1]}\right) - i'(x; P_X) + \log(x) + \frac{1}{x}. \tag{186}$$

■

APPENDIX C
PROOF OF LEMMA 6

We start from

$$\mathbb{E}\left[\frac{P_{Y|X}(\bar{Y}+1|\mathsf{A})}{P_{Y^\star}(\bar{Y}+1)}\mathbb{1}_{\bar{Y} < c}\right] = \sum_{k=0}^{c-1} P_{\bar{Y}}(k)\frac{\mathsf{A}^{k+1}e^{-\mathsf{A}}}{(k+1)!P_{Y^\star}(k+1)}. \tag{187}$$

We next find a lower bound on the $P_{Y^\star}(k+1)$ where we distinguish between two regimes $x_0 \le c$ and $c > x_0$.

First, for $1 \le k+1 \le \min\{x_0, c\}$, we have the following lower bound on $P_{Y^\star}(k+1)$:

$$(k+1)! P_{Y^\star}(k+1)$$
$$= \sum_{0 < x \le A} P_{X^\star}(x) x^{k+1} e^{-x} \tag{188}$$
$$\ge \sum_{1 \le x \le k+1} P_{X^\star}(x) x^{k+1} e^{-x} + \sum_{k+1 < x \le x_0} P_{X^\star}(x) x^{k+1} e^{-x} \tag{189}$$
$$\ge \sum_{1 \le x \le k+1} P_{X^\star}(x) e^{-1} + \sum_{k+1 < x \le x_0} P_{X^\star}(x) A^{k+1} e^{-A} \tag{190}$$
$$\ge \sum_{1 \le x \le k+1} P_{X^\star}(x) e^{-1} + \sum_{k+1 < x \le x_0} P_{X^\star}(x) A e^{-A} \tag{191}$$
$$\ge A e^{-A} \sum_{1 \le x \le x_0} P_{X^\star}(x) \tag{192}$$
$$= A e^{-A} \left(1 - \sum_{0 \le x < 1} P_{X^\star}(x) - P_{X^\star}(A) \right) \tag{193}$$
$$\ge A e^{-A} \left(1 - 3 e^{-\frac{C(A)}{1 - e^{-A}}} \right), \tag{194}$$

where in (190) we used $x^{k+1} e^{-x} \ge e^{-1}$, because $x \mapsto x^{k+1} e^{-x}$ is increasing for $x < k+1$, and in the second summation we used that $x \mapsto x^{k+1} e^{-x}$ is decreasing for $k+1 < x \le x_0 < A$; in (191), we used $A^{k+1} \ge A$ for $A \ge 1$ and $k \ge 0$; in (192), we used $A e^{-A} \le e^{-1}$ for $A \ge 1$; and in (194) we used that there is at most one support point in the open interval $(0, 1)$, and the upper bound on the probability mass in (14).

If $c > x_0$, we can follow analogous bounding steps to write

$$(k+1)! P_{Y^\star}(k+1)$$
$$\ge \sum_{1 \le x \le x_0} P_{X^\star}(x) e^{-1} \tag{195}$$
$$= e^{-1} \left(1 - \sum_{0 \le x < 1} P_{X^\star}(x) - P_{X^\star}(A) \right) \tag{196}$$
$$\ge e^{-1} \left(1 - 3 e^{-\frac{C(A)}{1 - e^{-A}}} \right), \qquad x_0 < k+1 < c. \tag{197}$$

If $c \le x_0$, plugging (194) into (187), we get

$$\mathbb{E} \left[\frac{P_{Y|X}(\bar{Y}+1|A)}{P_{Y^\star}(\bar{Y}+1)} \mathbb{1}_{\bar{Y} < c} \right]$$
$$\le \frac{1}{A e^{-A} \left(1 - 3 e^{-\frac{C(A)}{1 - e^{-A}}} \right)} \sum_{k=0}^{c-1} P_{\bar{Y}}(k) A^{k+1} e^{-A} \tag{198}$$
$$\le \frac{c A^{c-1}}{1 - 3 e^{-\frac{C(A)}{1 - e^{-A}}}} \tag{199}$$
$$\le \frac{x_0 A^{x_0-1}}{1 - 3 e^{-\frac{C(A)}{1 - e^{-A}}}}. \tag{200}$$

Finally, if $c > x_0$, we have

$$\mathbb{E} \left[\frac{P_{Y|X}(\bar{Y}+1|A)}{P_{Y^\star}(\bar{Y}+1)} \mathbb{1}_{\bar{Y} < c} \right]$$
$$= \mathbb{E} \left[\frac{P_{Y|X}(\bar{Y}+1|A)}{P_{Y^\star}(\bar{Y}+1)} (\mathbb{1}_{\bar{Y} \le x_0} + \mathbb{1}_{x_0 < \bar{Y} < c}) \right] \tag{201}$$
$$\le \frac{1}{1 - 3 e^{-\frac{C(A)}{1 - e^{-A}}}} \left[x_0 A^{x_0-1} + \sum_{k=\lceil x_0 \rceil}^{c-1} P_{\bar{Y}}(k) A^{k+1} e^{-A+1} \right] \tag{202}$$
$$\le \frac{1}{1 - 3 e^{-\frac{C(A)}{1 - e^{-A}}}} \left[x_0 A^{x_0-1} + \sum_{k=0}^{c-1} \frac{(x_0 A)^k}{k!} A e^{-A+1-x_0} \right] \tag{203}$$
$$= \frac{1}{1 - 3 e^{-\frac{C(A)}{1 - e^{-A}}}} \left[x_0 A^{x_0-1} + A e^{-A+1-x_0+x_0 A} \frac{\Gamma(c, x_0 A)}{c!} \right] \tag{204}$$
$$\le \frac{1}{1 - 3 e^{-\frac{C(A)}{1 - e^{-A}}}} \left[x_0 A^{x_0-1} + A e^{-A+1-x_0} \frac{(x_0 A)^c}{c!(x_0 A - c + 1)} \right] \tag{205}$$
$$\le \frac{1}{1 - 3 e^{-\frac{C(A)}{1 - e^{-A}}}} \left[x_0 A^{x_0-1} + \frac{A e^{-A+1-x_0+c} \left(\frac{x_0 A}{c} \right)^c}{\sqrt{2\pi c}(x_0 A - c + 1)} \right], \tag{206}$$

where (202) follows from the bound in (200); (204) and (205) follow by introducing the incomplete Gamma function $\Gamma(c, x_0 A)$, which is then upper-bounded as [45, Thm. 2.1]

$$\Gamma(c, x_0 A) \le \frac{(x_0 A)^c e^{-x_0 A}}{x_0 A - c + 1}, \qquad 0 < c < x_0 A + 1; \tag{207}$$

and (206) follows by using Stirling's lower bound $c! \ge \sqrt{2\pi c}\, c^c e^{-c}$.

APPENDIX D
PROOF OF LEMMA 7

By using the bound on the tail of the Poisson random variable in [46, Thm. 5.4], we have that

$$\mathbb{P}\left[\overline{Y} \ge c\right] \le \frac{(e x_0)^c e^{-x_0}}{c^c}, \quad c > x_0. \tag{208}$$

By choosing $c = e x_0 \log(A) > x_0$ for $A \ge e$, the bound in (208) implies that

$$\mathbb{P}\left[\overline{Y} \ge e x_0 \log(A)\right] \le \left(\frac{1}{\log(A)} \right)^{e x_0 \log(A)} e^{-x_0} \tag{209}$$
$$\le \left(\frac{1}{\log(A)} \right)^{e \log(A)} e^{-1}, \tag{210}$$

where the last inequality follows from the assumption that $A \ge e$ and $x_0 \ge 1$.

The proof is concluded by observing that $A \ge e$ implies

$$\left(\frac{1}{\log(A)} \right)^{e \log(A)} e^{-1} \le \frac{1}{A}. \tag{211}$$

REFERENCES

[1] H. Witsenhausen, "Some aspects of convexity useful in information theory," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 3, pp. 265–271, May 1980.

[2] J. G. Smith, "The information capacity of amplitude-and variance-constrained scalar Gaussian channels," *Inf. Control*, vol. 18, no. 3, pp. 203–219, 1971.

[3] A. Dytso, M. Goldenbaum, H. V. Poor, and S. S. Shitz, "When are discrete channel inputs optimal?—Optimization techniques and some new results," in *Proc. 52nd Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2018, pp. 1–6.

[4] J. P. Gordon, "Quantum effects in communications systems," *Proc. IRE*, vol. 50, no. 9, pp. 1898–1908, Sep. 1962.

[5] M. A. Khalighi and M. Uysal, "Survey on free space optical communication: A communication theory perspective," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 2231–2258, Nov. 2014.

[6] C. Komninakis, L. Vandenberghe, and R. D. Wesel, "Capacity of the binomial channel, or minimax redundancy for memoryless sources," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2001, p. 127.

[7] N. Farsad *et al.*, "Capacities and optimal input distributions for particle-intensity channels," *IEEE Trans. Mol., Biol. Multi-Scale Commun.*, vol. 6, no. 3, pp. 220–232, Dec. 2020.

[8] S. Verdú, "Poisson communication theory," *Int. Technion Commun. Day Honor Isr. Bar-David*, vol. 66, Mar. 1999.

[9] R. McEliece, E. Rodemich, and A. Rubin, "The practical limits of photon communication," *Jet Propuls. Lab. Deep Space Netw. Prog. Rep.*, vol. 42, pp. 63–67, Nov. 1979.

[10] J. Pierce, E. Posner, and E. Rodemich, "The capacity of the photon counting channel," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 61–77, Jan. 1981.

[11] S. Shamai (Shitz), "Capacity of a pulse amplitude modulated direct detection photon channel," *IEE Proc. I Commun., Speech Vis.*, vol. 137, no. 6, pp. 424–430, Dec. 1990.

[12] D. Brady and S. Verdú, "The asymptotic capacity of the direct detection photon channel with a bandwidth constraint," in *Proc. 28th Allerton Conf. Commun., Control Comp.*, Oct. 1990, pp. 691–700.

[13] J. Cao, S. Hranilovic, and J. Chen, "Capacity-achieving distributions for the discrete-time Poisson channel—Part I: General properties and numerical techniques," *IEEE Trans. Commun.*, vol. 62, no. 1, pp. 194–202, Jan. 2014.

[14] J. Cao, S. Hranilovic, and J. Chen, "Capacity-achieving distributions for the discrete-time Poisson channel—Part II: Binary inputs," *IEEE Trans. Commun.*, vol. 62, no. 1, pp. 203–213, Jan. 2014.

[15] M. Cheraghchi and J. Ribeiro, "Improved upper bounds and structural results on the capacity of the discrete-time Poisson channel," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4052–4068, Jul. 2019.

[16] A. Martinez, "Spectral efficiency of optical direct detection," *J. Opt. Soc. Amer. B, Opt. Phys.*, vol. 24, no. 4, pp. 739–749, 2007.

[17] A. Martinez, "Achievability of the rate $\frac{1}{2}\log(1 + \epsilon s)$ in the discrete-time Poisson channel," 2008, *arXiv:0809.3370*. [Online]. Available: https://arxiv.org/abs/0809.3370

[18] A. Lapidoth, J. H. Shapiro, V. Venkatesan, and L. Wang, "The discrete-time Poisson channel at low input powers," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3260–3272, Jun. 2011.

[19] L. Wang and G. W. Wornell, "A refined analysis of the Poisson channel in the high-photon-efficiency regime," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4299–4311, Jul. 2014.

[20] L. Wang and G. W. Wornell, "The impact of dark current on the wideband Poisson channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 2924–2928.

[21] M. Cheraghchi and J. Ribeiro, "Non-asymptotic capacity upper bounds for the discrete-time Poisson channel with positive dark current," 2020, *arXiv:2010.14858*. [Online]. Available: http://arxiv.org/abs/2010.14858

[22] A. Lapidoth and S. M. Moser, "On the capacity of the discrete-time Poisson channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 303–322, Jan. 2009.

[23] Y. Yu, Z. Zhang, L. Wu, and J. Dang, "Lower bounds on the capacity for Poisson optical channel," in *Proc. 6th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2014, pp. 1–5.

[24] F. W. J. Olver *et al.*, Eds., *NIST Digital Library of Mathematical Functions*. Accessed: Jun. 15, 2021. [Online]. Available: http://dlmf.nist.gov/

[25] M. Raginsky, "Strong data processing inequalities and ϕ-Sobolev inequalities for discrete channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3355–3389, Jun. 2016.

[26] Y. Polyanskiy and Y. Wu, "Strong data-processing inequalities for channels and Bayesian networks," in *Convexity and Concentration*. New York, NY, USA: Springer, 2017, pp. 211–249.

[27] F. du Pin Calmon, Y. Polyanskiy, and Y. Wu, "Strong data processing inequalities for input constrained additive noise channels," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1879–1892, Mar. 2017.

[28] A. Dytso, S. Yagli, H. V. Poor, and S. Shamai Shitz, "The capacity achieving distribution for the amplitude constrained additive Gaussian channel: An upper bound on the number of mass points," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2006–2022, Apr. 2019.

[29] S. Karlin, "Pólya type distributions, II," *Ann. Math. Statist.*, vol. 28, no. 2, pp. 281–308, 1957.

[30] S. Karlin, *Total Positivity*, vol. 1. Stanford, CA, USA: Stanford Univ. Press, 1968.

[31] A. Dytso and H. V. Poor, "Estimation in Poisson noise: Properties of the conditional mean estimator," *IEEE Trans. Inf. Theory*, vol. 66, no. 7, pp. 4304–4323, Jul. 2020.

[32] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and conditional mean estimation in Poisson channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1837–1849, May 2008.

[33] R. Atar and T. Weissman, "Mutual information, relative entropy, and estimation in the Poisson channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1302–1318, Mar. 2012.

[34] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 4, pp. 460–473, Jul. 1972.

[35] C.-I. Chang and L. D. Davisson, "On calculating the capacity of an infinite-input finite (infinite)-output channel," *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 1004–1010, Sep. 1988.

[36] J. Huang and S. P. Meyn, "Characterization and computation of optimal distributions for channel coding," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2336–2351, Jul. 2005.

[37] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[38] L. Barletta and A. Dytso, "Poisson noise channel with dark current: Numerical computation of the optimal input distribution," 2021. [Online]. Available: https://web.njit.edu/ ad94/PoissonNumerical.pdf

[39] I. Sason and S. Verdú, "Arimoto–Rényi conditional entropy and Bayesian m-ary hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 64, no. 1, pp. 4–25, Jan. 2017.

[40] T. M. Apostol, *Introduction to Analytic Number Theory*. New York, NY, USA: Springer, 1998.

[41] L. Baladová, "Minimum of average conditional entropy for given minimum probability of error," *Kybernetika*, vol. 2, no. 5, pp. 416–422, 1966.

[42] J. Cohen, J. H. Kempermann, and G. Zbaganu, *Comparisons of Stochastic Matrices With Applications in Information Theory, Statistics, Economics and Population*. New York, NY, USA: Springer, 1998.

[43] J. A. Adell and A. Lekuona, "Sharp estimates in signed Poisson approximation of Poisson mixtures," *Bernoulli*, vol. 11, no. 1, pp. 47–65, Jan. 2005.

[44] W. Wang and M. Á. Carreira-Perpiñán, "Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application," 2013, *arXiv:1309.1541*. [Online]. Available: http://arxiv.org/abs/1309.1541

[45] J. M. Borwein and O.-Y. Chan, "Uniform bounds for the complementary incomplete gamma function," *Math. Inequalities Appl.*, vol. 12, no. 1, pp. 115–121, 2009.

[46] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2017.