Non-Asymptotic Capacity Upper Bounds for the Discrete-Time Poisson Channel with Positive Dark Current

Mahdi Cheraghchi and João Ribeiro

***Abstract*—We derive improved and easily computable upper bounds on the capacity of the discrete-time Poisson channel under an average-power constraint and an arbitrary constant dark current term. This is accomplished by combining a general convex duality framework with a modified version of the digamma distribution considered in previous work of the authors.**

***Index Terms*—Optical communication, Discrete-time Poisson channel, Channel capacity, Dark current.**

I. Introduction

We consider a well-studied model of direct detection optical communication systems restricted to piecewise constant pulse-amplitude modulation (PAM) transmission techniques [1]. The transmitter is equipped with a photon-emitting source whose time-dependent piecewise constant intensity $X(t)$ can be modulated, and the receiver records the arrival times of photons, which follows a Poisson process with rate $X(t) + \lambda$ where $\lambda \geq 0$ is a *dark current* parameter modelling background interference. With practicality in mind, we are interested in determining the optimal transmission rate of this channel under average- and/or peak-power constraints on the input. As shown by Shamai [1], this is equivalent to determining the constrained capacity of the *Discrete-Time Poisson* (DTP) channel with dark current $\lambda \geq 0$, a memoryless channel which on input $x \in \mathbb{R}_0^+$ outputs Y_x following a Poisson distribution with mean $\lambda + x$, which we denote by $\mathsf{Poi}_{\lambda+x}$. Namely, it holds that

$$Y_x(y) = \mathsf{Poi}_{\lambda+x}(y) = e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}, \quad y = 0, 1, 2, \dots,$$

where $Y_x(y)$ denotes the probability that the output of the DTP channel on input x equals y. We are thus interested in the capacity of the DTP channel with dark current λ under an average- and/or peak-power constraint, given by

$$C(\lambda, \mu, A) = \sup_{X:\mathbb{E}[X]\leq\mu, 0\leq X\leq A} I(X; Y_X),$$

where Y_X denotes the output distribution of the DTP channel with dark current λ and input distribution X, and the supremum is taken over all distributions X supported on the non-negative real numbers and satisfying the given constraints. Throughout this work, we measure capacity in nats/channel use. For simplicity, when $A = \infty$, we denote the corresponding capacity of the DTP channel by $C(\lambda, \mu)$, and furthermore when $\lambda = 0$ we denote the corresponding capacity of the DTP channel by $C(\mu)$. For every λ, μ, and A, we have the chain of inequalities $C(\mu) \geq C(\lambda, \mu) \geq C(\lambda, \mu, A)$.

Remarkably, although the capacity of the Poisson channel without the restriction to PAM has been known for several decades [2], the exact value of $C(\lambda, \mu, A)$ remains unknown for any non-trivial choice of parameters, and we only have some upper and lower bounds on this quantity along with some asymptotic results. We discuss these in detail in Section I-A. Notably, whenever the average-power constraint μ is neither very small nor very large, and whenever the dark current λ is not very large compared to μ, the best known analytical upper bound on $C(\lambda, \mu)$ is actually an upper bound on $C(\mu)$.

A. Previous Work

Most previous work has focused on asymptotic settings where $\mu \to 0$ or $\mu \to \infty$, although there exist some capacity bounds applicable to non-asymptotic settings.

Brady and Verdú [3] were the first to study the asymptotic capacity of the DTP channel when $\mu \to \infty$. They derived bounds on $C(\lambda, \mu)$ when $\mu \to \infty$ and the ratio μ/λ stays constant. A characterization of the asymptotic behavior of $C(\lambda, \mu)$ when $\mu \to \infty$ and λ is constant was later obtained by Martinez [4] and Lapidoth and Moser [5]. Overall, the best upper bound on $C(\mu)$ for any μ outside the asymptotic regime $\mu \to 0$ was obtained by the authors in [6], improving on a previous upper bound of Martinez [4], and is given by

$$C(\mu) \leq \mu \ln\left(\frac{1 + \left(1 + e^{1+\gamma}\right)\mu + 2\mu^2}{e^{1+\gamma}\mu + 2\mu^2}\right) + \ln\left(1 + \frac{1}{\sqrt{2e}}\left(\sqrt{\frac{1 + (1 + e^{1+\gamma})\mu + 2\mu^2}{1 + \mu}} - 1\right)\right), \quad (1)$$

where $\gamma \approx 0.5772$ is the Euler-Mascheroni constant.

The setting where $\mu \to 0$ was first considered by Lapidoth, Shapiro, Venkatesan, and Wang [7], who determined the first order behavior of $C(\lambda, \mu, A)$ in this setting. In order to derive their results, the authors [7, Expression (114)] derive an explicit non-asymptotic upper bound on $C(\lambda, \mu)$ given by

$$C(\lambda, \mu) \leq F_1(\lambda, \mu) + F_2(\lambda, \mu) + F_3(\lambda, \mu) \quad (2)$$

M. Cheraghchi is with the EECS Department, University of Michigan – Ann Arbor, MI 48109, USA. Email: mahdich@umich.edu.

J. Ribeiro is with the CS Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Email: jlourenc@cs.cmu.edu.

This material is based upon work supported by the National Science Foundation under Grant No. CCF-2006455.

with F_1, F_2, and F_3 defined as

$$F_1(\lambda,\mu) = \left(\eta\ln\eta + \frac{1}{12\eta} + \frac{1}{2}\ln(2\pi\eta) + \lambda - \eta\ln\lambda - \ln(1-p)\right)e^{\eta+\eta\ln\lambda-\eta\ln\eta+\frac{\mu}{\eta-\sqrt{\eta}-\lambda}},$$

$$F_2(\lambda,\mu) = \max\left(0, \left(1+\ln(1/p)+\ln\lambda\right)\cdot \left(\mu + \frac{\lambda\mu}{\eta-\sqrt{\eta}-\lambda} + \lambda e^{\eta-1-\lambda+(\eta-1)\ln\lambda-(\eta-1)\ln(\eta-1)}\right)\right),$$

and

$$F_3(\lambda,\mu) = \mu\left(1+\frac{\lambda}{\eta-\lambda}\right)\max\left(0,\ln(1/\lambda)\right) + \mu\frac{\eta\ln(\eta/\lambda)}{\eta-\lambda},$$

where η is a free parameter that must be larger than some non-explicit constant $C_\lambda > 0$ depending on λ, and $p \in (0,1)$ is a free parameter. By inspection of [7, Section IV-B], it must at the very least be the case that $\eta - \sqrt{\eta} > \lambda$ for the bound to hold. Therefore, the upper bound in (2) is always significantly larger than

$$\mu(1+\max(0,1+\ln\lambda)+\max(0,\ln(1/\lambda))). \tag{3}$$

We will use this conservative underestimate of (2) when comparing the different bounds in Section III. Later, Wang and Wornell [8] determined the second-order asymptotics of the capacity of the DTP channel under an average-power constraint μ with dark current $\lambda = c\mu$ for an arbitrary constant $c \geq 0$.

The best capacity lower bounds on $C(\mu)$ and $C(\lambda,\mu)$ were obtained by Martinez [4], [9] and Cao, Hranilovic, and Chen [10], respectively, by considering the rate achievable by a gamma input distribution with arbitrary shape v for the DTP channel. We will compare our improved capacity upper bounds with the lower bound from [10, Expression (8)] in Section III.[1]

B. Notation

We denote random variables by uppercase letters such as X, Y, and Z. For a discrete random variable X, we denote by $X(x)$ the probability that X equals x, and the expected value of X is denoted by $\mathbb{E}[X]$. Moreover, we denote the Shannon entropy of a discrete random variable X by $H(X)$ and the Kullback-Leibler divergence between two discrete random variables X and Y by $D_{\mathsf{KL}}(X\|Y)$. The natural logarithm is denoted by ln, and we measure capacity in nats/channel use.

II. The Main Result

In this paper, we prove the following theorem, which yields a significantly improved non-asymptotic upper bound on the capacity of the DTP channel with constant dark current λ and average-power constraint μ in non-asymptotic regimes of μ.

[1] It should be noted that the formula in [10, Expression (8)] has a typo: In their notation, the first $v\log(\epsilon+\lambda)$ term should instead read $v\log(\epsilon+v)$.

Theorem 1: For every $\mu, \lambda \geq 0$ we have

$$C(\lambda,\mu) \leq \ln\left(\delta_\lambda + \frac{1}{\sqrt{2e}}\left(\frac{1}{\sqrt{1-q_{\lambda,\mu}}}-1\right)\right) - (\mu+\lambda)\ln q_{\lambda,\mu}, \tag{4}$$

with $\delta_\lambda = \exp(-\lambda e^\lambda E_1(\lambda))$, where $E_1(z) = \int_1^\infty \frac{e^{-zt}}{t}dt$ is the exponential integral function (with the convention that $0E_1(0)=0$), and $q_{\lambda,\mu}$ defined as

$$q_{\lambda,\mu} = 1 - \frac{1}{1+e^{1+\gamma}(\mu+\lambda)+\frac{2-e^{1+\gamma}}{1+\mu+\lambda}(\mu+\lambda)^2},$$

where $\gamma \approx 0.5772$ is the Euler-Mascheroni constant.

Remark 1: Although the upper bound from Theorem 1 does not have a closed-form expression since it features the exponential integral function (which is nevertheless easy to compute numerically), we can derive a good closed-form and elementary upper bound on $C(\lambda,\mu)$ by noting that [11, Section 5.1.20] and [12, Theorem 2] give the lower bound

$$e^\lambda E_1(\lambda) \geq \max\left(\frac{1}{2}\ln(1+2/\lambda), -e^\lambda\ln\left(1-e^{-\lambda e^\gamma}\right)\right)$$

for all $\lambda > 0$, and thus

$$\delta_\lambda \leq \min\left((1+2/\lambda)^{-\lambda/2}, (1-e^{-\lambda e^\gamma})^{\lambda e^\lambda}\right). \tag{5}$$

When λ is small, this elementary upper bound sharply approaches the upper bound from Theorem 1, and overall it improves on previously known bounds whenever μ is not small compared to λ.

As with most previous capacity upper bounds for the DTP channel, we derive Theorem 1 with the help of a general convex duality framework, which we state below in a specialized form for the DTP channel. This framework was originally derived in [13] and has also been used to derive the state-of-the-art upper bound on $C(\mu)$ [6]. As discussed in [6], it is equivalent to other existing frameworks (e.g., see [14], [4]).

Lemma 1 ([13], [6]): Suppose that there exist constants $a \in \mathbb{R}_0^+$, $b \in \mathbb{R}$, and a distribution Y supported on $\{0,1,2,\dots\}$ such that

$$D_{\mathsf{KL}}(\mathsf{Poi}_z\|Y) \leq az+b \tag{6}$$

for all $z \geq \lambda$. Then, we have

$$C(\lambda,\mu) \leq a(\mu+\lambda)+b$$

for all $\lambda,\mu \geq 0$.

An important quantity related to Lemma 1 is the *KL-gap* of the distribution Y with respect to the line $az+b$, which quantifies the sharpness with which the constraint (6) is satisfied, and is defined as

$$\Delta(z) = az+b-D_{\mathsf{KL}}(\mathsf{Poi}_z\|Y).$$

Previous applications of Lemma 1 [13], [6], [15] suggest that designing candidate distributions Y with overall smaller associated KL-gap leads to sharper capacity upper bounds. Although there is no known rigorous link between these two properties, we find this to be an effective design principle and we follow it in this work as well.

A. The Modified Digamma Distribution

The upper bounds presented are obtained by modifying the family of *digamma distributions* $Y^{(q)}$ studied in [13], [6] with $q \in (0,1)$, whose probability mass function satisfies

$$Y^{(q)}(y) = y_0 q^y \frac{e^{g(y)-y}}{y!}, \quad y = 0,1,2,\ldots,$$

where $g(0) = 0$ and $g(y) = y\psi(y)$ otherwise, with $\psi(\cdot)$ denoting the digamma function (which for integer $y > 0$ satisfies $\psi(y) = -\gamma + \sum_{i=1}^{y-1} 1/i$ [11, Section 6.3.2]), and y_0 is the normalizing factor. Using well-known properties of some special functions, it is possible to compute $D_{\mathsf{KL}}(\mathsf{Poi}_z \| Y^{(q)})$ exactly [13], [6], yielding

$$D_{\mathsf{KL}}(\mathsf{Poi}_z \| Y^{(q)}) = -\ln y_0 - z \ln q - z E_1(z)$$

for all $z \geq 0$ and $q \in (0,1)$, where we recall $E_1(\cdot)$ is the exponential integral (with the convention that $0E_1(0) = 0$). Therefore, every digamma distribution $Y^{(q)}$ has KL-gap $\Delta(z) = zE_1(z)$.

Our modification consists in changing the value of the digamma distribution $Y^{(q)}$ at $y = 0$ as a function of the dark current λ and renormalizing the distribution. The motivation behind this is that a careful choice of this function leads to a significantly smaller KL-gap overall. Based on the discussion above, we expect that this will lead to improved, easy-to-compute capacity upper bounds, which we show is the case. More precisely, for $\delta \in (0,1]$ and $q \in (0,1)$ we consider the *modified digamma distribution* $Y_\delta^{(q)}$ defined as

$$Y_\delta^{(q)}(y) = \begin{cases} \alpha\delta, & \text{if } y = 0, \\ \alpha Y^{(q)}(y)/y_0, & \text{if } y > 0, \end{cases}$$

where α is the new normalizing factor satisfying $1/\alpha = 1/y_0 + \delta - 1$, where we have used the fact that $Y^{(q)}(0) = y_0$. An analogous approach was used by the authors in [15] to derive improved capacity upper bounds on channels with synchronization errors. Moreover, we note that a related approach was employed by Martinez [4] in the special case where $\lambda = 0$ to improve the upper bound given by his candidate distribution, which originally had KL-gap bounded well away from 0 everywhere. However, no rigorous proof is given in [4] to show that this approach indeed works in that special case, with only numerical evidence being presented.

We begin by computing $D_{\mathsf{KL}}(\mathsf{Poi}_z \| Y_\delta^{(q)})$ for $q \in (0,1)$ and $\delta \in (0,1]$, which has a simple expression in terms of the original KL-gap $\Delta(z) = zE_1(z)$ of the digamma distribution $Y^{(q)}$. We have

$$\begin{aligned} D_{\mathsf{KL}}(\mathsf{Poi}_z \| Y_\delta^{(q)}) &= -H(\mathsf{Poi}_z) - \sum_{y=0}^{\infty} \mathsf{Poi}_z(y) \ln Y_\delta^{(q)}(y) \\ &= -\ln\alpha - z\ln q - e^{-z}\ln\delta - H(\mathsf{Poi}_z) \\ &\quad + \mathbb{E}_{y\sim\mathsf{Poi}_z}[\ln(y!) + y - g(y)] \\ &= -\ln\alpha - z\ln q - e^{-z}\ln\delta - zE_1(z). \quad (7) \end{aligned}$$

The last equality follows from the fact that

$$zE_1(z) = H(\mathsf{Poi}_z) - \mathbb{E}_{y\sim\mathsf{Poi}_z}[\ln(y!) + y - g(y)],$$

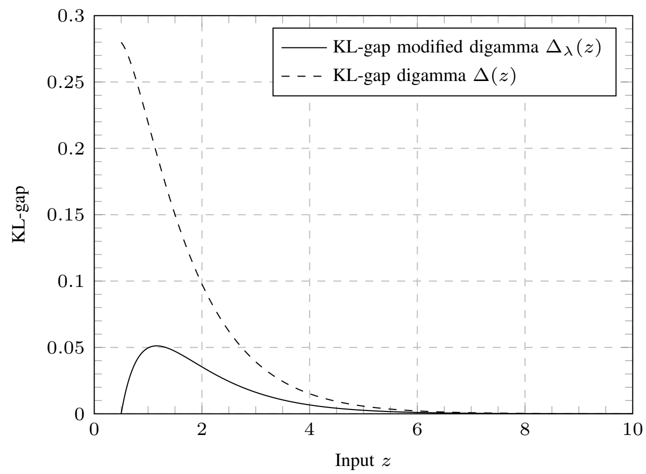

Fig. 1. Comparison between the KL-gap $\Delta(z)$ of the digamma distribution $Y^{(q)}$ and the KL-gap $\Delta_\lambda(z)$ of the modified digamma distribution $Y_\lambda^{(q)}$ for $z \in [\lambda, 10]$ when $\lambda = 0.5$.

as shown in [13], [6]. Given $\lambda \geq 0$, we consider the choice

$$\delta_\lambda = \exp(-\lambda e^\lambda E_1(\lambda)). \quad (8)$$

Consequently, by defining $Y_\lambda^{(q)} = Y_{\delta_\lambda}^{(q)}$ and using (7) we have

$$D_{\mathsf{KL}}(\mathsf{Poi}_z \| Y_\lambda^{(q)}) = -\ln\alpha - z\ln q + \lambda e^{\lambda - z} E_1(\lambda) - zE_1(z). \quad (9)$$

We now claim that the following result holds.

Theorem 2: For every $z \geq \lambda$ and $q \in (0,1)$ we have

$$D_{\mathsf{KL}}(\mathsf{Poi}_z \| Y_\lambda^{(q)}) \leq -\ln\alpha - z\ln q,$$

with KL-gap Δ_λ satisfying

$$0 \leq \Delta_\lambda(z) = zE_1(z) - \lambda e^{\lambda - z} E_1(\lambda) < zE_1(z) = \Delta(z).$$

Note that $\Delta_\lambda(\lambda) = 0$ and $\Delta_\lambda(z) \to 0$ exponentially fast when $z \to \infty$ (in general, Δ_λ is always smaller than Δ). Theorem 2 and the observations above justify our choice of δ_λ in (8); With this choice, we obtain a new family of modified digamma distributions $Y_\lambda^{(q)}$ with KL-gap Δ_λ that is always smaller than the original KL-gap Δ of the digamma distributions. Moreover, the KL-gap Δ_λ equals 0 at $z = \lambda$ and is significantly smaller than Δ around $z = \lambda$. Figure 1 illustrates the change in the KL-gap. Given the above, intuitively we expect to obtain a sharper upper bound on $C(\lambda, \mu)$ using this family of modified distributions.

Theorem 2 is an immediate consequence of (9) and the following lemma.

Lemma 2: For every $z \geq \lambda$ we have

$$\Delta_\lambda(z) = zE_1(z) - e^{\lambda - z}\lambda E_1(\lambda) \geq 0.$$

Proof: Multiplying both sides of the inequality above by e^z, we conclude that the desired inequality holds provided we can show that $ze^zE_1(z) \geq \lambda e^\lambda E_1(\lambda)$ for all $z \geq \lambda$. Equivalently, we must show that the function $f(z) = ze^zE_1(z)$ is non-decreasing when $z > 0$. Note that we have

$$f'(z) = (1+z)e^zE_1(z) - 1$$

for every $z > 0$, and we proceed to show that $f'(z) \geq 0$ for all $z > 0$. This implies the desired result. According to [11, Section 5.1.20], we can lower bound $e^z E_1(z)$ as

$$e^z E_1(z) > \frac{1}{2}\ln(1+2/z)$$

for all $z > 0$. Therefore, in order to show that $f'(z) \geq 0$ it is enough to note that

$$\frac{1+z}{2}\cdot\ln(1+2/z) \geq \frac{1+z}{2}\cdot\frac{4/z}{2+2/z} = 1$$

for all $z > 0$, which follows from the fact that $\ln(1+x) \geq \frac{2x}{2+x}$ for all $x \geq 0$. ■

B. Proof of Theorem 1

In this section, we prove our main result (Theorem 1) with the help of Lemma 1 and Theorem 2. First, by combining Lemma 1 and Theorem 2 we conclude that

$$C(\lambda,\mu) \leq \inf_{q\in(0,1)}[-\ln\alpha - (\mu+\lambda)\ln q]. \tag{10}$$

To obtain Theorem 1 from (10), we upper bound the term $-\ln\alpha$ by an easy-to-compute expression in terms of λ and q, and then choose q appropriately as a function of λ and μ.

Recalling that $1/\alpha = 1/y_0 - 1 + \delta_\lambda$ and the upper bound $1/y_0 \leq 1 + \frac{1}{\sqrt{2e}}\left(\frac{1}{\sqrt{1-q}} - 1\right)$ from [13, Corollary 16], we conclude that

$$-\ln\alpha \leq \ln\left(\delta_\lambda + \frac{1}{\sqrt{2e}}\left(\frac{1}{\sqrt{1-q}} - 1\right)\right) \tag{11}$$

for every $q \in (0,1)$. It remains now to choose $q = q_{\lambda,\mu}$ appropriately. For the case where $\lambda > 0$, we consider the direct extension of the choice of q for $\lambda = 0$ from [6], yielding

$$q_{\lambda,\mu} = 1 - \frac{1}{1 + e^{1+\gamma}(\mu+\lambda) + \frac{2-e^{1+\gamma}}{1+\mu+\lambda}(\mu+\lambda)^2}. \tag{12}$$

Combining (10), (11), and (12) leads to Theorem 1.

III. Comparison between Bounds

We present a comparison between the upper bound (4) that we have derived via the modified digamma distribution and previously known bounds on $C(\lambda,\mu)$ in Figure 2. As can be observed, the upper bound (4) improves on previous upper bounds whenever μ is not small compared to λ (more concretely, when $\mu > 0.01$ in the case pictured), and the elementary upper bound obtained by replacing δ_λ with the upper bound from (5) is extremely close to (4). It is interesting to note that our upper bound (4) behaves like $\frac{1}{2}\ln\mu$ when $\mu \to \infty$ for any $\lambda \geq 0$, which is optimal [5]. On the other hand, the bound does not approach 0 when $\mu \to 0$.

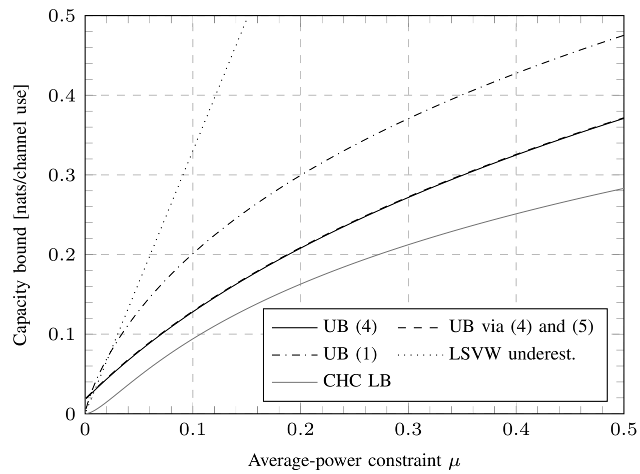

Fig. 2. Comparison between the upper bound (4), the elementary upper bound obtained by combining (4) and (5), the upper bound (1) for zero dark current, the upper bound (2) replaced by the LSVW underestimate (3), and the CHC lower bound from [10, Expression (8)] with $v = 0.05$ when $\lambda = 0.1$. UB stands for "upper bound", LB stands for "lower bound".

Acknowledgments

We thank Jun Chen for an insightful discussion about [16].

References

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

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

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

[4] A. Martinez, "Spectral efficiency of optical direct detection," *JOSA B*, vol. 24, no. 4, pp. 739–749, 2007.

[5] 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, 2009.

[6] 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, July 2019.

[7] 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, 2011.

[8] 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, 2014.

[9] A. Martinez, "A lower bound for the capacity of the discrete-time Poisson channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2009, pp. 2214–2215.

[10] J. Cao, S. Hranilovic, and J. Chen, "Lower bounds on the capacity of discrete-time Poisson channels with dark current," in *25th Biennial Symp. Commun. (QBSC), 2010*. IEEE, 2010, pp. 357–360.

[11] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover New York, 1965.

[12] H. Alzer, "On some inequalities for the incomplete gamma function," *Mathematics of Computation*, vol. 66, no. 218, pp. 771–778, 1997.

[13] M. Cheraghchi, "Capacity upper bounds for deletion-type channels," *J. ACM*, vol. 66, no. 2, pp. 9:1–9:79, Mar. 2019.

[14] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, 2003.

[15] M. Cheraghchi and J. Ribeiro, "Sharp analytical capacity upper bounds for sticky and related channels," *IEEE Trans. Inf. Theory*, vol. 65, no. 11, pp. 6950–6974, Nov 2019.

[16] 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, 2014.