Improved Capacity Upper Bounds for the Discrete-Time Poisson Channel

Mahdi Cheraghchi   João Ribeiro*

Abstract

We present new capacity upper bounds for the discrete-time Poisson channel with no dark current and an average-power constraint. These bounds are a simple consequence of techniques developed by one of the authors for the seemingly unrelated problem of upper bounding the capacity of binary deletion and repetition channels. Previously, the best known capacity upper bound in the regime where the average-power constraint does not approach zero was due to Martinez (JOSA B, 2007), which we re-derive as a special case of our framework. Furthermore, we instantiate our framework to obtain a closed-form bound that noticeably improves the result of Martinez everywhere.

1 Introduction

We study the capacity of the classical discrete-time Poisson (DTP) channel. Given an input $x \in \mathbb{R}^{\geq 0}$, the channel outputs a sample from Poisson distribution with mean $x + \lambda$, where $\lambda \geq 0$ is a channel parameter called the dark current. The DTP channel is motivated by applications in optical communication, involving a sender with a photon-emitting source and a receiver that observes the arrived photons (some of which may not have originated in the sender's source, hence the dark current parameter) [SS90].

The capacity of the DTP channel is infinite if there are no constraints on the input distributions. For this reason, a power constraint should be imposed on the input distribution. The most typical choice, that we consider in this work, is an average-power constraint $\mu \in \mathbb{R}^{\geq 0}$, under which only input distributions X satisfying $\mathbb{E}[X] \leq \mu$ are allowed. Several works also consider the case where a peak-power constraint is imposed on X, i.e., $X \leq A$ for some fixed $A \in \mathbb{R}^{>0}$ with probability 1 (e.g., [LM09, LSVW11, WW14, SSEL15, AAG+15]).

Currently, no expression for the capacity of the DTP channel under an average-power constraint is known. Consequently, there has been considerable interest in obtaining sharp bounds and in determining the asymptotic behavior of the DTP channel capacity in several settings, and in investigating properties of capacity-achieving distributions. We focus on upper bounds for the capacity of the DTP channel with $\lambda = 0$ under an average-power constraint μ. Note that any such upper bound is also a capacity upper bound for the DTP channel with $\lambda > 0$, as such a channel can be simulated from the DTP channel with $\lambda = 0$ by having the receiver add an independent Poisson random variable with parameter λ to the output.

1.1 Previous work

The two main regimes for studying the asymptotic behavior of the DTP channel capacity are when $\mu \to 0$ and $\mu \to \infty$. Brady and Verdú [BV90] studied the asymptotic behavior of the capacity under an average-power constraint μ when $\mu \to \infty$ and μ/λ is kept fixed. Later, Lapidoth and Moser [LM09] studied the same problem when λ is constant, with and without an additional peak-power constraint. When $\mu \to 0$, Lapidoth et al. [LSVW11] determined the first-order asymptotic behavior of the capacity when μ goes to zero, both when μ/λ is kept constant and when λ is fixed, with and without a peak-power constraint. Later, Wang and Wornell [WW14] improved their result when μ/λ is constant.

Obtaining capacity upper bounds for the DTP channel has been a major subject of interest. Explicit asymptotic capacity upper bounds for the DTP channel under an average-power constraint can be found

---

*Department of Computing, Imperial College London, UK. Emails: {m.cheraghchi, j.lourenco-ribeiro17}@imperial.ac.uk.

in [LM09, LSVW11, WW14]. The current best non-asymptotic upper bound, which is in fact the best capacity upper bound outside the limiting case $\mu \to 0$, was derived by Martinez [Mar07]. However, its proof contains a small gap, as mentioned in [LM09], and is not considered completely rigorous. A more detailed discussion of these upper bounds and of the asymptotic behavior of the capacity can be found in Section 3.

While we focus on capacity upper bounds, we mention that explicit (asymptotic and non-asymptotic) capacity lower bounds for several settings have been derived in [Mar07, LM09, CHC10, LSVW11, WW14, YZWD14]. Other aspects and settings of the DTP channel have also received attention recently. Cao, Hranilovic, and Chen [CHC14a, CHC14b] studied the shape of capacity-achieving distributions for the DTP channel, a topic which was first studied by Shamai [SS90]. A generalization of the DTP channel was studied by Aminian et al. [AAG+15], where a capacity upper bound in the presence of both average- and peak-power constraints is also given. Sutter et al. [SSEL15] studied numerical algorithms for approximating the capacity of the DTP channel in the presence of both average- and peak-power constraints, and obtain sharp capacity bounds in this setting.

1.2 Our contributions and techniques

In this work, we derive improved capacity upper bounds for the DTP channel with $\lambda = 0$ under an average-power constraint. Our technique is based on a natural convex duality formulation developed by Cheraghchi [Che17] for the seemingly unrelated problem of upper bounding the capacity of binary deletion and repetition channels.

We show that the result of Martinez [Mar07] can be effortlessly obtained as an immediate special (suboptimal) case of our results, thus giving a remarkably simple and rigorous proof for this bound. Furthermore, we extract two improved bounds from our more general result (Theorem 5); one involving the minimization of a smooth convex function over $(0, 1)$, as well as a closed-form bound (Theorem 7). Both of these bounds are strictly and significantly tighter than the bound by Martinez for all $\mu > 0$. Thus, we obtain the current best capacity upper bounds for the DTP channel with $\lambda = 0$ under an average-power constraint μ for all values of μ outside the limiting case $\mu \to 0$. An additional feature of our results is that they are remarkably simple to derive.

The rest of the article is organized as follows: In Section 2 we introduce our notation. Further discussion of the best previously known bounds, along with the asymptotic behavior of the capacity when $\lambda = 0$, appear in Section 3. The duality-based framework and the derivation of our upper bounds (including the bound by Martinez as a special case) are presented in Section 4. Finally, we compare the bounds from Section 4 with those from Section 3 in Section 5.

2 Notation

We denote the capacity of the DTP channel with average-power constraint μ and $\lambda = 0$ by $C(\mu)$. We measure capacity in nats per channel use and denote the natural logarithm by log. The Kullback-Leibler divergence between X and Y is denoted by $D_{\mathsf{KL}}(X \| Y)$.

3 Previously known upper bounds and asymptotic results

In this section, we survey the best previously known capacity upper bounds and the known results on the asymptotic behavior of $C(\mu)$. The asymptotic regimes considered in the literature are when $\mu \to 0$ and $\mu \to \infty$.

In the small μ regime, Lapidoth et al. [LSVW11] showed that

$$\lim_{\mu \to 0} \frac{C(\mu)}{\mu \log(1/\mu)} = 1.$$

Moreover, they gave the following upper bound matching the asymptotic behavior [LSVW11, expression (86)],

$$C(\mu) \leq -\mu \log p - \log(1 - p) + \frac{\mu}{\beta} + \mu \cdot \max\left(0, \frac{1}{2} \log \beta + \log\left(\frac{\bar{\Gamma}(1/2, 1/\beta)}{\sqrt{\pi}} + \frac{1}{2\beta}\right)\right), \tag{1}$$

where $p \in (0, 1)$ and $\beta > 0$ are free constants, and $\bar{\Gamma}$ is the upper incomplete gamma function. It is easy to see that the optimal choice for p is $p = \frac{\mu}{1+\mu}$.

Later, Wang and Wornell [WW14] determined the higher-order asymptotic behavior of $C(\mu)$ in the small μ regime, where it was shown that

$$C(\mu) = \mu \log(1/\mu) - \mu \log \log(1/\mu) + O(\mu)$$

when $\mu \to 0$. This was previously noted by Chung, Guha, and Zheng [CGZ11], although they only proved the result for a more restricted set of input distributions (as mentioned in [WW14]). Wang and Wornell [WW14, expression (180)] gave an upper bound (valid for small enough μ) matching this asymptotic behavior; namely,

$$C(\mu) \leq \mu + \mu \log \log\left(\frac{1}{\mu}\right) + \log\left(\frac{1}{1-\mu}\right) + \mu \log\left(\frac{1}{1 - \frac{1}{\log(1/\mu)}}\right) + \mu \cdot \sup_{x \geq 0} \phi_\mu(x), \tag{2}$$

where $\phi_\mu(x) := \frac{1-e^{-x}}{x} \log\left(\frac{x}{\mu \log(1/\mu)}\right)$.

In the large μ regime, Lapidoth and Moser [LM09] showed that

$$\lim_{\mu \to \infty} \frac{C(\mu)}{\log \mu} = \frac{1}{2}.$$

The best upper bound in this regime (and, in fact, anywhere outside the asymptotic limit $\mu \to 0$) was derived by Martinez [Mar07, expression (10)] and is given by

$$C(\mu) \leq \left(\mu + \frac{1}{2}\right) \log\left(\mu + \frac{1}{2}\right) - \mu \log \mu - \frac{1}{2} + \log\left(1 + \frac{\sqrt{2e} - 1}{\sqrt{1 + 2\mu}}\right). \tag{3}$$

It holds that (3) attains the first-order asymptotic behavior of $C(\mu)$ both when $\mu \to 0$ and when $\mu \to \infty$, and is strictly better than (1) for all $\mu > 0$. However, as noted in [LM09], the proof in [Mar07] is not considered to be completely rigorous as it contains a gap (a certain equality is only shown numerically).

4 The proposed upper bounds

In this section, we derive new upper bounds on $C(\mu)$. While previous upper bounds are mostly based on duality results from [LM03], our derivation (although still duality based) follows from the application of a framework recently developed in [Che17] in the context of binary deletion-type channels.

4.1 The convex duality formulation

In this section, we give a high-level overview of our approach towards obtaining improved capacity upper bounds.

Given a channel Ch with input and output alphabets contained in $\mathbb{R}$ and channel law $P_{Y|X}$, we denote by Ch_μ the channel having the same input and output alphabets and channel law $P_{Y|X}$ with the additional constraint that only input distributions whose corresponding output distributions Y satisfy $\mathbb{E}[Y] = \mu$ are admissible. We call such channels *mean-limited*, and denote the mean-limited version of the DTP channel by DTP_μ. A main component of our proof is the following natural duality result proved in [Che17]:

Theorem 1 ([Che17, Theorem 1], adapted). *Let Ch be a channel with input alphabet $\mathbb{R}^{\geq 0}$, output alphabet $\mathbb{Z}^{\geq 0}$, and channel law $P_{Y|X}$. Suppose that there exist a random variable Y, supported on the output alphabet, and parameters $\nu_1, \nu_2 \in \mathbb{R}$ such that*

$$D_{\mathsf{KL}}(Y_x \| Y) \leq \nu_1 \mathbb{E}[Y_x] + \nu_2$$

for every $x \in \mathbb{R}^{\geq 0}$, where Y_x denotes the output of Ch when x is given as input. Then, we have

$$C(\mathsf{Ch}_\mu) \leq \nu_1 \mu + \nu_2$$

3

for every $\mu \geq 0$. Moreover, if Y is the channel output distribution associated to some input distribution X and

$$D_{\mathsf{KL}}(Y_x\|Y) = \nu_1\mathbb{E}[Y_x] + \nu_2$$

for all $x \in \mathsf{supp}(X)$, then

$$C(\mathsf{Ch}_\mu) = \nu_1\mu + \nu_2.$$

For the DTP channel with $\lambda = 0$, we wish to find a distribution Y and parameters $\nu_1, \nu_2 > 0$ such that

$$D(Y_x\|Y) \leq \nu_1\mathbb{E}[Y_x] + \nu_2 = \nu_1 x + \nu_2$$

for all $x \in \mathbb{R}^{\geq 0}$, and the inequality gap as small as possible. Using Theorem 1, we readily obtain an upper bound for $C(\mathsf{DTP}_\mu)$, and subsequently, for $C(\mu)$.

4.2 The digamma distribution

The result of Martinez [Mar07] follows the common approach of a convex duality formulation that leads to capacity upper bounds given an appropriate distribution on the channel output alphabet. Indeed, this is also the approach that we take. The dual distribution chosen by [Mar07] is a negative binomial distribution, which is a natural choice corresponding to a gamma distribution for the channel input. However, lengthy manipulations and certain adjustments are needed to obtain a closed-form capacity upper bound for this choice. We use a slightly different duality formulation, as discussed in 4.1. Furthermore, for the dual output distribution, we use a distribution that we call the "digamma distribution" and is designed by Cheraghchi [Che17] precisely for the purpose of use in the duality framework of [Che17]. This distribution asymptotically behaves like the negative binomial distribution. However, it is constructed to automatically yield provable capacity upper bounds without need for any further manipulations or adjustments. This is the key to our refined bounds and dramatically simplified analysis[1].

For a parameter $q \in (0, 1)$, the digamma distribution $Y^{(q)}$ is defined over non-negative integers with probability mass function

$$P_{Y^{(q)}}(y) := y_0\frac{\exp(y\psi(y))(q/e)^y}{y!}, \quad y = 0, 1, \ldots, \tag{4}$$

where y_0 is a normalizing factor depending on q (we omit this dependence in the notation for brevity), $\psi(y) = \frac{d}{dy}\log\Gamma(y)$ is the digamma function (Γ denotes the gamma function), and $y\psi(y)$ is understood to be zero for $y = 0$. For positive integers y, we have $\psi(y) = -\gamma + \sum_{k=1}^{y-1} 1/k$, where $\gamma \approx 0.5772$ is the Euler-Mascheroni constant.

We will need to control the normalizing factor y_0, which is accomplished by the following result.

Lemma 2 ([Che17, Corollary 16]). *We have*

$$\log\left(1 + \frac{2}{e^{1+\gamma}}\left(\frac{1}{\sqrt{1-q}} - 1\right)\right) \leq -\log y_0 \leq \log\left(1 + \frac{1}{\sqrt{2e}}\left(\frac{1}{\sqrt{1-q}} - 1\right)\right)$$

for all $q \in (0, 1)$.

We will also be using the fact that the digamma distribution is closely related to the negative binomial distribution. We denote the negative binomial distribution with number of failures r (note that r is not necessarily an integer) and success probability p by $\mathsf{NB}_{r,p}$.

Lemma 3 ([Che17, Corollary 16]). *For all $y \geq 1$ and $q \in (0, 1)$,*

$$\frac{2}{e^{1+\gamma}}\mathsf{NB}_{1/2,q}(y) \leq \frac{\sqrt{1-q}P_{Y^{(q)}}(y)}{y_0} \leq \frac{1}{\sqrt{2e}}\mathsf{NB}_{1/2,q}(y).$$

---

[1]We note that the duality framework of [Che17] uses standard techniques and the dual-feasibility of the digamma distribution also has a simple proof.

4.3 A first capacity upper bound

In this section, we use the digamma distribution and the approach outlined in Section 4.1 in order to derive an upper bound for $C(\mu)$.

The Kullback-Leibler divergence between Y_x and $Y^{(q)}$ has a simple form for every x. Let $g(y) := y\psi(y)$. We have

$$
\begin{aligned}
D_{\mathsf{KL}}\left(Y_x\big|\big|Y^{(q)}\right) &= \sum_{y=0}^{\infty} P_{Y_x}(y)\log\left(\frac{P_{Y_x}(y)}{P_{Y^{(q)}}(y)}\right) \\
&= \sum_{y=0}^{\infty} P_{Y_x}(y)(-\log y_0 + y(1-\log q) - g(y) - x + y\log x) \\
&= -\log y_0 - x\log q + x\log x - \mathbb{E}_{Y_x}[g(Y_x)].
\end{aligned}
\tag{5}
$$

The following result lets us control $\mathbb{E}_{Y_x}[g(Y_x)]$.

Lemma 4 ([Che17, Corollary 9]). *We have*

$$
\mathbb{E}_{Y_x}[g(Y_x)] = x(E_1(x) + \log x),
$$

where $E_1(x) = \int_1^{\infty} e^{-xt}dt/t$ is the exponential integral function.

Using the fact that $E_1(x) > 0$ for all $x > 0$, Lemma 4 and (5) allow us to conclude that

$$
D_{\mathsf{KL}}\left(Y_x\big|\big|Y^{(q)}\right) \le -x\log q - \log y_0
\tag{6}
$$

for all $x \ge 0$. Applying Theorem 1, we conclude that

$$
C(\mathsf{DTP}_{\mu}) \le -\mu\log q - \log y_0.
\tag{7}
$$

Recall that we wish to convert an upper bound on $C(\mathsf{DTP}_{\mu})$ into an upper bound for $C(\mu)$. In order to do this, first note that if X and Y are the input and output distributions of the DTP channel with $\lambda = 0$, then $\mathbb{E}[Y] = \mathbb{E}[X]$. Therefore, (7) also gives an upper bound on the capacity of the DTP channel with the constraint that $\mathbb{E}[X] = \mu$. Moreover, since the RHS of (7) increases with μ for every fixed $q \in (0,1)$, we conclude that the upper bound still holds if we only require that $\mathbb{E}[X] \le \mu$. Therefore, we have

$$
C(\mu) \le -\mu\log q - \log y_0.
\tag{8}
$$

Finally, noting that (8) holds for every $q \in (0,1)$, we obtain the following result. Recall that y_0 is not a constant, but rather a normalizing factor that depends on q.

Theorem 5. *For all $\mu \ge 0$, we have*

$$
C(\mu) \le \inf_{q \in (0,1)}(-\mu\log q - \log y_0).
\tag{9}
$$

4.4 Improved bounds in a systematic way

While Theorem 5 gives an upper bound on $C(\mu)$, it involves minimizing a rather complicated function (for which we do not know an exact closed-form expression) over a bounded interval. Since it is of interest to have easy-to-compute but high quality upper bounds, we consider instantiating the parameter q inside the infimum in (9) with a simple function of μ. In this section, we present a systematic way of deriving such a good choice $q(\mu)$. Finally, we upper bound $-\log(y_0)$ using Lemma 2, obtaining an improved closed-form bound for $C(\mu)$.

We determine a good choice $q(\mu)$ for the parameter q in (9) indirectly by instead choosing $q(\mu)$ so that the associated distribution $Y^{(q(\mu))}$ (given by (4)) has expected value close to μ. The reasons for this are the following: First, a capacity-achieving distribution X yields a channel output distribution Y satisfying $\mathbb{E}[Y] = \mathbb{E}[X] \le \mu$, and, under the natural assumption that $C(\mu)$ is strictly increasing with μ, we must

actually have $\mathbb{E}[X] = \mathbb{E}[Y] = \mu$. While a capacity-achieving X does not necessarily induce a digamma distribution over the output, the digamma distribution seems to be close to optimal, since the gap between the two expressions in (6) is $xE_1(x)$, which decays exponentially with x. Second, numerical computation suggests that the distribution Y induced by the choice of q that minimizes the bound from Theorem 5 has expected value very close (or equal) to μ. While determining a choice $q(\mu)$ such that $\mathbb{E}[Y^{(q(\mu))}]$ is very close to μ for all $\mu > 0$ may be complicated, we settle for a choice $q(\mu)$ that behaves well when $\mu \to 0$ and $\mu \to \infty$.

We begin by studying how $q(\mu)$ should behave when $\mu \to \infty$. In this case, we should have $q(\mu) \to 1$. Lemma 2 implies that

$$\frac{2}{e^{1+\gamma}} + \left(1 - \frac{2}{e^{1+\gamma}}\right)\sqrt{1-q} \le \frac{\sqrt{1-q}}{y_0} \le \frac{1}{\sqrt{2e}} + \left(1 - \frac{1}{\sqrt{2e}}\right)\sqrt{1-q},$$

from which we can conclude that

$$\frac{2}{e^{1+\gamma}} \le \frac{\sqrt{1-q}}{y_0} \le \frac{1}{\sqrt{2e}} + o(1) \tag{10}$$

when $q \to 1$. Combining (10) with Lemma 3, we obtain

$$\frac{2\sqrt{2e}}{e^{1+\gamma}} - o(1) \le \frac{P_{Y^{(q)}}(y)}{\mathsf{NB}_{1/2,q}(y)} \le \frac{e^{1+\gamma}}{2\sqrt{2e}} \approx 1.038$$

for $y = 0, 1, \ldots$, when $q \to 1$, and so we conclude that the digamma distribution is well-approximated by $\mathsf{NB}_{1/2,q}$ when q is close to 1.

Recall that we want a choice of $q(\mu)$ such that $Y^{(q(\mu))}$ has expected value as close as possible to μ in the large μ regime. The choice of q which ensures that $\mathbb{E}[\mathsf{NB}_{1/2,q}] = \mu$ is $q = \frac{2\mu}{1+2\mu}$, and so we want $q(\mu)$ to satisfy $q(\mu) = \frac{2\mu}{1+2\mu} + o\left(\frac{1}{\mu}\right)$ when $\mu \to \infty$.

One could set $q(\mu) = \frac{2\mu}{1+2\mu}$ to obtain the desired behavior above, but we will show that we can correct this choice in order to achieve $\mathbb{E}[Y^{(q(\mu))}] = \mu + o(\mu)$ when $\mu \to 0$. To make the derivation simpler, we will instead work with the quantity $\frac{1}{1-q(\mu)}$.

Consider a choice $q(\mu)$ satisfying

$$\frac{1}{1-q(\mu)} = 1 + \alpha\mu + \frac{\beta\mu^2}{1+\mu}$$

for some constants α and β. It is easy to see that $\frac{1}{1-q(\mu)}$ behaves as $1 + \alpha\mu + o(\mu)$ when $\mu \to 0$ and as $1 + (\alpha+\beta)\mu + o(\mu)$ when $\mu \to \infty$, which means we can set its asymptotic behavior in both the small and large μ regimes independently of each other. Moreover, setting $\alpha + \beta = 2$ leads to the desired behavior $q(\mu) = \frac{2\mu}{1+2\mu} + o\left(\frac{1}{\mu}\right)$ when $\mu \to \infty$.

We now proceed to choose α. As mentioned before, we determine the choice of α which ensures that $\mathbb{E}[Y^{(q(\mu))}] = \mu + o(\mu)$ when $\mu \to 0$. It is straightforward to see that, by construction, $q(\mu) = \alpha\mu + o(\mu)$ when $\mu \to 0$. We will need the following result.

Lemma 6. *We have* $\mathbb{E}[Y^{(q)}] = e^{-(1+\gamma)}q + o(q)$ *as* $q \to 0$.

Proof. Recall that $g(y) = y\psi(y)$, and note that

$$\frac{\mathbb{E}[Y^{(q)}]}{q} = y_0 e^{-(1+\gamma)} + y_0 \sum_{y=2}^{\infty} y \cdot \frac{e^{g(y)-y}q^{y-1}}{y!}. \tag{11}$$

It is easy to see that y_0 approaches 1 (using Lemma 2, for example) and the second term in the RHS of (11) vanishes when $q \to 0$, and so the result follows. $\square$

The remarks above, combined with Lemma 6, imply that $\mathbb{E}[Y^{(q(\mu))}] = e^{-(1+\gamma)}\alpha\mu + o(\mu)$ when $\mu \to 0$. Therefore, it suffices to set $\alpha = e^{1+\gamma}$ to have $\mathbb{E}[Y^{(q(\mu))}] = \mu + o(\mu)$ when $\mu \to 0$. Based on this, we set $q(\mu)$ to be such that

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

Combining the previous discussion, Theorem 5, and Lemma 2, we immediately obtain the following result.

6

Theorem 7. *We have*

$$C(\mu) \le \inf_{q \in (0,1)} f(\mu, q), \tag{13}$$

where $f(\mu, q) := -\mu \log q + \log\left(1 + \frac{1}{\sqrt{2e}}\left(\frac{1}{\sqrt{1-q}} - 1\right)\right)$.

In particular, by instantiating q with $q(\mu)$ defined in (12),

$$C(\mu) \le \mu \log\left(\frac{1 + (1 + e^{1+\gamma})\mu + 2\mu^2}{e^{1+\gamma}\mu + 2\mu^2}\right) + \log\left(1 + \frac{1}{\sqrt{2e}}\left(\sqrt{\frac{1 + (1 + e^{1+\gamma})\mu + 2\mu^2}{1 + \mu}} - 1\right)\right). \tag{14}$$

Note that $f(\mu, \cdot)$ is an elementary, smooth, and convex function for every fixed $\mu \ge 0$. Therefore, (13) can be easily approximated to any desired degree of accuracy.

Remark 8. *The reasons why we base our choice of $q(\mu)$ on* (9) *instead of* (13) *are the following: First, $q(\mu)$ is still close to optimal when used in* (13) *(see Figure 1). Second, the choice is independent of the upper bound on* $-\log y_0$, *and so can be reutilized if a better bound is used.*

4.5 The result of Martinez as a special case

In this section, we show that the bound by Martinez (3) can be quite easily recovered through our techniques. More precisely, we show that this bound is a special case of (13) with a sub-optimal choice of $q = 2\mu/(1+2\mu)$. In particular, this implies that (13) is strictly tighter than (3). In this section, we define $m(\mu)$ to be the RHS of (3). Recall that $f(\mu, q) = -\mu \log q + \log\left(1 + \frac{1}{\sqrt{2e}}\left(\frac{1}{\sqrt{1-q}} - 1\right)\right)$.

Theorem 9. *We have $f\left(\mu, \frac{2\mu}{1+2\mu}\right) = m(\mu)$ for all $\mu \ge 0$. Moreover, for every $\mu > 0$ there is $q_\mu^* \in (0,1)$ such that $f(\mu, q_\mu^*) < m(\mu)$.*

Proof. To prove the first statement of the theorem, we compute

$$m(\mu) - f\left(\mu, \frac{2\mu}{1+2\mu}\right)$$

$$= \left(\mu + \frac{1}{2}\right)\log\left(\mu + \frac{1}{2}\right) - \mu\log\mu - \frac{1}{2} + \log\left(1 + \frac{\sqrt{2e} - 1}{\sqrt{1 + 2\mu}}\right)$$

$$- \mu\log\left(\frac{1 + 2\mu}{2\mu}\right) - \log\left(1 + \frac{1}{\sqrt{2e}}\left(\sqrt{1 + 2\mu} - 1\right)\right)$$

$$= \frac{1}{2}\log\left(\mu + \frac{1}{2}\right) + \mu\left(\log\left(\mu + \frac{1}{2}\right) - \log\mu\right) - \frac{1}{2} + \log\left(\frac{\sqrt{1 + 2\mu} + \sqrt{2e} - 1}{\sqrt{1 + 2\mu}}\right)$$

$$- \mu\log\left(\frac{1 + 2\mu}{2\mu}\right) - \log\left(\frac{\sqrt{1 + 2\mu} + \sqrt{2e} - 1}{\sqrt{2e}}\right)$$

$$= \frac{1}{2}\log\left(\mu + \frac{1}{2}\right) - \frac{1}{2} + \log\left(\sqrt{\frac{2e}{1 + 2\mu}}\right) = 0.$$

To see that the second statement holds, it suffices to show that $\frac{\partial f}{\partial q}\left(\mu, \frac{2\mu}{1+2\mu}\right) \ne 0$ for all $\mu > 0$. We have

$$\frac{\partial f}{\partial q}(\mu, q) = -\frac{\mu}{q} + \frac{1}{2\left(\sqrt{2e} + \frac{1}{\sqrt{1-q}} - 1\right)(1-q)^{3/2}}. \tag{15}$$

Instantiating with $q = \frac{2\mu}{1+2\mu}$ yields

$$\frac{\partial f}{\partial q}\left(\mu, \frac{2\mu}{1+2\mu}\right) = -\frac{1 + 2\mu}{2} + \frac{(1 + 2\mu)^{3/2}}{2\left(\sqrt{2e} + \sqrt{1 + 2\mu} - 1\right)},$$

and now it is enough to note that

$$-(1 + 2\mu)\Big(\sqrt{2e} + \sqrt{1 + 2\mu} - 1\Big) + (1 + 2\mu)^{3/2}$$
$$= (1 + 2\mu)(1 - \sqrt{2e}) < 0$$

for all $\mu \geq 0$. $\qquad\square$

Finally, we show that the explicit choice $q(\mu)$ from Section 4.4 yields a strictly better upper bound than the Martinez bound (3).

Theorem 10. *We have $f(\mu, q(\mu)) < f\Big(\mu, \frac{2\mu}{1+2\mu}\Big) = m(\mu)$ for all $\mu > 0$.*

Proof. We first show that the statement holds whenever $\mu \geq 1.61$ and when μ is sufficiently small. Let $d(\mu) := m(\mu) - f(\mu, q(\mu))$.

We begin by noting that $q(\mu) > \frac{2\mu}{1+2\mu}$ for all $\mu > 0$. Then, due to the convexity of $f(\mu, \cdot)$, the statement holds for a given μ if $\frac{\partial f}{\partial q}(\mu, q(\mu)) < 0$ (recall (15)). This can be seen to hold for $\mu \geq 1.61$ by standard algebraic manipulations.

It can be easily verified that $d(\mu)$ behaves as

$$\left(1 + \gamma + \frac{1}{\sqrt{2e}} - \log 2 - \frac{e^{\frac{1}{2}+\gamma}}{2\sqrt{2}}\right)\mu + o(\mu) \approx 0.27\mu + o(\mu)$$

when $\mu \to 0$, which implies that $m(\mu) > f(\mu, q(\mu))$ for μ small enough.

For the remaining cases, one can use a computer algebra system to formally show that $d(\mu) > 0$. However, $d(\mu)$ is a complicated transcendental expression, and so cannot be fed directly to such a system. We avoid this issue in the following way: For $\mu \in [0.3, 1.61]$, we lower bound $d(\mu)$ by positive rational functions. This is done by replacing the logarithmic and square root terms of the expression by appropriate bounds which are themselves rational functions. Then, the question of whether $d(\mu) > 0$ is reduced to showing that a certain polynomial is positive in the given interval, which can be formally checked by a computer algebra system with little effort. For $\mu < 0.3$, our lower bounds for $d(\mu)$ are not good enough, and so we use the same reasoning to show that its second derivative $d''(\mu)$ is negative for $\mu < 0.3$. This implies that $d(\mu)$ is concave in $[0, 0.3]$, which, combined with the previous results, concludes the proof.

We do not explicitly write down the relevant lower bounds for $d(\mu)$ and upper bounds for the second derivative, as they feature high-degree polynomials. Instead, we describe the relevant bounds on the logarithmic and square root terms. Then, determining the corresponding rational function and formally checking whether it is positive/negative in a given interval is a straightforward (albeit tedious) process.

The expression $d(\mu)$ features logarithmic terms, along with square root terms of the form $\sqrt{1 + 2\mu}$ and $\sqrt{(1 + (1 + e^{1+\gamma})\mu + 2\mu^2)/(1 + \mu)}$ (recall (3) and (14)). For every $x \geq 1$, we have the bounds

$$\frac{(x - 1)(6 + 5(x - 1))}{2(3 + 2(x - 1))} \leq \log x \leq \frac{(x - 1)(x + 5)}{2x(2 + x)},$$

which can be found in [Top06]. Furthermore, we can upper bound $\sqrt{1 + 2\mu}$ and $\sqrt{(1 + (1 + e^{1+\gamma})\mu + 2\mu^2)/(1 + \mu)}$ by their Taylor series of degree 5 and 3, respectively, around $\mu = 1$. Replacing the relevant terms in $d(\mu)$ by their respective bounds described above yields a rational function lower bound which can be easily shown to be positive for $\mu \in [0.3, 1.61]$ by a standard computer algebra system.

For $\mu < 0.3$, the bounds above are not tight enough to show that $d(\mu)$ is positive, and so we focus on its second derivative $d''(\mu)$. However, $d''(\mu)$ cannot be fed directly to a computer algebra system either, and so we follow the same reasoning as before. The only terms of $d''(\mu)$ that need to be bounded are of the form $\sqrt{1 + 2\mu}$ and $\sqrt{(1 + (1 + e^{1+\gamma})\mu + 2\mu^2)/(1 + \mu)}$. It suffices to upper bound (resp. lower bound) $\sqrt{1 + 2\mu}$ by its Taylor series of degree 1 (resp. 2) around $\mu = 0$. However, we need to be more careful when dealing with $\sqrt{(1 + (1 + e^{1+\gamma})\mu + 2\mu^2)/(1 + \mu)}$. We split the interval $[0, 0.3]$ into two intervals: First, in $(0, 0.25]$ we lower bound it by its Taylor series of degree 2 around $\mu = 0$. Second, in $(0.25, 0.3]$ we lower bound it by its Taylor series of degree 2 around $\mu = 0.25$.
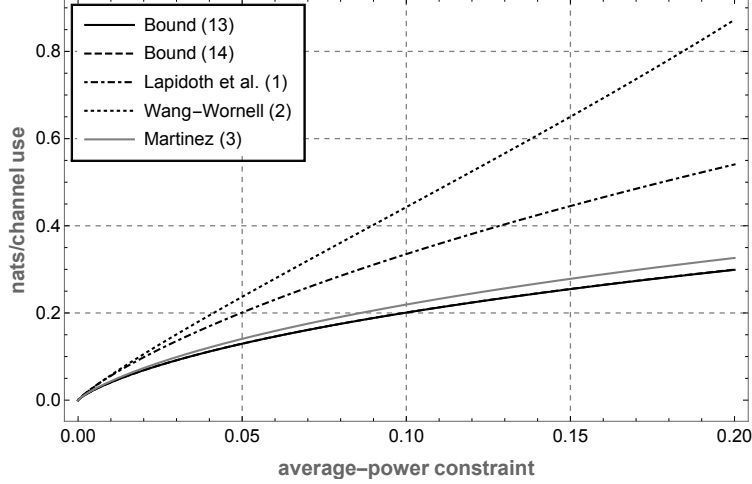
Figure 1: Comparison of known upper bounds for $\mu \in [0, 0.2]$.

Replacing the relevant terms of $d''(\mu)$ by their respective bounds, we obtain a negative rational function upper bounding $d''(\mu)$ in each of $(0, 0.25]$ and $(0.25, 0.3]$, which can be formally checked with a computer algebra system. This implies that $d(\mu)$ is concave in $(0, 0.3]$, and so, combined with the facts that $d(\mu) > 0$ for μ small enough and $d(\mu) > 0$ for $\mu \geq 0.3$, we conclude that $d(\mu) > 0$ for all $\mu > 0$. □

5   Comparison with previously known upper bounds

In this section, we compare the bounds from Theorem 7 with the previously known bounds described in Section 1. Moreover, we investigate the loss incurred by using (14) instead of (9).

Figure 1 showcases a plot comparing the bounds from Theorem 7 to the previously known upper bounds. The curve corresponding to the bound of Lapidoth et al. (1) is actually the plot of $\mu \log\left(\frac{1+\mu}{\mu}\right) + \log(1+\mu)$, which lower bounds the RHS of (1). There is a noticeable improvement over the Martinez bound (3) when μ is not very small, and one can see that (14) is very close to (13) (with significant overlap), which confirms that the choice $q(\mu)$ from Section 4.4 is close to optimal.

Due to the fact that our bounds are tighter than Martinez's bound, both of them satisfy the first-order asymptotic behavior of $C(\mu)$ when $\mu \to 0$ and when $\mu \to \infty$. However, they do not exhibit the correct second order asymptotic term when $\mu \to 0$. In fact, the second-order asymptotic term of our bounds when $\mu \to 0$ is $-O(\mu)$, while the correct term is $-\mu \log\log(1/\mu)$. For this reason, our bounds do not improve on the Wang-Wornell bound (2) when μ is sufficiently small (numerically, when $\mu < 10^{-6}$), while they noticeably improve on every previous bound when μ is not too small.

Figure 2 showcases the distance of the Martinez bound (3) to (9) and (14). The plotted curves have similar shapes and are close to each other, which shows that we do not lose much by replacing $-\log y_0$ by the upper bound of Lemma 2 and instantiating q with the sub-optimal explicit choice $q(\mu)$ from Section 4.4.

Figure 3 showcases the relative distance of the Martinez bound (3) to (9) and (14). Observe that, using (14), we obtain an improvement of up to 8.2% over (3), while we can get improvements close to 9.5% using (9). Once again, note that the two curves are close to each other and similar shape, reinforcing the fact that the loss incurred by using (14) instead of (9) is small.

References

[AAG+15] Gholamali Aminian, Hamidreza Arjmandi, Amin Gohari, Masoumeh Nasiri-Kenari, and Urbashi Mitra. Capacity of diffusion-based molecular communication networks over LTI-Poisson channels. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, 1(2):188–201, 2015.
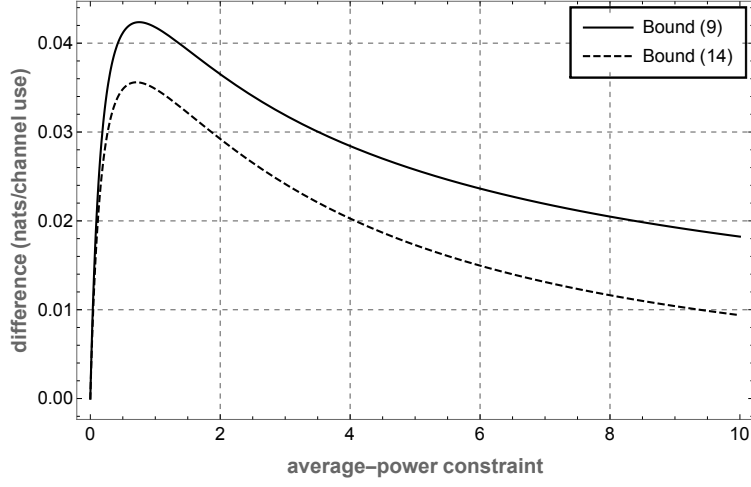
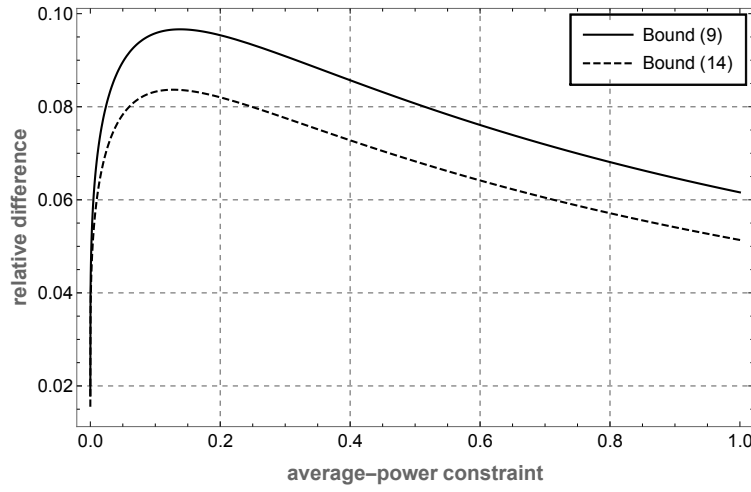Figure 2: Comparison of difference between (3) and (9), and between (3) and (14) for $\mu \in [0, 10]$.

Figure 3: Relative difference between (3) and (9), and between (3) and (14) for $\mu \in [0, 1]$.

[BV90]   David Brady and Sergio Verdú. The asymptotic capacity of the direct detection photon channel with a bandwidth constraint. In *28th Allerton Conference on Communication, Control and Computing*, pages 691–700, 1990.

[CGZ11]   Hye Won Chung, Saikat Guha, and Lizhong Zheng. On capacity of optical channels with coherent detection. In *49th Annual Allerton Conference on Communication, Control, and Computing, 2011*, pages 879–885. IEEE, 2011.

[CHC10]   Jihai Cao, Steve Hranilovic, and Jun Chen. Lower bounds on the capacity of discrete-time Poisson channels with dark current. In *25th Biennial Symposium on Communications (QBSC), 2010*, pages 357–360. IEEE, 2010.

[CHC14a]   Jihai Cao, Steve Hranilovic, and Jun Chen. Capacity-achieving distributions for the discrete-time Poisson channelpart i: General properties and numerical techniques. *IEEE Transactions on Communications*, 62(1):194–202, 2014.

[CHC14b]   Jihai Cao, Steve Hranilovic, and Jun Chen. Capacity-achieving distributions for the discrete-time Poisson channelpart ii: Binary inputs. *IEEE Transactions on Communications*, 62(1):203–213, 2014.

[Che17]   Mahdi Cheraghchi. Capacity upper bounds for deletion-type channels. *CoRR*, abs/1711.01630, 2017.

[LM03]   Amos Lapidoth and Stefan M Moser. Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels. *IEEE Transactions on Information Theory*, 49(10):2426–2467, 2003.

[LM09]   Amos Lapidoth and Stefan M Moser. On the capacity of the discrete-time Poisson channel. *IEEE Transactions on Information Theory*, 55(1):303–322, 2009.

[LSVW11]   Amos Lapidoth, Jeffrey H Shapiro, Vinodh Venkatesan, and Ligong Wang. The discrete-time Poisson channel at low input powers. *IEEE Transactions on Information Theory*, 57(6):3260–3272, 2011.

[Mar07]   Alfonso Martinez. Spectral efficiency of optical direct detection. *JOSA B*, 24(4):739–749, 2007.

[SS90]   S Shamai (Shitz). Capacity of a pulse amplitude modulated direct detection photon channel. *IEE Proceedings I (Communications, Speech and Vision)*, 137(6):424–430, 1990.

[SSEL15]   Tobias Sutter, David Sutter, Peyman Mohajerin Esfahani, and John Lygeros. Efficient approximation of channel capacities. *IEEE Transactions on Information Theory*, 61(4):1649–1666, 2015.

[Top06]   Flemming Topsøe. Some bounds for the logarithmic function. *Inequality theory and applications*, 4:137–151, 2006.

[WW14]   Ligong Wang and Gregory W Wornell. A refined analysis of the Poisson channel in the high-photon-efficiency regime. *IEEE Transactions on Information Theory*, 60(7):4299–4311, 2014.

[YZWD14]   Yingying Yu, Zaichen Zhang, Liang Wu, and Jian Dang. Lower bounds on the capacity for Poisson optical channel. In *Sixth International Conference on Wireless Communications and Signal Processing (WCSP), 2014*, pages 1–5. IEEE, 2014.