Capacity-Achieving Distributions for the Discrete-Time Poisson Channel–Part II: Binary Inputs

Jihai Cao, Steve Hranilovic, *Senior Member, IEEE*, and Jun Chen, *Member, IEEE*

***Abstract*—Discrete-time Poisson (DTP) channels exist in many scenarios including space laser communication systems which operate over long distances and which can be corrupted by reflected and scattered light. Through simulation, binary-input distributions have been observed to be optimal in many cases, however, little analytical work exists on conditions for optimality or the form of optimal signalling.**

In this second part, the general properties of Part I are extended to the case of DTP channels where binary-inputs are optimal. Necessary and sufficient conditions on the optimality of binary (i.e. two mass point) distributions are presented by leveraging the general properties of DTP capacity-achieving distributions. Closed-form expressions of the capacity-achieving distributions are derived in several important special cases including zero dark current and for high dark current. Numerical results are presented to elucidate the developed analytical work.

***Index Terms*—Discrete-time Poisson channel, capacity-achieving distributions, binary inputs.**

I. Introduction

FEW analytical results exist for optimal signalling on discrete-time Poisson (DTP) channels. The only analytically derived capacity-achieving distribution for the DTP channel is given by Shamai [1], where a binary input distribution is shown to be optimal in the case of zero dark current and background light and with only a peak power constraint ($A < 3.3679$). A majority of results derive analytical upper and lower bounds on the DTP channel under peak and average constraints [2], [3], [4], [5], however, no insights on optimal signalling design can be drawn from such approaches.

In this continuation of our earlier paper [6], we place particular emphasis on binary signalling in the DTP channel. This is motivated primarily by the need for a better understanding of the long range optical channels, which exists in long-range optical communications such as intersatellite laser links that operate over ranges of tens of thousands kilometres with limited transmit power. Simulation results show that, in the low power regime, the capacity-achieving distributions typically consist of two mass points [6], [7]. To gain a theoretical understanding of this phenomenon, using the general results of Part I, necessary and sufficient conditions on the optimality of binary distributions are derived and closed-form expressions of the capacity-achieving distributions are presented in several special cases.

The paper is organized as follows. For completeness, Sec. II briefly reviews the channel model. Necessary and sufficient conditions on the optimality of binary distributions are established in Sec. III. These conditions are used to derive a closed-form expression of the capacity-achieving distributions in several special cases, i.e., when there is no dark current and when the dark current is relatively large compared with the peak power constraint. To provide context for the theoretical results, some numerical results are presented in Sec. IV. Section V contains concluding remarks and directions for future work.

II. Channel Model

In the DTP channel, the data are represented by modulating the output intensity of an optical source, x [photons/second]. It is assumed that the intensity is fixed in discrete time intervals, which, without loss of generality, are taken to be of unity duration. At the receiver, a photon counting detector is used which outputs a measurement, y [photons], impinging on the detector in any time interval which is Poisson distributed. Note that the received count is corrupted by background light as well as dark current of the detector, modelled by a rate λ [photons/second]. The channel law of the DTP channel is thus,

$$P_{Y|X}(y|x) = \frac{(x+\lambda)^y}{y!}e^{-(x+\lambda)}, \quad x \in \mathbb{R}^+,\ y \in \mathbb{Z}^+. \tag{1}$$

Two fundamental constraints on the emitted intensity are imposed due to practical issues in implementation. An average intensity constraint

$$\mathbb{E}(X) \le \varepsilon, \qquad \text{[Average Power Constraint]}, \tag{2}$$

arises due to the limited energy storage on spacecraft while a peak optical intensity constraint

$$0 \le X \le A, \qquad \text{[Peak Power Constraint]}, \tag{3}$$

is required due to limitations in the laser and driving electronics. For the balance of the paper, it is assumed that $0 \le \varepsilon \le A$ for finite ε and A.

As illustrated by Shamai [1], the optimal input distribution is discrete with a finite number of mass points (with peak constraint). The input distribution is defined by the

Manuscript received February 19, 2013; revised August 2 and October 31, 2013. The editor coordinating the review of this paper and approving it for publication was E. Agrell.

The authors are with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada (e-mail: caojih@grads.ece.mcmaster.com, {hranilovic, junchen}@mcmaster.ca).

Digital Object Identifier 10.1109/TCOMM.2013.112513.130143

constellation of amplitude points $\psi_x = \{x_1, x_2, \dots, x_n\}$, $0 \le x_1 < x_2 < \dots < x_n \le A$, with accompanying probabilities $\psi_p = \{p_1, p_2, \cdots, p_n\}$ yielding the distribution,

$$dF_x = p_1\delta(x - x_1) + p_2\delta(x - x_2) + \dots + p_n\delta(x - x_n),$$

where $\delta(\cdot)$ denotes the Dirac impulse functional. By convention, the optimal inputs under peak constraint A and mean ε are $\psi_x^*(A, \varepsilon)$, $\psi_p^*(A, \varepsilon)$ and $F_x^*(A, \varepsilon)$. Following the convention in [8], define

$$\begin{aligned} i(x, F_x) &\triangleq -\sum_{y=0}^{\infty} P_{Y|X}(y|x) \log \frac{P_Y(y)}{P_{Y|X}(y|x)} \\ &= (x+\lambda)\log(x+\lambda) - x - \sum_{y=0}^{\infty} e^{-(x+\lambda)} \frac{(x+\lambda)^y}{y!} \\ &\quad \times \log\left(\sum_{j=1}^{n} p_j e^{-x_j}(x_j+\lambda)^y\right), \quad (4) \end{aligned}$$

where the mutual information induced by input F_x takes the form

$$I(F_x) = \int_0^A i(x, F_x) dF_x = \sum_{j=1}^{n} p_j i(x_j, F_x).$$

III. Binary Capacity-Achieving Distributions

Intuition from earlier simulation studies [6], [7] suggests that binary signalling is often optimal for low-power DTP channels. In order to further study the nature of these optimal distributions, for completeness, consider the following general properties of all capacity-achieving distributions for the DTP channel.

Corollary 1 (Mass point at zero [6])**.** *Under average and peak power constraints, the capacity-achieving distribution for the DTP channel always contains a point at zero, i.e.,* $0 \in \psi_x^*(A, \varepsilon)$.

Corollary 2 (Point at peak amplitude [6])**.** *The capacity-achieving distribution for a DTP channel with only peak power constraint always contains a mass point at* A*, i.e.,* $A \in \psi_x^*(A, A)$.

In this section, the general results of Corollary 1 and Corollary 2 are applied to binary signalling to develop necessary and sufficient conditions on the optimality of binary distributions and to derive analytical capacity-achieving distributions in several special cases.

Theorem 3 (Conditions for the capacity-achieving distribution being binary)**.** *For a DTP channel with peak and average power constraints, the capacity-achieving distribution* dF_x^* *is binary, i.e.* $|\psi_x^*(A, \varepsilon)| = 2$*, iff one of the following two conditions holds:*

1) *[Average power constraint active] There exists a* $B \in (0, A]$ *such that, for all* $x \in [0, A]$,

$$\frac{x}{B}\left(i(B, F_x^*) - i(0, F_x^*)\right) + i(0, F_x^*) - i(x, F_x^*) \ge 0, \quad (5)$$

and

$$i(B, F_x^*) - i(0, F_x^*) \ge 0, \quad (6)$$

where

$$dF_x^* = \left(1 - \frac{\varepsilon}{B}\right)\delta(x) + \frac{\varepsilon}{B}\delta(x - B). \quad (7)$$

2) *[Slack in average power constraint] For all* $x \in [0, A]$,

$$i(0, F_x^*) - i(x, F_x^*) \ge 0, \quad (8)$$

and

$$i(0, F_x^*) = i(A, F_x^*), \quad (9)$$

where

$$dF_x^* = (1 - \beta)\delta(x) + \beta\delta(x - A). \quad (10)$$

In this case, β *is the solution of*

$$\begin{aligned} &\sum_{y=0}^{\infty}\left(e^{-\lambda}\frac{\lambda^y}{y!} - e^{-(\lambda+A)}\frac{(\lambda+A)^y}{y!}\right) \\ &\qquad \times \log\left((1-\beta)\lambda^y + \beta e^{-A}(\lambda+A)^y\right) \\ &= \lambda\log\lambda - (A+\lambda)\log(A+\lambda) + A. \quad (11) \end{aligned}$$

Proof: This theorem follows from the KKT conditions [6, Theorem 5] and Corollaries 1 and 2. Note that Corollary 1 implies the existence of a mass point at 0. Now consider a binary input distribution $dF_x = (1 - \beta)\,\delta(x) + \beta\delta(x - B)$, $0 < B \le A$ and $0 < \beta < 1$. If dF_x satisfies the KKT conditions, then $dF_x^* = dF_x$. According to [6, Theorem 5] this can happen in two possible ways depending on whether the average power constraint is active:

- $\mathbb{E}_{F_x}\{X\} = \varepsilon$ implies $\mu \ge 0$, where μ is the Lagrange multiplier defined in [6, Eq. (10)]. This further implies $\beta = \varepsilon/B$. Equations (5) and (6) follow directly from [6, Theorem 5].
- $\mathbb{E}_{F_x}\{X\} < \varepsilon$ (i.e., there is slack in the average constraint) implies $\mu = 0$. Thus, if F_x is optimal, $B = A$ by Corollary 2. Now invoking [6, Theorem 5] yields (8) and (9). The optimal β is the solution of (9), which can be expanded to (11).

■

Theorem 3 can be used to determine whether a binary distribution is capacity-achieving. If none of the conditions are satisfied then the optimal distribution is not binary. In general, it is difficult to use Theorem 3 to obtain a closed-form solution for the binary capacity-achieving distribution for a general λ. In what follows, we consider several special cases in which more explicit results can be obtained.

A. When $\lambda = 0$

In the case of $\lambda = 0$, Theorem 3 can be simplified to yield some insights. This condition corresponds to the case when there is no background illumination falling on the receiver and the dark current is zero. The $\lambda = 0$ condition models cases when the satellite receiver aperture is pointed away from light scatters, has narrow band optical filters, and the photoreceiver is at low temperature making the dark current negligible [9, pp.96-7].

Theorem 4 (Conditions for $|\psi_x^*(A, \varepsilon)| = 2$ when $\lambda = 0$)**.** *When* $\lambda = 0$*, the capacity-achieving distribution is binary iff the values of* ε *and* A *satisfy one of the following conditions:*

1) *(Peak power constraint active only)* $0 < A < 3.3679$ *and*

$$\varepsilon > f_1(A) \triangleq \frac{A}{e^{\frac{A}{e^A-1}} + 1 - e^{-A}}. \tag{12}$$

In this case, the capacity achieving distribution is given by (10).

2) *(Peak and average power constraints active)*

$$\varepsilon \leq f_1(A), \tag{13}$$

and for all $x \in [0, A]$,

$$\eta(A, x) \geq 0, \tag{14}$$

where

$$\eta(A, x) = \left(\frac{x}{A} - 1 - \frac{x}{A}e^{-A} + e^{-x}\right) \log \frac{1 - \frac{\varepsilon}{A} + \frac{\varepsilon}{A}e^{-A}}{\frac{\varepsilon}{A}e^{-A}} - x - x \log \frac{x}{eA}. \tag{15}$$

In this case, the capacity-achieving distribution is given by (7) with $B = A$.

3) *(Average power constraint active only) There exists* $B \in (\varepsilon, A)$ *such that*

$$\varepsilon = f_2(B) \triangleq B \Big/ \left(e^{\frac{B^2+B}{e^B-B-1}} - e^{-B} + 1\right), \tag{16}$$

and for all $x \in [0, A]$,

$$\eta(B, x) \geq 0. \tag{17}$$

In this case, the capacity-achieving distribution is given by (7).

Remark: Cond. 1 of Theorem 4 corresponds to Cond. 2 of Theorem 3 while Cond. 2 and Cond. 3 of Theorem 4 correspond to Cond. 1 of Theorem 3.

Proof: Note that for $\lambda = 0$ and $dF_x^* = (1 - \beta)\delta(x) + \beta\delta(x - B)$ for some $\beta \in (0, 1)$ and $B \in (0, A]$, we have

$$i(x, F_x^*) = x \log \frac{x}{eB} - e^{-x} \log\left(1 - \beta + \beta e^{-B}\right) - (1 - e^{-x}) \log\left(\beta e^{-B}\right).$$

Also note that $\eta(A, x)$ in (15) is simply the *multiplier function* in [6, (6)] with $\lambda = 0$.

1) The first condition has been treated in [1] and corresponds to Cond. 2 in Theorem 3. With $\lambda = 0$ and with peak power constraint only, Shamai proved that the capacity-achieving distribution is (10) with $\beta^{-1} = e^{\frac{A}{e^A-1}} + 1 - e^{-A}$ when $A < 3.3679$ and correspondingly the average power is bounded as (12).
2) The second condition follows from Cond. 1 in Theorem 3 with $B = A$. Rearranging (6) with $\lambda = 0$ gives the first inequality while substituting $\lambda = 0$ into (5) yields the second.
3) The third condition also follows from Cond. 1 in Theorem 3, however, with $B < A$. In particular, the inequality corresponds to (5) in Theorem 3. Note that $\frac{x}{B}\left(i(B, F_x^*) - i(0, F_x^*)\right) + i(0, F_x^*) - i(x, F_x^*)$ must attain the minimum at $x = B$. As a consequence, we have

$$\frac{\partial}{\partial x}\left(\frac{x}{B}\left(i(B, F_x^*) - i(0, F_x^*)\right) + i(0, F_x^*) - i(x, F_x^*)\right)\Big|_{x=B} = 0,$$

which gives

$$\varepsilon = B \Big/ \left(e^{\frac{B^2+B}{e^B-B-1}} - e^{-B} + 1\right) = f(B). \tag{18}$$

It can be verified that (18) implies (6). ∎

B. When λ is large

Consider the case where there is intense shot-noise as a result of background illumination, i.e., λ large. This situation can physically arise in intersatellite communication links when high intensity background light from solar irradiation can cause large λ. When λ is large, high intensity Poisson approaches a Gaussian distribution. For this reason, the use of a Gaussian channel law in this regime is popular in the literature on optical wireless communications (see, e.g., [10], [11]).

The binary distribution which satisfies both the average and the peak power constraints with equality, i.e.,

$$dF_x^\dagger = \left(1 - \frac{\varepsilon}{A}\right)\delta(x) + \frac{\varepsilon}{A}\delta(x - A), \tag{19}$$

has been shown numerically to be capacity-achieving in the optical intensity channel with Gaussian noise in the low signal-to-noise regime [10]. Note that (19) is in fact the binary *maxentropic distribution* [6, Eq. (26)] introduced in Part I when $A \geq 2\varepsilon$. It has also been shown numerically that (19) is capacity-achieving in DTP channel when the input power is small enough. Here, Theorem 3 is leveraged to give a rigorous proof of the optimality of (19) when λ is large enough.

Theorem 5 (Capacity-achieving distribution when λ is large)**.** *For a DTP channel with* $\varepsilon < A/2$*, the capacity-achieving distribution* $dF_x^* = dF_x^\dagger$ *given by* (19) *when* λ *is sufficiently large.*

Proof: It suffices to show that the distribution $dF_x^\dagger$ in (19) satisfies (5) and (6) in Theorem 3 when λ is sufficiently large. By definition,

$$i(x, dF_x^\dagger) = (x + \lambda)\log(x + \lambda) - x - (x + \lambda)\log\lambda - \left(\sum_{y=0}^{\infty} e^{-(x+\lambda)} \frac{(x + \lambda)^y}{y!} \log\left(1 + re^{-A}(1 + A/\lambda)^y\right) - \log(1 + r)\right),$$

$$M(\mu, x, F_x^\dagger) = \frac{x}{A}\left(i(A, F_x^\dagger) - i(0, F_x^\dagger)\right) + i(0, F_x^\dagger) - i(x, F_x^\dagger),$$

where $r \triangleq \frac{\varepsilon/A}{1-\varepsilon/A}$ and $M(\cdot)$ is the multiplier function defined in [6, (6)] and corresponds to left-hand side of (5). Note that $r < 1$ when $\varepsilon < A/2$.

To stress the dependence of $i(x, F_x^\dagger)$ and $M(\mu, x, F_x^\dagger)$ on λ, denote them by $i(x, F_x^\dagger, \lambda)$ and $M(\mu, x, F_x^\dagger, \lambda)$, respectively. Let $z = \frac{1}{\lambda}$ and define $\hat{M}(\mu, x, F_x^\dagger, z) \triangleq M(\mu, x, F_x^\dagger, \lambda)$ and $\hat{M}'(\mu, x, F_x^\dagger, z) \triangleq \frac{\partial}{\partial x}M(\mu, x, F_x^\dagger, \lambda)$.

For any $\tilde{z} > 0$, it follows from Taylor's theorem that

$$\hat{M}(\mu, x, F_x^\dagger, \tilde{z}) = \hat{M}(\mu, x, F_x^\dagger, 0) + \frac{\partial}{\partial z}\hat{M}(\mu, x, F_x^\dagger, z)\Big|_{z=0}\tilde{z} + \frac{\frac{\partial^2}{\partial z^2}\hat{M}(\mu, x, F_x^\dagger, z)\Big|_{z=\theta_{\tilde{z}}}}{2}\tilde{z}^2,$$

$$\hat{M}'(\mu, x, F_x^\dagger, \tilde{z}) = \hat{M}'(\mu, x, F_x^\dagger, 0) + \frac{\partial}{\partial z}\hat{M}'(\mu, x, F_x^\dagger, z)\Big|_{z=0}\tilde{z} + \frac{\frac{\partial^2}{\partial z^2}\hat{M}'(\mu, x, F_x^\dagger, z)\Big|_{z=\theta'_{\tilde{z}}}}{2}\tilde{z}^2,$$

where $\theta_{\tilde{z}} \in [0, \tilde{z}]$ and $\theta'_{\tilde{z}} \in [0, \tilde{z}]$. It can be shown that

$$\hat{M}(\mu, x, F_x^\dagger, 0) = 0, \tag{20}$$

$$\frac{\partial}{\partial z}\hat{M}(\mu, x, F_x^\dagger, z)\Big|_{z=0} = \frac{x(A-x)}{2}, \tag{21}$$

$$\hat{M}'(\mu, x, F_x^\dagger, 0) = 0, \tag{22}$$

$$\frac{\partial}{\partial z}\hat{M}'(\mu, x, F_x^\dagger, z)\Big|_{z=0} = \frac{A}{2} - x, \tag{23}$$

where derivation details are presented in Appendices A–D respectively.

Let δ be an arbitrary number in the interval $(0, \frac{A}{2})$. Note that

$$\frac{\partial}{\partial z}\hat{M}(\mu, x, F_x^\dagger, z)\Big|_{z=0} \geq \frac{\delta(A-\delta)}{2}, \quad x \in [\delta, A-\delta],$$

$$\frac{\partial}{\partial z}\hat{M}'(\mu, x, F_x^\dagger, z)\Big|_{z=0} \geq \frac{A}{2} - \delta, \quad x \in [0, \delta],$$

$$\frac{\partial}{\partial z}\hat{M}'(\mu, x, F_x^\dagger, z)\Big|_{z=0} \leq -\frac{A}{2} + \delta, \quad x \in [A-\delta, A].$$

In view of the fact that $\frac{\partial^2}{\partial z^2}\hat{M}(\mu, x, F_x^\dagger, z)$ and $\frac{\partial^2}{\partial z^2}\hat{M}'(\mu, x, F_x^\dagger, z)$ are continuous functions of (x, z) over the compact set $[0, A] \times [0, 1]$, there exists a constant Γ such that $|\frac{\partial^2}{\partial z^2}\hat{M}(\mu, x, F_x^\dagger, z)| \leq \Gamma$ and $|\frac{\partial^2}{\partial z^2}\hat{M}'(\mu, x, F_x^\dagger, z)| \leq \Gamma$. As a consequence, one can readily show that

$$\hat{M}(\mu, x, F_x^\dagger, \tilde{z}) > 0, \quad x \in [\delta, A-\delta],$$

$$\hat{M}'(\mu, x, F_x^\dagger, \tilde{z}) > 0, \quad x \in [0, \delta],$$

$$\hat{M}'(\mu, x, F_x^\dagger, \tilde{z}) < 0, \quad x \in [A-\delta, A],$$

when $\tilde{z}$ is sufficiently close to 0. These inequalities together with the fact that $\hat{M}(\mu, 0, F_x^\dagger, \tilde{z}) = \hat{M}(\mu, A, F_x^\dagger, \tilde{z}) = 0$ imply that

$$\hat{M}(\mu, x, F_x^\dagger, \tilde{z}) \geq 0, \quad x \in [0, A],$$

when $\tilde{z}$ is sufficiently close to 0. This means (5) is satisfied for sufficiently large λ.

To complete the proof, it remains to verify that the distribution $F_x^\dagger$ given by (19) satisfies (6) in Theorem 3 when λ is sufficiently large. For $z = 1/\lambda$, define

$$\hat{\mu}(z) = \frac{i(A, F_x^\dagger, \lambda) - i(0, F_x^\dagger, \lambda)}{A}.$$

For any $\tilde{z} > 0$, it follows from Taylor's theorem that

$$\hat{\mu}(\tilde{z}) = \hat{\mu}(0) + \frac{\partial}{\partial z}\hat{\mu}(z)\Big|_{z=0}\tilde{z} + \frac{\frac{\partial^2}{\partial z^2}\hat{\mu}(z)\Big|_{z=\theta_{\tilde{z}}}}{2}\tilde{z}^2,$$

where $\theta_{\tilde{z}} \in [0, \tilde{z}]$. In view of (24) in Appendix A, $\hat{\mu}(0) = 0$. Moreover, it can be shown that

$$\begin{aligned}\frac{\partial}{\partial z}\hat{\mu}(z)\Big|_{z=0} &= \lim_{\lambda\to\infty}\left(\left(1+\frac{\lambda}{A}\right)\log\left(1+\frac{A}{\lambda}\right) - 1\right)\lambda + \frac{(\Lambda(0) - \Lambda(A))}{A} \\ &= \frac{A}{2} - \frac{r}{1+r}A \\ &> 0,\end{aligned}$$

where the last inequality is due to the fact that $r < 1$ and

$$\begin{aligned}\Lambda(x) &\triangleq \lim_{\lambda\to\infty}\left\{\sum_{y=0}^{\infty} e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\log\left(1 + re^{-A}(1+A/\lambda)^y\right) - \log(1+r)\right\}\lambda \\ &= \left(Ax - \frac{A^2}{2}\right)\frac{r}{1+r} + \frac{A^2}{2}\frac{r}{(1+r)^2}\end{aligned}$$

through (37) in Appendix B.

Since $\frac{\partial^2}{\partial z^2}\hat{\mu}(z)$ is a continuous function of z over the interval $[0, 1]$, there exists a constant Υ such that $|\frac{\partial^2}{\partial z^2}\hat{\mu}(z)| < \Upsilon$ for all $z \in [0, 1]$. Now one can readily see that $\hat{\mu}(\tilde{z})$ must be nonnegative when $\tilde{z}$ is sufficiently close to 0 (i.e, $i(A, F_x^\dagger, \lambda) - i(0, F_x^\dagger, \lambda) \geq 0$ when λ is sufficiently large). This completes the proof. ■

Theorem 5 shows that under the high background light condition the optimal signalling is binary and satisfies both peak and average power constraints.

IV. Numerical Examples

In order to provide some insight on the analytical results, several numerical examples are presented in this section.

Figure 1 considers DTP channels with $\lambda = 0$ and visualizes regions where the capacity-achieving distribution is binary. In addition, following the three conditions in Theorem 4, areas where each constraint is active are clearly shown. Note that by [6, Lemma 3], stretching the constellation at the input of a DTP channel increases mutual information. Thus, at least one of peak or average constraint must hold in the capacity-achieving distribution.

According to Cond. 1 of Theorem 4, the horizontal hatched region in Fig. 1 bounded between $A < 3.3679$, $\varepsilon = A$, and $f_1(A)$ in (12) corresponds to the set of all DTP channels which have binary capacity-achieving distributions where the average constraint is inactive. The vertical hatched region including the boundaries corresponds to the set of DTP channels in which Cond. 2 holds. Both of the constraints are active in this condition and the optimal distribution is binary. This region was plotted by sampling points (A, ε) and repeatedly verifying the inequalities (13) and (14). An interesting observation from Fig. 1 is that, for $\lambda = 0$, if the capacity-achieving distribution is binary with mass point at A, then $A < 3.3679$. This numerical observation complements Shamai's earlier analytical results [1] for DTP channels with only peak constraint. The solid filled region in Fig. 1 corresponds to the set of DTP channels ($\lambda = 0$) in which Cond. 3 of Theorem 4 is satisfied, i.e., inactive peak constraint. Notice that $A > B = f_2^{-1}(\varepsilon)$

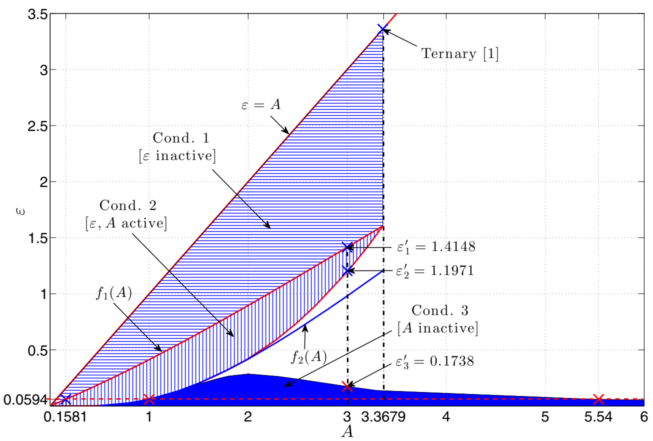


Fig. 1. Regions of (A, ε) where binary distributions are capacity-achieving for $\lambda = 0$. Areas corresponding to the activity of peak and average constraints are highlighted and defined in Theorem 4.

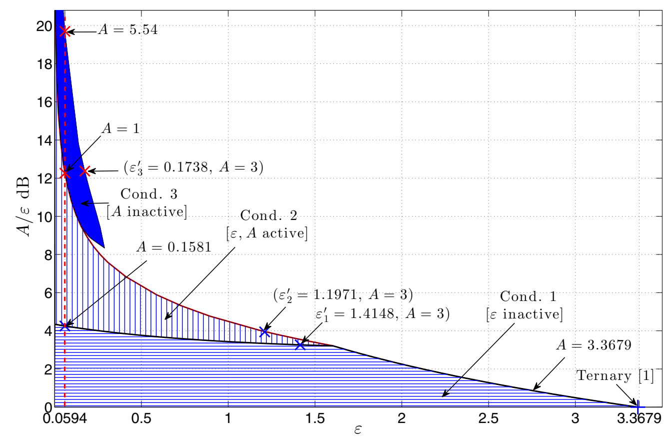


Fig. 2. Identical results to Figure 1, plotted for peak-to-average ratio (A/ε) versus ε $(\lambda = 0)$.

in (16), also plotted in Fig. 1, is covered by Cond. 3 in Theorem 4. This region is obtained by sampling in ε and computing the corresponding B via (16). Then, Cond. (17) is checked by plotting $\eta(B, x)$ for $x \in [0, A]$. As mentioned earlier, an inactive peak constraint in the capacity-achieving distribution implies that the mean constraint is active.

For DTP channels in the area outside the highlighted regions in Fig. 1, the capacity-achieving distribution is non-binary. In particular, Shamai [1] showed that, for a DTP channel with peak power constraint only, $A = 3.3679$ is the transition point between the capacity-achieving distribution being binary and ternary.

Using Fig. 1, the example in [6, Fig. 2] can be expanded by observing the behaviour of the capacity-achieving distribution for $\lambda = 0$, fixed $\varepsilon = 0.0594$ and increasing A. When $\varepsilon \leq A < f_1^{-1}(\varepsilon) = 0.1581$, Cond. 1 of Theorem 4 is satisfied and the optimal input distribution is binary and the average constraint is inactive. In [6, Fig. 2], the same interval of A for inactive mean is found numerically. For larger A, i.e., $f_1^{-1}(\varepsilon) \leq A \leq f_2^{-1}(\varepsilon) = 1$, Cond. 2 Theorem 4 is satisfied and the capacity-achieving distributions are binary with both constraints active. Notice that the same conclusion is found in [6, Fig. 2] through numerical computation. Cond. 3 of Theorem 4 is satisfied for $f_2^{-1}(\varepsilon) = 1 < A \leq 5.54$ and the capacity-achieving distribution is binary with inactive peak power constraint, as in [6, Fig. 2] . For $A > 5.54$, none of the conditions of Theorem 4 are satisfied and the resulting capacity-achieving distribution is non-binary. Our numerical study [6, Fig. 2] shows that the optimal distribution is in fact ternary. Thus, our results in Fig. 1 describe the phenomenon of oscillating activity of peak constraint observed numerically in [6, Fig. 2]. In this example, using Theorem 4, the corner points for the transition of inactivity of the constraints are described analytically via $f_1(A)$ and $f_2(A)$.

Figure 2 visualizes the results in Fig. 1 in a different way. Here regions of binary capacity-achieving distributions are plotted on a peak-to-average ratio (A/ε) versus average constraint axis. This visualization is particularly useful in cases where A/ε is fixed, say on a launched spacecraft, while the mean power ε can change due to varying range. Roughly speaking, for A/ε less than about 4 dB for binary capacity-achieving distributions the peak constraint is solely active. However, for larger peak-to-average ratios, binary capacity-achieving distributions satisfy both peak and average constraints. For the largest values of A/ε, only the average constraint is active for binary capacity-achieving signalling.

Figure 3 plots the capacity-achieving distributions for fixed $A = 3$, $\lambda = 0$ and increasing ε. The capacity-achieving distributions of the DTP channel are computed using the deterministic annealing algorithm described in [6, Sec. IV.A]. The behaviour of the constraints can be understood by following the vertical line $A = 3$ in Fig. 1. For $\varepsilon < \varepsilon_3' = 0.1738$ or equivalently -7.5 dB, it is evident from Fig. 3 that the capacity-achieving distribution is binary with inactive peak-constraint. Notice in Fig. 1, that this corresponds to the region where Cond. 3 of Theorem 4 is satisfied. Increasing ε yields a region where the capacity-achieving distribution is ternary and none of the conditions of Theorem 4 are satisfied. Further increase in $\varepsilon_2' = 1.1971 \leq \varepsilon \leq \varepsilon_1'$ yields a return to binary capacity-achieving distributions, however, now with active peak and average constraints, i.e., Cond. 2 of Theorem 4 is satisfied. For $\varepsilon_1' = 1.4148 < \varepsilon \leq A$ the capacity-achieving distribution is binary with inactive average constraint (Cond. 1 in Theorem 4). This threshold can also be observed in Fig. 1 and can be computed as $\varepsilon_1' = f_1(3)$.

Figure 4 illustrates the capacity-achieving distributions with fixed ε and A, $\varepsilon/A < 1/2$, and increasing λ. Theorem. 5 states that for λ large enough the capacity achieving distribution is

$$dF_x^* = dF_x^\dagger = \frac{3}{4}\delta(x) + \frac{1}{4}\delta(x - 10).$$

Notice that in Fig. 4, for $\lambda \geq 5.525$ that the computed capacity-achieving distribution corresponds to dF_x^* and is fixed for larger λ.

V. Conclusions

The DTP channel has been widely used in modelling many channels such as long range space optical systems. This paper provides insight into the capacity-achieving distributions for DTP channels, especially in the low power regime where binary signalling is often optimal. Necessary and sufficient

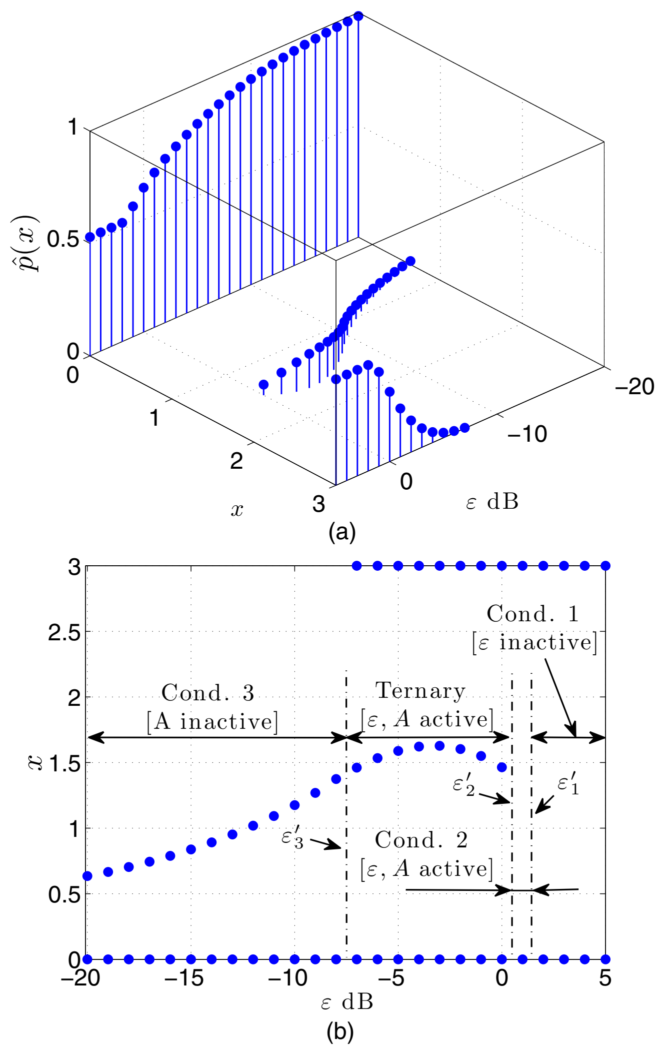


Fig. 3. Capacity-achieving distributions for fixed $A = 3$, $\lambda = 0$ and increasing ε: (a) distributions and (b) position of mass points. The values for thresholds ε'_1, ε'_2 and ε'_3 can be visualized in Fig. 1.

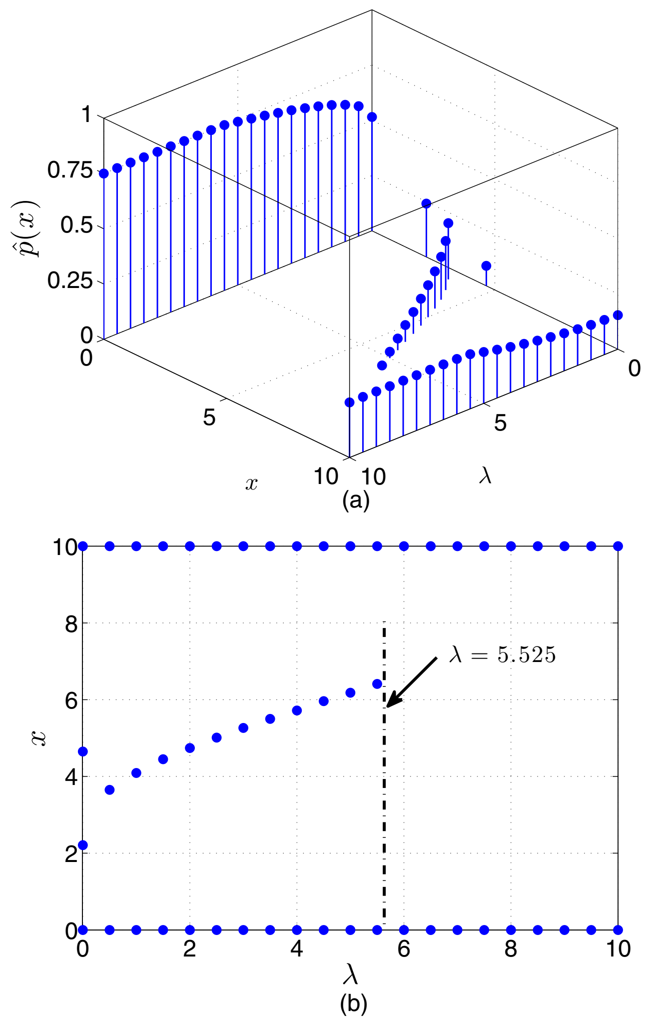


Fig. 4. Capacity-achieving distributions for fixed $A = 10$, $\varepsilon = 2.5$ with increasing λ: (a) distributions and (b) positions of mass points.

conditions on the optimality of binary distributions are established, which are further leveraged to obtain closed-form expressions of the capacity-achieving distributions in several special cases.

In particular, for $\lambda = 0$, three conditions on ε and A, corresponding to the activity of peak and average constraints, are given and corresponding forms of capacity-achieving distributions are provided. In the case of shot-noise limited DTP channels, we show that the binary maxentropic distribution is in fact capacity-achieving for λ large enough. Numerical simulations are provided to verify the analytical claims and provide insight on their application.

The motivation for this work is to gain insight not only into the capacity of DTP channels but also on signalling strategies to approach the capacity. Our ongoing work centres on the design of non-uniform signalling and efficient encoding/decoding structures suitable for multi-gigabit per second long-range optical intersatellite links.

Appendix A
Proof of (20)

We shall show that

$$\lim_{\lambda\to\infty} i(x, F_x^{\dagger}, \lambda) = 0, \quad x \in [0, A], \tag{24}$$

from which (20) follows immediately.

It is easy to verify that

$$\lim_{\lambda\to\infty} (x+\lambda)\log(x+\lambda) - x - (x+\lambda)\log\lambda = 0.$$

Therefore, it suffices to show

$$\begin{aligned}&\lim_{\lambda\to\infty}\sum_{y=0}^{\infty} e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\log\left(1+re^{-A}(1+A/\lambda)^y\right)\\&=\log(1+r).\end{aligned} \tag{25}$$

Let L be some positive odd integer. Based on Taylor's theorem,

$$\begin{aligned}&\log\left(1+re^{-A}(1+A/\lambda)^y\right)\\&=\sum_{l=1}^{l=L}(-1)^{(l+1)}\frac{r^l e^{-lA}(1+A/\lambda)^{yl}}{l}-\frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}},\end{aligned}$$

where

$$\xi \triangleq re^{-A}(1+A/\lambda)^y \tag{26}$$

and $\theta_\xi \in [0,\xi]$. Note that

$$\begin{aligned}
&\lim_{\lambda\to\infty}\sum_{y=0}^{\infty} e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\log\left(1+re^{-A}(1+A/\lambda)^y\right)\\
&=\lim_{L\to\infty(L\text{ odd})}\lim_{\lambda\to\infty}\sum_{y=0}^{\infty} e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\\
&\left(\sum_{l=1}^{l=L}(-1)^{(l+1)}\frac{r^l e^{-lA}(1+A/\lambda)^{yl}}{l}-\frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}}\right)\\
&=\lim_{L\to\infty(L\text{ odd})}\lim_{\lambda\to\infty}\sum_{l=1}^{l=L}(-1)^{(l+1)}\frac{r^l}{l}e^{-lA}e^{-(x+\lambda)}e^{\frac{(\lambda+x)(\lambda+A)^l}{\lambda^l}}\\
&\quad-\sum_{y=0}^{\infty} e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}}\\
&=\lim_{L\to\infty(L\text{ odd})}\lim_{\lambda\to\infty}\sum_{l=1}^{l=L}(-1)^{(l+1)}\frac{r^l}{l}e^{\frac{lAx}{\lambda}+(\lambda+x)\left(\binom{l}{2}\frac{A^2}{\lambda^2}+\cdots+\frac{A^l}{\lambda^l}\right)}\\
&\quad-\sum_{y=0}^{\infty} e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}}\\
&=\lim_{L\to\infty(L\text{ odd})}\sum_{l=1}^{l=L}(-1)^{(l+1)}\frac{r^l}{l}\\
&\quad-\lim_{\lambda\to\infty}\sum_{y=0}^{\infty} e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}}\\
&=\log(1+r)\\
&\quad-\lim_{L\to\infty(L\text{ odd})}\lim_{\lambda\to\infty}\sum_{y=0}^{\infty} e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\\
&\qquad\frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}}.
\end{aligned}\tag{27}$$

Since $\theta_\xi \geq 0$, it follows that

$$\begin{aligned}
&\lim_{L\to\infty(L\text{ odd})}\lim_{\lambda\to\infty}\left|\sum_{y=0}^{\infty} e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}}\right|\\
&\leq\lim_{L\to\infty(L\text{ odd})}\lim_{\lambda\to\infty}\sum_{y=0}^{\infty} e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\frac{\xi^{L+1}}{L+1}\\
&=\lim_{L\to\infty(L\text{ odd})}\lim_{\lambda\to\infty}\frac{r^{L+1}}{L+1}e^{-(L+1)A}e^{-(x+\lambda)}e^{\frac{(\lambda+x)(\lambda+A)^{L+1}}{\lambda^{L+1}}}\\
&=\lim_{L\to\infty(L\text{ odd})}\frac{r^{L+1}}{L+1}\\
&=0.
\end{aligned}\tag{28}$$

Therefore, we have

$$\begin{aligned}
&\lim_{L\to\infty(L\text{ odd})}\lim_{\lambda\to\infty}\sum_{y=0}^{\infty} e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}}\\
&=0,
\end{aligned}$$

which together with (27) proves (25).

Appendix B
Proof of (21)

Note that

$$\begin{aligned}
&\frac{\partial}{\partial z}\hat{M}(\mu,x,F_x^\dagger,z)\Big|_{z=0}\\
&=\lim_{z\to 0}\frac{\hat{M}(\mu,x,F_x^\dagger,z)-\hat{M}(\mu,x,F_x^\dagger,0)}{z}\\
&=\lim_{\lambda\to\infty}M(\mu,x,F_x^\dagger,\lambda)\lambda\\
&=\lim_{\lambda\to\infty}x\lambda\Bigg(\left(1+\frac{\lambda}{A}\right)\log\left(1+\frac{A}{\lambda}\right)\\
&\qquad-\left(1+\frac{\lambda}{x}\right)\log\left(1+\frac{x}{\lambda}\right)\Bigg)\\
&\quad+\Lambda(x)-\Lambda(0)+\frac{x}{A}\left(\Lambda(0)-\Lambda(A)\right),
\end{aligned}\tag{29}$$

where

$$\begin{aligned}
&\Lambda(x)\\
&\triangleq\lim_{\lambda\to\infty}\left\{\sum_{y=0}^{\infty} e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\log\left(1+re^{-A}(1+A/\lambda)^y\right)\right.\\
&\qquad\left.-\log(1+r)\right\}\lambda.
\end{aligned}$$

It is easy to prove that

$$\begin{aligned}
&\lim_{\lambda\to\infty}x\left(\left(1+\frac{\lambda}{A}\right)\log\left(1+\frac{A}{\lambda}\right)-\left(1+\frac{\lambda}{x}\right)\log\left(1+\frac{x}{\lambda}\right)\right)\lambda\\
&=\frac{x(A-x)}{2}.
\end{aligned}\tag{30}$$

Let

$$L \triangleq 2\lfloor\frac{\log\lambda+1}{2}\rfloor-1.$$

Invoking Taylor's theorem and changing the order of summation gives

$$\begin{aligned}
&\Lambda(x)\\
&=\lim_{\lambda\to\infty}\left\{\sum_{y=0}^{\infty} e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\sum_{l=1}^{l=L}(-1)^{(l+1)}\frac{r^l e^{-lA}(1+A/\lambda)^{yl}}{l}\right.\\
&\qquad\left.-\sum_{l=1}^{l=L}(-1)^{(l+1)}\frac{r^l}{l}\right\}\lambda-\lim_{\lambda\to\infty}\left\{\sum_{y=0}^{\infty} e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\right.\\
&\qquad\left.\left(\frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}}-\frac{r^{L+1}}{(L+1)(1+\theta_r)^{(L+1)}}\right)\right\}\lambda\\
&=\lim_{\lambda\to\infty}\left\{\sum_{l=1}^{l=L}(-1)^{(l+1)}\frac{r^l}{l}\lambda\left(e^{-lA}e^{-(x+\lambda)}e^{\frac{(\lambda+x)(\lambda+A)^l}{\lambda^l}}-1\right)\right\}\\
&\quad-\lim_{\lambda\to\infty}\left\{\sum_{y=0}^{\infty} e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\left(\frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}}\right.\right.\\
&\qquad\left.\left.-\frac{r^{L+1}}{(L+1)(1+\theta_r)^{(L+1)}}\right)\right\}\lambda,
\end{aligned}\tag{31}$$

where $\theta_r \in [0,r]$.

Now consider the first limit in (31). Let

$$\begin{aligned}
\phi \triangleq&\ \frac{lAx}{\lambda}+\frac{l(l-1)A^2}{2\lambda}+\frac{l(l-1)A^2x}{2\lambda^2}\\
&+(\lambda+x)\left(\binom{l}{3}\frac{A^3}{\lambda^3}+\cdots+\frac{A^l}{\lambda^l}\right).
\end{aligned}\tag{32}$$

We have

$$\lim_{\lambda\to\infty}\sum_{l=1}^{l=L}(-1)^{(l+1)}\frac{r^l}{l}\lambda\left(e^{-lA}e^{-(x+\lambda)}e^{\frac{(\lambda+x)(\lambda+A)^l}{\lambda^l}}-1\right)$$

$$
\begin{aligned}
&= \lim_{\lambda\to\infty} \sum_{l=1}^{l=L} (-1)^{(l+1)} \frac{r^l}{l} \lambda \left(e^{\phi} - 1\right) \\
&= \lim_{\lambda\to\infty} \sum_{l=1}^{l=L} (-1)^{(l+1)} \frac{r^l}{l} \lambda \left(1 + \phi + \theta_\phi \phi^2 - 1\right) \qquad (33) \\
&= \lim_{\lambda\to\infty} \sum_{l=1}^{l=L} (-1)^{(l+1)} \frac{r^l}{l} \left(lAx + \frac{l(l-1)A^2}{2}\right) \\
&\quad + \lim_{\lambda\to\infty} \sum_{l=1}^{l=L} (-1)^{(l+1)} \frac{r^l}{l} \lambda \Bigg(\frac{l(l-1)A^2 x}{2\lambda^2} \\
&\qquad + (\lambda + x)\left(\binom{l}{3}\frac{A^3}{\lambda^3} + \cdots + \frac{A^l}{\lambda^l}\right) + \theta_\phi \phi^2 \Bigg) \\
&= \left(Ax - \frac{A^2}{2}\right)\frac{r}{1+r} + \frac{A^2}{2}\frac{r}{(1+r)^2} \\
&\quad + \lim_{\lambda\to\infty} \sum_{l=1}^{l=L} (-1)^{(l+1)} \frac{r^l}{l} \lambda \Bigg(\frac{l(l-1)A^2 x}{2\lambda^2} \\
&\qquad + (\lambda + x)\left(\binom{l}{3}\frac{A^3}{\lambda^3} + \cdots + \frac{A^l}{\lambda^l}\right) + \theta_\phi \phi^2 \Bigg), \qquad (34)
\end{aligned}
$$

where (33) is due to Taylor's theorem and $\theta_\phi = \frac{1}{2}e^{\phi'}$ for some $\phi' \in [0, \phi]$. For $l \in (3, L]$,

$$
\binom{l}{i}\frac{A^i}{\lambda^i} < l^i \frac{A^i}{\lambda^i} \leq \left(\frac{A\log\lambda}{\lambda}\right)^i \leq \left(\frac{A\log\lambda}{\lambda}\right)^3, \quad i = 3, \ldots, l,
$$

when λ is sufficiently large. Therefore,

$$
\begin{aligned}
\phi \leq \frac{Ax\log\lambda}{\lambda} &+ \frac{\log\lambda(\log\lambda - 1)A^2}{2\lambda} \\
&+ \frac{\log\lambda(\log\lambda - 1)A^2 x}{2\lambda^2} + \frac{A^3(\log\lambda)^4(\lambda + x)}{\lambda^3}
\end{aligned}
$$

for large λ. As a consequence, $\phi \to 0$ and $\theta_\phi \to \frac{1}{2}$ uniformly for $l \in [3, L]$ and $x \in [0, A]$ as $\lambda \to \infty$. Furthermore, we have

$$
\begin{aligned}
&\lim_{\lambda\to\infty} \Bigg| \sum_{l=1}^{l=L} (-1)^{(l+1)} \frac{r^l}{l} \lambda \Bigg(\frac{l(l-1)A^2 x}{2\lambda^2} \\
&\qquad + (\lambda + x)\left(\binom{l}{3}\frac{A^3}{\lambda^3} + \cdots + \frac{A^l}{\lambda^l}\right) + \theta_\phi \phi^2 \Bigg) \Bigg| \\
&\leq \lim_{\lambda\to\infty} \sum_{l=1}^{l=L} \frac{r^l}{l} \lambda \Bigg(\frac{l(l-1)A^2 x}{2\lambda^2} \\
&\qquad + (\lambda + x)\left(\binom{l}{3}\frac{A^3}{\lambda^3} + \cdots + \frac{A^l}{\lambda^l}\right) + \theta_\phi \phi^2 \Bigg) \\
&\leq \lim_{\lambda\to\infty} \sum_{l=1}^{l=L} \frac{r^l}{l} \lambda \Bigg(\frac{l(l-1)A^2 x}{2\lambda^2} \\
&\qquad + (\lambda + x)\left(\frac{A\log\lambda}{\lambda}\right)^3 (\log\lambda - 2) + \theta_\phi \phi^2 \Bigg) \\
&\leq \lim_{\lambda\to\infty} \sum_{l=1}^{l=L} \frac{r^l}{l} \lambda \left(\frac{l(l-1)A^2 x}{2\lambda^2} + \frac{A^3(\log\lambda)^4}{\lambda^2} + \theta_\phi \phi^2 \right) \\
&\leq \lim_{\lambda\to\infty} \sum_{l=1}^{l=L} \left(\frac{(\log\lambda - 1)A^2 x}{2\lambda} + \frac{A^3(\log\lambda)^4}{\lambda} + \lambda\theta_\phi \phi^2 \right) \\
&= \lim_{\lambda\to\infty} \log\lambda \Bigg(\frac{(\log\lambda - 1)A^2 x}{2\lambda} + \frac{A^3(\log\lambda)^4}{\lambda} \\
&\qquad + \frac{\lambda}{2}\Bigg(\frac{Ax\log\lambda}{\lambda} + \frac{\log\lambda(\log\lambda - 1)A^2}{2\lambda} \\
&\qquad + \frac{\log\lambda(\log\lambda - 1)A^2 x}{2\lambda^2} + \frac{A^3(\log\lambda)^4(\lambda + x)}{\lambda^3} \Bigg)^2 \Bigg) \\
&= 0,
\end{aligned}
$$

which together with (34) implies

$$
\begin{aligned}
&\lim_{\lambda\to\infty} \sum_{l=1}^{l=L} (-1)^{(l+1)} \frac{r^l}{l} \lambda \left(e^{-lA} e^{-(x+\lambda)} e^{\frac{(\lambda+x)(\lambda+A)^l}{\lambda^l}} - 1 \right) \\
&\qquad = \left(Ax - \frac{A^2}{2}\right)\frac{r}{1+r} + \frac{A^2}{2}\frac{r}{(1+r)^2}. \qquad (35)
\end{aligned}
$$

Next consider the second limit in (31). We have

$$
\begin{aligned}
&\lim_{\lambda\to\infty} \Bigg| \Bigg\{ \sum_{y=0}^{\infty} e^{-(\lambda+x)} \frac{(\lambda+x)^y}{y!} \Bigg(\frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}} \\
&\qquad - \frac{r^{L+1}}{(L+1)(1+\theta_r)^{(L+1)}} \Bigg) \Bigg\} \lambda \Bigg| \\
&\leq \lim_{\lambda\to\infty} \Bigg\{ \sum_{y=0}^{\infty} e^{-(\lambda+x)} \frac{(\lambda+x)^y}{y!} \Bigg(\frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}} \\
&\qquad + \frac{r^{L+1}}{(L+1)(1+\theta_r)^{(L+1)}} \Bigg) \Bigg\} \lambda \\
&\leq \lim_{\lambda\to\infty} \Bigg\{ \sum_{y=0}^{\infty} e^{-(\lambda+x)} \frac{(\lambda+x)^y}{y!} \Bigg(\frac{\xi^{L+1}}{L+1} \\
&\qquad + \frac{r^{L+1}}{(L+1)(1+\theta_r)^{(L+1)}} \Bigg) \Bigg\} \lambda \\
&= \lim_{\lambda\to\infty} \Bigg\{ \frac{r^{L+1}}{L+1} e^{-(L+1)A} e^{-(x+\lambda)} e^{\frac{(\lambda+x)(\lambda+A)^{L+1}}{\lambda^{L+1}}} \\
&\qquad + \frac{r^{L+1}}{(L+1)(1+\theta_r)^{(L+1)}} \Bigg\} \lambda \\
&= \lim_{\lambda\to\infty} \frac{r^{L+1}\lambda}{L+1} \Bigg\{ e^{\frac{(L+1)Ax}{\lambda} + (\lambda+x)\left(\left(\binom{L+1}{2}\frac{A^2}{\lambda^2} + \binom{L+1}{3}\frac{A^3}{\lambda^3} + \cdots + \frac{A^{L+1}}{\lambda^{L+1}}\right)} \\
&\qquad + \frac{1}{(1+\theta_r)^{(L+1)}} \Bigg\} \\
&\leq \lim_{\lambda\to\infty} \frac{r^{L+1}\lambda}{L+1} \Bigg\{ e^{\frac{(L+1)Ax}{\lambda} + (\lambda+x)(L+1)^2 \frac{A^2}{\lambda^2} L} \\
&\qquad + \frac{1}{(1+\theta_r)^{(L+1)}} \Bigg\} \\
&= \lim_{\lambda\to\infty} \frac{r^{\log\lambda+1}\lambda}{\log\lambda+1} \Bigg\{ e^{\frac{(\log\lambda+1)Ax}{\lambda} + (\lambda+x)(\log\lambda+1)^2 \frac{A^2}{\lambda^2} \log\lambda} \\
&\qquad + \frac{1}{(1+\theta_r)^{(\log\lambda+1)}} \Bigg\} \\
&= 0. \qquad (36)
\end{aligned}
$$

Combining (35) and (36) with (31) gives

$$\Lambda(x) = \left(Ax - \frac{A^2}{2}\right)\frac{r}{1+r} + \frac{A^2}{2}\frac{r}{(1+r)^2}. \quad (37)$$

Substitute (30) and (37) into (29) proves (21).

Appendix C
Proof of (22)

Note that

$$\begin{aligned}
&\hat{M}'(\mu, x, F_x^\dagger, z)\\
&= \frac{\partial M(\mu, x, F_x^\dagger, \lambda)}{\partial x}\\
&= -\frac{\partial i(x, F_x^\dagger, \lambda)}{\partial x} + \frac{1}{A}(i(A, F_x^\dagger, \lambda) - i(0, F_x^\dagger, \lambda))\\
&= -\log(x+\lambda) + \frac{\partial S(x, F_x^\dagger, \lambda)}{\partial x}\\
&\quad + \frac{1}{A}(i(A, F_x^\dagger, \lambda) - i(0, F_x^\dagger, \lambda)), \quad (38)
\end{aligned}$$

where

$$\begin{aligned}
&S(x, F_x^\dagger, \lambda)\\
&= \sum_{y=0}^{\infty} e^{-(x+\lambda)}\frac{(x+\lambda)^y}{y!}\log\left((1-\frac{\varepsilon}{A})\lambda^y + \frac{\varepsilon}{A}e^{-A}(A+\lambda)^y\right).
\end{aligned}$$

It can be verified that

$$\begin{aligned}
\frac{\partial S(x, F_x^\dagger, \lambda)}{\partial x} &= -e^{-(x+\lambda)}\sum_{y=0}^{\infty}\frac{(x+\lambda)^{y-1}(x+\lambda-y)}{y!}\\
&\left[-\log(1+r) + y\log\lambda + \log\left(1+re^{-A}(1+A/\lambda)^y\right)\right].
\end{aligned}$$

One can readily prove that

$$e^{-(x+\lambda)}\sum_{y=0}^{\infty}\frac{(x+\lambda)^{y-1}(x+\lambda-y)}{y!}\log(1+r) = 0,$$

$$-e^{-(x+\lambda)}\sum_{y=0}^{\infty}\frac{(x+\lambda)^{y-1}(x+\lambda-y)}{y!}y\log\lambda = \log\lambda.$$

Thus,

$$\begin{aligned}
&\frac{\partial S(x, F_x^\dagger, \lambda)}{\partial x}\\
&= \log\lambda - e^{-(x+\lambda)}\sum_{y=0}^{\infty}\frac{(x+\lambda)^{y-1}(x+\lambda-y)}{y!}\\
&\qquad\qquad \log\left(1+re^{-A}(1+A/\lambda)^y\right)\\
&= \log\lambda - \sum_{y=0}^{\infty}e^{-(x+\lambda)}\frac{(x+\lambda)^y}{y!}\log\left(1+re^{-A}(1+A/\lambda)^y\right)\\
&\quad + \sum_{y=1}^{\infty}e^{-(x+\lambda)}\frac{(x+\lambda)^{y-1}}{(y-1)!}\log\left(1+re^{-A}(1+A/\lambda)^y\right).
\end{aligned}$$

Following the derivation of (25), one can show that

$$\begin{aligned}
&\lim_{\lambda\to\infty}\sum_{y=1}^{\infty}e^{-(x+\lambda)}\frac{(x+\lambda)^{y-1}}{(y-1)!}\log\left(1+re^{-A}(1+A/\lambda)^y\right)\\
&= \log(1+r),
\end{aligned}$$

which together with (24) and (25) proves (22).

Appendix D
Proof of (23)

Note that

$$\begin{aligned}
&\frac{\partial}{\partial z}\hat{M}'(\mu, x, F_x^\dagger, z)\Big|_{z=0}\\
&= \lim_{z\to 0}\frac{\hat{M}'(\mu, x, F_x^\dagger, z) - \hat{M}'(\mu, x, F_x^\dagger, 0)}{z}\\
&= \lim_{\lambda\to\infty}\frac{\partial}{\partial x}M(\mu, x, F_x^\dagger, \lambda)\lambda\\
&= \lim_{\lambda\to\infty}\left(\left(1+\frac{\lambda}{A}\right)\log\left(1+\frac{A}{\lambda}\right) - \log\left(1+\frac{x}{\lambda}\right) - 1\right)\lambda\\
&\quad - \frac{1}{A}\left(\Lambda(A) - \Lambda(0)\right)\\
&\quad - \lim_{\lambda\to\infty}\left\{\sum_{y=0}^{\infty}e^{-(x+\lambda)}\frac{(x+\lambda)^y}{y!}\log\left(1+re^{-A}(1+A/\lambda)^y\right)\right.\\
&\quad \left.-\sum_{y=1}^{\infty}e^{-(x+\lambda)}\frac{(x+\lambda)^{y-1}}{(y-1)!}\log\left(1+re^{-A}(1+A/\lambda)^y\right)\right\}\lambda.
\end{aligned} \quad (39)$$

It can be easily verified that

$$\begin{aligned}
\lim_{\lambda\to\infty}\left(\left(1+\frac{\lambda}{A}\right)\log\left(1+\frac{A}{\lambda}\right) - \log\left(1+\frac{x}{\lambda}\right) - 1\right)\lambda&\\
= \frac{A}{2} - x,& \quad (40)
\end{aligned}$$

$$\lim_{\lambda\to\infty}\frac{1}{A}\left(\Lambda(A) - \Lambda(0)\right) = \frac{Ar}{1+r}. \quad (41)$$

Let ξ be defined according to (26). By Taylor's theorem, we have

$$\begin{aligned}
&\lim_{\lambda\to\infty}\left\{\sum_{y=0}^{\infty}e^{-(x+\lambda)}\frac{(x+\lambda)^y}{y!}\log\left(1+re^{-A}(1+A/\lambda)^y\right)\right.\\
&\quad \left.-\sum_{y=1}^{\infty}e^{-(x+\lambda)}\frac{(x+\lambda)^{y-1}}{(y-1)!}\log\left(1+re^{-A}(1+A/\lambda)^y\right)\right\}\lambda\\
&= \lim_{L\to\infty, L\text{ odd}}\lim_{\lambda\to\infty}\left\{\sum_{y=0}^{\infty}e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\right.\\
&\left(\sum_{l=1}^{l=L}(-1)^{(l+1)}\frac{r^l e^{-lA}(1+A/\lambda)^{yl}}{l} - \frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}}\right)\\
&-\sum_{y=1}^{\infty}e^{-(x+\lambda)}\frac{(x+\lambda)^{y-1}}{(y-1)!}\left(\sum_{l=1}^{l=L}(-1)^{(l+1)}\frac{r^l e^{-lA}(1+A/\lambda)^{yl}}{l}\right.\\
&\qquad\qquad \left.\left.-\frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}}\right)\right\}\lambda\\
&= \lim_{L\to\infty, L\text{ odd}}\lim_{\lambda\to\infty}\left\{\sum_{l=1}^{l=L}(-1)^{(l+1)}\frac{r^l}{l}\lambda e^{-(\lambda+x+lA)}e^{\frac{(\lambda+x)(\lambda+A)^l}{\lambda^l}}\right.\\
&\quad \left.-\sum_{l=1}^{l=L}(-1)^{(l+1)}\frac{r^l}{l}\lambda\left(1+\frac{A}{\lambda}\right)^l e^{-(\lambda+x+lA)}e^{\frac{(\lambda+x)(\lambda+A)^l}{\lambda^l}}\right\}\\
&- \lim_{L\to\infty, L\text{ odd}}\lim_{\lambda\to\infty}\left\{\sum_{y=0}^{\infty}e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}}\right.\\
&\qquad \left.-\sum_{y=1}^{\infty}e^{-(x+\lambda)}\frac{(x+\lambda)^{y-1}}{(y-1)!}\frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}}\right\}\lambda,
\end{aligned}$$

where $\theta_\xi \in [0,\xi]$. Let ϕ be defined according to (32). It is easy to see that

$$\begin{aligned}
&\lim_{L\to\infty, L \text{ odd}} \lim_{\lambda\to\infty} \sum_{l=1}^{l=L} (-1)^{(l+1)} \frac{r^l}{l} \lambda e^{-(\lambda+x+lA)} e^{\frac{(\lambda+x)(\lambda+A)^l}{\lambda^l}} \\
&\qquad\qquad \left[1 - \left(1+\frac{A}{\lambda}\right)^l\right] \\
&= \lim_{L\to\infty, L \text{ odd}} \lim_{\lambda\to\infty} \sum_{l=1}^{l=L} (-1)^{(l+1)} \frac{r^l}{l} \lambda e^{\phi} \left[1 - \left(1+\frac{A}{\lambda}\right)^l\right] \\
&= \lim_{L\to\infty, L \text{ odd}} \lim_{\lambda\to\infty} \sum_{l=1}^{l=L} (-1)^{(l+1)} \frac{r^l}{l} \lambda \left(1+\theta'_\phi \phi\right) \\
&\qquad\qquad \left(1 - 1 - \left(\frac{lA}{\lambda} + \binom{l}{2}\frac{A^2}{\lambda^2} + \cdots + \frac{A^l}{\lambda^l}\right)\right) \\
&= \lim_{L\to\infty, L \text{ odd}} \lim_{\lambda\to\infty} \sum_{l=1}^{l=L} (-1)^{(l+2)} \frac{r^l}{l} \left(1+\theta'_\phi \phi\right) \\
&\qquad\qquad \left(lA + \binom{l}{2}\frac{A^2}{\lambda} + \cdots + \frac{A^l}{\lambda^{l-1}}\right) \\
&= \lim_{L\to\infty, L \text{ odd}} \lim_{\lambda\to\infty} \sum_{l=1}^{l=L} (-1)^{l} \frac{r^l}{l} \\
&\quad \left[\theta'_\phi \phi \left(lA + \binom{l}{2}\frac{A^2}{\lambda} + \cdots + \frac{A^l}{\lambda^{l-1}}\right)\right. \\
&\qquad\qquad \left. + \binom{l}{2}\frac{A^2}{\lambda} + \cdots + \frac{A^l}{\lambda^{l-1}}\right] \\
&\quad + \lim_{L\to\infty, L \text{ odd}} \sum_{l=1}^{l=L} (-1)^l r^l A \\
&= \frac{-Ar}{1+r} + \lim_{L\to\infty, L \text{ odd}} \lim_{\lambda\to\infty} \sum_{l=1}^{l=L} (-1)^{l} \frac{r^l}{l} \\
&\quad \left[\theta'_\phi \phi \left(lA + \binom{l}{2}\frac{A^2}{\lambda} + \cdots + \frac{A^l}{\lambda^{l-1}}\right)\right. \\
&\qquad\qquad \left. + \binom{l}{2}\frac{A^2}{\lambda} + \cdots + \frac{A^l}{\lambda^{l-1}}\right],
\end{aligned}$$

where $\theta'_\phi = e^{\phi'}$ for some $\phi' \in [0,\phi]$ and $\theta'_\phi \to 1$ uniformly for $l \in [1,L]$ and $x \in [0,A]$ as $\lambda \to \infty$. One can readily verify that

$$\begin{aligned}
&\lim_{\lambda\to\infty} \left| \sum_{l=1}^{l=L} (-1)^l \frac{r^l}{l} \left(\theta_\phi \phi \left(lA + \binom{l}{2}\frac{A^2}{\lambda} + \cdots + \frac{A^l}{\lambda^{l-1}}\right)\right.\right. \\
&\qquad\qquad \left.\left. + \binom{l}{2}\frac{A^2}{\lambda} + \cdots + \frac{A^l}{\lambda^{l-1}}\right)\right| \\
&\le \lim_{\lambda\to\infty} \sum_{l=1}^{l=L} \frac{r^l}{l} \left[\theta_\phi \left(\frac{lA}{\lambda} + l^2 \frac{A^2}{\lambda^2}(l-1)\right)\right. \\
&\qquad\qquad \left. \left(lA + l^2\frac{A^2}{\lambda}(l-1)\right) + l^2 \frac{A^2}{\lambda}(l-1)\right] \\
&\le \lim_{\lambda\to\infty} L \left[\left(\frac{LA}{\lambda} + L^2\frac{A^2}{\lambda^2}(L-1)\right)\right. \\
&\qquad\qquad \left. \left(LA + L^2\frac{A^2}{\lambda}(L-1)\right) + L^2\frac{A^2}{\lambda}(L-1)\right] \\
&= 0
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\lim_{L\to\infty, L \text{ odd}} \lim_{\lambda\to\infty} \sum_{l=1}^{l=L} (-1)^{(l+1)} \frac{r^l}{l} \lambda e^{-(\lambda+x+lA)} e^{\frac{(\lambda+x)(\lambda+A)^l}{\lambda^l}} \\
&\qquad\qquad \left[1 - \left(1+\frac{A}{\lambda}\right)^l\right] = \frac{-Ar}{1+r}.
\end{aligned}$$

Moreover, following the derivation of (28), one can show that

$$\begin{aligned}
&\lim_{L\to\infty, L \text{ odd}} \lim_{\lambda\to\infty} \left| \left\{ \sum_{y=0}^{\infty} e^{-(\lambda+x)} \frac{(\lambda+x)^y}{y!} \frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}} \right.\right. \\
&\qquad \left.\left. - \sum_{y=1}^{\infty} e^{-(x+\lambda)} \frac{(x+\lambda)^{y-1}}{(y-1)!} \frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}} \right\} \lambda \right| \\
&= 0.
\end{aligned}$$

As a consequence, we have

$$\begin{aligned}
&\lim_{\lambda\to\infty} \left\{ \sum_{y=0}^{\infty} e^{-(x+\lambda)} \frac{(x+\lambda)^y}{y!} \log\left(1 + re^{-A}(1+A/\lambda)^y\right) \right. \\
&\qquad \left. - \sum_{y=1}^{\infty} e^{-(x+\lambda)} \frac{(x+\lambda)^{y-1}}{(y-1)!} \log\left(1 + re^{-A}(1+A/\lambda)^y\right) \right\} \lambda \\
&= \frac{-Ar}{1+r}.
\end{aligned} \tag{42}$$

Substituting (40), (41), and (42) into (39) proves (23).

References

[1] S. Shamai (Shitz), "Capacity of a pulse amplitude modulated direct detection photon channel," *Proc. Inst. Elec. Eng.*, vol. 137, no. 6, pp. 424–430, Dec. 1990.

[2] S. Shamai (Shitz), "On the capacity of a direct-detection photon channel with intertransition-constrained binary input," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1540–1550, Nov. 1991.

[3] 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.

[4] 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.

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

[6] 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.* (TCOM-TPS-13-0142).

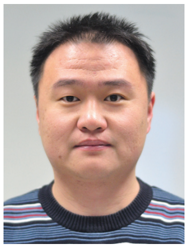
[7] J. Cao, S. Hranilovic, and J. Chen, "Channel capacity and non-uniform signalling for discrete-time Poisson channels," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 5, no. 4, pp. 329–337, Apr. 2013.

[8] J. G. Smith, "The information capacity of amplitude and variance-constrained scalar Gaussian channels," *Inf. Contr.*, vol. 18, pp. 203–219, 1971.

[9] H. Hemmati (Ed.), *Deep Space Optical Communication*. Wiley Inter-Science, 2006.

[10] A. A. Farid and S. Hranilovic, "Channel capacity and non-uniform signalling for free-space optical intensity channels," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 9, pp. 1553–1563, Dec. 2009.

[11] J. M. Kahn and J. R. Barry, "Wireless infrared communications," *Proc. IEEE*, vol. 85, no. 2, pp. 265–298, Feb. 1997.

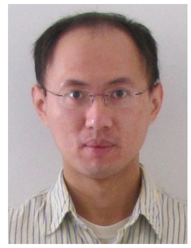


Jihai Cao received the B.A.Sc. and M.A.Sc. degrees with honours in electrical engineering from Harbin Institute of Technology, China in 2006 and 2008 respectively and Ph.D. from McMaster University, Canada in 2013. His field of interest includes wireless optical communication, information theory, and coding.

Steve Hranilovic (S'94-M'03-SM'07) received the B.A.Sc. degree with honours in electrical engineering from the University of Waterloo, Canada in 1997 and M.A.Sc. and Ph.D. degrees in electrical engineering from the University of Toronto, Canada in 1999 and 2003 respectively.

He is currently an Associate Professor in the Department of Electrical and Computer Engineering, McMaster University, (Hamilton, ON, Canada), where he also serves as Associate Chair for undergraduate studies. During 2010-2011 he spent his research leave as Senior Member, Technical Staff in Advanced Technology for Research in Motion, Waterloo, Canada. His research interests are in the areas of free-space and wired optical communications, digital communication algorithms, and electronic and photonic implementation of coding and communication systems. He is the author of the book *Wireless Optical Communications Systems* (New York:Springer, 2004).

Dr. Hranilovic is a licensed Professional Engineer in the Province of Ontario and was awarded the Government of Ontario Early Researcher Award in 2006. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS in the area of Optical Wireless Communications.



Jun Chen (S'03-M'06) received the B.E.degree with honors in communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001 and the M.S. and Ph.D. degrees in electrical and computer engineering from Cornell University, Ithaca, NY, in 2004 and 2006, respectively.

He was a Postdoctoral Research Associate in the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign, Urbana, IL, from 2005 to 2006, and a Postdoctoral Fellow at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, from 2006 to 2007. He is currently an Associate Professor of Electrical and Computer Engineering at McMaster University, Hamilton, ON, Canada. His research interests include information theory, wireless communications, and signal processing.

He received several awards for his research, including the Josef Raviv Memorial Postdoctoral Fellowship in 2006, the Early Researcher Award from the Province of Ontario in 2010, and the IBM Faculty Award in 2010.