DISCRETE-TIME POISSON CHANNEL: CAPACITY AND SIGNALLING DESIGN

DISCRETE-TIME POISSON CHANNEL: CAPACITY AND SIGNALLING DESIGN

By

JIHAI CAO, B.S. (Electrical Engineering),
M.S. (Electrical & Communication Engineering),
Harbin Institute of Technology, Harbin, China

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree
Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2013) MCMASTER UNIVERSITY
(Electrical and Computer Engineering) Hamilton, Ontario

TITLE: **Discrete-Time Poisson Channel: Capacity and Signalling Design**

AUTHOR: Jihai Cao
B.S. (Electrical Engineering),
M.S. (Electrical & Communication Engineering),
Harbin Institute of Technology, Harbin, China

SUPERVISOR: Dr. Steve Hranilovic and Dr. Jun Chen

NUMBER OF PAGES: xxi, 138

Dedications

To my parents and my wife.

Abstract

The discrete-time Poisson (DTP) channel models a wide range of optical communication channels. The channel capacity and capacity-achieving distributions are generally unknown. This thesis addresses system design of DTP channels and presents novel contributions to the capacity of DTP channel, properties and closed-form expression of the capacity-achieving distribution under peak and average constraints, signalling design, and sum-capacity-achieving distributions of DTP multiple access channel (MAC) with peak amplitude constraints.

Two algorithms are developed to compute the channel capacity of DTP channel as well as the capacity-achieving distribution with average and peak amplitude constraints. Tight lower bounds based on input distributions with simple forms are presented. Non-uniform signalling algorithms to achieve the channel capacity are also demonstrated. Fundamental properties of capacity-achieving distributions for DTP channels are established. Furthermore, necessary and sufficient conditions on the optimality of binary distributions are presented. Analytical expressions for the capacity-achieving distributions of the DTP channel are derived when there is no dark current and when the dark current is large enough. A two-user DTP multiple access channel model is proposed and it is shown that the sum-capacity-achieving distributions under peak amplitude constraints are discrete with a finite number of mass points.

Acknowledgements

I would like to express my heartfelt gratitude to my advisors, Prof. Steve Hranilovic and Prof. Jun Chen, for their consistent guidance and insightful discussions during my Ph.D. study. I will never forget the mornings that Prof. Steve Hranilovic and I revised the papers at 7:30 AM, the nights that Prof. Jun Chen and I exchanged email at 2 AM. I was lucky to work closely with them and their infectious passions in complementing areas inspired this work in many different ways. I am especially grateful to their unparalleled support during this self-development journey of my Ph.D. study.

I would also like to express my appreciation to Prof. Jiankang Zhang and Prof. Tim Davidson for serving in my committee and defence, and for their valuable comments and discussions. In addition, I want to thank the staff in the ECE department and in particular Mrs. Cheryl Gies and Mr. Dan Manolescu who always helped me.

I would like to sincerely thank my professors at Harbin Institute of Technology for all that I have learned from them during my B.S. and M.S. studies. Specifically, I would like to express my gratitude to Prof. Guanghui Ren, for his guidance and support both on academic and non-academic matters.

Finally, I would like to thank for my parents, my old sisters and my wife, for supporting me. This work would not happen without your love.

List of Acronyms

AU	Astronomical Unit
BICM	Bit-interleaved coded modulation
BER	Bit-error rate
BPSK	Binary phase-shift keying
CDMA	Code division multiple access
CTP	Discrete-time Poisson
DA	Deterministic annealing
DD	Direct detection
DTP	Discrete-time Poisson
ESA	European space agency
FOV	Field-of-View
FSO	Free-space optical
Gbps	Gigabits per second
GEO	Geostationary Earth orbit
HEO	High earth orbit
IM	Intensity Modulation
KKT	Karush-Kuhn-Tucker
K-L	Kullback-Leibler
LDPC	Low density parity check
LEO	Low earth orbit
MAC	Multiple access channel
Mbps	Megabits per second
MLC	Multi-level coding
MSD	Multi-stage decoding
OISL	Optical intersatellite links
PAM	Pulse amplitude modulation
pdf	Probability density function
PMT	Photomultiplier tube
PPM	Pulse position modulation
RF	Radio frequency

List of Notations and Symbols

c	Speed of light in a vacuum
f	Mapper function
h	Planck's constant
$\|x\|$	Magnitude of x when x is a complex number
z	Optical link distance
$\|B\|$	Cardinality of set B
C	Channel capacity
$\mathrm{E}[\cdot]$	Expectation operator
$\mathcal{F}$	Set of input distributions
H	Discrete source entropy
K	Number of mass points
L	Lower bound on the capacity
U	Upper bound on the capacity
M	Multiplier function
N	Quantization level
$1/T$	Data rate
W	Spectral radiance of Earth
$I(X;Y)$	Mutual information
ε	Average optical power
λ	Dark current

d_t	Transmitter Aperture diameter
d_r	Receiver Aperture diameter
λ_w	Laser wavelength
η_T	Transmitter optical efficiency
η_R	Receiver optical efficiency
η	Detector quantum efficiency
γ	Closed contour
ψ_x	Input constellation
ψ_p	Input probability
F_x	Cumulative distribution function of the input
θ_T	Pointing error
G_T	Transmit antenna gain
P_T	Average transmitter power
L_T	Pointing loss factor associated with the pointing error
Ω	Receiver Field-of-View
$\Delta\lambda$	Bandwidth of receiver filter

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

The Poisson photon counting channel models the quantum effects in optical detection, in which a light field is converted to electric current [1, 2]. At the receiver, a semi-classical approach for evaluating optical communication systems treats the received radiation field as a wave [1], and explains its interaction with the released electrons through a probabilistic relation. Given the input light intensity (i.e., power), it was demonstrated that the number of the released electrons in any time interval obeys a Poisson distribution [1, 3].

This model is typical when the photodetectors are used with deterministic receiver fields [1]. For example, it exists in wide classes of optical communication systems, especially those with low input intensity, high background noise levels and high gain photomultiplier tube (PMT) [1]. An interesting recent application of practical interest is intersatellite laser communication links. At the receiver, the optical intensity is low due to the limited power and the long distance transmission which is usually over thousands kilometres [4]. Further, optical intersatellite communication is free of atmospheric turbulence since the links are in space. The only noise is the shot noise from the received signal which is dominated by background light either scattered from the earth or directly from other planets.

Practically, due to energy and component limitations, both the mean, ε, and peak, A, emitted intensities (i.e. power) are constrained for the optical channel. Moreover, all signals must assume non-negative amplitudes since the physical quantity modulated is a normalized power. A *continuous-time* Poisson (CTP) channel is one in which no bandwidth limitation is imposed and the input intensity is an arbitrary continuous waveform. It was shown that to achieve the capacity of the CTP channel the bandwidth of the input signal is necessarily infinite, which is not practical. In contrast, the *discrete-time* Poisson (DTP) channel considers only pulse amplitude modulated (PAM) signals where the transmitted intensity varies across discrete time slots and is fixed in each slot. Thus, there is an inherent bandwidth limitation in the DTP channel. The receiver outputs the photon counts received in each time slot [5,6].

The current understanding of the DTP channel is limited. One of the main challenges in the DTP channel is to find the channel capacity and the corresponding optimal input distribution. Another major issue of the DTP channel is in signalling design. It is worth noting that signalling schemes developed for other channels may not be suitable for the DTP channel. This is partly because the DTP channels are fundamentally different from widely studied radio frequency (RF) channels or the optical intensity channels with additive Gaussian noise and partly because the signal must satisfy peak, average and non-negativity constraints. These open problems in the DTP channel motivate the need for an innovative study of these systems — a task which is carried out in this thesis.

Table 1.1: Comparison of the Three Communication Systems at Two Different Data Volumes for a Mars Mission [9]

Data Volume (Gb/day)	**Communication Band**	**Mass (kg)**	**Power (W)**
1.0	X-Band	16.7	29
	Ka-Band	11.3	32
	Optical	7.4	22.8
10	X-Band	23.2	62
	Ka-Band	22.55	46
	Optical	9	30

1.2 Background

1.2.1 Satellite Experiments

Optical intersatellite communication has witnessed a rapid increase in data rate and decreasing in mass during the last few years. For satellite applications, free-space optical (FSO) communications provide larger bandwidth, higher data rates, smaller beam divergence and higher antenna gains from smaller apertures as compared to traditional radio frequency (RF) transceivers [7, 8]. This high gain translates into a significant reduction in the required transceiver power, volume and mass, which are heavily constrained in any aerospace application and even more so in deep space communication. For example, Table 1.1 compares optical and RF communication systems for a Mars mission, which is required to return data with daily volumes of 1 or 10 Gbits over the distance around 2.7 Astronomical Unit (AU). It can be seen directly that for the same data volume, optical transceivers can save more than 50% mass and consuming power as compared with a RF system.

Figure 1.1 presents an illustrative diagram for the concept of optical intersatellite communication including earth relay satellite. The Earth relay satellite is satellite placed in a geostationary Earth orbit (GEO) at a distance of 36000 km from the Earth. It relays information to and from low-earth orbit (LEO) satellites, aircraft, and ground stations that otherwise would not be able to communicate for long periods of time or would not be able to communicate at all [10]. In earth observation or military scenarios, high speed and low latency (high availability) data transmission are critical for exploration or scientific missions, and large volumes of data collected by the LEO satellite or aircraft must be returned before the next orbit. Intersatellite laser links between the earth relay satellites and spacecraft, which are usually over a range of 1000-45000 km, increase the amount of time that a LEO is in communication contact with the ground station, and thus would increase the amount of data that could be transferred.

Table 1.2 presents a collection of optical intersatellite links (OISL) implemented or to be implemented. For these LEO-GEO missions, Fig. 1.1 can be applied directly. The semi-conductor intersatellite link experiment (SILEX) is the first civilian optical communications program of the European space agency (ESA). It consists of two optical communication payloads embedded on spacecrafts ARTEMIS and SPOT-IV. It allows the optical transmission at data rate of 50 Mbps from LEO to GEO with a total mass of 160 kg [4]. In 2005, the Artemis-OICETS (inter-orbit communications engineering test satellite) mission implemented a bi-directional optical link with the data rate of 50 Mbps from OICETS to Artemis and 2 Mbps in the opposite direction with a total mass of 140 kg. The AlphaSat-TanDEM mission is a LEO-GEO mission. These satellites were launched in the end of 2012, which are aimed to demonstrate an optical link at rate of 1.8-2.8 Gbps over a range up to 45000 km. The JAXA future OISL mission is aimed to implement an optical link at the data rate of 2.5 Gbps with a total LEO mass of 35 kg [11]. More LEO-GEO missions are available in [4, 12].

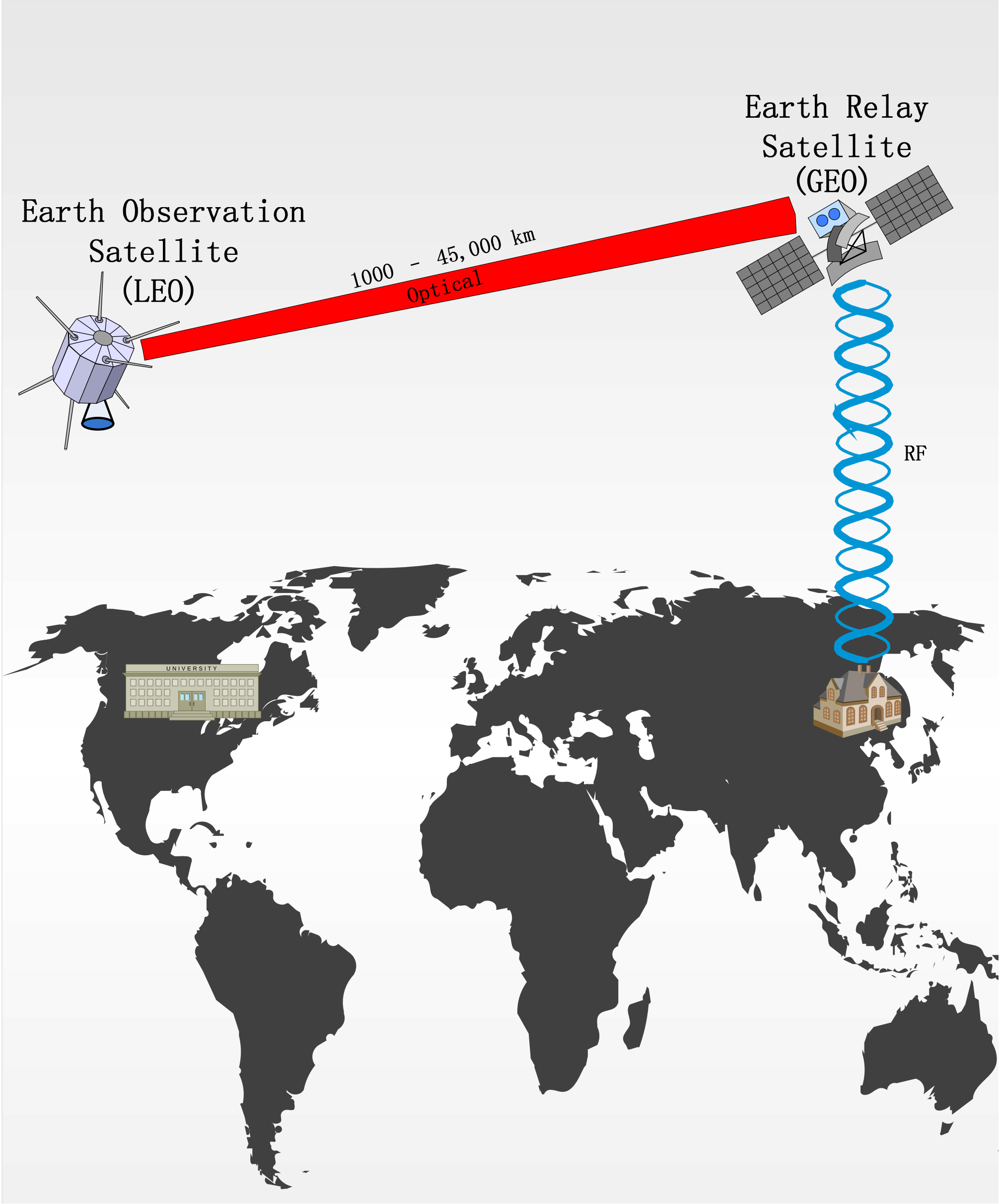


Figure 1.1: Concept of optical intersatellite communication including the Earth relay satellite.

The OISL also include LEO-LEO optical links. For example, in 1998, Motorola announced the Celestri mission, which is a LEO-LEO optical inter-satellite system and has a high average data throughput with the transmission data rate of 7.5 Gbps over a range of 1800-6000 km [13]. Recently, an optical link between two LEO satellites, TerraSAR-X and NFIRE, has been demonstrated at a rate of 5.625 Gbps over a range of 3800-4900 km with a telescope diameter of 12.5 cm, a total mass of 32 kg and a power consumption of less than 120 W for the entire transceiver [14].

1.2.2 Capacity Results

One of the most fundamental steps to design any efficient communication system is to compute the channel capacity. Channel capacity is defined as the maximum rate at which reliable communication can take place [15]. The capacity of a CTP channel is known under different constraints and has been studied extensively. Kabanov derived the capacity of the CTP channel under a peak power constraint in [16], and Davis obtained the channel capacity with a peak power constraint and average power constraint [17]. Frey extended these constraints to time-varying functions [18] and he also considered the case with an L-norm constraint [19]. Shamai in [20] considered the capacity of the CTP channel with constraints on the transmitted pulse width and in [21], Shamai and Lapidoth developed capacity bounds for general spectral constraints. It should be noticed that there is no bandwidth limitation for the input signal of the CTP channel, and to achieve the capacity of the CTP channel the bandwidth is necessarily infinite.

Currently no analytic expression is known for the capacity of the DTP channel. However, many capacity bounds have been derived. In [5,6], McEliece defined the DTP channel with zero dark current and developed several upper

Table 1.2: OISL Experiments [7].

Mission	**Date & Range**	**Data Rate / Architecture**	**Power & Mass**
SILEX	1997 LEO to GEO (SPOT IV to Artemis) 45,000 km	50 Mbps (LEO to GEO); 2 Mbps (GEO to LEO); 850 nm, telescope 25 cm, pointing accuracy 0.3-0.8 μrad, BER$< 10^{-6}$, 60 mW continuous transmit power.	180 W 160 kg
Celestri	1998 Specs only LEO-LEO 1800-6000 km	7.5 Gbps, throughput after overhead is 4.5 Gbps; BER$< 4 \times 10^{-10}$.	$<$100 W $<$ 25 kg
Artemis-OICETS (LUCE payload)	2005 LEO-GEO First Bi-directional link	50 Mbps (to Artemis) 847 nm, 200 mW laser diode, 26 cm telescope; 2 Mpbs (from Artemis); 819 nm, 2PPM.	220 W 140 kg (LUCE)
NFIRE-TerraSAR-X	2008 LEO-LEO 3800-4900 km	5.626 Gbps, 1.064 μm, 700 mW transmit power, FOV= 2 mrad, BER$< 10^{-9}$, 12.5 cm aperture (diameter)	120 W 32 kg
AlphaSat–Sentinel 1A and Sentinel 2A, and TanDEM-X	2012 LEO-GEO 45,000 km	1.8-2.8 Gbps; design BER$< 10^{-8}$, 13.5 cm telescope, 1064 nm, 5 W transmit power.	140 W 45 kg
JAXA future OISL mission	In Development LEO-GEO	2.5 Gbps, 1.064 μm, 4 W transmit power. Aperture diameter 10 cm (LEO) and 20 cm (GEO).	150 W 35 kg (LEO); 50 kg (GEO)

bounds under average and peak constraints. In [20], Shamai developed lower and upper bounds on the channel capacity of the DTP channel with binary inputs that are located at $\{0, A\}$ with average power constraint only. Lapidoth and Moser [22] derived analytical lower and upper bounds on the capacity of the DTP channel with dark current. These bounds are asymptotically tight as the peak and average powers tend to infinity, although they are often loose in the low input power range. In [23] lower and upper bounds on the DTP channel capacity are given asymptotically as the average power tends to zero with a fixed peak power. In [24], Martinez derived a class of tight lower bounds based on the mutual information induced by the Gamma distribution and non-asymptotic upper bounds via a duality method for the DTP channel with only an average power constraint and no dark current.

Comparatively little work has been done on the characterization of capacity-achieving distributions for the DTP channels. Note that the channel capacity can be easily computed once the capacity-achieving distribution is given; more importantly, the capacity-achieving distribution provides important guidelines for practical signalling design. In his seminal work [25], Smith studied the capacity-achieving distribution for peak amplitude and average power constrained additive Gaussian channels and provided a method to check the optimality of an input distribution via the Karush-Kuhn-Tucker (KKT) conditions. Following Smith's approach, Shamai [26] found that the capacity-achieving distributions of the DTP channel consist of finite number of mass points and proved that the capacity achieving-distribution is binary when dark current is absent and the peak power constraint in a certain range. Related work extending Smith's early approach to different but related channels can be found in [27–29].

1.2.3 Signalling Design

Direct detection (DD) receivers measure the intensity of the received field while ignoring the phase of the received field on the detector. Therefore, intensity modulation is necessary for optical systems utilizing DD. Pulse amplitude modulation (PAM) is one of the most popular intensity modulation schemes developed in optical communication systems [29]. In particular, on-off keying (OOK), which is a binary-level version of PAM, is widely used in optical systems [30, 31]. An alternative modulation method is called pulse position modulation (PPM), which has gained much attention due to its power efficiency [30, 32, 33]. Note that PPM can be considered as coded OOK.

Almost all existing PAM-based constructions are based on uniform signalling (i.e., the input symbols are used with the same frequency). However, it is well known [34, 35] that uniform signalling is only able to achieve the mutual information across the DTP channel induced by the uniform input distribution, which is in general strictly below the true channel capacity according to the existing capacity results [22, 23, 26, 36, 37]. On the other hand, non-uniform signalling design often involves sophisticated encoding and decoding operations and several different strategies have proposed. In [38], low-density parity check (LDPC) codes over GF(q) are designed together with deterministic mappers to induce the desired (non-uniform) distribution. The disadvantage of using non-binary LDPC codes is the higher decoding complexity, which limits the use of this technique. Alternatively, in this case where the capacity-achieving distribution is non-uniform, the channel capacity can be approached by adopting multi-level coding (MLC) and multi-stage decoding (MSD) with a mapper applied to multiple binary linear codes (e.g., binary LDPC codes) [39] . In particular, this method has been used to approach the capacity of terrestrial FSO channels with Gaussian noise [29]. However, in general, MLC/MSD suffers from error propagation, latency in decoding and requires multiple encoders and decoders.

1.2.4 Poisson Multiple Access Channel

Figure 1.2 demonstrates an illustrative diagram for the concept of a two-user multiple access optical communication between two LEOs and one GEO. Suppose both the LEOs are Earth observation satellites collecting data. These LEOs are only in the view of ground station a few minutes per orbit, therefore, the GEO aids in relaying the downlink data from LEO to ground. A multiple access channel model is suitable for use in the design of optical links in such scenarios.

Although no such system exists in practice, many LEO satellites currently operate within small distances of each other and could be imaged in the field-of-view (FOV) of the GEO to realize an optical multiple access channel (MAC). Table 1.3 presents a list of spacecrafts missions where several LEOs operate in close proximity. In 2007, the Orbital Express spacecraft successfully demonstrated autonomous rendezvous with a passive spacecraft with the minimum distance less than 120 m [40]. The TanDEM-X mission consists of 2 TerraSAR-X radar satellites flying in close formation with the distance less than 1 km [41, 42]. The recently launched PRISMA mission [43] is a demonstration mission which enables LEO formation flight with a passive target spacecraft within a minimum distance less than 10 km and an error less than 10 m. The ASPIICS/PROBA-3 mission [44] will demonstrate two high earth orbit (HEO) satellites formation flight within a distance of 150 m. The JC2SAT mission [45] will demonstrate two nano-satellites formation flying over a range of 100 m to 5 km. The Micro-Arc-second X-ray Imaging Mission (MAXIM) is a project to investigate the feasibility of achieving astronomical

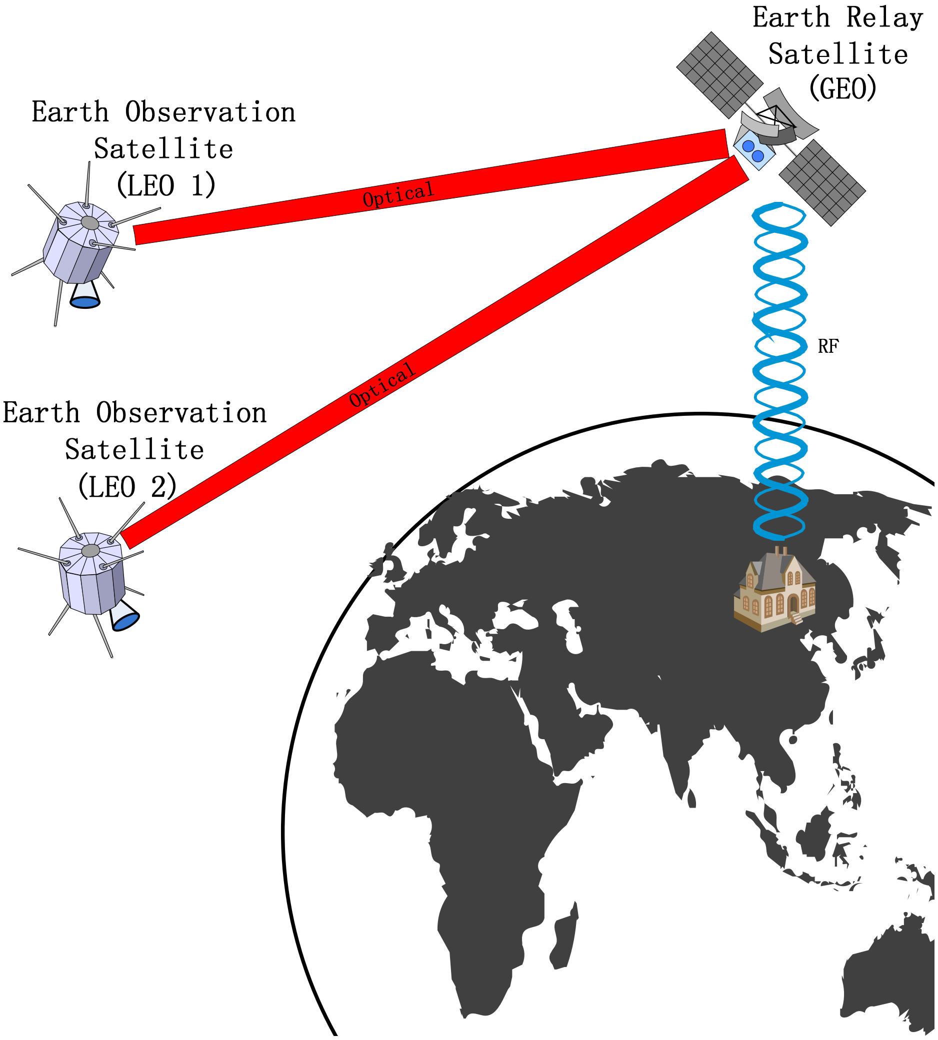


Figure 1.2: Concept of multiple access optical intersatellite communication between two LEOs and one GEO.

Table 1.3: Satellites & Spacecrafts Formation Flying Missions [12]

Mission	**Satellites & Spacecrafts amount**	**Orbit radius**	**Minimal Operation range**
Orbital Express Program, 2007 [40]	2, ASTRO and NextSat	492 km	<120 m
TanDEM-X, 2010 [41, 42]	2 TerraSAR-X satellites	514 km	<1000m
PRISMA, 2010 [43]	2, Mango and Tango	710 km	10 km
ASPIICS/ PROBA-3, T.B.D [44]	2 spacecraft	Perigee 800 km, apogee 60524 km,	150 m
JC2SAT, T.B.D [45]	2 Nanosatellites	650 km	100 m to 5 km
CanX-4/-5, T.B.D [46]	2 Nanosatellites	Perigee 550 km, apogee 900 km	50 m to 1000 m
The MAXIM X-Ray Interferometer Concept (Shepherd Satellite), T.B.D [47]	33 Nanosatellites	10 km	40 m

images using a formation flying x-ray interferometer. It consists of 33 x-ray mirror satellites, a trailing collector satellite, and a detector spacecraft [47]. These mirror satellites form a disc-shaped constellation with diameter 200 m and the distance between the detector spacecraft and the constellation is about 5000 km.

This thesis proposes a DTP MAC model, which could be taken as a modification of the existing CTP MAC. All these experiments listed in Table 1.3 operate within small minimum distances, therefore could adopt the DTP MAC model in their optical links. Currently, all the existing results relating to Poisson MAC are based on the CTP MAC model. The channel capacity for two-user CTP MAC was found in [48] and it is shown that code division multiple access (CDMA) is optimal in this case. The error exponent of continuous-time, two-user Poisson MAC model was found in [49]. The problem of the continuous model is that there is no bandwidth limitation on the input signal, and it is known that infinite bandwidth is required to achieve the maximum sum-rate (i.e, the mutual information between the two users and the output), which is not practical in use.

1.3 Thesis Contributions

The main contributions of this thesis are as follows: 1) analytical results on the characterization of the capacity-achieving distributions of the DTP channel, 2) algorithms for computing the capacity and the corresponding optimal input distribution, 3) signalling design, and 4) discreteness of the sum-capacity-achieving distributions of the DTP MAC under a peak amplitude constraint.

In particular, two algorithms are developed for computing the channel capacity of the DTP channel as well as the capacity-achieving distribution in the presence of dark current and under average and peak amplitude constraints [36, 37]. Although the capacity of this channel is unknown, numerical calculation of the channel capacity is implemented using a particle method and deterministic annealing (DA) algorithm. The particle method is typically used to compute the capacity value under fixed peak power and average power constraints, i.e., a single point on the capacity curve. In contrast, the DA method is capable of producing a segment of the capacity curve with fixed peak power constraint and with increasing average power constraint. Tight lower bounds based on input distributions with simple forms are developed [37]. These bounds provide extremely good approximations to the channel capacity, especially in the low power regime.

Motivated by the classical construction by Gallager [50], a non-uniform mapper is coupled to a binary LDPC code to generate the desired input distribution and a joint demapper/decoder is designed based on the sum-product algorithm [36]. This scheme has a lower complexity compared with to constructions based on non-binary LDPC codes. Unlike MLC/MSD which requires multiple encoders and decoders, this scheme requires a single encoder and decoder, and is therefore free of error propagation and has less latency in decoding. Simulations indicate near capacity performance of the proposed coding system and significant gains over information rates using traditional uniform signalling. A key observation of our work is that significant gains in rate can be achieved for the same average power consumption by using optical transceivers with non-uniform signalling and a modest increase in peak power.

Several fundamental properties of capacity-achieving distributions for the DTP channels are established [37]. In particular, it is demonstrated that all capacity-achieving distributions of the DTP channel have zero as a mass point. In the case of only a peak constraint, it is further shown that the optimal distribution always has a mass point at the maximum amplitude. Moreover, under solely an average power constraint, it is shown that a finite number of mass points are insufficient to achieve the capacity.

Though binary (i.e. two mass point) distributions have been observed through simulation to be optimal in many cases, very few analytical results exist on the optimality conditions or the form of optimal signalling. In this thesis, necessary and sufficient conditions on the optimality of binary distributions are derived by leveraging the general properties of the DTP capacity-achieving distributions. Furthermore, closed-form analytical expressions of the capacity-achieving distributions are derived in several important special cases including the zero dark current case and the high dark current case [51].

This thesis also proposes a two-user (additive-rate) DTP multiple access channel model, which can be viewed as a modification of the CTP MAC and as an extension of the single-user DTP channel. It is assumed that input signals for both users should be PAM with fixed intensity in each time interval. Also, due to practical limits, peak amplitude constraints are imposed on the input signals. It is shown that the sum-capacity-achieving distributions of the DTP MAC under peak amplitude constraints are discrete with a finite number of mass points [52].

1.4 Thesis Structure

The balance of this thesis is organized as follows.

Chapter 2 contains a detailed description of the DTP channel model, including the sources of the constraints and a typical example based on a satellite laser communication link. Two numerical methods to compute both the capacity and the capacity-achieving distributions for DTP channel are also presented, i.e., the particle method and the DA method. The particle method yields the optimal input distribution under given constraints for a single point on the capacity curve, while the DA method results in a segment of the capacity curve. Both of these methods rely on the Blahut-Arimoto algorithm [56] to update the probabilities of the mass points. A simple family of maxentropic input distributions is defined and used to develop tight lower bounds on the channel capacity through the evaluation of the envelope of information rates for different constellation sizes. A novel joint demapper/decoder using the sum-product algorithm is developed which requires a single encoder and decoder to implement the non-uniform signalling scheme. It is demonstrated that the achievable rates closely follow the channel capacity over a range of input powers.

Chapter 3 provides general results of the capacity-achieving and approaching distributions for DTP channels. It is shown that for an arbitrary discrete input distribution defined over a finite number of mass points, both shifting the constellation leftward and stretching the constellation with the corresponding probability unchanged increase the mutual information. Consequently, all capacity-achieving distributions of the DTP channel have zero as a mass point. Moreover, in the case of only a peak constraint, the optimal distribution always has a mass point at the maximum amplitude. It is also proved that a finite number of mass points are not sufficient to achieve the capacity of DTP channel under a solely average constraint. Necessary and sufficient conditions on the optimality of binary distributions are also established in this chapter, which are further leveraged to obtain closed-form expressions of the capacity-achieving distributions in several special cases, e.g., for zero dark current and for large dark current case. It is demonstrated that the binary maxentropic distribution is capacity-achieving when the dark current is large enough.

In Chapter 4, the two-user DTP-MAC model is precisely defined, involving the constraints and the channel law. It is shown that to achieve the maximum sum rate of a two-user DTP MAC with peak constraints, discrete input distributions are required.

Finally, Chapter 5 presents concluding remarks and suggestions for future work.

Chapter 2

Capacity Computation and Non-Uniform Signalling for Discrete-Time Poisson Channels

2.1 Introduction

Currently, closed-form analytical expressions for the capacity of DTP channel under peak and average power constraints remain unknown in general. It is of practical importance to develop numerical methods to compute such values. The numerical computation of both the channel capacity and the optimal distribution are useful for system implementation such as code design. These numerical results, as well as tight lower and upper bounds, serve as guidelines for system design.

In this chapter, two numerical algorithms to compute both the channel capacity and the capacity-achieving distribution are developed. Tight lower bounds based on simple closed-form input distributions are also presented.

These bounds are close to the capacity especially during the low input power range. Non-uniform signalling techniques, which achieve higher rates compared to uniform schemes, are shown based on the capacity and the developed bounds.

2.2 Channel Model

2.2.1 Channel Description

In discrete-time Poisson channels, data are transmitted by sending PAM intensity signals which are constant in discrete time slots. In contrast to continuous-time Poisson channels, which admit arbitrary waveforms, the discrete-time Poisson model imposes a bandwidth limit by constraining transmitted signals to be rectangular PAM. The PAM amplitudes are limited to $\mathbb{R}^+$ since the underlying quantity modulated is the optical intensity.

Although intersatellite links operate above the atmosphere, unintended light scattered from the Earth as well as from other planets and stars will impinge on the receiver [8]. Dark current represents the detector non-ideality and corrupts the received counts even in the absence of illumination [53, Chap. 5]. Dark current arises in all photodetectors and is a fundamental limitation on the performance of any optical receiver.

The receiver is modelled as a photon counter which generates an integer representing the number of received photons. Specifically, in each time slot, ΔT, a channel input intensity x [photons/second] is corrupted by the combined impact of background radiation and photodetector dark current at a rate of λ [photons/second]. The channel output y [photons], is a random value related to the number of received photons in ΔT and obeys a Poisson distribution with mean $(x+\lambda)\Delta T$ [1]. Since the intensities, i.e., photon arrival rates, sum at the receiver, the model is termed *additive rate*. Notice that this model is only appropriate for IM/DD optical intensity channels since the phase information at the receiver is ignored.

Without loss of generality, assume $\Delta T = 1$ and accordingly

$$P_{Y|X}(y|x) = \frac{(x+\lambda)^y}{y!} e^{-(x+\lambda)}, \quad x \in \mathbb{R}^+,\ y \in \mathbb{Z}^+. \tag{2.1}$$

Furthermore, due to device constraints and limited energy storage at the transmitter (e.g., a satellite in deep-space laser communications), there is an average power constraint,

$$\mathbb{E}(X) \leq \varepsilon, \qquad \text{[Average Power Constraint]}. \tag{2.2}$$

Also, due to the dynamic range limitation of the transmitter, the peak intensity must also be constrained,

$$0 \leq X \leq A, \qquad \text{[Peak Power Constraint]}. \tag{2.3}$$

Without loss of generality, it is assumed that $0 \leq \varepsilon \leq A$.

Notice that the constraints in this model can also be relaxed to yield DTP channels with only average power constraint (i.e., $A \to \infty$) or only peak power constraint (i.e., $\varepsilon = A$).

The channel capacity, C, of a discrete-time Poisson channel is the maximum mutual information over input distributions satisfying channel constraints, namely,

$$\begin{aligned} C &\triangleq \max_{F_x \in \mathcal{F}} I(X;Y) \\ &= \max_{F_x \in \mathcal{F}} \int_x \left[\sum_y P_{Y|X}(y|x) \log \frac{P_{Y|X}(y|x)}{P_Y(y)} \right] dF_x, \end{aligned} \tag{2.4}$$

where

$$\mathcal{F} \triangleq \left\{ F_x(x) : \int_0^A dF_x = 1, \mathbb{E}_{F_x}\{X\} \le \varepsilon \right\}.$$

It is well known that the capacity-achieving distributions are discrete and in general have a finite number of mass points for finite A and ε [26]. Therefore, there is no loss of generality in considering an input distribution defined over constellation $\psi_x = \{x_1, x_2, \ldots, x_n\}$, $0 \le x_1 < x_2 < \ldots < x_n \le A$, with corresponding probability masses $\psi_p = \{p_1, p_2, \cdots, p_n\}$. Let F_x denote the cdf of the input, that is

$$dF_x = p_1\delta(x - x_1) + p_2\delta(x - x_2) + \ldots + p_n\delta(x - x_n), \tag{2.5}$$

where $\delta(\cdot)$ denotes the Dirac impulse functional. $\mathbb{E}$ is the expectation operator, $P_Y(\cdot)$ denotes the corresponding distribution on the channel output and $P_{Y|X}(\cdot|\cdot)$ denotes the channel law.

Finally, define $\psi_x^*(A, \varepsilon)$, $\psi_p^*(A, \varepsilon)$, $F_x^*(A, \varepsilon)$ to be the corresponding optimal values under constraints A and ε.

2.2.2 A Typical Example: LEO Intersatellite Link

Consider an example of a LEO laser communication link demonstrated between TerraSAR-X and NFIRE satellites [14]. This link operates at a data rate 5.625 Gbps over a link distance 3800-4900 km. Table 2.4 provides a list of parameters for these terminals [14] and realistic values for the link.

A simplified link budget analysis can be used to estimate the average number of received signal photons for a link at wavelength λ_w over a range z with signalling interval T as [54]

$$\begin{aligned} \varepsilon &= P_T \eta_T \eta_R \eta \left(\frac{\lambda_w}{4\pi z} \right)^2 G_T L_T G_R T \frac{\lambda_w}{hc} \\ &= 7.5027, \end{aligned} \tag{2.6}$$

Table 2.4: Terminal Characteristics

Wavelength, λ_w	1064 nm
Data Rate, $1/T$	5.625 Gbps
Link Distance, z	4900 km
Peak Transmit Power, $2P_T$	700 mW
Transmitter Aperture Diameter, d_t	125 mm
Transmitter Optical Efficiency, η_T [1]	0.5
Receiver Aperture Diameter, d_r	125 mm
Receiver Optical Efficiency, η_R [2]	0.35
Detector Quantum Efficiency, η [3]	0.7
Pointing Error, θ_T	5 μradian (rms)
Spectral Radiance of Earth, W [4]	5×10^{-3} Watt/cm$^2 - \mu$m$-sr$
Receiver Field-of-View, Ω [5]	2 mrad
Bandwidth of Receiver Filter, $\Delta\lambda$	2 nm

[1] Assume 0.5, [54]

[2] Assume 0.35, [54]

[3] Assume 0.7, [54]

[4] This value is corresponds to the wavelength 1μm [1]

[5] Typically, this value is between $1.7 - 2.2$ mrad [14]

where P_T is the average transmitter power, η_T and η_R are efficiencies of the transmitter and receiver optics, η is the quantum efficiency of the detector, d_r and d_t are the aperture diameters of transmitter and receiver, h is Planck's constant and c is the speed of light in a vacuum.

The receiver antenna gain is

$$G_R = \left(\frac{\pi d_R}{\lambda_w}\right)^2, \tag{2.7}$$

while the gain of the transmit antenna relative to an isotropic emitter is

$$G_T = \left(\frac{\pi d_t}{\lambda_w}\right)^2. \tag{2.8}$$

The loss due to pointing error θ_T is estimated by

$$L_T = \exp(-G_T \theta_T^2). \tag{2.9}$$

The primary noise source of the receiver is assumed to arise from the sun and scattered light from bodies such as the moon and Earth. The average number of background photons received per signalling interval can be estimated as [1]

$$\begin{aligned} \lambda &= W(\lambda_w)\pi \left(\frac{d_r}{2}\right)^2 (\Delta\lambda)\, \Omega^2 \frac{\pi}{4} T \frac{\lambda_w}{hc} \\ &= 3.6649, \end{aligned} \tag{2.10}$$

where $\Delta\lambda$ the bandwidth of the receiver filter and Ω is the receiver field-of-view. The product, $\Omega^2\pi/4$, can be taken as the range of receiver arrival angles observed by the detector area [1]. The factor $W(\lambda_w)$ is the spectral radiance of Earth, which is mainly due to atmospheric scattering excited by the moon and galactic sources [1, 55].

2.3 Particle Method

2.3.1 Particle Method for DTP Capacity Computation

The Blahut-Arimoto algorithm [56] can be used to find the channel capacity and input distribution for constrained channels where input and output are chosen from discrete finite sets. In [57], the algorithm is extended to channels with continuous input distributions by discretizing them into a list of points termed particles.

In this section, the techniques in [56, 57] are adapted to the discrete-time Poisson channel with peak and average amplitude constraints to compute tight bounds on the capacity and to find the capacity-achieving input distribution. An advantage of this approach is that it is able to produce accurate estimates of the channel capacity even for large values of ε and A. Previous approaches using brute-force optimization techniques suffer from very large dimensionality for large ε and A and take excessive amounts of computing time.

Consider the input probability density for X in (2.5) ,

$$dF_x = \sum_{i=1}^{n} p_i \delta(x - x_i),$$

where p_i, $i = 1, \ldots, n$, are real and non-negative with $\sum_{i=1}^{n} p_i = 1$ and $x_i \in [0, A]$. The value of n must be chosen large enough to ensure the convergence of the algorithm as discussed in [57].

The optimization problem (2.4) is solved iteratively where $\{(x_i^{(k)}, p_i^{(k)})\}$ denotes the list of particles at the k-th step.

The list of particles is alternately updated using the following two steps:

$$p^{(k)} \triangleq \arg\max_{p} I(\{(x^{(k-1)}, p)\}), \qquad \text{(W-step)} \tag{2.11}$$

$$x^{(k)} \triangleq \arg\max_{x} I(\{(x, p^{(k)})\}). \qquad \text{(X-step)} \tag{2.12}$$

The W-step (2.11) optimizes the weights p with the positions $x^{(k-1)}$ fixed and can be accomplished by the constrained Blahut-Arimoto algorithm [56] with average constraint ε. The X-step (2.12) maximizes $I(\{(x_i, p_i)\})$ by optimizing the positions with the weights fixed. Practically the X-step is accomplished by means of a steepest ascent technique [57].

After k iterations, a lower bound on the capacity C can be shown to be

$$C \geq L^{(k)} = I(\{(x_i^{(k)}, p_i^{(k)})\}), \tag{2.13}$$

while an upper bound on C is given by

$$C \leq U^{(k)} = \max_{x \in [0,A]} \left[D(P_{Y|X}(y|x) || \hat{P}(y)^{(k)}) - s^{(k)} x \right] + s^{(k)} \sum_{i=1}^{n} p_i^{(k)} x_i^{(k)}, \tag{2.14}$$

where $s^{(k)}$ is a parameter set to ensure convergence [56].

The convergence of this particle method is guaranteed by the W-step and X-step, since both yield non-decreasing mutual information. However, the convergence point may not be globally optimal. This is because in the X-step, the result of the gradient descent method might be only locally optimal. In addition, empirically n is selected to be large enough for convergence. Although currently the global convergence of this particle method is still unsolved theoretically, the optimality of the final result could be verified through the KKT conditions, which are introduced in Theorem 5 of Section 3.2.

2.3.2 Numerical Results and Analytical Bounds

Figure 2.1 shows $L^{(k)}$ and $U^{(k)}$ for the discrete-time Poisson channel with $A/\varepsilon = 4$ and $\lambda = 3$ as a function of the average input power. The value of n should be large enough to ensure the convergence of the algorithm, and in these simulations $n = 200$. It can be seen that both bounds nearly coincide over a wide range of powers. The gap between $L^{(k)}$ and $U^{(k)}$ is nearly 10^{-5}

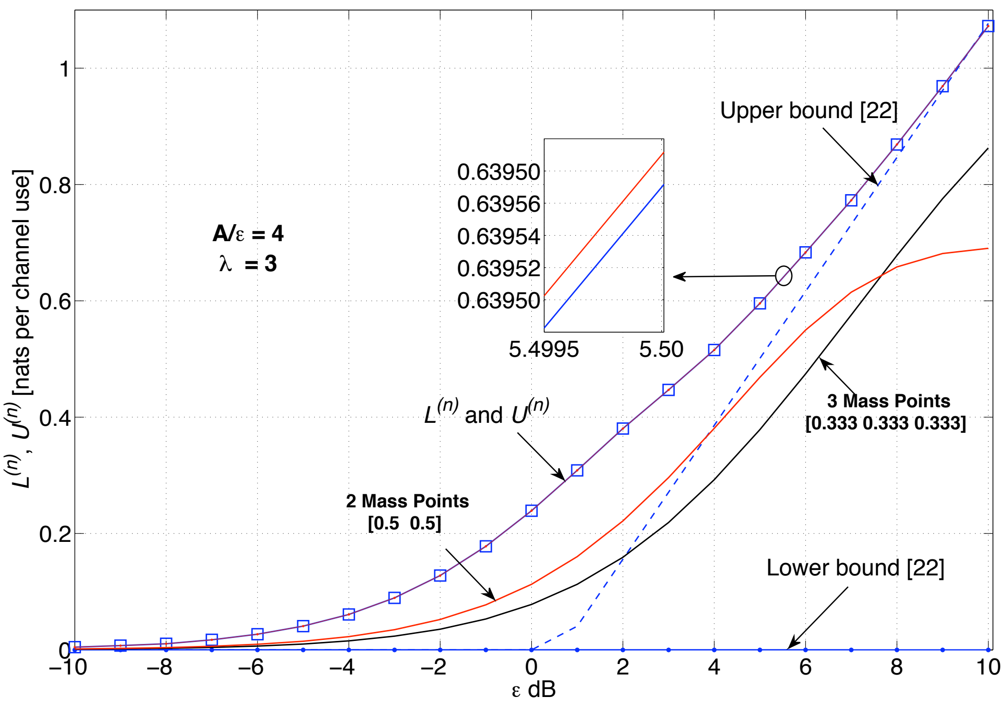


Figure 2.1: Bounds $L^{(k)}$, $U^{(k)}$ on channel capacity (on top of each other) and closed form bounds from [22] with $A/\varepsilon = 4$ and $\lambda = 3$.

nats per channel use after 100 iterations and the accuracy can be improved further by increasing the number of iterations. The lower and upper bounds of Lapidoth and Moser [22, Eq. (12), Eq. (13)] are also presented for comparison. Notice that due to their asymptotic nature, these bounds yield no insight at lower power levels. In Fig. 2.1 the mutual information for uniform two and three point constellations are also plotted. There is a large gap in the mutual information using equiprobable constellations and the channel capacity, which demonstrates the importance of non-uniform signalling for discrete-time Poisson channels.

Figure 2.2 shows the capacity-achieving distributions over ε with $A/\varepsilon = 4$ and $\lambda = 3$. As noted in [26], the capacity-achieving distribution for the

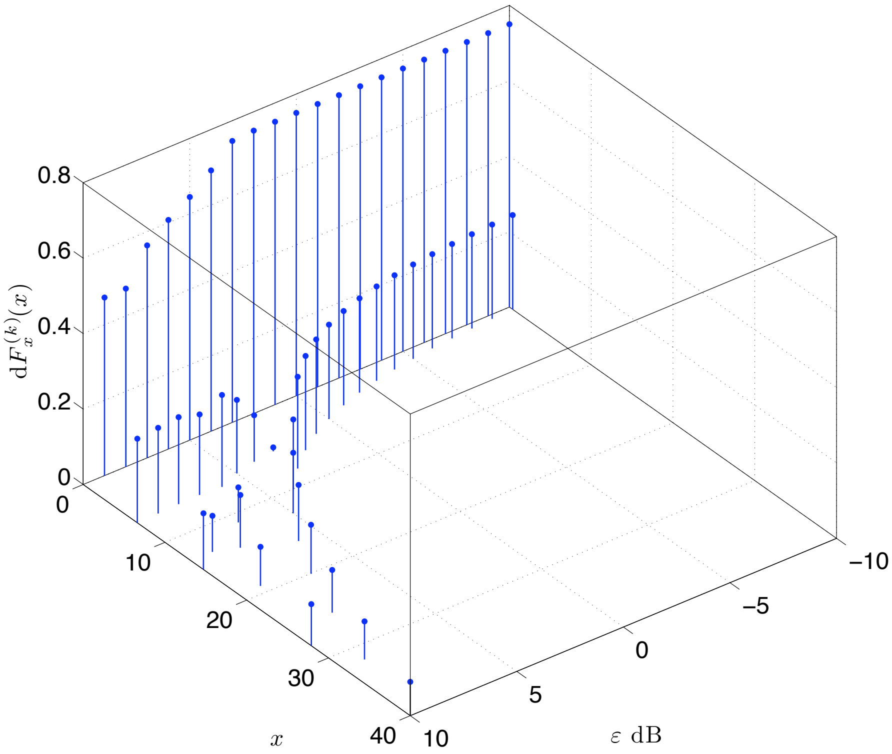


Figure 2.2: Capacity achieving input distributions for $A/\varepsilon = 4$ and $\lambda = 3$.

discrete-time Poisson channel with average power and peak power constraint is discrete. Notice also that there are always probability mass points at $x_i = 0$. In Corollary 2 of next chapter, it is demonstrated that the capacity-achieving distribution for the DTP channel under average and peak power constraints always contains a mass point located at 0. Additionally notice that when $\varepsilon < 3$ dB, the capacity-achieving distributions are non-uniform binary distribution.

Figure 2.3 plots similar mutual information curves with peak-to-average ratio $A/\varepsilon = 2$ which is common in many optical transceivers. The analytical upper and lower bounds of Lapidoth and Moser [22, Eq. (18), Eq. (19)] under the same conditions are also presented. Again, the analytical bounds only yield insight for very high values of ε. Notice that equiprobable signalling achieves rates close to the capacity. Indeed, for $\varepsilon < 6.4$ dB the capacity-achieving

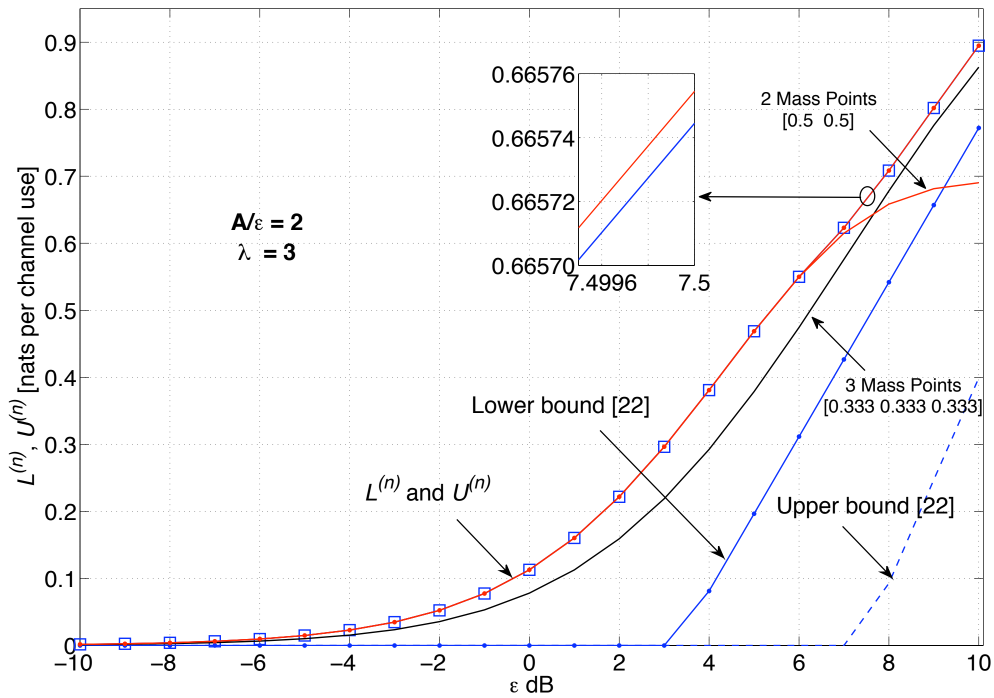


Figure 2.3: Bounds $L^{(k)}$, $U^{(k)}$ on channel capacity (on top of each other) and closed form bounds from [22] with $A/\varepsilon = 2$ and $\lambda = 3$.

distribution is binary and nearly uniform. Thus, non-uniform signalling is not essential in the case of $A/\varepsilon = 2$ to approach capacity. However, comparing Figs. 2.1 and 2.3 illustrates that for a given average power consumption, large increases in channel capacity are available by increasing the peak emitted power. For spacecraft applications, ε is a metric of the lifetime of the batteries. Thus, building optical transceivers with higher peak powers and non-uniform signalling can deliver far higher rates for the same average power consumption.

2.4 Deterministic Annealing Algorithm

2.4.1 Deterministic Annealing Algorithm for DTP Capacity Computation

The particle method is able to compute the capacity of the DTP channel under fixed average and peak power constraints as well as the associated optimal input distribution, however, it is computationally intensive as it requires an initial discrete distribution with large enough cardinality (usually hundreds) to ensure convergence.

In this section, a more computationally efficient algorithm is proposed based on the *deterministic annealing* (DA) method. The DA method was originally developed to compute the rate-distortion function [58], which provides a strong motivation for our work due to the similarity between the computation of channel capacity and rate-distortion function. Unlike the particle method which only generates a single point on the capacity curve, the DA method results in a discretized segment of the capacity curve. Therefore, it would be helpful to adapt this DA method to analyse how the capacity evolving with increasing average power.

According to [56, Corollary 9], a parametric expression of channel capacity in terms of s for a given A is given by

$$C(\varepsilon) = s\varepsilon + \max_{Q(x|y)} \sum_{x} \exp\left(\sum_{y} P_{Y|X}(y|x) \log Q(x|y) - sc(x) \right), \tag{2.15}$$

where $C(\varepsilon)$ is some point on the capacity curve parameterized by s, $Q(x|y)$ is the conditional distribution of $X = x$ given $Y = y$, with $c(x)$ being the cost associated with symbol x. For the DTP channel, we have $c(x) = x$. The parameter s can be interpreted as the slope of the capacity curve at a given average power, ε and a fixed peak amplitude and has a similar physical meaning as β in [58].

Define the cost function κ as

$$\kappa \triangleq \max_{Q(x|y)} \sum_{x} \exp\left(\sum_{y} P_{Y|X}(y|x)\log Q(x|y) - sx\right). \tag{2.16}$$

In contrast with the case of computing the rate-distortion function where the cost function is minimized [58], the cost function (2.16) is maximized in the computation of channel capacity.

Figure 2.4 presents a sketch of the DA algorithm. The input to the algorithm are A and λ. The algorithm initializes with a binary distribution ($n = 2$) and s set to some positive value. Through the annealing process, the value of s decreases to 0 gradually. For every fixed s, the cost function (2.16) is deterministically maximized through updating the probability masses and positions of the input symbols recursively. The input probability is updated based on the Blahut-Arimoto rule (2.17)-(2.19), while the positions of mass points are updated via a gradient descent technique (2.20). A similar gradient descent method has been used to improve clustering performance through deterministic annealing [60] and in our previous work [36]. It should be mentioned that, although simple to implement, the gradient descent technique is not guaranteed to converge to the global optimal solution.

More mass points may be needed during the annealing process as s decreases. To change the value of n, a *two-symbols-one-location strategy* is adapted from [58, Sec. VI]. In this technique two mass points are assigned to each location, each carrying half the probability mass. During the iteration of algorithm, the points may stay merged at the same location or diverge depending on the phase transition condition. Constellation points that approach to within a small distance of each other (e.g. 10^{-3}) are merged into a two co-located mass points with the same total probability. Notice that some of the duplicate mass points may also diverge away from each other during the gradient descent phase. For example, assume for a given s, $n = 2$. However, if the corresponding optimal distribution is ternary, then one of the mass points will diverge away from its pair during step (2.20). The algorithm will then store 6 points before next iteration. For further details on this technique, the reader is referred to [58].

The result of the DA algorithm has not been proven to necessarily converge to the global optimum (i.e., capacity-achieving) distribution. Convergence of the Blahut-Arimoto algorithm is guaranteed and the convergence of the gradient descent algorithm to a local optimum can be realized through the proper selection of the step size θ_k in (2.20) (e.g., according to the Armijo rule [59]). Although not necessarily optimum, the output of the DA algorithm can be tested for optimality against the KKT conditions in Theorem 12. In the numerical results that follow, in practice all of the outputs of the DA algorithm satisfy Theorem 12 are are thus *capacity-achieving*.

The convergence of this DA method to a local optimum is guaranteed by similar reasoning as the particle method introduced in Section 2.3. However, the global optimality of the numerical results of the DA algorithm can be verified through the KKT conditions (Theorem 5) in Chap. 3.

2.4.2 Simulation Results Using DA

Figure 2.5 plots the channel capacity curves for increasing ε, with the peak power constraint $A = 100$ and for $\lambda = 0$ and $\lambda = 10$. For comparison, the results obtained via the particle method [36] are also presented. From the simulation results, it is apparent that both of these algorithms yield the channel capacity as well as the capacity-achieving distribution. Notice, however, that

```
Input: $A$, $\lambda$
Output: A discretized segment of the capacity curve
Initialization (typical values): $s = 1$, $n = 2$, $x_1 = 0$, $x_2 = 0.01$, $p(x_1) = p(x_2) = 0.5$, $\omega = 10^{-5}$,
$\delta = 10^{-2}$, $N_{BA} = N_{GD} = 20$ ;
repeat
    Initialize: $k = 1$, $\varepsilon^{(0)} = 0$;
    repeat
        $k = k + 1$;
        [Blahut-Arimoto] for $N_{BA}$ iterations do
            For every mass point, update $p(x_i)$ according to rules:
            $Q^{(k)}(x_i|y_j) = \frac{p(x_i^{(k-1)})P(y_j|x_i^{(k-1)})}{\sum_{i'} p(x_{i'}^{(k-1)})P(y_j|x_{i'}^{(k-1)})},$ (2.17)
            $p^{(k)}(x_i) = \frac{\exp\left(\sum_j P(y_j|x_i^{(k-1)}) \log Q(x_i^{(k-1)}|y_j) - s x_i^{(k-1)}\right)}{\sum_i \exp\left(\sum_j P(y_j|x_i^{(k-1)}) \log Q(x_i^{(k-1)}|y_j) - s x_i^{(k-1)}\right)},$ (2.18)
            $p^{(k)}(y_j) = \sum_i p(x_i^{(k-1)})P(y_j|x_i^{(k-1)}),$ (2.19)
        end
        [Gradient Descent] for $N_{GD}$ iterations do
            Update all $x_i$ with appropriate choice for step-size $\theta_k$
            $x_i^{(k)} = x_i^{(k-1)} + \theta_k \frac{\partial}{\partial x_i} \left( \sum_y P_{Y|X}(y|x_i) \log Q(x_i|y) - s x_i \right) \Big|_{x_i = x_i^{(k-1)}}$ (2.20)
        end
        Compute $\varepsilon^{(k)} = \sum_i x_i^{(k)} p(x_i^{(k)})$ ;
    until $|\varepsilon^{(k)} - \varepsilon^{(k-1)}| \leq \omega \vee k \geq 100$ ;
    $C = \sum_i p(x_i) \sum_j P(y_j|x_i) \log \left( \frac{P(y_j|x_i)}{p(y_j)} \right).$
    Apply the "two-symbols-one-location strategy" ;
    Re-initialize: $s = (1 - \delta)s$ ;
until $s \leq 10^{-5}$ ;
```

Figure 2.4: Algorithm for the deterministic annealing algorithm

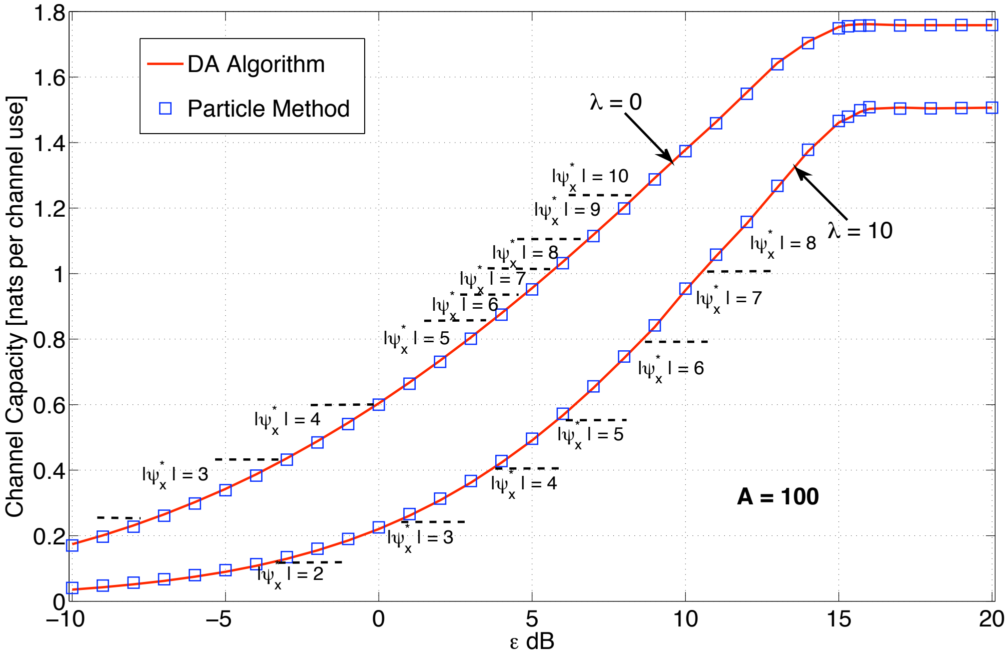


Figure 2.5: Channel capacity when $A = 100$, $\lambda = 0$ and 10 with various ε. The dash lines indicate the transition points of $|\psi_x^*|$, the number of mass points in the capacity-achieving distribution.

the DA algorithm generates a discretized segment of the capacity curve rather than discrete points.

Consider the case of $A = 100$ and $\lambda = 10$ in Fig. 2.5 where $|\psi_x^*|$ denotes the total number of the mass points in the capacity-achieving distribution. For $\varepsilon < -4$ dB, the optimal input distribution is binary and as ε increases, so too does $|\psi_x^*|$. However, for $\varepsilon > 11$ dB, the capacity-achieving distribution has $|\psi_x^*| = 8$ points. Notice also that for $\varepsilon > 16$ dB, neither the capacity-achieving distribution nor the channel capacity changes with increasing ε. This saturation in capacity with increasing ε indicates that the average power constraint becomes ineffective and the capacity is limited by the peak power constraint. Notice that this phenomenon occurs for $\varepsilon >\approx A/2 = 17$ dB.

The case of peak power constraint only and $\lambda = 0$ was treated by Shamai in [26, Eq. (14)], where it is claimed that the capacity-achieving distribution is

$$dF_x^* = (1-\beta_1-\beta_2)\delta(x) + \beta_1\delta(x-0.3839A) + \beta_2\delta(x-A) \tag{2.21}$$

in the region $3.3679 \leq A < \phi$, for some $\phi > 3.3679$. In other words, the constellation of the capacity-achieving distribution scales linearly with A. However, after extensive simulation, it can be shown that for ternary capacity-achieving distributions, (2.21) is only capacity-achieving with $A = 3.3679$ but not for larger values of A. In particular, for $\lambda = 0$, $A = 4$ and only peak power constraint, the capacity-achieving distribution is

$$dF_x^* = 0.4927\delta(x) + 0.0768\delta(x-1.4033) + 0.4305\delta(x-4),$$

where the middle point is not located at $x = 4 \times 0.3839 = 1.5356$.

2.4.3 Comparison of DA and Particle Method

There are many similarities between the particle method and the DA method. For example, both algorithms update the input probabilities through the Blahut-Arimoto algorithm for fixed constellations, and both adapt the gradient decent technique to update the positions of the input constellations under the resulting probability. Simulations indicate that both the particle method and the DA method are capable of providing precise capacity results under given constraints for DTP channels.

It is also beneficial to compare the computational complexity of these two algorithms. Suppose for any discrete input distribution defined as (2.5), let $B(n)$ and $G(n)$ denote the number of operations that are required to implement the Blahut-Arimoto method and the gradient decent method in one iteration, respectively. It is not difficult to see that $B(n)$ and $G(n)$ increase linearly in n. For computing a single capacity point corresponding to (A, ε), the particle method approximately needs $K(B(n) + G(n))$ operations, where n is the number of mass points used and K is the number of iterations, while the number of steps the DA method needs to run is approximately upperbounded by $NK(B(2n') + G(2n'))$ and lowerbounded by $\max\{\sum_{i=2}^{n'} K(B(2i) + G(2i)), NK(B(4) + G(4))\}$, where n' is the number of constellation points used in the final step of the DA method (which is in general significantly smaller than n) and N is the number of points on the capacity curve computed by the DA method before the given average power constraint is reached (note that the number of mass points in the DA method is twice the number of constellation points due to the "two-symbols-one-location strategy").

Generally speaking, compared with the particle method, the DA method has the advantage of computing a discretized segment of the capacity curve of the DTP channels with a fixed peak power constraint and increasing average power constraints. However, for computing a single point on the capacity curve (i.e., for fixed ε and A), the DA method may not be a suitable choice since it still needs to initialize with some small $\varepsilon' < \varepsilon$, for which the capacity-achieving distribution is binary, then increases ε' and the size of the input constellation gradually until $\varepsilon' = \varepsilon$. Therefore, in this setting, the particle method, which is designed specifically for computing a single point on the capacity curve, is in general more advantageous. However, the DA method is still a desirable choice when $2n'N$ is smaller than n.

2.5 Maxentropic Capacity-Approaching Distributions

2.5.1 Maxentropic Capacity-Approaching Distributions

Although there are many algorithms for computing both the capacity and the optimal input distributions for DTP channels (e.g., the particle method in Sec. 2.3 and DA method in Sec. 2.4), often it is instructive to develop simple closed-form input distributions which give tight lower bounds on the capacity.

Define a family of distributions, termed *maxentropic*, which have equally spaced mass points in $[0, A]$ and which maximize entropy subject to average and peak power constraints. In particular, the maxentropic distributions are given as [29]

$$dF_x^{\dagger}(K) = \sum_{k=0}^{K} \bar{p}_k \delta\left(x - k\frac{A}{K}\right),$$

where

$$\bar{p}_k = \frac{1}{K+1}, \qquad A \leq 2\varepsilon. \tag{2.22}$$

$$\bar{p}_k = \frac{t^k}{1+t+t^2+\ldots+t^K}, \qquad A \geq 2\varepsilon. \tag{2.23}$$

and t is some number between 0 and 1. These distributions are in fact *Boltzmann distributions* which are widely applied in statistical physics and clustering [61] and have also been applied in Gaussian noise-corrupted optical channels [29].

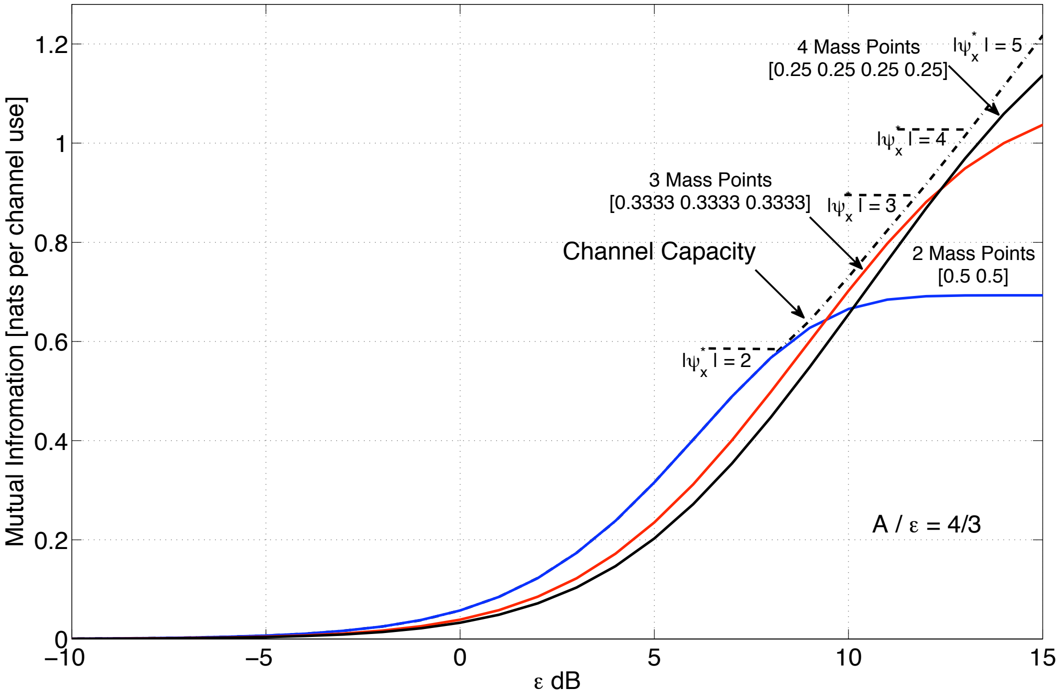


Figure 2.6: Mutual information induced by the maxentropic input distributions and the number of mass points versus ε when $A/\varepsilon = 4/3$ and $\lambda = 3$. For comparison, the channel capacity and the number of mass points in the capacity-achieving distribution, $|\psi_x^*|$, are also provided.

2.5.2 Performance of Maxentropic Distributions on DTP Channels

The information rates induced by the maxentropic distributions are shown in Figs. 2.6 and 2.7 for different A/ε and $\lambda = 3$. The channel capacity in both cases are included for comparison and computed through the deterministic annealing method of Section 2.4.1.

It can seen from the figures that the performance of the maxentropic distributions is very close to the channel capacity and even capacity-achieving in some cases. At low ε, the binary maxentropic distribution

$$dF_x^\dagger(1) = \left(1 - \frac{\varepsilon}{A}\right)\delta(x) + \frac{\varepsilon}{A}\delta(x - A) \tag{2.24}$$

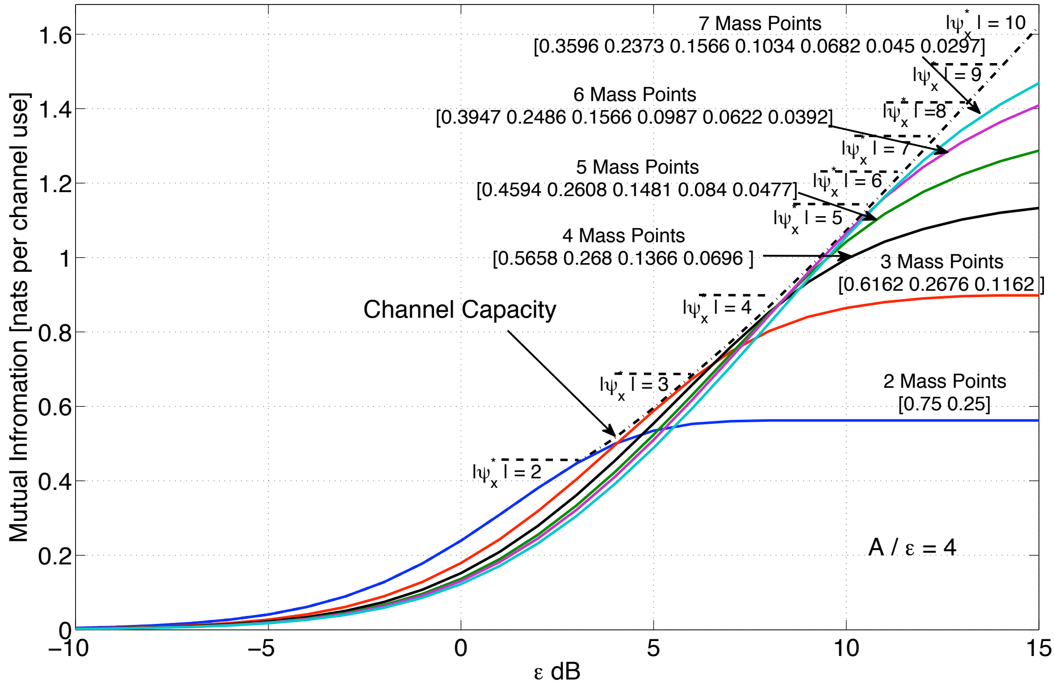


Figure 2.7: Mutual information induced by the maxentropic input distributions and the number of mass points versus ε when $A/\varepsilon = 4$ and $\lambda = 3$. For comparison, the channel capacity and the number of mass points in the capacity-achieving distribution, $|\psi_x^*|$, are also provided.

is capacity achieving in both figures. As ε increases, the number of mass points, K, required to approach the capacity also increases. In fact, the mutual information induced by the maxentropic distributions is often negligibly far from the capacity for high ε with less mass points than the optimal case. For example, in Fig. 2.6 when $\varepsilon > 12.5$ dB, a $K = 3$ maxentropic distribution is close to the channel capacity while $|\psi_x^*| = 5$ points are in the optimal distribution. Similarly, in Fig. 2.7, a $K = 6$ maxentropic distribution approaches the capacity where up to 8 mass points are needed when $\varepsilon > 12.5$ dB.

Therefore, the envelope of the information rates of maxentropic distributions forms a close approximation to the channel capacity of DTP channels. The simple form of the maxentropic distributions also makes them a useful first step in non-uniform signalling design for DTP channels and often has the practical benefit of having near-capacity performance with less mass points.

2.6 Non-Uniform Signalling Design for Discrete-Time Poisson Channels

2.6.1 Motivation for Non-Uniform Signalling

As discussed, the example in Fig. 2.1 in Sec. 2.3 indicates a large gap between the mutual information using uniform signalling and the channel capacity, and demonstrates the importance of non-uniform signalling for DTP channel. In addition, based on the practical link budget in Sec. 2.2.2, the capacity of the LEO-LEO intersatellite link is computed following Sec. 2.3 using the estimated $\varepsilon = 7.5027$ and $\lambda = 3.6649$ and shown in Fig. 2.8 versus the peak constraint A. As a metric of comparison, the mutual information rate achieved by conventional uniform binary signalling scheme with the same ε

and λ is computed to be $C_0 = 0.6718$ nats per channel. Fig. 2.8 also presents the rate gain available by using non-uniform signalling over C_0.

Notice that even for $A/\varepsilon = 2$ there is a gain in rate of 12% over the baseline uniform binary scheme C_0[1]. Thus, non-uniform signalling is useful in improving the data rate in all cases in Fig. 2.8. Increasing the available peak power by 50% to about 1 W gives an $A/\varepsilon = 3$ and yields a 30% gain in rate over conventional binary uniform signalling for the same average power. Further increasing the peak constraint improves the capacity with smaller relative increases in rate. Thus, using laser emitters with larger peak powers and non-uniform signalling can yield impressive gains in the channel capacity of the LEO intersatellite link while keeping the average power consumption constant. That is, this improvement in rate does not come at the expense of increased average energy usage.

The remainder of this section considers practical algorithms to approach these large gains in rate by employing non-uniform signalling.

2.6.2 Practical Constellation Design: Constrained Particle Method

The resulting capacity-achieving distributions obtained in Sec. 2.3 and Sec. 2.4 have arbitrary probability mass values which may not be practical for code design. Simply quantizing the optimal distribution does not take full advantage of the average power constraint.

Consider a *constrained particle method* which incorporates the quantization levels into distribution design and is defined as follows:

1. Choose a large enough n and run the W-step and X-step of Sec. 2.3

---

[1] In fact, in this case the capacity-achieving distribution is non-uniform and ternary.

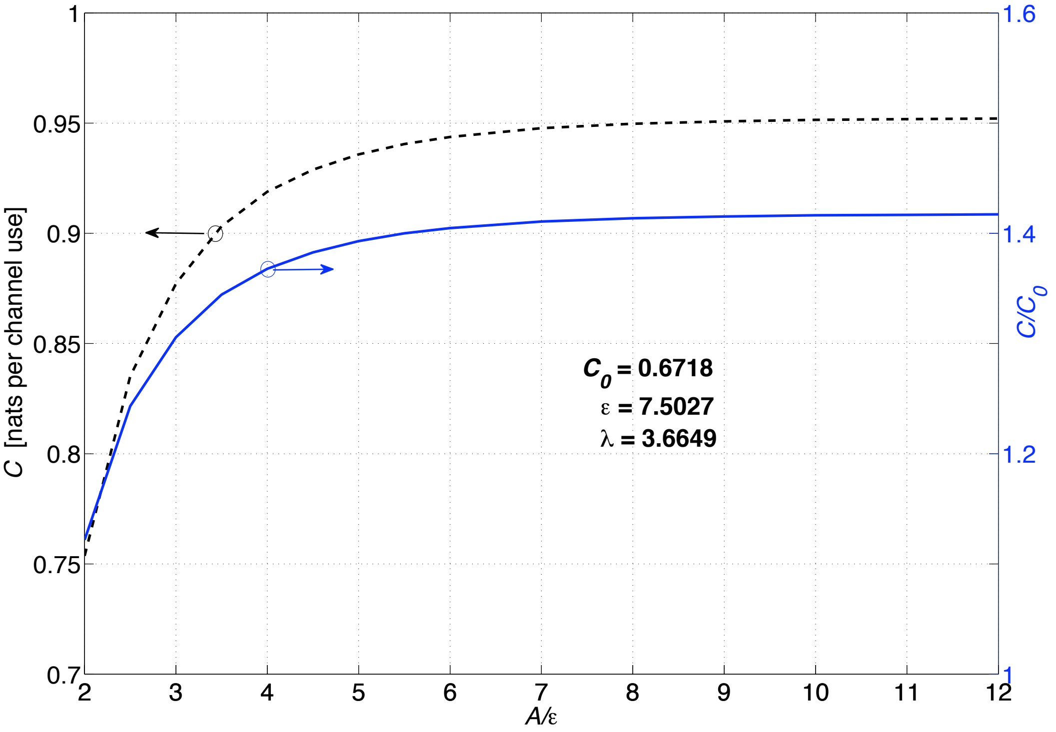


Figure 2.8: Channel capacity for LEO intersatellite link versus A/ε with $\varepsilon = 7.5027$ and $\lambda = 3.6649$. For comparison, the gain in rate versus uniform binary signalling, C_0, is also presented.

iteratively until convergence to yield the capacity-achieving distribution $\{(x_i^*, p_i^*)\}$.

2. Select $N \in \mathbb{Z}^+$ according to the system requirements. In general, larger N provide more precise quantization results at the expense of complexity.
3. Enumerate all possible distributions of the form $\{(x_i^*, \hat{p}_i)\}$ where $\hat{p}_i = k/2^N$ for $k \in \{0, 1, \ldots, 2^N - 1\}$. Notice that $\sum \hat{p}_i = 1$. Denote the collection of all such distributions by $\hat{\mathcal{P}}$ and $\hat{\mathcal{P}}_\varepsilon \subseteq \hat{\mathcal{P}}$ as the collection which satisfy the average amplitude constraint.
4. If $|\hat{\mathcal{P}}_\varepsilon| > 0$ (i.e., at least one combination satisfies the average power constraint), choose the element in $\hat{\mathcal{P}}_\varepsilon$ which has the smallest Kullback-Leibler (K-L) divergence to $\{(x_i^*, p_i^*)\}$. Then, run the X-step (2.12) under the average constraint.
5. Else, no elements in $\hat{\mathcal{P}}$ satisfy the average constraint. Choose the distribution in $\hat{\mathcal{P}} - \hat{\mathcal{P}}_\varepsilon$ which has the smallest K-L divergence to $\{(x_i^*, p_i^*)\}$ and denote it $\{(\tilde{x}_i, \tilde{p}_i)\}$. Scale this distribution as $\{(\alpha\tilde{x}_i, \tilde{p}_i)\}$, where $\alpha = \varepsilon / \sum \tilde{x}_i \tilde{p}_i$ to ensure the average constraint is satisfied.

Notice that the resulting source distribution satisfies all channel constraints and has quantized probability mass values. Steps 4 and 5 are the most costly in terms of time complexity but guarantee that the algorithm will converges to an optimal point.

The mutual information of the constrained particle method with $N = 2$ and 3 as a function of $1/\lambda$ for fixed ε and A is shown in Fig. 2.9. For comparison, the channel capacity computed by the particle method of Sec. 2.3 is also presented. The largest gap between the mutual information and channel capacity is approximately 0.02 nats/channel use. Notice that as λ decreases,

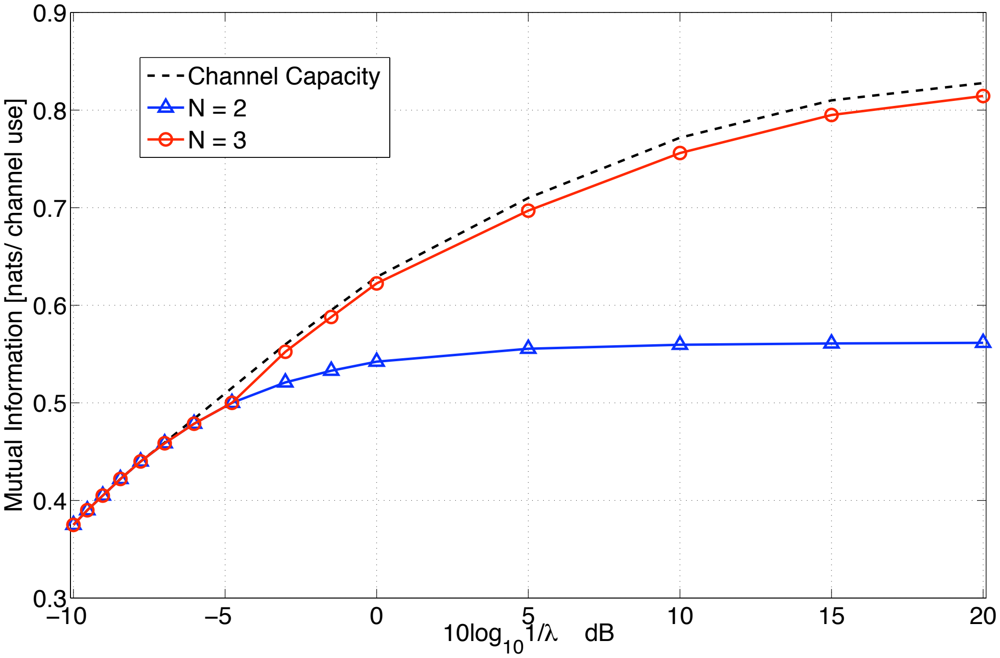


Figure 2.9: Channel capacity and mutual information for constellations from the constrained particle method for $\varepsilon = 4$ dB and $A/\varepsilon = 4$.

constellations with more quantization levels are required to approach the channel capacity.

Fig. 2.10 plots the capacity-achieving input distributions and the results of the constrained particle method for $\varepsilon = 4$ dB and $A/\varepsilon = 4$ with $\lambda = 10, 4, 0.1, 0.01$, respectively. When $\lambda = 10$, the capacity-achieving distribution and the constrained signal constellation coincide. Notice also that for $\lambda = 4$ the output of the constrained method results in fewer mass points than the capacity achieving distribution. Thus, the constrained technique often produces a less complex transmitter with fewer output amplitudes while remaining very close to the channel capacity.

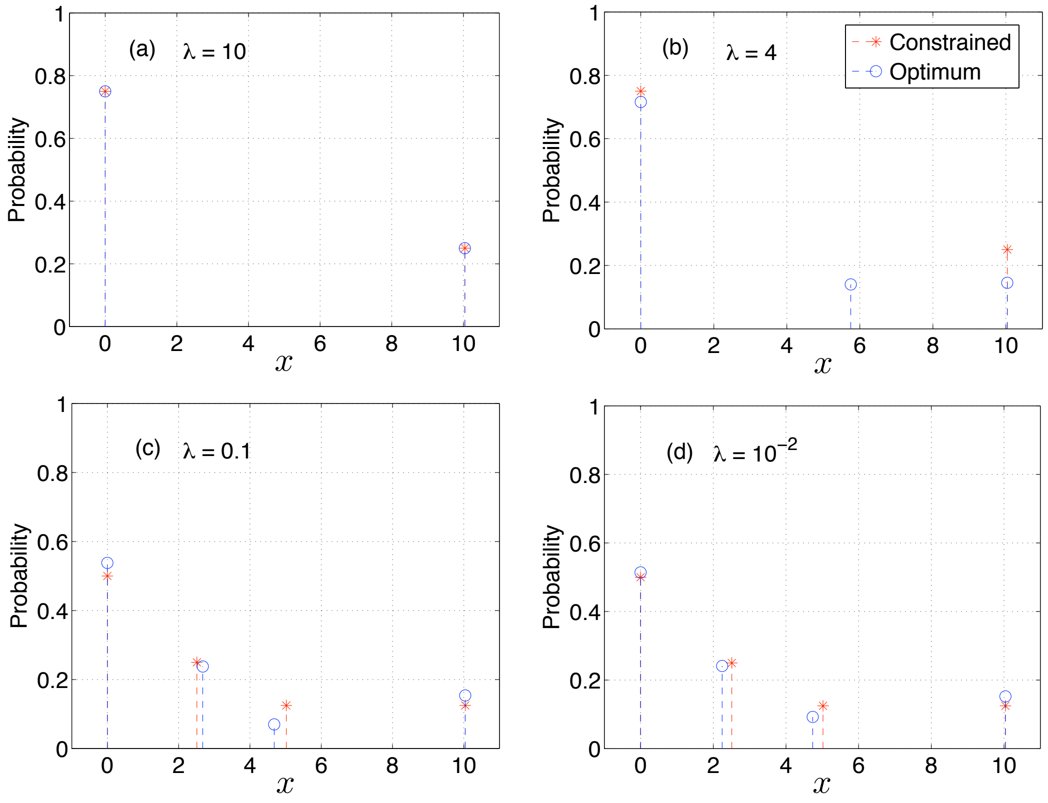


Figure 2.10: The optimum and the proposed input distribution for different λ when $\varepsilon = 4$ dB and $A/\varepsilon = 4$. For (a) and (b) $N = 2$ and for (c) and (d) $N = 3$.

2.6.3 Coding and Non-Uniform Signalling

As seen in earlier sections, to approach the capacity of the discrete-time Poisson channel, signalling at discrete amplitudes with non-uniform probabilities is necessary. In previous work on related channels [29, 39], a mapper is used to induce the correct distribution and coupled with multilevel coding (MLC) and multi-stage decoding (MSD) to approach capacity. In general, however, MLC/MSD suffer from error propagation, latency in decoding and require multiple encoders and decoders.

In this work, a single code is used to encode all bits and the mapper obtained from the constrained particle method is implemented to induce the correct distribution. At the receiver demapping and decoding are done jointly via the sum-product algorithm.

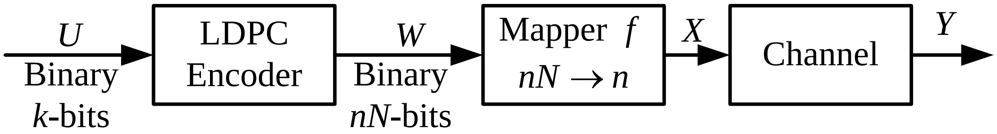


Figure 2.11: System model for the developed encoding method and mapping scheme

Fig. 2.11 presents a block diagram of the encoder. Let the message U be composed of k bits assumed to be uniformly distributed and input to the low-density parity check (LDPC) encoder. Define the length of the LDPC code to be nN, where 2^N is the number of quantization steps in the constrained particle method. The value n is an integer selected so that the capacity $C > k/n$. Additionally, group the output coded bits as $\left(W_1^{(i)}, W_2^{(i)}, \ldots, W_N^{(i)}\right)$, for $i = 1, 2, \ldots, n$. Notice that since the LDPC code is a linear block code, the output distribution of the symbols in W can be assumed to be uniform. Let $f : \{0, 1\}^N \to [0, A]$ be a deterministic mapper which induces the desired distribution as determined by the constrained particle method in Sec. 2.6.2. This mapper is straightforward to implement since all probability masses are constrained to be of the form $k/2^N$.

Thus, each block of N coded bits, indexed by i, is input to the mapper to yield a single channel input X_i.

2.6.4 Code Design: Example I

Consider channel constraints $A/\varepsilon = 4$ and $\lambda = 3$. From Fig. 2.2, it is apparent that for the range $-10 \leq \varepsilon \leq 2$ dB that the capacity-achieving distribution has two mass points at $\{0, A\}$ and $p_0 = 3/4$. In this example, the encoding, mappping and joint demapping/decoding processes are described in detail and their performance simulated.

2.6.4.1 Encoding and Mapping

For $N = 2$ and assuming uniformly distributed input bits, the mapper f induces the desired distribution

$$\left(W_1^{(i)} W_2^{(i)}\right) \xmapsto{f} X : X = \begin{cases} A, & W_1^{(i)} = W_2^{(i)} = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{2.25}$$

The equivalent channel seen by bit W_1 (and W_2 due to the symmetry of the mapper) can be found by marginalizing the conditional probability

$$\begin{aligned} P_{Y|W}(y|w_1 = 1) &= \sum_{w_2} P_{Y|X}(y, w_2|w_1 = 1) \\ &= \frac{1}{2} P_{Y|X}(y|A) + \frac{1}{2} P_{Y|X}(y|0), \\ P_{Y|W}(y|w_1 = 0) &= P_{Y|X}(y|0), \end{aligned}$$

where $P_{Y|X}(\cdot|\cdot)$ is the channel law.

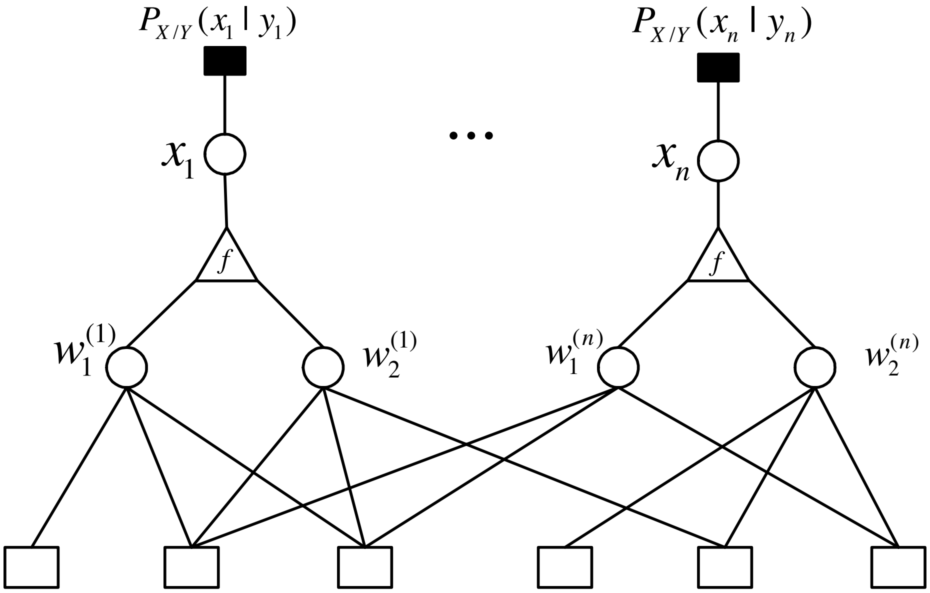


Figure 2.12: Developed factor graph for joint demapping and decoding

2.6.4.2 Joint Demapping and Decoding

Consider representing the LDPC code and the mapper together in a factor graph. An example for $N = 2$ with the mapper (2.25) is presented in Fig. 2.12. Message passing on this graph using the sum-product algorithm can demap and decode the bits jointly.

Notice that the lower part of the graph represents a traditional LDPC code and the mapping function f is represented by the triangular nodes. Furthermore, both $w^{(i)}$ and x_i are binary in this example. Following the standard sum-product algorithm [62], the message from the mapper to the message bit $w_1^{(i)}$ is

$$\begin{aligned}\mu_{f\to w_1^{(i)}}(w_1^{(i)}=1) &= \mu_{x_i\to f}(x_i=A)\mu_{w_2^{(i)}\to f}(w_2^{(i)}=1)\\ &\qquad + \mu_{x_i\to f}(x_i=0)\mu_{w_2^{(i)}\to f}(w_2^{(i)}=0),\\ \mu_{f\to w_1^{(i)}}(w_1^{(i)}=0) &= \mu_{x_i\to f}(x_i=0)\mu_{w_2^{(i)}\to f}(w_2^{(i)}=1)\end{aligned}$$

$$+ \mu_{x_i \to f}(x_i = 0)\mu_{w_2^{(i)} \to f}(w_2^{(i)} = 0).$$

An analogous message from f to $w_2^{(i)}$ is also simple to derive based on (2.25). Notice that the mapper may induce loops with small girth. In practice, 50 iterations of message passing on the LDPC graph are done between each iteration of messages from the mapper, $\mu_{f \to w}$.

For this example, the message from x_i to f can be written as the log-likelihood ratio

$$m_{x_i \to f} = \ln \frac{\mu_{x_i \to f}(x_i = 0)}{\mu_{x_i \to f}(x_i = 1)} = \ln \frac{P(x_i = 0|y_i)}{P(x_i = 1|y_i)} = \ln \frac{3P(y_i|x_i = 0)}{P(y_i|x_i = A)}. \tag{2.26}$$

All other message passing for the LDPC code takes place in the standard manner [63]. Due to the symmetry of the mapper in $w_1^{(i)}$ and $w_2^{(i)}$, the update rules for both are the same. After several rounds of message passing, a hard decision is made for each $w^{(i)}$.

2.6.4.3 Simulation on BER Performance

Notice that the previous discussion of the joint demapping/decoding technique depends only on the particular mapper chosen. In order to have a concrete example, referring to Fig. 2.1, the channel capacity when $\varepsilon = -1.21$ dB, $A/\varepsilon = 4$ and $\lambda = 3$ is approximately 0.2438 bits (0.169 nats) per channel use. As shown in Fig. 2.2, the capacity-achieving distribution in this case has two amplitude points at 0 and A and has probability mass $p_0 = 0.75$ corresponding to the previously developed mapper (for $N = 2$).

In order to realize the code design for this system, an LDPC code with rate $R = 0.12$ bits/channel use is required since two encoded symbols are mapped to a channel symbol. An LDPC code with rate 0.12 bits/channel use is designed using [64] for an AWGN to yield the degree distributions

$$\lambda(x) = 0.5513x + 0.2031x^2 + 0.0917x^4 + 0.0045x^6$$

$$
\begin{aligned}
& +0.017x^7 + 0.0995x^8 + 0.033x^9, \\
\rho(x) &= x^2.
\end{aligned}
$$

The total length of the code is set to 10000 bits, which corresponds to 5000 transmitted channel symbols.

The bit error rate (BER) performance of the system is shown in Fig. 2.13 versus $1/\lambda$ for ε and A fixed. The figure indicates the point corresponding to $\lambda = 3$ which was used for design. Notice that the BER drops as $1/\lambda$ increases. For comparison, a uniform distribution which satisfies the same average power constraint is also considered. At $1/\lambda = 1.31$ the information rate using uniform signalling is 0.24 bits/channel use which is identical to the designed rate. Clearly, uniform signaling is quite far from the channel capacity and a non-uniform signaling scheme, such as the one presented here, is a required to take full advantage of discrete-time Poisson channels.

2.6.5 Code Design: Example II

Consider the design of a non-uniform mapper and coding scheme under the conditions for a LEO intersatellite link described in Sec. 2.2.2 (i.e., $\varepsilon = 7.5027$ and $\lambda = 3.6649$). A value of $A/\varepsilon = 2.62$ is selected since, from Fig. 2.8, the channel capacity is approximately 25% greater than the uniform signalling case (i.e., 1.22 bits/channel use). Notice that this is a mild increase in the peak-to-average ratio over the uniform system which inherently has $A/\varepsilon = 2$. The goal of this example is to quantify the practical gains in rate which can be realized by exploiting the small increase in peak amplitude for the same ε and λ.

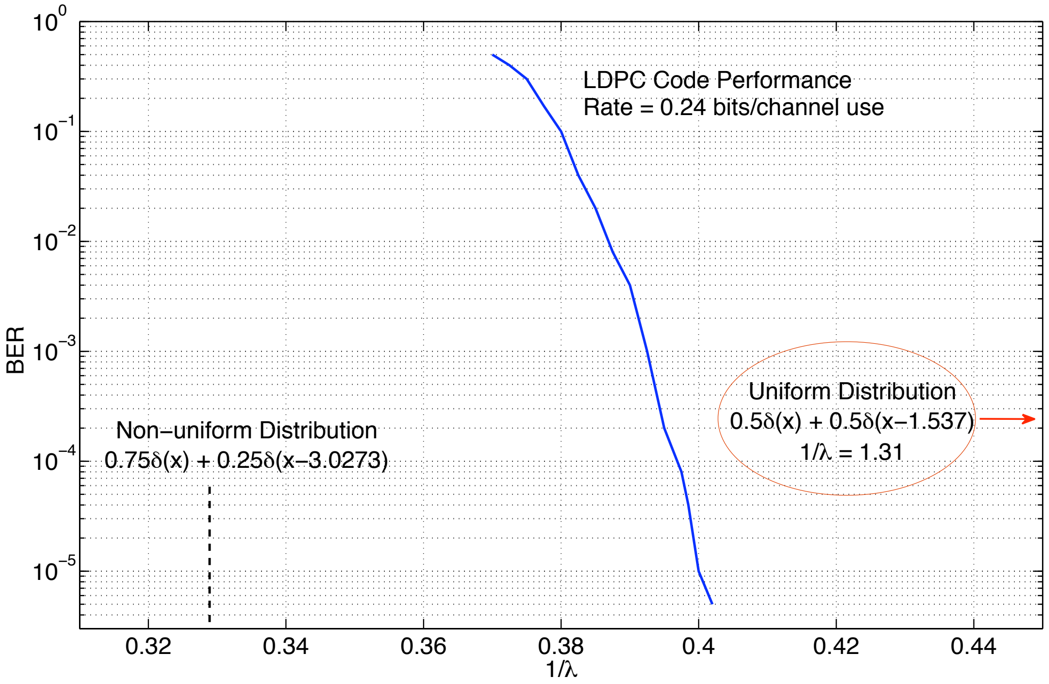


Figure 2.13: BER versus $1/\lambda$ for the non-uniform signalling using finite length LDPC codes for $\varepsilon = -1.21$ dB , $A/\varepsilon = 4$. The value of $1/\lambda$ corresponding to optimal uniform (out of range) and non-uniform signalling at 0.24 bits/channel use is presented for comparison.

2.6.5.1 Encoding and Mapping

For $\varepsilon = 7.5027$, $\lambda = 3.6649$ and $A/\varepsilon = 2.62$, the channel capacity is 1.22 bits/channel use, which is achieved by the input distribution

$$\begin{aligned} dF^*(x) = 0.4444\delta(x) + 0.2001\delta(x - 7.1835) \\ + 0.0883\delta(x - 9.2315) + 0.2672\delta(x - 19.6503). \end{aligned}$$

Setting $N = 2$ in the constrained particle method results in the following distribution

$$\hat{p}^*_X(x) = 0.5\delta(x) + 0.25\delta(x - 8.76) + 0.25\delta(x - 19.6505), \tag{2.27}$$

with the mutual information rate 1.20 bits/channel use.

Notice that this distribution can be induced through the simple mapper f

$$\left(W_1^{(i)} W_2^{(i)}\right) \overset{f}{\longmapsto} X : X = \begin{cases} A_0, & W_1^{(i)} = 0 \\ A_1, & W_1^{(i)} = 1,\ W_2^{(i)} = 0 \\ A_2, & W_1^{(i)} = 1,\ W_2^{(i)} = 1 \end{cases} \tag{2.28}$$

where $A_0 = 0$, $A_1 = 8.76$ and $A_2 = 19.6505$.

The equivalent channel seen by bit W_1 and W_2 as well as the message passing rules can be found by simple extension of the results in Sec. 2.6.4.

2.6.5.2 Simulation on BER Performance

Since the information rate with the input (2.27) when $\lambda = 3.6649$ is 1.20 bits/channel use, an LDPC code with rate $R = 0.568$ bits/channel use and degree distributions [65]

$$\begin{aligned} \lambda(x) &= 0.181804x + 0.197579x^2 + 0.011671x^3 \\ &\quad +0.098834x^4 + 0.063856x^5 + 0.239152x^{24} \end{aligned}$$

$$
\begin{aligned}
& +0.207105x^{25}, \\
\rho(x) \quad = \quad & 0.839350x^{10} + 0.160650x^{11},
\end{aligned}
$$

is applied in the this system with a code length of 10000 bits.

Fig. 2.14 plots the BER of the coding system with mapper (2.28) versus $1/\lambda$. Notice that the BER of the joint coding/mapping system is less than 10^{-5} for the target $\lambda = 3.6649$ computed in the LEO link budget (Sec. 2.2.2). Thus, for same ε and λ the resulting system has a rate which is 17% larger than that of uniform signalling (i.e., $C_0 = 0.969$ bits/channel use). This corresponds to a data rate of 6.6 Gbps. It must be noted, however, that this increase in rate is achieved at the expense of an increase in peak-to-average ratio from 2 to 2.62.

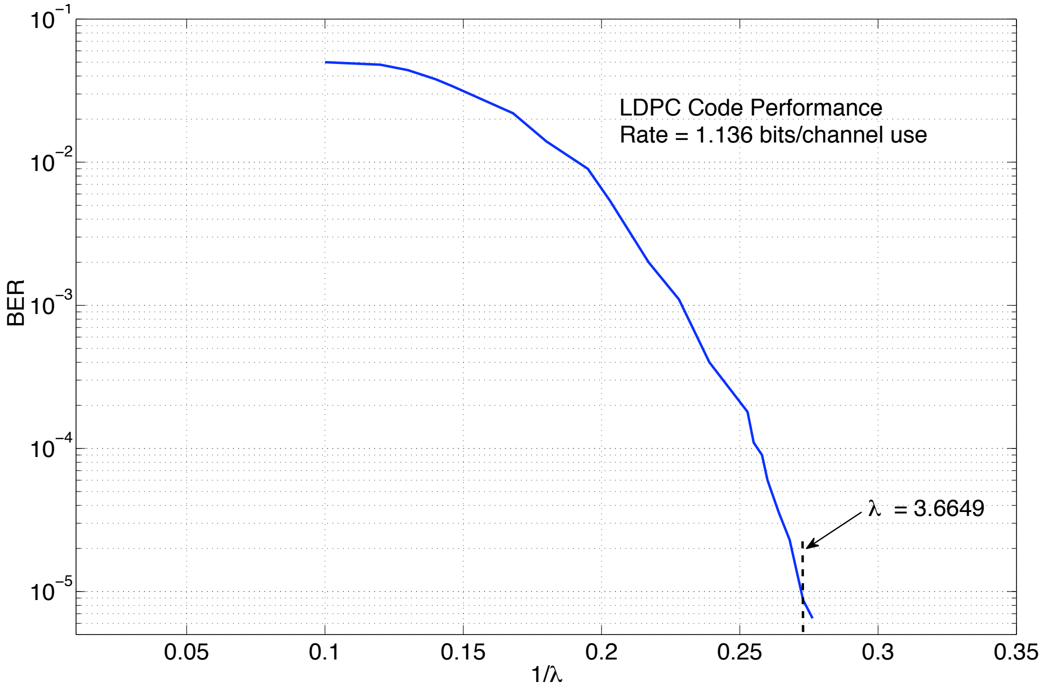


Figure 2.14: BER versus $1/\lambda$ for the non-uniform signalling using finite length LDPC codes with $R = 1.136$ bits/channel use as $\varepsilon = 7.5027$, $A/\varepsilon = 2.62$.

2.6.6 Code Design: Example III

In some cases to approach the channel capacity an $N = 3$ encoder and decoder are necessary. For example, from Fig. 2.9, for $\varepsilon = 2.51$, $A/\varepsilon = 4$ and $\lambda = 2.59$, the channel capacity is 0.7668 bits per channel use. Applying the $N = 3$ constrained particle method yields the following distribution

$$p\left(0\right) = \frac{5}{8},\ p\left(\frac{A}{2}\right) = \frac{2}{8},\ p\left(A\right) = \frac{1}{8}, \tag{2.29}$$

and the corresponding mutual information is 0.75 bits/channel use.

Symbols are drawn three at a time from an LDPC code with rate $R = 0.25$ bits/channel use and applied to the mapping function f, defined in (2.30), to yield a channel symbol. An LDPC code of length 12000 bits with rate $R = 0.25$ bits per channel use and the degree distributions [65]

$$\begin{aligned}\lambda(x) &= 0.602x + 0.238x^2 + 0.0309x^3 \\ &\quad +0.021x^4 + 0.0491x^5 + 0.0141x^6 \\ &\quad +0.0208x^7 + 0.023x^8 + 0.001x^9, \\ \rho(x) &= 0.0001x + 0.1017x^2 + 0.8982x^3,\end{aligned}$$

is combined with the mapper f in (2.30). The equivalent channel law as well as the messages passed during the sum product algorithm can be derived in a similar fashion to those in the example in Sec. 2.6.4.

$$W = [W_1 W_2 W_3] \overset{f}{\longmapsto} X : X = \begin{cases} A_0, & W_1 = 0. \\ A_0, & W_1 = 1, W_2 = 1, W_3 = 0, \\ A_1, & W_1 = 1, W_2 = 0, W_3 = 0, \\ A_1, & W_1 = 1, W_2 = 0, W_3 = 1, \\ A_2, & W_1 = W_2 = W_3 = 1. \end{cases} \tag{2.30}$$

The BER performance of the system is shown in Fig. 2.15 versus $1/\lambda$ for ε and A fixed. For the non-uniform distribution (2.29), at $1/\lambda = 0.389$, the information rate is identical to the design rate, however, in order to implement the rate 0.75 bits/channel use with the BER less than 10^{-5}, $1/\lambda = 0.89$ is needed for this system. The information rate with the input (2.29) and $1/\lambda = 0.89$ is about 0.886 bits/channel use. A uniform distribution with the same average power constraint has information rate equal to 0.75 bits/channel use when $1/\lambda = 1.068$. Therefore, the practical coding scheme illustrated here is reliable at a higher value of λ than the optimal uniform signalling scheme satisfying the same average optical power constraint.

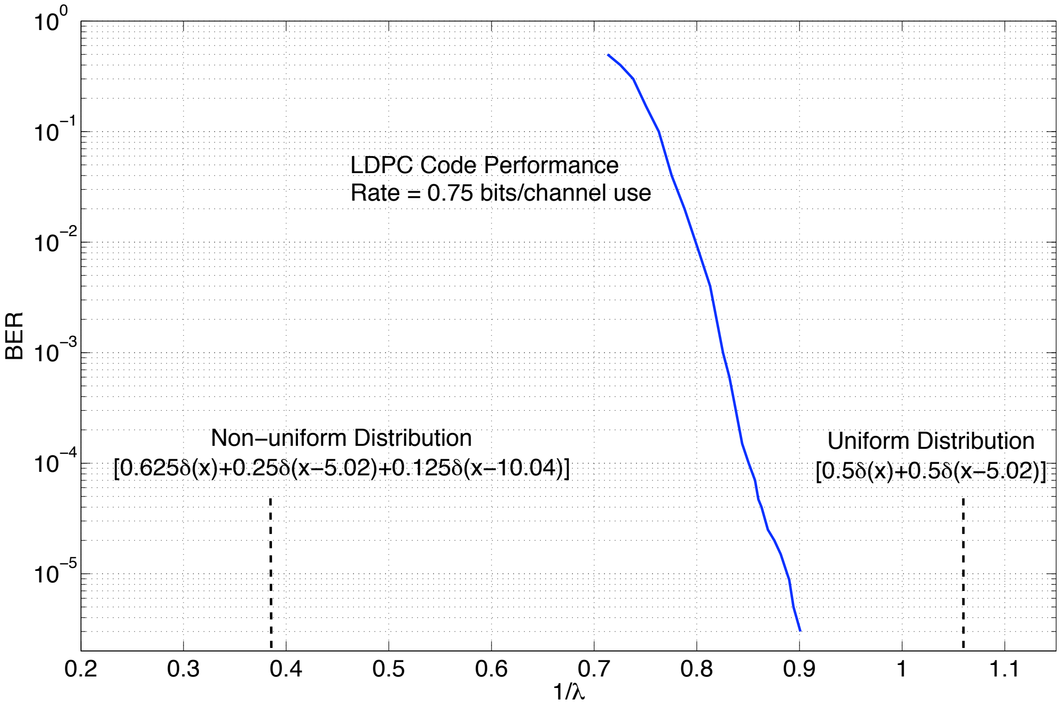


Figure 2.15: BER versus $1/\lambda$ for the non-uniform signalling using finite length LDPC codes as $\varepsilon = 4$ dB , $A/\varepsilon = 4$. The $1/\lambda$ corresponding to optimal uniform signalling and non-uniform signalling with $R = 0.75$ bits/channel use is presented for comparison.

2.7 Conclusions

This chapter presents capacity calculations and non-uniform signalling design for intersatellite discrete-time Poisson channels corrupted by dark current under peak and average power constraints. Based on a realistic link budget of a LEO satellite communication link, for a given average power, significant gains in rate can be achieved using non-uniform signalling with a modest increase in peak power. Thus, non-uniform signalling is necessary and important to extract the maximum rate from such intersatellite communication links.

The channel capacity and the capacity-achieving distribution are calculated by adapting a particle-based Blahut-Arimoto algorithm and a deterministic annealing based algorithm. Although both algorithms adapt Blahut-Arimoto algorithm to update the probability and adapt gradient decent method to update the constellations, the essential ideas of these two algorithms are different. The particle method focuses on the capacity result under fixed peak power and average power constraints, i.e., a single point on the capacity curve. In contrast, the DA method, computes a segment of the capacity curve for a fixed peak power constraint and with increasing average power constraint. Convergence of these algorithms are guaranteed by the Blahut-Arimoto method and the gradient decent method. The global optimality of the result cannot be proved theoretically while can be verified through the KKT conditions presented in Chap. 3.

Practically, these two methods can be applied in different situations. The particle method, which initializes with some large number of particles, can provide precise results for DTP channel with fixed peak and average constraints. In contrast, the DA method needs to initialize with some small ε, has advantage in drawing a discretized segment of the capacity curve for fixed peak power constraint. In other situations, i.e., with both constraints fixed or with fixed average power constraint, the particle method is more convenient to use.

A simple family of maxentropic input distributions is defined and used to develop tight lower bounds on the channel capacity through the evaluation of the envelope of information rates for different K.

An interesting insight of this chapter is that at low input powers, when the received power is smaller or at most on the same order as the the dark current, that binary inputs are often optimal. Indeed, our simulations have shown that at low input powers the binary maxentropic distribution in (2.24) is optimal over a wide range of A.

In next chapter, analytical results are expanded to develop necessary and sufficient conditions for binary inputs being optimal as well as presenting a closed-form for the capacity-achieving distribution for large λ under both peak and average constraints.

A constrained particle method is also developed which leads to practical signal constellations that can be applied to code design directly. A joint demapper/decoder using the sum-product algorithm is developed and requires a single encoder and decoder. This technique is similar in spirit to bit-interleaved coded modulation (BICM) [66] in the use of mappers and demappers, however, in this work random bit-interleavers and de-interleavers are not employed but are designed to induce the correct input distributions. Three code design examples, including the one based on the practical parameters of the LEO satellite link, are presented to quantify performance.

Simulation results show that the rate performance is close to the capacity with the BER less than 10^{-5} and far outperforms uniform signalling schemes in all scenarios. For the LEO example, for typical values of ε and λ a gain in rate of 17% over uniform signalling is realized with practical codes at a cost of moderate peak amplitude increase.

Chapter 3

Capacity-Achieving Distributions for the DTP Channel

3.1 Introduction

In contrast to work in last chapter which concentrates on numerical results of the capacity and capacity-achieving distribution, this chapter investigates the closed-from expressions of the capacity-achieving distributions of the DTP channel. This approach not only provides insight on the channel capacity but is also a useful tool to guide signalling design. It is shown that the capacity-achieving distributions always have a mass point at zero and, in the case of only a peak power constraint, also a point at the maximum amplitude. It is also proved that distributions that consist of a finite number of mass points are not capacity-achieving for the DTP channel with only average power constraint.

This chapter also places particular emphasis on the *low power regime* of the DTP channel, defined in Section 2.2, which refers to the case when the received power is smaller or at most on the same order as the the dark current. This is motivated primarily by the need for a better understanding of the lower power optical channel, 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.

Numerical results in Sec 2.3 and Sec. 2.4 show that, in the low power regime, the capacity-achieving distributions typically consist of two mass points. To gain a theoretical understanding of this phenomenon, 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.

To remind the reader, the channel model used in the chapter is defined rigorously in Sec. 2.2.

3.2 General Results for Capacity-Achieving Distributions

In this section, three general properties of the capacity-achieving distribution are demonstrated. It is shown that $x = 0$ is always contained in the constellation of the capacity-achieving distribution (i.e., $0 \in \psi_x^*(A, \varepsilon)$). Furthermore, when there is only a peak power constraint A, $x = A$ is also contained in the constellation of the capacity-achieving distribution (i.e., $A \in \psi_x^*(A, A)$). Finally, it is proved that distributions consist of a finite number of mass points are not capacity-achieving for the DTP channel with only average power constraint.

3.2.1 Properties of the Capacity-Achieving Distribution

Lemma 1 (Shifting downward increases mutual information)**.** *Let* X *denote a random variable defined over constellation* $\psi_x = \{x_1, x_2, \ldots, x_n\}$*,* $0 < x_1 < x_2 < \ldots < x_n \leq A$*, with corresponding probability masses* $\psi_p = \{p_1, p_2, \cdots, p_n\}$*,* $\forall i\ p_i \neq 0$*. Let* Y *denote the output of a DTP channel generated by* X*. Define another input* X_Δ *to be identically distributed as* X *but with a shifted constellation* $\psi_{x_\Delta} = \{x_1 - \Delta, x_2 - \Delta, \cdots, x_n - \Delta\}$*, and let* Y_Δ *denote the corresponding output. For any* $\Delta \in (0, x_1]$*,*

$$I(X;Y) \leq I(X_\Delta; Y_\Delta),$$

with equality if and only if $|\psi_x| = 1$*.*

Proof. A detailed proof can be found in Appendix A.1. □

Corollary 2 (Mass point at zero)**.** *The capacity-achieving distribution for the DTP channel under average and peak power constraints always contains a mass point located at 0. That is,* $0 \in \psi_x^*(A, \varepsilon)$ *for any constraints* A *and* ε*.*

Proof. This is a direct consequence of Lemma 1. □

Lemma 3 (Squeezing decreases mutual information)**.** *Let* X *denote a random variable defined over constellation* $\psi_x = \{x_1, x_2, \ldots, x_n\}$*,* $0 \leq x_1 < x_2 < \ldots < x_n \leq A$*, with corresponding probability masses* $\psi_p = \{p_1, p_2, \cdots, p_n\}$*,* $\forall i\ p_i \neq 0$*. Let* Y *denote the output of a DTP channel generated by* X*. Define another input* $X_\alpha = \alpha X$ *with squeezed constellation* $\{\alpha x_1, \alpha x_2, \cdots, \alpha x_n\}$*, and let* Y_α *denote the corresponding output. For* $\alpha \in [0, 1)$*,*

$$I(X;Y) \geq I(X_\alpha; Y_\alpha),$$

with equality if and only if $|\psi_x| = 1$*.*

Proof. This result can be proved by invoking [67, Theorem 2]. A more elementary proof is provided in Appendix A.2. $\square$

Corollary 4 (Point at peak amplitude)**.** *For the DTP channel, when only the peak power constraint is imposed, the capacity-achieving distribution always contains a mass point located at* A*, i.e.,* $A \in \psi_x^*(A, A)$*.*

Proof. This follows from Lemma 3 directly. $\square$

Another interpretation of Lemma 1, is that the mutual information is monotonically decreasing with λ. By Proposition 15 in Appendix A, increasing λ is equivalent to shifting the constellation to the right. This interpretation can also be shown via [67, Theorem 1].

3.2.2 On the Capacity-Achieving Distribution under only Average Power Constraint

Analogous to the definition made by Smith [25] and Shamai [26], for input distribution F_x on the DTP channel (see Section 2.2), let

$$\begin{aligned} i(x, F_x) &\triangleq -\sum_{y=0}^{\infty} P_{Y|X}(y|x) \log \frac{P_Y(y)}{P_{Y|X}(y|x)} \\ &= (x+\lambda)\log(x+\lambda) - x - \sum_{y=0}^{\infty} e^{-(x+\lambda)} \frac{(x+\lambda)^y}{y!} \log\left(\int_0^A e^{-x}(x+\lambda)^y dF_x\right). \end{aligned} \tag{3.1}$$

The mutual information induced by F_x can be written as

$$I(F_x) = \int_0^A i(x, F_x) dF_x.$$

Finally, define the *multiplier function* with Lagrange multiplier $\mu \geq 0$ as

$$M(\mu, x, F_x) = I(F_x) + \mu(x - \varepsilon) - i(x, F_x). \tag{3.2}$$

The following theorem from [26] is of fundamental importance for this chapter. It works as a basic tool to verify the optimality of *any* distribution for DTP channel under peak power and average constraints capacity-achieving or not. For example, the numerical results from the particle method (Section 2.3) and DA method (Section 2.4).

Theorem 5 (KKT conditions [26])**.** *$F_x(A, \varepsilon)$ is capacity-achieving iff the following conditions are satisfied for some $\mu \geq 0$,*

$$M(\mu, x, F_x(A, \varepsilon)) \geq 0, \quad x \in [0, A]. \tag{3.3}$$

$$M(\mu, x, F_x(A, \varepsilon)) = 0. \quad x \in \psi(F_x(A, \varepsilon)). \tag{3.4}$$

where $\psi(F_x(A, \varepsilon))$ is the set of points of increase of $F_x(A, \varepsilon)$.

The following result provides a partial characterization of the capacity-achieving distribution for the DTP channel when the peak power constraint is relaxed (i.e., only an average power constraint is imposed).

Theorem 6 (Insufficiency of distributions with bounded support under average power constraint)**.** *Distributions with bounded support are not capacity-achieving for the DTP channel under only an average power constraint.*

Proof. Suppose instead that the capacity-achieving distribution dF_x^* under average power constraint ε has a bounded support $\Omega \subseteq [0, A^*]$. It follows from Theorem 12 that

$$\begin{aligned} M(\mu, x, F_x^*) &\geq 0, \quad x \in [0, A], \\ M(\mu, x, F_x^*) &= 0, \quad x \in \psi(F_x^*), \end{aligned} \tag{3.5}$$

for any $A \geq A^*$.

Suppose x_1 and x_2 are two arbitrary points of increase of F_x^* with $x_1 < x_2$. In view of the fact that $M(\mu, x_1, F_x^*) = M(\mu, x_2, F_x^*) = 0$, we can rewrite the multiplier function as

$$M(\mu, x, F_x^*) = \frac{x - x_1}{x_2 - x_1} \left(i(x_2, F_x^*) - i(x_1, F_x^*)\right) + i(x_1, F_x^*) - i(x, F_x^*).$$

Note that

$$\begin{aligned}
&M(\mu, A, F_x^*)\\
&= \frac{A - x_1}{x_2 - x_1} \left(i(x_2, F_x^*) - i(x_1, F_x^*)\right) + i(x_1, F_x^*) - i(A, F_x^*)\\
&= \frac{A - x_1}{x_2 - x_1} \left(i(x_2, F_x^*) - i(x_1, F_x^*)\right) + i(x_1, F_x^*) - (A+\lambda)\log(A+\lambda) + A\\
&\quad + \sum_{y=0}^{\infty} e^{-(A+\lambda)} \frac{(A+\lambda)^y}{y!} \log\left(\int_0^A e^{-x}(x+\lambda)^y dF_x^*\right)\\
&\leq \frac{A - x_1}{x_2 - x_1} \left(i(x_2, F_x^*) - i(x_1, F_x^*)\right) + i(x_1, F_x^*) - (A+\lambda)\log(A+\lambda) + A\\
&\quad + \sum_{y=0}^{\infty} e^{-(A+\lambda)} \frac{(A+\lambda)^y}{y!} \log\left((A^*+\lambda)^y\right) \qquad (3.6)\\
&= \frac{A - x_1}{x_2 - x_1} \left(i(x_2, F_x^*) - i(x_1, F_x^*)\right) + i(x_1, F_x^*) - (A+\lambda)\log(A+\lambda) + A\\
&\quad + (A+\lambda)\log(A^*+\lambda), \qquad (3.7)
\end{aligned}$$

where (3.6) is due to the fact that the support of dF_x^* is contained in $[0, A^*]$. It can be seen from (3.7) that $M(\mu, A, F_x^*) < 0$ when A is large enough since the term $-A\log A$ prevails ($i(x_2, F_x^*)$ and $i(x_1, F_x^*)$ can be viewed as constants), which is contradictory with (3.5). Thus, under solely an average power constraint, distributions with bounded support are not capacity-achieving. □

It was shown by Shamai in [26] that, with peak power constraint and with or without average power constraint, the capacity-achieving distribution for the DTP channel must consist of a finite number of mass points. Theorem

6 indicates that this conclusion does not hold if the peak power constraint is removed.

3.3 Binary Capacity-Achieving Distributions

Intuition from earlier simulation studies in Sec. 2.3 and Sec. 2.4 suggests that binary signalling is often optimal for low-power DTP channels. In this section, the general results of Corollary 2 and Corollary 4 in Sec. 3.2 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 7 (Conditions for the capacity-achieving distribution being binary)**.** *For a DTP channel with peak and average power constraints, the capacity-achieving distribution 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* $x \in [0, A]$,

$$\frac{x}{B}\left(i(B, F_x^*) - i(0, F_x^*)\right) + i(0, F_x^*) - i(x, F_x^*) \geq 0, \tag{3.8}$$

and

$$i(B, F_x^*) - i(0, F_x^*) \geq 0, \tag{3.9}$$

where

$$dF_x^* = \left(1 - \frac{\varepsilon}{B}\right)\delta(x) + \frac{\varepsilon}{B}\delta(x - B). \tag{3.10}$$

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

$$i(0, F_x^*) - i(x, F_x^*) \geq 0, \tag{3.11}$$

and

$$i(0, F_x^*) = i(A, F_x^*), \tag{3.12}$$

where

$$dF_x^* = (1-\beta)\delta(x) + \beta\delta(x-A). \tag{3.13}$$

In this case, β is the solution of

$$\sum_{y=0}^{\infty} \left(e^{-\lambda}\frac{\lambda^y}{y!} - e^{-(\lambda+A)}\frac{(\lambda+A)^y}{y!} \right) \log\left((1-\beta)\lambda^y + \beta e^{-A}(\lambda+A)^y\right)$$
$$= \lambda\log\lambda - (A+\lambda)\log(A+\lambda) + A. \tag{3.14}$$

Proof. This theorem follows from Theorem 5 and Corollaries 2 and 4. Note that Corollary 2 implies the existence of a mass point at 0. Now consider a binary input distribution $dF_x^* = (1-\beta)\,\delta(x) + \beta\delta(x-B)$, $0 < B \leq A$ and $0 < \beta < 1$. The KKT conditions lead to two possible cases depending on whether the average power constraint is active:

- $\mathbb{E}_{F_x^*}\{X\} = \varepsilon$ implies $\mu \geq 0$, where μ is the Lagrange multiplier defined in (3.2). This further implies $\beta = \varepsilon/B$. Equations (3.8) and (3.9) follow directly from Theorem 5.

- $\mathbb{E}_{F_x^*}\{X\} < \varepsilon$ (i.e., there is slack in the average constraint) implies $\mu = 0$. Thus, given that F_x^* is capacity-achieving, $B = A$ by Corollary 4. Now invoking Theorem 5 yields (3.11) and (3.12). The optimal β is the solution of (3.12), which can be expanded to (3.14).

□

Theorem 7 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 7 to obtain a closed-form solution for the binary capacity-achieving distribution for a general λ. In what follows, several special cases are considered in which more explicit results can be obtained.

3.3.1 When $\lambda = 0$

In the case of $\lambda = 0$, Theorem 7 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 [8, pp.96-7].

Theorem 8 (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*

$$\varepsilon > f_1(A) \triangleq \frac{A}{e^{\frac{A}{e^A-1}} + 1 - e^{-A}}. \tag{3.15}$$

In this case, the capacity achieving distribution is given by (3.13).

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

$$\varepsilon \le f_1(A), \tag{3.16}$$

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

$$\eta(A, x) \ge 0, \tag{3.17}$$

where

$$\eta(A,x) = \left(\frac{x}{A} - 1 - \frac{x}{A}e^{-A} + e^{-x}\right)\log\frac{1 - \frac{\varepsilon}{A} + \frac{\varepsilon}{A}e^{-A}}{\frac{\varepsilon}{A}e^{-A}} - x - x\log\frac{x}{eA}. \tag{3.18}$$

In this case, the capacity-achieving distribution is given by (3.10) with $B = A$.

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

$$\varepsilon = f_2(B) \triangleq B \bigg/ \left(e^{\frac{B^2+B}{e^B - B - 1}} - e^{-B} + 1\right), \tag{3.19}$$

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

$$\eta(B,x) \geq 0. \tag{3.20}$$

In this case, the capacity-achieving distribution is given by (3.10).

Remark: Cond. 1 of Theorem 8 corresponds to Cond. 2 of Thm. 7 while Cond. 2 and Cond. 3 of Thm. 8 correspond to Cond. 1 of Thm. 7.

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

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

Also note that $\eta(A,x)$ in (3.18) is simply the *multiplier function* in (3.2) with $\lambda = 0$.

1. The first condition has been treated in [26] and corresponds to Cond. 2 in Theorem 7. With $\lambda = 0$ and with peak power constraint only, Shamai proved that the capacity-achieving distribution is (3.13) with $\beta^{*(-1)} = e^{\frac{A}{e^A - 1}} + 1 - e^{-A}$ when $A < 3.3679$ and correspondingly the average power is bounded as (3.15).

2. The second condition follows from Cond. 1 in Theorem 7 with $B = A$. Rearranging (3.9) with $\lambda = 0$ gives the first inequality while substituting $\lambda = 0$ into (3.8) yields the second.

3. The third condition also follows from Cond. 1 in Theorem 7, however, with $B < A$. In particular, the inequality corresponds to (3.8) in Theorem 7. 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,

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

   which gives

   $$\varepsilon = B \Big/ \left(e^{\frac{B^2+B}{e^B - B - 1}} - e^{-B} + 1\right) = f(B). \tag{3.21}$$

   It can be verified that (3.21) implies (3.9). Moreover, both f_1 and f_2 are numerically verified to be monotone increasing functions in the range of $[0, 3.3679]$, therefore, in Cond. 3, $B = f_2^{-1}(\varepsilon)$.

$\square$

3.3.2 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 where high intensity background light from solar irradiation causes large λ. When λ is large, a high intensity Poisson distribution 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., [29], [68]).

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

$$dF_x^{\dagger} = \left(1 - \frac{\varepsilon}{A}\right)\delta(x) + \frac{\varepsilon}{A}\delta(x - A), \tag{3.22}$$

has been shown numerically to be capacity-achieving in the optical intensity channel with Gaussian noise in the low signal-to-noise regime [29]. Note that (3.22) is in fact the binary maxentropic distribution (2.24) introduced in Sec. 2.5 when $A \geq 2\varepsilon$. It has also been shown numerically that (3.22) is capacity-achieving in DTP channel when the input power is relatively small enough. Here, Theorem 7 is leveraged to give a rigorous proof of the optimality of (3.22) when λ is large enough.

Theorem 9 (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* (3.22) *when λ is sufficiently large.*

Proof. It suffices to show that the distribution $dF_x^{\dagger}$ in (3.22) satisfies (3.8) and (3.9) in Theorem 7 when λ is sufficiently large.

By definition,

$$\begin{aligned} i(x, dF_x^{\dagger}) &= (x+\lambda)\log(x+\lambda) - x - (x+\lambda)\log\lambda \\ &\quad - \left(\sum_{y=0}^{\infty} e^{-(x+\lambda)}\frac{(x+\lambda)^y}{y!}\log\left(1 + re^{-A}(1 + A/\lambda)^y\right) - \log(1+r)\right), \\ M(\mu, x, F_x^{\dagger}) &= \frac{x}{A}\left(i(A, F_x^{\dagger}) - i(0, F_x^{\dagger})\right) + i(0, F_x^{\dagger}) - i(x, F_x^{\dagger}), \end{aligned}$$

where $r \triangleq \frac{\varepsilon/A}{1-\varepsilon/A}$ and $M(\cdot)$ is the multiplier function defined in [37, (6)] and corresponds to left-hand side of (3.8). 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

$$\hat{M}(\mu, x, F_x^\dagger, 0) = 0, \tag{3.23}$$

$$\frac{\partial}{\partial z}\hat{M}(\mu, x, F_x^\dagger, z)\Big|_{z=0} = \frac{x(A-x)}{2}, \tag{3.24}$$

$$\hat{M}'(\mu, x, F_x^\dagger, 0) = 0, \tag{3.25}$$

$$\frac{\partial}{\partial z}\hat{M}'(\mu, x, F_x^\dagger, z)\Big|_{z=0} = \frac{A}{2} - x, \tag{3.26}$$

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

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

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

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

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 (3.8) is satisfied for sufficiently large λ.

To complete the proof, it remains to verify that the distribution $F_x^\dagger$ given by (3.22) satisfies (3.9) in Theorem 7 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 (A.11) 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 (A.24) 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 non-negative 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). Therefore $dF_x^* = dF_x^\dagger$ for λ large enough. This completes the proof. $\square$

Theorem 9 shows that under the high background light condition the optimal signalling is binary and satisfies both peak and average power constraints.

3.4 Numerical Examples

3.4.1 Example: Inactive Peak Power Constraint

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

It is easy to see that the average power constraint is inactive if it is greater than the mean of the capacity-achieving distribution under the peak power constraint only. However, it is natural to expect that the peak power constraint is always active since separating constellation points maximally should improve performance. Somewhat surprisingly, we shall show that the peak power constraint can be inactive (i.e., $A \notin \psi_x^*$) in some cases. It should be pointed out that $A \notin \psi_x^*$ does not mean the peak power constraint is superfluous. Indeed, according to Theorem 6, distributions with bounded support are not capacity-achieving if the peak power constraint is removed.

Figure 3.1 plots the capacity-achieving distributions for $\lambda = 0$, $\varepsilon = 0.0594439$ and various A. In both sub-figures, the dashed curve represents the peak power constraint A while the dots represent the positions of mass points in the optimal input distributions. All simulations are carried out using the particle method developed in Sec. 2.3.1. In Fig. 3.1(a) the capacity-achieving distributions are plotted for different peak-to-average ratios, A/ε with ε and λ fixed. Fig. 3.1(b) highlights how the positions of the mass points in F_x^* evolve with increasing A. Notice that, when $A \leq 1$, the peak power constraint is active and F_x^* has a mass point at A. When $1 < A < 5.54$, as noted earlier, there is slack in the peak power constraint and $\psi_x^* = \{0, 1\}$. Once $A \geq 5.54$, the capacity achieving distribution is ternary and again the peak power constraint is active. Thus, in this example, it is observed that for a fixed ε and λ the peak power constraint alternates between being active and inactive as A increases.

To investigate further, consider the specific case for the same λ and ε as above and $A = 50\varepsilon = 2.972195$. Using particle method from Sec. 2.3.1, the capacity-achieving distribution is found to be binary and of the form

$$dF_x^* = (1-\varepsilon)\delta(x) + \varepsilon\delta(x-1). \tag{3.27}$$

Thus, the average constraint is active while there is slack in the peak constraint. Further, the multiplier function in (3.20) is computed to be

$$\eta(1, x) = 3.7844e^{-x} - x\log x + 2.392x - 3.7844, \tag{3.28}$$

and plotted in Fig. 3.2. Considering Theorem 5, it is clear from Fig. 3.2 that dF_x^* in (3.27) is in fact capacity-achieving for *any* $A \in (1, 5.54)$ and the peak power constraint is inactive in this range. Notice also that, when $A = 5.54$, both peak and average constraints are met with equality and the capacity-achieving distribution is ternary with $\psi_x^* = \{0, 1, 5.54\}$.

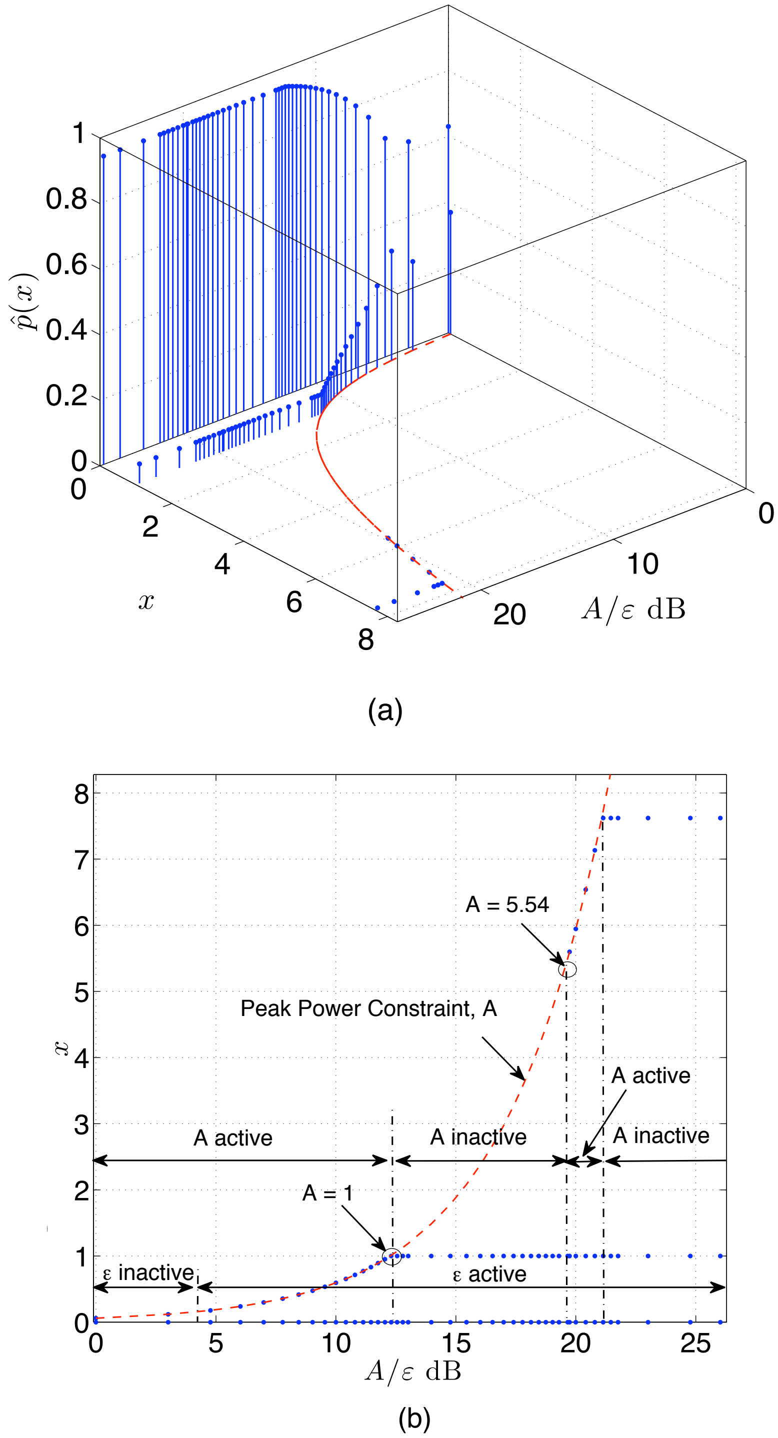


Figure 3.1: Capacity-achieving distributions when $\lambda = 0$ and $\varepsilon = 0.0594439$ with various A: (a) distributions and (b) positions of mass points.

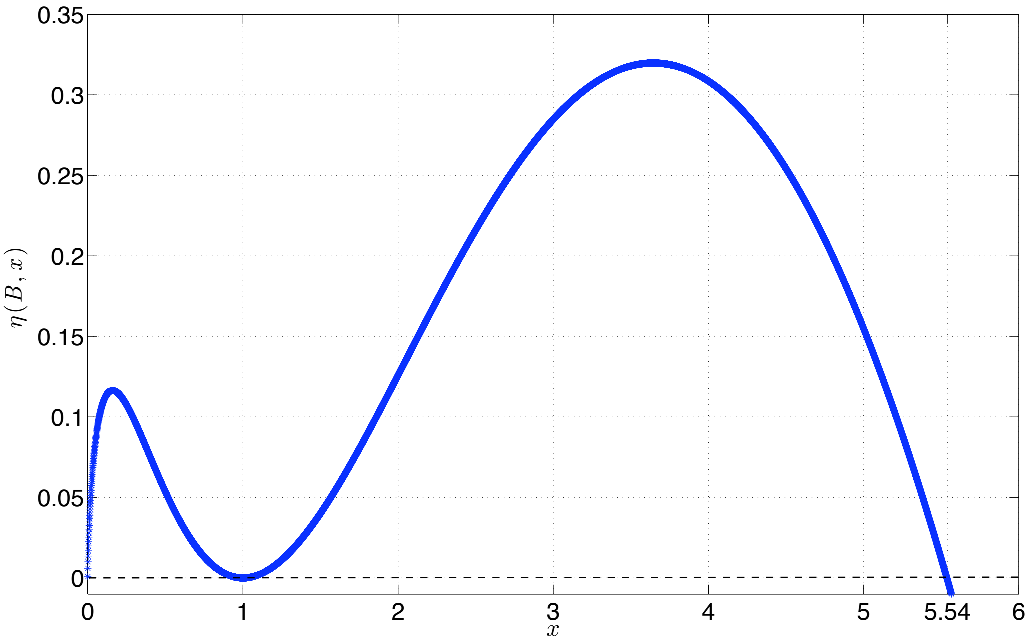


Figure 3.2: Multiplier Function for $\varepsilon = 0.0594439$, $B = 1$, $\lambda = 0$ and $A \in (1, 5.54)$.

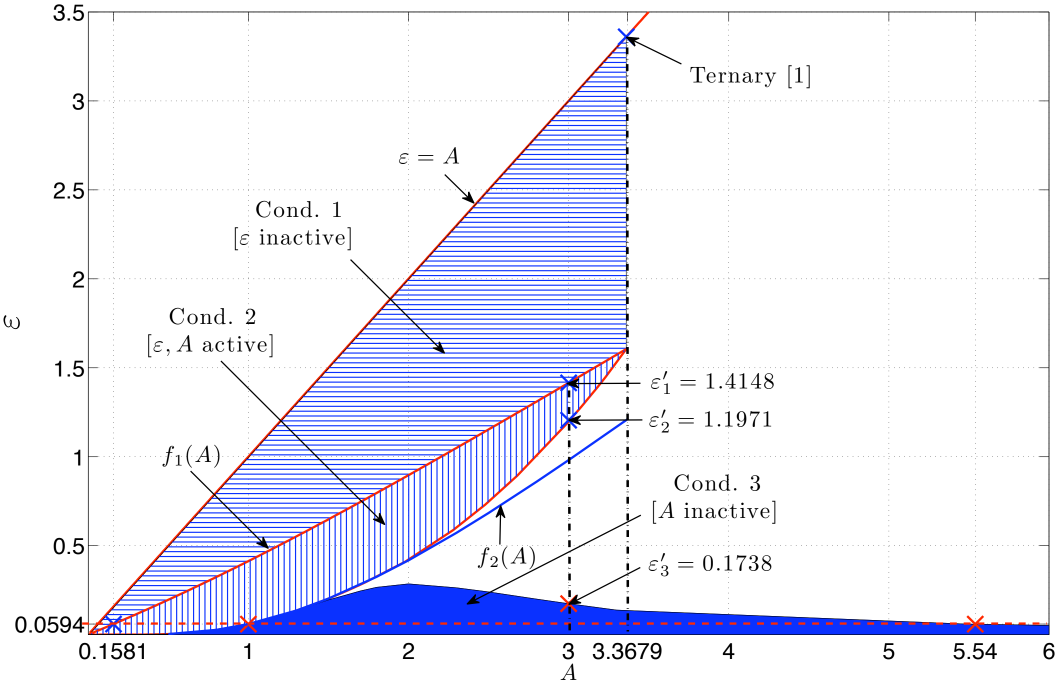


Figure 3.3: Regions of (A, ε) where binary distributions are capacity-achieving for $\lambda = 0$. Areas corresponding to the activity of peak and average constraints are highlighted and defined in Theorem 8.

3.4.2 Channels with Binary Capacity-Achieving Distribution when $\lambda = 0$

To summarize Theorem 8 for $\lambda = 0$, Fig. 3.3 considers DTP channels with $\lambda = 0$ and visualizes regions of ε and A where the capacity-achieving distribution is binary. In addition, following the three conditions in Theorem 8, areas where each constraint is active are clearly shown. Note, that by Lemma 3 of Sec. 3.2, 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 8, the horizontal hatched region in Fig. 3.3 bounded between $A < 3.3679$, $\varepsilon = A$, and $f_1(A)$ in (3.15) corresponds

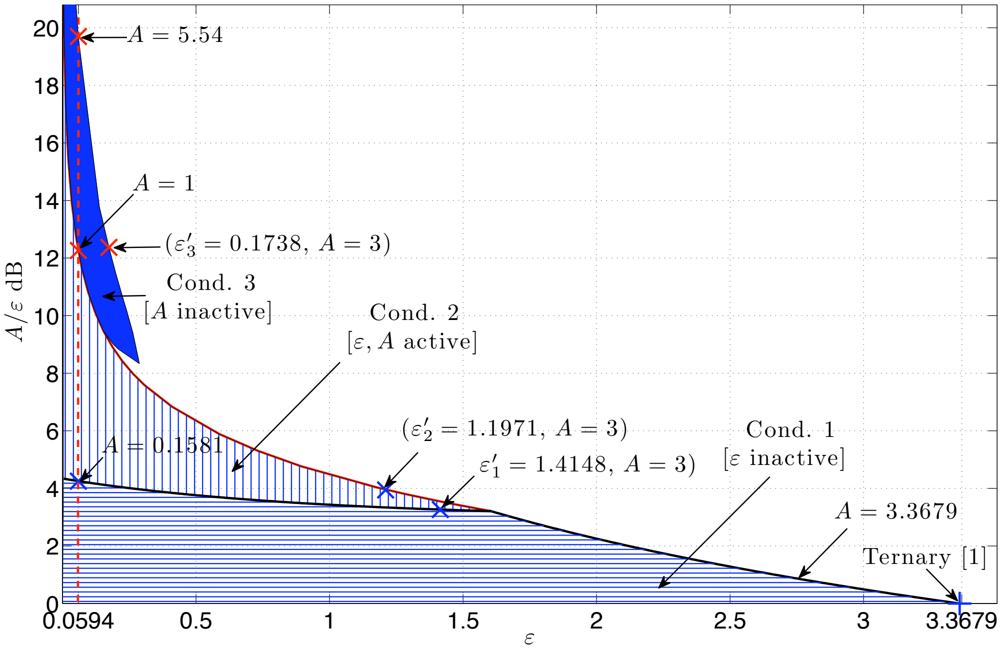


Figure 3.4: Identical results to Figure 1, plotted for peak-to-average ratio (A/ε) versus ε $(\lambda = 0)$.

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 (3.16) and (3.17). An interesting observation from Fig. 3.3 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 [26] for DTP channels with only peak constraint. The solid filled region in Fig. 3.3 corresponds to the set of DTP channels $(\lambda = 0)$ in which Cond. 3 of Theorem 8 is satisfied, i.e., inactive peak constraint. Notice that $A > B = f_2^{-1}(\varepsilon)$ in (3.19), also plotted in

Fig. 3.3, is covered by Cond. 3 in Theorem 8. This region is obtained by sampling in ε and computing the corresponding B via (3.19). Then, Cond. (3.20) 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. 3.3, the capacity-achieving distribution is non-binary. In particular, for the point $A = \varepsilon = 3.3679$, i.e., a DTP channel with peak power constraint only, Shamai showed that the capacity-achieving distribution is ternary [26].

Using Fig. 3.3, the numerical example in Fig. 3.1 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 8 is satisfied and the optimal input distribution is binary and the average constraint is inactive. In Fig. 3.3, 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 8 is satisfied and the capacity-achieving distributions are binary with both constraints active. Notice that the same conclusion is found in Fig. 3.3 through numerical computation. Cond. 3 of Theorem 8 is satisfied for $f_2^{-1}(\varepsilon) = 1 < A < 5.54$ and the capacity-achieving distribution is binary with inactive peak power constraint, as in Fig. 3.3. For $A \geq 5.54$, none of the conditions of Theorem 8 are satisfied and the resulting capacity-achieving distribution is non-binary. The numerical study in Fig. 3.1 shows that the optimal distribution is in fact ternary. Thus, the results in Fig. 3.3 describe the phenomenon of oscillating activity of peak constraint observed numerically in Fig. 3.1. In this example, using Theorem 8, the corner points for the transition of inactivity of the constraints are described analytically via $f_1(A)$ and $f_2(A)$.

Figure 3.4 visualizes the results in Fig. 3.3 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 3.5 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 Fig. 3.3. The behaviour of the constraints can be understood by following the vertical line $A = 3$ in Fig. 3.3. For $\varepsilon < \varepsilon'_3 = 0.1738$ or equivalently -7.5 dB, it is evident from Fig. 3.5 that the capacity-achieving distribution is binary with inactive peak-constraint. Notice in Fig. 3.3, that this corresponds to the region where Cond. 3 of Theorem 8 is satisfied. Increasing ε yields a region where the capacity-achieving distribution is ternary and none of the conditions of Theorem 8 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 8 is satisfied. For $\varepsilon'_1 = 1.4148 < \varepsilon \leq A$ the capacity-achieving distribution is binary with inactive average constraint (Cond. 1 in Theorem 8). This threshold can also be observed in Fig. 3.3 and can be computed as $\varepsilon'_1 = f_1(3)$.

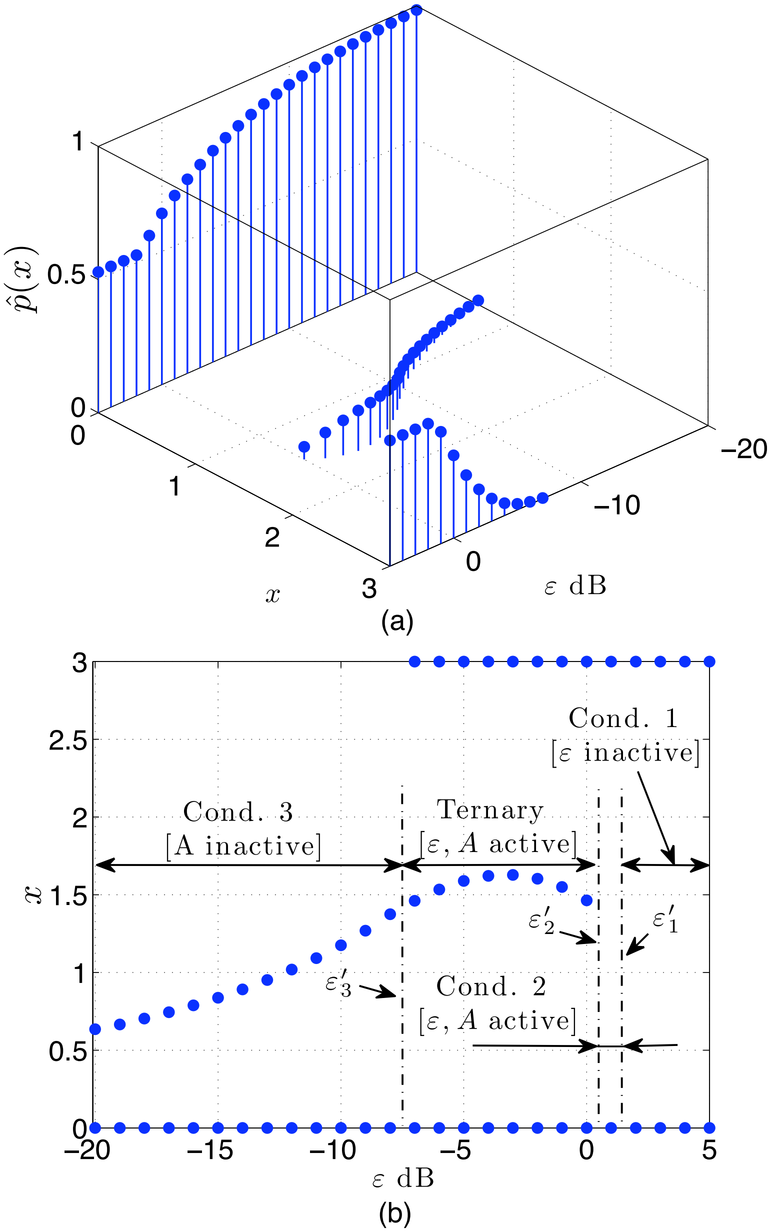


Figure 3.5: Capacity-achieving distributions for fixed $A = 3$, $\lambda = 0$ and increasing ε: (a) distributions and (b) position of mass points. The values for thresholds ε'_1, ε'_2 and ε'_3 can be visualized in Fig. 3.3.

3.4.3 Capacity-Achieving Distributions for Increasing λ

Figure 3.6 illustrates the capacity-achieving distributions with fixed ε and A, $\varepsilon/A < 1/2$, and increasing λ. Theorem. 9 states that for λ large enough the capacity achieving distribution is

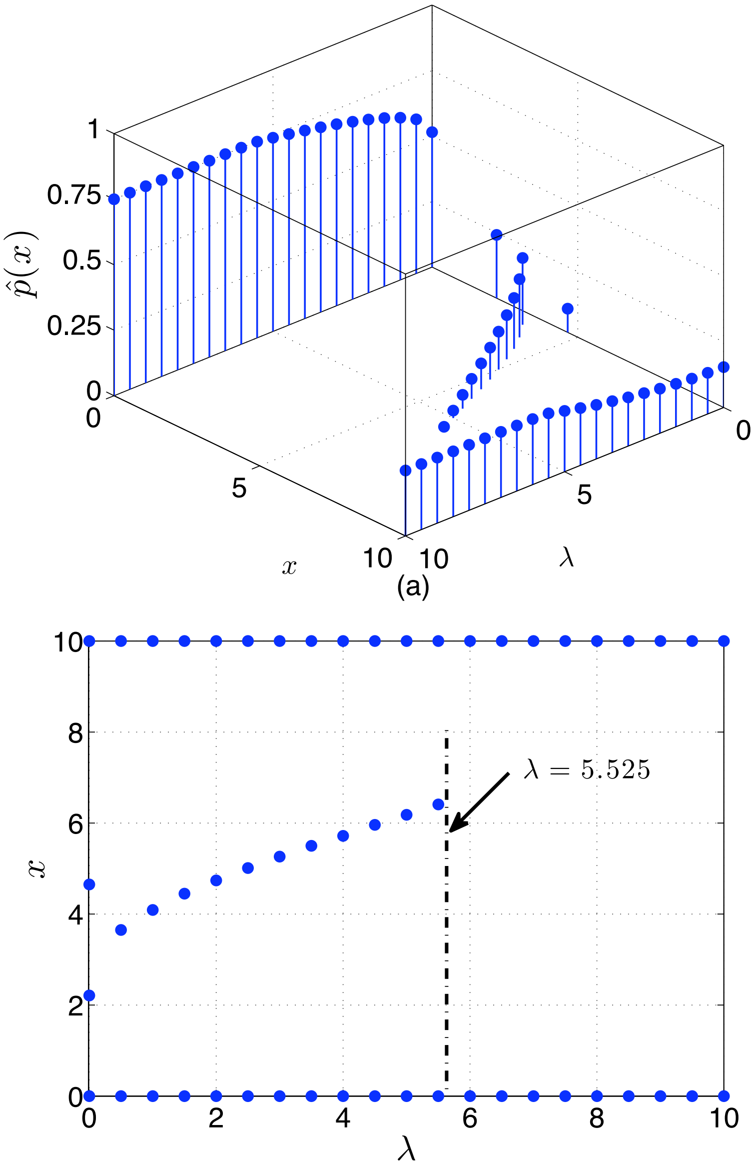
$$dF_x^* = \frac{3}{4}\delta(x) + \frac{1}{4}\delta(x-10).$$

Notice that in Fig. 3.6, for $\lambda \geq 5.525$ that the computed capacity-achieving distribution corresponds to dF_x^* and is fixed for increasing λ.

3.5 Conclusions

This chapter provides insight into the capacity-achieving and approaching distributions for DTP channels, including some fundamental properties of the capacity-achieving distributions. Moreover, particular emphasis is placed on the low power regime where binary signalling is often optimal. Necessary and sufficient 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. All these theoretical conclusions provide significant support to the numerical results in Chapter 2 and provide guidance for practical system design.

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 binary capacity-achieving distributions are provided. In the case of shot-noise limited DTP channels, it is shown that the binary maxentropic distribution is in fact capacity-achieving for λ large enough.



(a)

(b)

Figure 3.6: Capacity-achieving distributions for fixed $A = 10$, $\varepsilon = 2.5$ with increasing λ: (a) distributions and (b) positions of mass points.

Numerical simulations are provided to verify the analytical claims and provide insight on their application. For example, for optical intersatellite communication, when the receiver is corrupted by high background irradiation, binary signalling with the form of (3.22) is capacity-achieving. Also, under ideal conditions of negligible dark current, the regions visualized in Fig. 3.3 show whether the capacity-achieving distribution is binary and what constraints are active.

Chapter 4

On the Sum-Capacity of DTP MAC with Peak Constraint

4.1 Introduction

Previous chapters focus on single user DTP channels, e.g., LEO-LEO or LEO-GEO optical intersatellite links. In particular, LEO-GEO optical links are useful to downlink earth observation data to a ground station, as shown in Fig. 1.1. However, to increase the total data delivered per orbit, it is possible to consider multiple LEOs simultaneously downlinking their observation data via a single GEO, as illustrated in Fig. 1.2. In this context, a new channel model is required.

In this chapter, a multi-user DTP MAC model is proposed, which can be taken as modification of the traditional CTP MAC and an extension of the single-user DTP channel. The DTP MAC imposes a bandwidth limit by constraining transmitted signals to be rectangular PAM (i.e., with fixed intensities in discrete time intervals), therefore, it should be contrasted with its continuous-time counterpart [48, 49], in which there is no bandwidth limit for any user and the input can be arbitrary waveform. Additionally, this DTP MAC considers at least two independent inputs in each channel realization. The channel output, which operates as a similar way of the one-user DTP, obeys the Poisson distribution with mean proportional to the sum of all the input intensities with dark current.

Currently, one of the most fundamental results known regarding single user DTP channel is that when under peak power constraint, the capacity-achieving distribution consists of finite number of mass points [26]. In this chapter, it will be demonstrated that similar conclusion holds, i.e., for the DTP MAC under peak power constraints, the optimal input distribution consists of finite number of mass points. This conclusion provides significant support for any numerical method computing the optimal input distributions which achieve the sum-capacity under peak power constraints and leads guidance for code design in practical use.

The proof involving the analytical properties and the KKT conditions relies heavily on the methods developed in [26], which is in turn based on Smith's seminal work [25, 69]. The derivations in [25, 26, 69] cannot be adapted directly to this work since multiple users are considered at the input. It is also instructive to compare our work with related work on the Gaussian MAC [70], in which a two-user discrete-time Gaussian MAC under identical peak power constraints was studied.

The rest of this chapter is organized as follows. The two-user DTP MAC model is introduced in Section 4.2. The main result is proved in Section 4.3. Section 4.4 contains some concluding remarks. The derivation of the KKT conditions and analyticity are presented in Appendix B.1 and Appendix B.2, respectively.

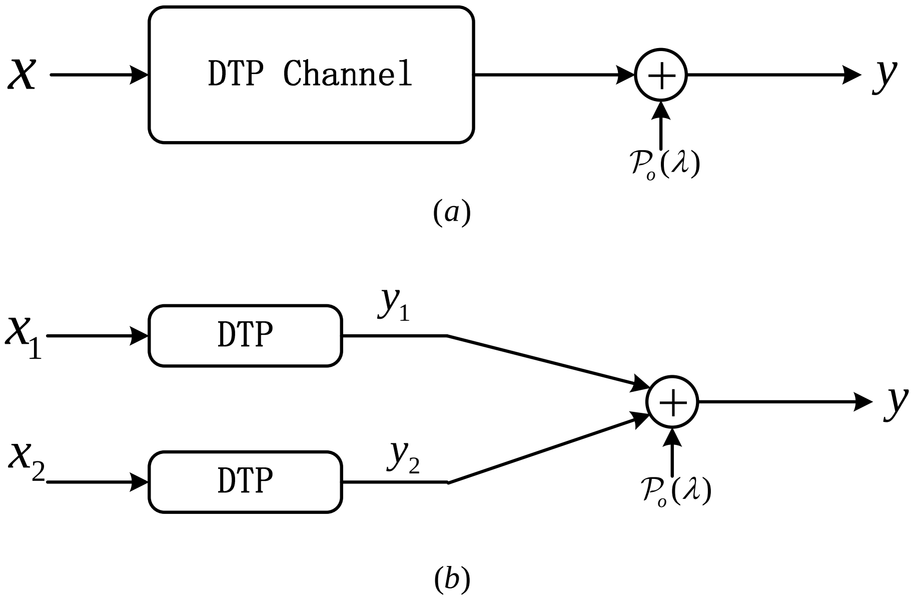


Figure 4.1: Schematic diagram of the single- and two-user DTP multiple access channel. (a) The single-user channel. y is Poisson distributed with rate $x + \lambda$. (b) The two user channel DTP MAC. y is Poisson distributed with rate $x_1 + x_2 + \lambda$.

4.2 Channel Model

This chapter considers an additive-rate Poisson MAC. It is assumed that the users are sufficiently separated so that the images of the incident optical fields for each user on the surface of the detector do not overlap. Correspondingly, the output of the receiver is taken as resulting from the additive intensities of each independent user [71]. An additive-rate CTP MAC model was employed in the analysis of single-user detectors in multiple-access free space applications [71, 72].

Figure 4.1(b) demonstrates a two-user additive-rate DTP MAC. In this model, the detector outputs a single Poisson distributed variable which counts photoelectrons arising from both users as well as shot-noise. Equivalently, as illustrated in Fig. 4.1(b) the output count, y, is the sum of the output of two independent DTP channels for each user plus the shot noise, $\mathcal{P}_o(\lambda)$, which is also Poisson distributed with rate λ. The single user DTP model is drawn in Fig. 4.1(a) for comparison. Similar to the single user DTP channel, the receiver of a two-user DTP MAC is modelled as a photon counter which generates an integer representing the number of receiver photons. Additionally, it is assumed that there is perfect synchronization for the two users both on frame and symbol levels. Moreover, it is assumed the two users are independent, i.e., there is no collaboration between the two LEOs.

Specifically, in each time interval ΔT, the two channel input intensity x [photons/second] and w [photons/second] are corrupted by the combined impact of dark current and background radiation at a rate of λ [photons/second]. The channel output y [photons], represents the number of received photons in ΔT and obeys a Poisson distribution with mean $(x+w+\lambda)\Delta T$. Since the intensities, i.e. photon arrival rates, sum at the receiver, this model is termed as *rate-additive*. Notice also that this assumes an intensity modulation, direct detection (IM/DD) optical intensity channel with incoherent light sources where data is imposed solely on the intensity of the transmitted fields. Without loss of generality, assume $\Delta T = 1$, that is,

$$P_{Y|X_1,X_2}(y|x_1,x_2) = \frac{(x_1+x_2+\lambda)^y}{y!}e^{-(x_1+x_2+\lambda)},$$

where $x, w \in \mathbb{R}^+,\ y \in \mathbb{Z}^+$ and $\lambda > 0$.

Peak amplitude constraints A_i are imposed on the input signals from those two users, i.e.,

$$0 \le x_1 \le A_1 \quad \text{and} \quad 0 \le x_2 \le A_2,$$

due to the dynamic range limitation of each transmitter.

The sum-capacity C is defined as the maximum mutual information over all independent input distribution pairs satisfying peak amplitude constraints,

$$C \triangleq \max_{X_1 \perp X_2, F_{X_1} \in \mathcal{F}_1, F_{X_2} \in \mathcal{F}_2} I(X_1, X_2; Y) \tag{4.1}$$

where F_{X_1} and F_{X_2} are the distributions of X_1 and X_2, respectively, and $\mathcal{F}_i$ denotes the set of distributions over $[0, A_i]$, $i = 1, 2$.

4.3 The Main Result

4.3.1 The Main Result

The main result of this chapter is summarized in the following theorem.

Theorem 10. *For a two-user DTP MAC under peak amplitude constraints, the sum-capacity-achieving input distributions are discrete with a finite number of mass points.*

Proof. This theorem is a direct consequence of the following theorem. Let X_1^*, X_2^* be the optimal inputs of the DTP MAC. For fixed X_1^*, according to Theorem 11, X_2^* is unique and discrete with a finite number of mass points. Similarly, for fixed X_2^*, according to Theorem 11, X_1^* is also unique and discrete with a finite number of mass points.

$\square$

Theorem 11. *Let X and W be two independent random variables. For any fixed $F_W \in \mathcal{F}_2$, the optimal solution to the following maximization*

$$\max_{F_X \in \mathcal{F}_1} I(X, W; Y) \tag{4.2}$$

is unique and is discrete with a finite number of mass points.

4.3.2 Proof of Theorem 11

As in [69], define

$$\begin{aligned}i(x,F_x) \triangleq -\sum_{y=0}^{\infty}\int_0^{A_2} &\frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!}\\ &\times \log\frac{\int_0^{A_1}\int_0^{A_2} e^{-(x'+w'+\lambda)}(x'+w'+\lambda)^y dF_{W'}dF_{X'}}{e^{-(x+w+\lambda)}(x+w+\lambda)^y} dF_W,\end{aligned}$$

Then,

$$I(X,W;Y) = \int_0^{A_1} i(x,F_X)dF_X.$$

Notice that

$$\begin{aligned}&\sum_{y=0}^{\infty}\int_0^{A_2}\frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!}\log e^{-(x+w+\lambda)}(x+w+\lambda)^y dF_W\\ &=\int_0^{A_2}\sum_{y=0}^{\infty}\frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!}\log e^{-(x+w+\lambda)}(x+w+\lambda)^y dF_W\\ &=\int_0^{A_2}(x+w+\lambda)\log\frac{x+w+\lambda}{e}dF_W.\end{aligned}$$

Thus,

$$\begin{aligned}i(x,F_x) = -\sum_{y=0}^{\infty}\int_0^{A_2} &\frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!}\\ &\times\log\int_0^{A_1}\int_0^{A_2} e^{-(x'+w'+\lambda)}(x'+w'+\lambda)^y dF_{W'}dF_{X'}dF_W\\ &+\int_0^{A_2}(x+w+\lambda)\log\frac{x+w+\lambda}{e}dF_W\\ =-\sum_{y=0}^{\infty}\int_0^{A_2} &\frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!}\gamma_y dF_W\\ &+\int_0^{A_2}(x+w+\lambda)\log\frac{x+w+\lambda}{e}dF_W,\end{aligned}$$

where

$$\gamma_y \triangleq \log\int_0^{A_1}\int_0^{A_2} e^{-(x+w+\lambda)}(x+w+\lambda)^y dF_W dF_X.$$

In the rest of this chapter it is assumed that F_W is fixed. As a consequence, one can define a mapping $\chi : \mathcal{F}_1 \to \mathbb{R}^+$ such that $\chi(F_X) = I(X, W; Y)$ for $F_X \in \mathcal{F}_1$. Note that $\mathcal{F}_1$ is a convex and compact space in the Lévy metric d_L [69, p. 21]. (See [73] for the definition of Lévy metric.)

The proofs of the following two technical results are relegated to Appendix B.1 and B.2, respectively.

Proposition 12 (KKT conditions)**.** *Let F_X^* be the optimal solution to the maximization problem (4.2). Then*

$$\begin{aligned} i(x, F_X^*) &\le \chi(F_X^*), \quad x \in [0, A_1] \\ i(x, F_X^*) &= \chi(F_X^*), \quad x \in \psi_X^* \end{aligned} \tag{4.3}$$

where ψ_X^* is the set of points points of increase of F_X^*.

Remark: By Proposition 17 in Appendix B.1, the optimal solution to (4.2) is unique.

Proposition 13 (Analyticity of $i(x, F_X)$)**.** *Given $F_X \in \mathcal{F}_1$, $i(x, F_X)$ (as a function of x) is analytic on $\mathcal{R}_\lambda \triangleq \{x : Re(x) + \lambda > 0\}$ for $\lambda > 0$.*

Now it is ready to prove Theorem 11. Suppose ψ_x^* contains infinite number of mass points in the interval $[0, A_1]$, then it follows from the Bolzano-Weierstrass theorem that ψ_x^* has a limit point in $[0, A_1]$. Therefore, by Proposition 13 and the identity theorem, Equation (4.3), or equivalently,

$$\begin{aligned} &\sum_{y=0}^{\infty} \int_0^A \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} \gamma_y dF_W \\ &\qquad = \int_0^A (x+w+\lambda) \log \frac{x+w+\lambda}{e} dF_W - \chi(F_X^*), \end{aligned} \tag{4.4}$$

must hold for all $x \in \mathcal{R}_\lambda$. It will be shown that this leads to a contradiction.

Note that

$$\begin{aligned}\gamma_y &= \log \int_0^{A_1} \int_0^{A_2} e^{-(x+w+\lambda)}(x+w+\lambda)^y dF_W dF_X \\ &\leq \log \int_0^{A_1} \int_0^{A_2} (A_1 + A_2 + \lambda)^y dF_W dF_X \\ &= y \log(A_1 + A_2 + \lambda).\end{aligned}$$

For the LHS of (4.4),

$$\begin{aligned}&\sum_{y=0}^{\infty} \int_0^{A_2} \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} \gamma_y dF_W \\ &\leq \sum_{y=0}^{\infty} \int_0^{A_2} \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} y \log(A_1 + A_2 + \lambda) dF_W \\ &= \int_0^{A_2} e^{-(x+w+\lambda)} \log(A_1 + A_2 + \lambda) \sum_{y=0}^{\infty} \frac{(x+w+\lambda)^{y+1}}{y!} dF_W \\ &= \int_0^{A} (x+w+\lambda) \log(A_1 + A_2 + \lambda) dF_w \\ &= (x + E[W] + \lambda) \log(A_1 + A_2 + \lambda).\end{aligned}$$

As a consequence, the LHS of (4.4) cannot grow faster than $x \log(A_1 + A_2 + \lambda)$. On the other hand, the RHS of (4.4) grows as $x \log x$, which leads to a contradiction. Therefore, $|\psi_x^*|$ must be finite.

4.4 Conclusion

A DTP MAC model is proposed in this chapter. In contrast with the CTP MAC, in which there is no bandwidth limit and the input can be arbitrary waveform, the DTP MAC imposes a bandwidth limit by transmitting rectangular PAM signals. Therefore, this DTP MAC has an implicit bandwidth limitation making it closer to practice.

This DTP MAC model is suited for these spacecraft missions where at least two LEO satellites operate within small distances and communicate with GEO simultaneously. Such constellations of LEOs for various Earth-observation and scientific missions exist and several are presented in Table 1.3. The DTP MAC model developed here, provides an alternative scheme to establish reliable communication links between the LEOs and GEO.

It is shown that the sum-capacity-achieving distributions under peak amplitude constraints are discrete with a finite number of mass points. This conclusion provides support for numerical algorithms computing the sum capacity and leads guidance for code design. Additionally, with minor modification from this approach, it can be found that the proof can be extended to the maximum weighted sum rate case, i.e., an arbitrary boundary point. Moreover, the proof does not rely on the number of the users, i.e., the result can be extended to the general multiuser case in a straightforward manner.

Chapter 5

Conclusions and Future Directions

5.1 Conclusions

This thesis presents techniques of system design for the DTP channel . Four major contributions are presented in the following areas: 1) Computation of the channel capacity and the capacity-achieving distributions; 2) signaling design and implementation algorithm; 3) properties and closed form expressions of the capacity-achieving distributions, especially the binary distributions; 4) discreteness of the capacity-achieving distributions of DTP MAC with peak amplitude constraints.

The channel capacity and the capacity-achieving distribution are computed by adapting the particle method and DA algorithm. Both are capable of computing capacity and optimal input distribution for DTP channels. The particle method is suited for computing capacity of a given DTP channel with fixed constraints, while the DA method is more appropriate when demonstrating a segment of the capacity curve of DTP channel with increasing average power constraint and with fixed peak power constraint. The global optimality of the results produced by these algorithms can be verified by the KKT conditions. Tight lower bounds on the channel capacity are developed through the evaluation of the envelope of the information rates of the maxentropic input distributions. These bounds are close to the channel capacity especially during the low input power range. A constrained particle method is adapted from [57] and shows better performance compared the traditional quantization method. The novel non-uniform signalling scheme demonstrates near capacity performance in the DTP channel and shows significant gains over the traditional uniform signalling.

Over a dozen of satellite missions are mentioned in this thesis. Through the techniques introduced before, this thesis provides assistance in system design for optical satellite communications, from channel capacity to code design. A practical optical intersatellite link budget is considered in thesis, based on which the channel capacity is computed through the particle method and non-uniform signalling is implemented. The simulation shows the proposed non-uniform signalling scheme achieve a rate 17% larger than that of uniform signalling.

This thesis also presents several fundamental properties of the capacity-achieving distributions of DTP channels. It is shown that the mass point, zero, is always contained in the capacity-achieving distributions of DTP channel. When there is peak amplitude constraint solely, it is further shown that the optimal input distributions always have a mass point at the peak. Moreover, in the case of only average amplitude constraint, it is demonstrated that a finite number of mass points are insufficient to achieve the capacity. These general properties of the DTP channel are applied to the case where binary input are optimal. Necessary and sufficient conditions on the optimality of binary distributions are also demonstrated. Closed-form analytical expressions of the capacity-achieving distributions are derived in several important cases including zero dark current and for high dark current.

A two-user DTP MAC model is proposed, in which only PAM signalling is considered. This is more practical compared with the traditional continuous-time Poisson MAC, where no bandwidth limits are imposed on the inputs. It is shown that the sum-capacity-achieving distributions of the DTP MAC under peak amplitude constraints are discrete with a finite number of mass points. The result still holds for general multiuser case. This conclusion provides insight for numerical algorithm design to compute the capacity region and code design for these multiple LEOs to one GEO optical communication channels.

5.2 Future Work

In Chapter 2, although the proposed non-uniform scheme illustrates near capacity performance, the code itself is not optimized for DTP channel. Since the DTP channel is in fact non-symmetric channel, some codes which operate well in symmetric channels may not be suitable for the DTP channel. Therefore, more work should be completed related to code optimizing by the density evolution or EXIT chart methods. In addition, in the simulation of Section 2.6.5, one assumption made is that the channel does not change during the communication. Practically, this may not always be true. Rate-adaptive procedures for intersatellite optical communication are necessary and should be investigated for practical use.

In Chapter 3, some general properties of capacity-achieving distributions are presented. Meanwhile, the closed-form expressions of the optimal input distributions are also shown in some special cases. However, in general, for arbitrary λ, the analytical expressions are still unknown. Future work may include solving the KKT equations to get the optimal solutions. Also, it would be interesting to see if there are any other tools besides the KKT conditions to find or to verify the optimal input distribution.

In Chapter 4, it is demonstrated that for multi-user DTP MAC channel under amplitude constraints only, the sum-capacity-achieving distributions consist of finite number of mass points. However, the closed-form of these optimal input distribution are unknown. It would be more practical to consider a model which considers both average power constraint and peak peak power constraint. Based on the satellite formation flying missions listed in Table 1.3, it would be interesting to develop algorithms to compute the achievable rate region. Furthermore, it would be interesting to find out both of the numerical results and analytical results for the sum-capacity-achieving or weighted-sum-capacity-achieving distributions. Future work also includes code design to implement points in the rate region.

As discussed, non-uniform signalling has significant gains in rate compared with the traditional uniform signalling. Additionally, free-space optical links have the potential to establish secure high data rate intersatellite communications links. Therefore, an exciting future direction is to consider implementation issues for non-uniform coding schemes in intersatellite optical links, in both the single user and multi-user scenarios.

Appendices

Appendix A

Proofs of Chapter 3

A.1 Proof of Lemma 1

Proposition 14. *If both* $X \to Y \to Z$ *and* $X \to Z \to Y$ *form Markov chains, then* $P_{Y|X}(y|x)/P_{Z|X}(z|x)$ *does not depend on* x*.*

Proof. If both $X \to Y \to Z$ and $X \to Z \to Y$ form Markov chains, then

$$\begin{aligned}
P_{Y|Z}(y|z) &= P_{Y|Z,X}(y|z,x) \\
&= \frac{P_{Y,Z|X}(y,z|x)}{P_{Z|X}(z|x)} \\
&= \frac{P_{Y|X}(y|x)P_{Z|Y}(z|y)}{P_{Z|X}(z|x)},
\end{aligned}$$

which implies that $P_{Y|X}(y|x)/P_{Z|X}(z|x)$ does not depend on x. □

Now proceed to prove Lemma 1. Let Pois(Δ) denote a Poisson distributed random variable with mean Δ. Let $W \sim \text{Pois}(\Delta)$ be independent of X_Δ and Y_Δ. By the data processing inequality,

$$I(X_\Delta; Y_\Delta + W) \leq I(X_\Delta; Y_\Delta). \tag{A.1}$$

Note that the conditional distribution of $Y_\Delta + W$ given $X_\Delta = x - \Delta$ is the same as that of Y given $X = x$, which implies $H(Y_\Delta + W | X_\Delta = x - \Delta) = H(Y|X = x)$ for any $x \geq \Delta$. Moreover, $Y_\Delta + W$ and Y are identically distributed; as a consequence, $H(Y_\Delta + W) = H(Y)$. Now one can readily show that

$$I(X_\Delta; Y_\Delta + W) = I(X; Y), \tag{A.2}$$

which, together with (A.1), yields the desired inequality.

Note that the equality in (A.1) holds if and only if $X_\Delta \rightarrow Y_\Delta + W \rightarrow Y_\Delta$ form a Markov chain. Since $X_\Delta \rightarrow Y_\Delta \rightarrow Y_\Delta + W$ form a Markov chain, it can be shown by leveraging Proposition 14 that $X_\Delta \rightarrow Y_\Delta + W \rightarrow Y_\Delta$ also form a Markov chain if and only if $|\psi_x| = 1$. This completes the proof of Lemma 1.

A.2 Proof of Lemma 3

For a DTP channel with dark current of rate λ and an input distribution specified by constellation ψ_x and probability masses ψ_p, let $I_{\lambda,\psi_x,\psi_p}$ denote the resulting mutual information between the channel input and the channel output. One can prove the following proposition by following the derivation of (A.2).

Proposition 15. $I_{\lambda,\psi_x,\psi_p} = I_{0,\psi_x+\lambda,\psi_p}$

Now proceed to prove Lemma 3. It suffices to consider the case $\alpha \in (0, 1)$ since the degenerate case $\alpha = 0$ is trivially true. Define $\text{Binom}(y, \alpha)$ as a Binomial distribution with $y \in \mathbb{Z}^+$ trials and with success probability α in each trial.

First consider the special case $\lambda = 0$. Introduce a random variable Z such that $X \rightarrow Y \rightarrow Z$ form a Markov chain, where the conditional distribution of

Z given $Y = y$ is $\mathrm{Binom}(y, \alpha)$ for all y. By the data processing inequality,

$$I(X; Z) \leq I(X; Y). \tag{A.3}$$

Note that the conditional distribution of Z given $X = x$ is the same as that of Y_α given $X_\alpha = \alpha x$, which implies $H(Z|X = x) = H(Y_\alpha|X_\alpha = \alpha x)$ for any $x \geq 0$. Moreover, Z and Y_α are identically distributed; as a consequence, $H(Z) = H(Y_\alpha)$. Now one can readily show that

$$I(X; Z) = I(X_\alpha; Y_\alpha). \tag{A.4}$$

which, together with (A.3), implies

$$I_{0,\alpha\psi_x,\psi_p} \leq I_{0,\psi_x,\psi_p}. \tag{A.5}$$

Note that the equality in (A.3) holds if and only if $X \to Z \to Y$ form a Markov chain. Since $X \to Y \to Z$ form a Markov chain, it can be shown by leveraging Proposition 14 that $X \to Z \to Y$ also form a Markov chain if and only if $|\psi_x| = 1$. Therefore, the equality in (A.5) holds if and only if $|\psi_x| = 1$.

For the general case $\lambda \geq 0$, it can be verified that

$$\begin{aligned} I_{\lambda,\psi_x,\psi_p} &= I_{0,\psi_x+\lambda,\psi_p} && \text{(A.6)} \\ &\geq I_{0,\psi_x+\frac{\lambda}{\alpha},\psi_p} && \text{(A.7)} \\ &\geq I_{0,\alpha\psi_x+\lambda,\psi_p}, && \text{(A.8)} \\ &= I_{\lambda,\alpha\psi_x,\psi_p}, && \text{(A.9)} \end{aligned}$$

where (A.6) and (A.9) are due to Proposition 15, (A.7) is due to Lemma 1, and (A.8) is due to (A.5). Clearly, $I_{\lambda,\psi_x,\psi_p} \geq I_{\lambda,\alpha\psi_x,\psi_p}$ is equivalent to the desired inequality

$$I(X; Y) \geq I(X_\alpha, Y_\alpha). \tag{A.10}$$

To complete the proof of Lemma 3, it suffices to show that the equality in (A.10) holds if and only if $|\psi_x| = 1$. The "if" part is trivially true. The "only if" part is a simple consequence of the fact that the equality in (A.5) holds only if $|\psi_x| = 1$.

A.3 Proof of (3.23)

We shall show that

$$\lim_{\lambda\to\infty} i(x, F_x^\dagger, \lambda) = 0, \quad x \in [0, A], \tag{A.11}$$

from which (3.23) 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

$$\lim_{\lambda\to\infty} \sum_{y=0}^{\infty} e^{-(\lambda+x)} \frac{(\lambda+x)^y}{y!} \log\left(1 + re^{-A}(1+A/\lambda)^y\right) = \log(1+r). \tag{A.12}$$

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

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

where

$$\xi \triangleq re^{-A}(1+A/\lambda)^y, \tag{A.13}$$

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

$$\lim_{\lambda\to\infty} \sum_{y=0}^{\infty} e^{-(\lambda+x)} \frac{(\lambda+x)^y}{y!} \log\left(1 + re^{-A}(1+A/\lambda)^y\right)$$

$$\begin{aligned}
&= \lim_{L\to\infty (L \text{ odd})} \lim_{\lambda\to\infty} \sum_{y=0}^{\infty} e^{-(\lambda+x)} \frac{(\lambda+x)^y}{y!} \left(\sum_{l=1}^{L} (-1)^{(l+1)} \frac{r^l e^{-lA}(1+A/\lambda)^{yl}}{l} - \frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}} \right) \\
&= \lim_{L\to\infty (L \text{ odd})} \lim_{\lambda\to\infty} \sum_{l=1}^{L} (-1)^{(l+1)} \frac{r^l}{l} e^{-lA} e^{-(x+\lambda)} e^{\frac{(\lambda+x)(\lambda+A)^l}{\lambda^l}} - \sum_{y=0}^{\infty} e^{-(\lambda+x)} \frac{(\lambda+x)^y}{y!} \frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}} \\
&= \lim_{L\to\infty (L \text{ odd})} \lim_{\lambda\to\infty} \sum_{l=1}^{L} (-1)^{(l+1)} \frac{r^l}{l} e^{\frac{lAx}{\lambda} + (\lambda+x)\left(\binom{l}{2}\frac{A^2}{\lambda^2} + \binom{l}{3}\frac{A^3}{\lambda^3} + \cdots + \frac{A^l}{\lambda^l}\right)} \\
&\quad - \sum_{y=0}^{\infty} e^{-(\lambda+x)} \frac{(\lambda+x)^y}{y!} \frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}} \\
&= \lim_{L\to\infty (L \text{ odd})} \sum_{l=1}^{L} (-1)^{(l+1)} \frac{r^l}{l} - \lim_{\lambda\to\infty} \sum_{y=0}^{\infty} e^{-(\lambda+x)} \frac{(\lambda+x)^y}{y!} \frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}} \\
&= \log(1+r) - \lim_{L\to\infty (L \text{ odd})} \lim_{\lambda\to\infty} \sum_{y=0}^{\infty} e^{-(\lambda+x)} \frac{(\lambda+x)^y}{y!} \frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}}.
\end{aligned} \tag{A.14}$$

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

$$\begin{aligned}
&\lim_{L\to\infty (L \text{ odd})} \lim_{\lambda\to\infty} \left| \sum_{y=0}^{\infty} e^{-(\lambda+x)} \frac{(\lambda+x)^y}{y!} \frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}} \right| \\
&\leq \lim_{L\to\infty (L \text{ odd})} \lim_{\lambda\to\infty} \sum_{y=0}^{\infty} e^{-(\lambda+x)} \frac{(\lambda+x)^y}{y!} \frac{\xi^{L+1}}{L+1} \\
&= \lim_{L\to\infty (L \text{ odd})} \lim_{\lambda\to\infty} \frac{r^{L+1}}{L+1} e^{-(L+1)A} e^{-(x+\lambda)} e^{\frac{(\lambda+x)(\lambda+A)^{L+1}}{\lambda^{L+1}}} \\
&= \lim_{L\to\infty (L \text{ odd})} \frac{r^{L+1}}{L+1} \\
&= 0.
\end{aligned} \tag{A.15}$$

Therefore, we have

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

which together with (A.14) proves (A.12).

A.4 Proof of (3.24)

Note that

$$\begin{aligned}\frac{\partial}{\partial z}\hat{M}(\mu,x,F_x^\dagger,z)\Big|_{z=0} &= \lim_{z\to 0}\frac{\hat{M}(\mu,x,F_x^\dagger,z)-\hat{M}(\mu,x,F_x^\dagger,0)}{z}\\ &= \lim_{\lambda\to\infty} M(\mu,x,F_x^\dagger,\lambda)\lambda\\ &= \lim_{\lambda\to\infty} x\left(\left(1+\frac{\lambda}{A}\right)\log\left(1+\frac{A}{\lambda}\right)-\left(1+\frac{\lambda}{x}\right)\log\left(1+\frac{x}{\lambda}\right)\right)\lambda\\ &\quad+\Lambda(x)-\Lambda(0)+\frac{x}{A}\left(\Lambda(0)-\Lambda(A)\right). \qquad \text{(A.16)}\end{aligned}$$

It is easy to prove that

$$\lim_{\lambda\to\infty} x\left(\left(1+\frac{\lambda}{A}\right)\log\left(1+\frac{A}{\lambda}\right)-\left(1+\frac{\lambda}{x}\right)\log\left(1+\frac{x}{\lambda}\right)\right)\lambda=\frac{x(A-x)}{2}. \qquad \text{(A.17)}$$

Let

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

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

$$\begin{aligned}\Lambda(x) &= \lim_{\lambda\to\infty}\left\{\sum_{y=0}^{\infty}e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\sum_{l=1}^{L}(-1)^{(l+1)}\frac{r^l e^{-lA}(1+A/\lambda)^{yl}}{l}-\sum_{l=1}^{L}(-1)^{(l+1)}\frac{r^l}{l}\right\}\lambda\\ &\quad-\lim_{\lambda\to\infty}\left\{\sum_{y=0}^{\infty}e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\left(\frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}}-\frac{r^{L+1}}{(L+1)(1+\theta_r)^{(L+1)}}\right)\right\}\lambda\\ &= \lim_{\lambda\to\infty}\left\{\sum_{l=1}^{L}(-1)^{(l+1)}\frac{r^l}{l}\lambda\left(e^{-lA}e^{-(x+\lambda)}e^{\frac{(\lambda+x)(\lambda+A)^l}{\lambda^l}}-1\right)\right\}\\ &\quad-\lim_{\lambda\to\infty}\left\{\sum_{y=0}^{\infty}e^{-(\lambda+x)}\frac{(\lambda+x)^y}{y!}\left(\frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}}-\frac{r^{L+1}}{(L+1)(1+\theta_r)^{(L+1)}}\right)\right\}\lambda,\end{aligned} \qquad \text{(A.18)}$$

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

Now consider the first limit in (A.18). Let

$$\phi \triangleq \frac{lAx}{\lambda} + \frac{l(l-1)A^2}{2\lambda} + \frac{l(l-1)A^2x}{2\lambda^2} + (\lambda + x)\left(\binom{l}{3}\frac{A^3}{\lambda^3} + \cdots + \frac{A^l}{\lambda^l}\right). \tag{A.19}$$

We have

$$\begin{aligned}
&\lim_{\lambda\to\infty}\sum_{l=1}^{L}(-1)^{(l+1)}\frac{r^l}{l}\lambda\left(e^{-lA}e^{-(x+\lambda)}e^{\frac{(\lambda+x)(\lambda+A)^l}{\lambda^l}}-1\right)\\
&=\lim_{\lambda\to\infty}\sum_{l=1}^{L}(-1)^{(l+1)}\frac{r^l}{l}\lambda\left(e^{\phi}-1\right)\\
&=\lim_{\lambda\to\infty}\sum_{l=1}^{L}(-1)^{(l+1)}\frac{r^l}{l}\lambda\left(1+\phi+\theta_\phi\phi^2-1\right) && \text{(A.20)}\\
&=\lim_{\lambda\to\infty}\sum_{l=1}^{L}(-1)^{(l+1)}\frac{r^l}{l}\left(lAx+\frac{l(l-1)A^2}{2}\right)\\
&\quad+\lim_{\lambda\to\infty}\sum_{l=1}^{L}(-1)^{(l+1)}\frac{r^l}{l}\lambda\left(\frac{l(l-1)A^2x}{2\lambda^2}+(\lambda+x)\left(\binom{l}{3}\frac{A^3}{\lambda^3}+\cdots+\frac{A^l}{\lambda^l}\right)+\theta_\phi\phi^2\right)\\
&=\left(Ax-\frac{A^2}{2}\right)\frac{r}{1+r}+\frac{A^2}{2}\frac{r}{(1+r)^2}\\
&\quad+\lim_{\lambda\to\infty}\sum_{l=1}^{L}(-1)^{(l+1)}\frac{r^l}{l}\lambda\left(\frac{l(l-1)A^2x}{2\lambda^2}+(\lambda+x)\left(\binom{l}{3}\frac{A^3}{\lambda^3}+\cdots+\frac{A^l}{\lambda^l}\right)+\theta_\phi\phi^2\right),
\end{aligned}$$

$$\text{(A.21)}$$

where (A.20) is due to Taylor's theorem and $\theta_\phi = \frac{1}{2}e^{\phi'}$ for some $\phi' \in [0, \phi]$. For $l \in (3, L]$,

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

when λ is sufficiently large. Therefore,

$$\phi \le \frac{Ax\log\lambda}{\lambda} + \frac{\log\lambda(\log\lambda-1)A^2}{2\lambda} + \frac{\log\lambda(\log\lambda-1)A^2x}{2\lambda^2} + \frac{A^3(\log\lambda)^4(\lambda+x)}{\lambda^3}$$

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

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

which together with (A.21) implies

$$\lim_{\lambda\to\infty}\sum_{l=1}^{L}(-1)^{(l+1)}\frac{r^l}{l}\lambda\left(e^{-lA}e^{-(x+\lambda)}e^{\frac{(\lambda+x)(\lambda+A)^l}{\lambda^l}}-1\right)=\left(Ax-\frac{A^2}{2}\right)\frac{r}{1+r}+\frac{A^2}{2}\frac{r}{(1+r)^2}. \tag{A.22}$$

Next consider the second limit in (A.18). We have

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

$$
\begin{aligned}
&= \lim_{\lambda\to\infty} \left\{ \frac{r^{L+1}}{L+1} e^{-(L+1)A} e^{-(x+\lambda)} e^{\frac{(\lambda+x)(\lambda+A)^{L+1}}{\lambda^{L+1}}} + \frac{r^{L+1}}{(L+1)(1+\theta_r)^{(L+1)}} \right\} \lambda \\
&= \lim_{\lambda\to\infty} \frac{r^{L+1}\lambda}{L+1} \left\{ e^{\frac{(L+1)Ax}{\lambda} + (\lambda+x)\left(\binom{L+1}{2}\frac{A^2}{\lambda^2} + \binom{L+1}{3}\frac{A^3}{\lambda^3} + \cdots + \frac{A^{L+1}}{\lambda^{L+1}}\right)} + \frac{1}{(1+\theta_r)^{(L+1)}} \right\} \\
&\leq \lim_{\lambda\to\infty} \frac{r^{L+1}\lambda}{L+1} \left\{ e^{\frac{(L+1)Ax}{\lambda} + (\lambda+x)(L+1)^2\frac{A^2}{\lambda^2}L} + \frac{1}{(1+\theta_r)^{(L+1)}} \right\} \\
&= \lim_{\lambda\to\infty} \frac{r^{\log\lambda+1}\lambda}{\log\lambda+1} \left\{ e^{\frac{(\log\lambda+1)Ax}{\lambda} + (\lambda+x)(\log\lambda+1)^2\frac{A^2}{\lambda^2}\log\lambda} + \frac{1}{(1+\theta_r)^{(\log\lambda+1)}} \right\} \\
&= 0. \qquad \text{(A.23)}
\end{aligned}
$$

Combining (A.22) and (A.23) with (A.18) gives

$$\Lambda(x) = \left(Ax - \frac{A^2}{2}\right)\frac{r}{1+r} + \frac{A^2}{2}\frac{r}{(1+r)^2}. \qquad \text{(A.24)}$$

Substitute (A.17) and (A.24) into (A.16) proves (3.24).

A.5 Proof of (3.25)

Note that

$$
\begin{aligned}
\hat{M}'(\mu, x, F_x^\dagger, z) &= \frac{\partial M(\mu, x, F_x^\dagger, \lambda)}{\partial x} \\
&= -\frac{\partial i(x, F_x^\dagger, \lambda)}{\partial x} + \frac{1}{A}(i(A, F_x^\dagger, \lambda) - i(0, F_x^\dagger, \lambda)) \\
&= -\log(x+\lambda) + \frac{\partial S(x, F_x^\dagger, \lambda)}{\partial x} + \frac{1}{A}(i(A, F_x^\dagger, \lambda) - i(0, F_x^\dagger, \lambda)),
\end{aligned} \qquad \text{(A.25)}
$$

where

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

It can be verified that

$$\frac{\partial S(x, F_x^\dagger, \lambda)}{\partial x} = -e^{-(x+\lambda)}\sum_{y=0}^{\infty}\frac{(x+\lambda)^{y-1}(x+\lambda-y)}{y!}\Big[-\log(1+r) + y\log\lambda$$

$$+ \log\left(1 + re^{-A}(1+A/\lambda)^y\right)\Big].$$

One can readily prove that

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

Thus,

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

Following the derivation of (A.12), one can show that

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

which together with (A.11) and (A.12) proves (3.25).

A.6 Proof of (3.26)

Note that

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

$$
\begin{aligned}
& - \lim_{\lambda\to\infty} \left\{ \sum_{y=0}^{\infty} e^{-(x+\lambda)} \frac{(x+\lambda)^y}{y!} \log\left(1 + re^{-A}(1+A/\lambda)^y\right) \right. \\
& \quad \left. - \sum_{y=1}^{\infty} e^{-(x+\lambda)} \frac{(x+\lambda)^{y-1}}{(y-1)!} \log\left(1 + re^{-A}(1+A/\lambda)^y\right) \right\} \lambda. \qquad \text{(A.26)}
\end{aligned}
$$

It can be easily verified that

$$
\lim_{\lambda\to\infty} \left(\left(1 + \frac{\lambda}{A}\right) \log\left(1 + \frac{A}{\lambda}\right) - \log\left(1 + \frac{x}{\lambda}\right) - 1 \right) \lambda = \frac{A}{2} - x, \qquad \text{(A.27)}
$$

$$
\lim_{\lambda\to\infty} \frac{1}{A} \left(\Lambda(A) - \Lambda(0)\right) = \frac{Ar}{1+r}. \qquad \text{(A.28)}
$$

Let ξ be defined according to (A.13). By Taylor's theorem, we have

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

where $\theta_\xi \in [0, \xi]$. Let ϕ be defined according to (A.19). It is easy to see that

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

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

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

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

Therefore, we have

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

Moreover, following the derivation of (A.15), one can show that

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

$$- \sum_{y=1}^{\infty} e^{-(x+\lambda)} \frac{(x+\lambda)^{y-1}}{(y-1)!} \frac{\xi^{L+1}}{(L+1)(1+\theta_\xi)^{(L+1)}} \Bigg\} \lambda \Bigg| = 0.$$

As a consequence, we have

$$\begin{aligned} &\lim_{\lambda\to\infty} \Bigg\{ \sum_{y=0}^{\infty} e^{-(x+\lambda)} \frac{(x+\lambda)^{y}}{y!} \log\left(1 + re^{-A}(1+A/\lambda)^y\right) \\ &\qquad - \sum_{y=1}^{\infty} e^{-(x+\lambda)} \frac{(x+\lambda)^{y-1}}{(y-1)!} \log\left(1 + re^{-A}(1+A/\lambda)^y\right) \Bigg\} \lambda \\ &= \frac{-Ar}{1+r}. \end{aligned} \tag{A.29}$$

Substituting (A.27), (A.28), and (A.29) into (A.26) proves (3.26)

Appendix B

Proofs of Chapter 4

B.1 The KKT Conditions

In the view of [69, Corollary 1] (see also [25,26]), for the purpose of proving Proposition 12, it suffices to show that χ is a strictly concave, continuous, and weakly differentiable mapping from $\mathcal{F}_1$ to $\mathbb{R}^+$.

Lemma 16. *$P_Y(y; F_X)$ is bounded and continuous in F_X.*

Let $\{F_n\}_{n\geq 1}$ be a sequence in $\mathcal{F}_1$ such that F_n converges to F in the Lévy metric for some $F \in \mathcal{F}_1$. Note that

$$\begin{aligned}
\lim_{n\to\infty} P_Y(y; F_n) &= \lim_{n\to\infty} \int_0^{A_1}\int_0^{A_2} \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} dF_W dF_n \\
&\overset{(a)}{=} \int_0^{A_1}\int_0^{A_2} \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} dF_W dF \\
&= P_Y(y; F),
\end{aligned}$$

where (a) is due to the Helly-Bray theorem [73]. Now it remains to prove that

$$\int_0^{A_2} \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} dF_W$$

is bounded and continuous in x. The "bounded" part follows from the fact that

$$\int_0^{A_2} \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} dF_W \le \frac{e^{-\lambda}(A_1+A_2+\lambda)^y}{y!}.$$

For the "continuous" part,

$$\begin{aligned}&\lim_{n\to\infty} \int_0^{A_2} \frac{e^{-(x_n+w+\lambda)}(x_n+w+\lambda)^y}{y!} dF_W \\ &= \lim_{n\to\infty} \int_{-\infty}^{\infty} \text{rect}\left(\frac{w}{A_2}\right) \frac{e^{-(x_n+w+\lambda)}(x_n+w+\lambda)^y}{y!} dF_W,\end{aligned}$$

where

$$\text{rect}(u) = \begin{cases} 1 & \text{if } 0 \le u \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

and

$$\left|\text{rect}\left(\frac{w}{A_2}\right) \frac{e^{-(x_n+w+\lambda)}(x_n+w+\lambda)^y}{y!}\right| \le \left|\text{rect}\left(\frac{w}{A_2}\right) \frac{(A_1+w+\lambda)^y}{y!}\right|,$$

$$\begin{aligned}\int_{-\infty}^{\infty} \text{rect}\left(\frac{w}{A_2}\right) \frac{(A_1+w+\lambda)^y}{y!} dF_W &\le \int_{-\infty}^{\infty} \text{rect}\left(\frac{w}{A_2}\right) \frac{(A_1+A_2+\lambda)^y}{y!} dF_W \\ &= \frac{(A_1+A_2+\lambda)^y}{y!}.\end{aligned}$$

Therefore, by the *Dominated convergence theorem*,

$$\lim_{n\to\infty} \int_0^{A_2} \frac{e^{-(x_n+w+\lambda)}(x_n+w+\lambda)^y}{y!} dF_W = \int_0^{A_2} \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} dF_W. \tag{B.1}$$

Proposition 17. *$\chi(F_X)$ is a strictly concave function of F_X.*

Proof. Suppose two arbitrary input distribution $F_1, F_2 \in \mathcal{F}_1$. Let X_1 and X_2 be distributed according to F_1 and F_2, respectively. Also assume $X \sim \theta F_1 + (1-\theta)F_2$, where $\theta \in [0,1]$.

Define S as some binary variable independent with X and W, and the distribution is

$$P(S=0)=\theta, \quad P(S=1)=1-\theta.$$

Then,

$$\begin{aligned} I(X,W,S;Y) &= I(X,W;Y) + \underbrace{I(S;Y|X,W)}_{=0} \\ &= I(S;Y) + I(W,X;Y|S) \\ &= I(S;Y) + \theta I(X_1,W;Y) + (1-\theta) I(X_2,W;Y). \end{aligned}$$

Since $I(S;Y) \geq 0$,

$$I(X,W;Y) \geq \theta I(X_1,W;Y) + (1-\theta) I(X_2,W;Y). \tag{B.2}$$

For the "strict" part, note in (B.2), the equality holds iff $I(S;Y)=0$, which means S is independent with Y, i.e., $P_Y(y;F_1)=P_Y(y;F_2)$. Therefore, it suffices to show that

$$P_Y(y;F_1)=P_Y(y;F_2) \Rightarrow d_L(F_1,F_2)=0.$$

Define $V_i = X_i + W$, $i=1,2$. Note that

$$\Phi_{V_i} = \Phi_{X_i}\Phi_W, \quad i=1,2,$$

where Φ_U denotes the characteristic function of U for any random variable U. By [26, Lemma 2],

$$P_Y(y;F_1)=P_Y(y;F_2) \Rightarrow \Phi_{V_1}=\Phi_{V_2}.$$

Therefore, we have

$$\Phi_W(\Phi_{X_1}-\Phi_{X_2})=0.$$

Since F_W is a distribution over an compact support $[0, A_2]$, it follows from Schwartz's Paley-Wiener theorem [74] that Φ_W is an entire function (when extended to the whole complex plane) and consequently its zeros are isolated, which, together with the (uniform) continuity of characteristic functions, implies that

$$\Phi_{X_1} - \Phi_{X_2} = 0,$$

i.e., $F_1 = F_2$. $\square$

Proposition 18. $\chi(F_X)$ *is continuous in* F_X*.*

This means the convergence of a sequence of distribution in the Lévy metric implies the convergence of the corresponding mutual information.

Proof. Note that

$$\chi(F_X) = H_Y(F_X) - H_{Y|X,W}(F_X),$$

where

$$H_Y(F_X) = \sum_{y=0}^{\infty} -P_Y(y; F_X) \log(P_Y(y; F_X)),$$

and

$$H_{Y|X,W}(F_X) = \int_0^{A_1} \left[\sum_{y=0}^{\infty} \int_0^{A_2} \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} \right.$$
$$\left. \times \log \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} dF_W \right] dF_X.$$

It suffices to prove that both $H_Y(F_X)$ and $H_{Y|X,W}(F_X)$ are continuous in F_X.

1. $H_Y(F_X)$ is continuous in F_X. Let $\{F_n\}_{n\geq 1}$ be a sequence in $\mathcal{F}_1$ such that F_n converges to F in the Lévy metric for some $F \in \mathcal{F}_1$. Note that

$$\lim_{n\to\infty} \sum_{y=0}^{\infty} -P_Y(y; F_n) \log(P_Y(y; F_n))$$

$$\overset{(b)}{=} \sum_{y=0}^{\infty} \lim_{n\to\infty} -P_Y(y;F_n)\log(P_Y(y;F_n))$$
$$\overset{(c)}{=} \sum_{y=0}^{\infty} -P_Y(y;F)\log(P_Y(y;F)),$$

where (c) is due to Lemma 16. For (b), we need to invoke the dominated convergence theorem [73]. In view of the fact that

$$\frac{e^{-(A_1+A_2+\lambda)}\lambda^y}{y!} \leq P_Y(y;F_n) \leq \frac{e^{-\lambda}(A_1+A_2+\lambda)^y}{y!},$$

therefore,

$$\Big| - P_Y(y;F_n)\log P_Y(y;F_n)\Big| \leq \frac{e^{-\lambda}(A_1+A_2+\lambda)^y}{y!} \times \Big((A_1+A_2+\lambda) - y\log\lambda + \log y!\Big).$$

Note that

$$\begin{aligned}
&\sum_{y=0}^{\infty} \frac{e^{-\lambda}(A_1+A_2+\lambda)^y}{y!}\Big((A_1+A_2+\lambda) - y\log\lambda + \log y!\Big)\\
&\leq (A_1+A_2+\lambda)e^{A_1+A_2}\\
&\quad + (A_1+A_2+\lambda)e^{A_1+A_2}\log\lambda + \sum_{y=0}^{\infty}\frac{e^{-\lambda}(A_1+A_2+\lambda)^y}{y!}\log y!\\
&\leq (A_1+A_2+\lambda)e^{A_1+A_2} + (A_1+A_2+\lambda)e^{A_1+A_2}\log\lambda\\
&\quad + \sum_{y=0}^{\infty}\frac{e^{-\lambda}(A_1+A_2+\lambda)^y}{y!} y\log y\\
&= (A_1+A_2+\lambda)e^{A_1+A_2} + (A_1+A_2+\lambda)e^{A_1+A_2}\log\lambda\\
&\quad + \sum_{y=0}^{\infty}\frac{e^{-\lambda}(A_1+A_2+\lambda)^{y+1}}{y!}\log(y+1)\\
&\leq (A_1+A_2+\lambda)e^{A_1+A_2} + (A_1+A_2+\lambda)e^{A_1+A_2}\log\lambda\\
&\quad + \sum_{y=0}^{\infty}\frac{e^{-\lambda}(A_1+A_2+\lambda)^{y+1}}{y!} y
\end{aligned}$$

$$\begin{aligned}
&= (A_1 + A_2 + \lambda)e^{A_1+A_2} \\
&\quad + (A_1 + A_2 + \lambda)e^{A_1+A_2} \log \lambda + (A_1 + A_2 + \lambda)^2 e^{A_1+A_2} \\
&< +\infty. \qquad \text{(B.3)}
\end{aligned}$$

Therefore, the conditions of the dominated convergence theorem [73] are indeed satisfied.

2. $H_{Y|X,W}(F_X)$ is continuous in F_X. It is obvious that

$$\sum_{y=0}^{\infty} \int_0^{A_2} \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} \log \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} dF_W$$

is continuous in x. Moreover,

$$\begin{aligned}
&\left| \sum_{y=0}^{\infty} \int_0^{A_2} \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} \log \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} dF_W \right| \\
&\leq \sum_{y=0}^{\infty} \int_0^{A_2} \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} \\
&\qquad\qquad \times \left[y \log(x+w+\lambda) + (x+w+\lambda) + \log y! \right] dF_W \\
&\leq \sum_{y=0}^{\infty} \left[\frac{e^{-\lambda}(A_1 + A_2 + \lambda)^y}{y!} y \log(A_1 + A_2 + \lambda) \right. \\
&\quad \left. + \frac{e^{-\lambda}(A_1 + A_2 + \lambda)^y}{y!} (A_1 + A_2 + \lambda) + \frac{e^{-\lambda}(A_1 + A_2 + \lambda)^y}{y!} \log y! \right] \\
&< +\infty,
\end{aligned}$$

where the last step follows by an argument similar to that for (B.3). Now one can readily complete the proof by invoking the Helly-Bray theorem.

□

Proposition 19. $\chi(F_X)$ *is weakly differentiable in* $\mathcal{F}_1$*, i.e., the limit*

$$\lim_{\theta \to 0^+} \frac{\chi((1-\theta)F_X + \theta F) - \chi(F_X)}{\theta}$$

exists for any $F_X, F \in \mathcal{F}_1$*.*

Proof. Let

$$F_\theta \triangleq (1-\theta)F_X + \theta F, \quad J(\theta, F_X, F) = \frac{\chi(F_\theta) - \chi(F_X)}{\theta}.$$

Recall the definition of $\chi(F_X)$,

$$\begin{aligned}
&\chi(F_X) \\
&= -\sum_{y=0}^{\infty} P_Y(y;F_X) \log P_Y(y;F_X) \\
&\quad + \sum_{y=0}^{\infty} \int_0^{A_1} \int_0^{A_2} \underbrace{\frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} \log \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!}}_{f_\lambda(x,w,y)} dF_W dF_X.
\end{aligned}$$

Define

$$\begin{aligned}
P_Y(y;F_\theta) &= \int_0^{A_1} \int_0^{A_2} \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} dF_W \Big[(1-\theta)dF_X + \theta dF\Big] \\
&= (1-\theta)P_Y(y;F_X) + \theta P_Y(y;F) \\
&= P_Y(y;F_X) + \theta(P_Y(y;F) - P_Y(y;F_X)).
\end{aligned}$$

Notice that

$$\begin{aligned}
&J(\theta, F_X, F) \\
&= \frac{\chi(F_\theta) - \chi(F_X)}{\theta} \\
&= -\frac{1}{\theta} \sum_{y=0}^{\infty} P_Y(y;F_\theta) \log P_Y(y;F_\theta) + \frac{1}{\theta} \sum_{y=0}^{\infty} \int_0^{A_1} \int_0^{A_2} f_\lambda(x,w,y) dF_W dF_\theta \\
&\quad + \frac{1}{\theta} \sum_{y=0}^{\infty} P_Y(y;F_X) \log P_Y(y;F_X) - \frac{1}{\theta} \sum_{y=0}^{\infty} \int_0^{A_1} \int_0^{A_2} f_\lambda(x,w,y) dF_W dF_X \\
&= -\frac{1}{\theta} \sum_{y=0}^{\infty} P_Y(y;F_\theta) \log P_Y(y;F_\theta) + \frac{1}{\theta} \sum_{y=0}^{\infty} P_Y(y;F_X) \log P_Y(y;F_X) \\
&\quad + \int_0^{A_1} \int_0^{A_2} f_\lambda(x,w,y) dF_W (dF - dF_X).
\end{aligned}$$

Also,

$$
\begin{aligned}
&-\frac{1}{\theta}\sum_{y=0}^{\infty} P_Y(y;F_\theta)\log P_Y(y;F_\theta) + \frac{1}{\theta}\sum_{y=0}^{\infty} P_Y(y;F_X)\log P_Y(y;F_X) \\
&= -\frac{1}{\theta}\sum_{y=0}^{\infty}\Big[P_Y(y;F_X) + \theta(P_Y(y;F) - P_Y(y;F_X))\Big] \\
&\qquad\qquad\qquad \times \log\Big[P_Y(y;F_X) + \theta(P_Y(y;F) - P_Y(y;F_X))\Big] \\
&\quad + \frac{1}{\theta}\sum_{y=0}^{\infty} P_Y(y;F_X)\log P_Y(y;F_X) \\
&= -\sum_{y=0}^{\infty}\frac{P_Y(y;F_X)}{\theta}\log\Big[P_Y(y;F_X) + \theta(P_Y(y;F) - P_Y(y;F_X))\Big] \\
&\quad - \sum_{y=0}^{\infty}(P_Y(y;F) - P_Y(y;F_X))\log\Big[P_Y(y;F_X) + \theta(P_Y(y;F) - P_Y(y;F_X))\Big] \\
&\quad + \frac{1}{\theta}\sum_{y=0}^{\infty} P_Y(y;F_X)\log P_Y(y;F_X). \qquad \text{(B.4)}
\end{aligned}
$$

By Taylor expansion theorem,

$$
\begin{aligned}
&\log\Big[P_Y(y;F_X) + \theta(P_Y(y;F) - P_Y(y;F_X))\Big] \\
&= \log P_Y(y;F_X) + \log\left[1 + \theta\frac{P_Y(y;F) - P_Y(y;F_X)}{P_Y(y;F_X)}\right] \\
&= \log P_Y(y;F_X) + \theta\frac{P_Y(y;F) - P_Y(y;F_X)}{P_Y(y;F_X)} - \frac{1}{2!}\frac{\left[\theta\frac{P_Y(y;F) - P_Y(y;F_X)}{P_Y(y;F_X)}\right]^2}{\left[1 + \xi\theta\frac{P_Y(y;F) - P_Y(y;F_X)}{P_Y(y;F_X)}\right]^2} \\
&= \log P_Y(y;F_X) + \theta\frac{P_Y(y;F) - P_Y(y;F_X)}{P_Y(y;F_X)} \\
&\qquad\qquad - \frac{\theta^2}{2!}\left[\frac{P_Y(y;F) - P_Y(y;F_X)}{P_Y(y;F_X) + \xi\theta(P_Y(y;F) - P_Y(y;F_X))}\right]^2, \qquad \text{(B.5)}
\end{aligned}
$$

where $\xi = \xi(y) \in (0,1)$.

Substitute (B.5) into (B.4),

$$
-\sum_{y=0}^{\infty}\frac{P_Y(y;F_X)}{\theta}\log\Big[P_Y(y;F_X) + \theta(P_Y(y;F) - P_Y(y;F_X))\Big]
$$

$$
\begin{aligned}
&- \sum_{y=0}^{\infty}(P_Y(y;F) - P_Y(y;F_X)) \log \Big[P_Y(y;F_X) + \theta(P_Y(y;F) - P_Y(y;F_X))\Big] \\
&+ \frac{1}{\theta}\sum_{y=0}^{\infty} P_Y(y;F_X) \log P_Y(y;F_X) \\
&= -\sum_{y=0}^{\infty} \frac{P_Y(y;F_X)}{\theta}\Bigg\{ \log P_Y(y;F_X) + \theta \frac{P_Y(y;F) - P_Y(y;F_X)}{P_Y(y;F_X)} \\
&\qquad\qquad - \frac{\theta^2}{2!}\left[\frac{P_Y(y;F) - P_Y(y;F_X)}{P_Y(y;F_X) + \xi\theta(P_Y(y;F) - P_Y(y;F_X))}\right]^2 \Bigg\} \\
&\quad - \sum_{y=0}^{\infty}(P_Y(y;F) - P_Y(y;F_X))\Bigg\{ \log P_Y(y;F_X) + \theta \frac{P_Y(y;F) - P_Y(y;F_X)}{P_Y(y;F_X)} \\
&\qquad\qquad - \frac{\theta^2}{2!}\left[\frac{P_Y(y;F) - P_Y(y;F_X)}{P_Y(y;F_X) + \xi\theta(P_Y(y;F) - P_Y(y;F_X))}\right]^2 \Bigg\} \\
&\quad + \frac{1}{\theta}\sum_{y=0}^{\infty} P_Y(y;F_X) \log P_Y(y;F_X) \\
&= -\sum_{y=0}^{\infty}(P_Y(y;F) - P_Y(y;F_X)) \log P_Y(y;F_X) \\
&\quad + \theta\sum_{y=0}^{\infty}\Bigg\{ \frac{(P_Y(y;F) - P_Y(y;F_X))^2(P_Y(y;F_X) + \theta(P_Y(y;F) - P_Y(y;F_X))}{2\left[P_Y(y;F_X) + \xi\theta(P_Y(y;F) - P_Y(y;F_X))\right]^2} \\
&\qquad\qquad - \frac{\left[(P_Y(y;F) - P_Y(y;F_X)\right]^2}{P_Y(y;F_X)} \Bigg\}. \qquad \text{(B.6)}
\end{aligned}
$$

The infinite summation involved θ need to be proved bounded.

$$
\begin{aligned}
\theta\sum_{y=0}^{\infty} \frac{\left[(P_Y(y;F) - P_Y(y;F_X)\right]^2}{P_Y(y;F_X)} &= \theta\sum_{y=0}^{\infty}\left(\frac{P_Y(y;F)^2}{P_Y(y;F_X)} - 2P_Y(y;F) + P_Y(y;F)\right) \\
&= \theta\sum_{y=0}^{\infty}\frac{P_Y(y;F)^2}{P_Y(y;F_X)} - \theta \\
&\le \theta\sum_{y=0}^{\infty}\frac{P_Y(y;F)}{P_Y(y;F_X)} e^{-\lambda}\frac{(2A+\lambda)^y}{y!} - \theta \\
&\le \theta\sum_{y=0}^{\infty} e^{2A}\left(1 + \frac{2A}{\lambda}\right)^y e^{-\lambda}\frac{(2A+\lambda)^y}{y!} - \theta
\end{aligned}
$$

$$= \theta \left(e^{2A-\lambda} e^{\frac{(2A+\lambda)^y}{\lambda}} - 1 \right).$$

Also,

$$\begin{aligned}
&P_Y(y; F_X) + \xi\theta(P_Y(y; F) - P_Y(y; F_X)) \\
&\geq (1 - \xi\theta) P_Y(y; F_X) \\
&\geq (1 - \theta) P_Y(y; F_X) \\
&\geq \frac{1}{2} P_Y(y; F_X).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\left| \sum_{y=0}^{\infty} \frac{(P_Y(y; F) - P_Y(y; F_X))^2 (P_Y(y; F_X) + \theta(P_Y(y; F) - P_Y(y; F_X))}{2 \left[P_Y(y; F_X) + \xi\theta(P_Y(y; F) - P_Y(y; F_X)) \right]^2} \right| \\
&\leq \frac{1}{8} \sum_{y=0}^{\infty} \frac{(P_Y(y; F) - P_Y(y; F_X))^2}{P_Y(y; F_X)} \left[(1 - \theta) + \theta \frac{P_Y(y; F)}{P_Y(y; F_X)} \right] \\
&< \infty. \qquad \text{(B.7)}
\end{aligned}$$

Combine (B.4), (B.6) and (B.7),

$$\begin{aligned}
&\lim_{\theta \to 0^+} J(\theta, F_X, F) \\
&= -\sum_{y=0}^{\infty} (P_Y(y; F) - P_Y(y; F_X)) \log P_Y(y; F_X) + \int_0^{A_1} \int_0^{A_2} f_\lambda(x, w, y) dF_W (dF - dF_X) \\
&= -\sum_{y=0}^{\infty} \int_0^{A_1} \int_0^{A_2} \left\{ \frac{e^{-(x+w+\lambda)} (x + w + \lambda)^y}{y!} \right. \\
&\qquad\qquad \left. \times \log \int_0^{A_1} \int_0^{A_2} \frac{e^{-(x'+w'+\lambda)} (x' + w' + \lambda)^y}{y!} dF_{W'} dF_{X'} \right\} dF_W dF \\
&\quad + \sum_{y=0}^{\infty} \int_0^{A_1} \int_0^{A_2} \left\{ \frac{e^{-(x+w+\lambda)} (x + w + \lambda)^y}{y!} \right. \\
&\qquad\qquad \left. \times \log \int_0^{A_1} \int_0^{A_2} \frac{e^{-(x'+w'+\lambda)} (x' + w' + \lambda)^y}{y!} dF_{W'} dF_{X'} \right\} dF_W dF_X
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{y=0}^{\infty}\int_0^{A_1}\int_0^{A_2}\frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!}\\
&\qquad\qquad\times\log\frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!}dF_W(dF-dF_X)\\
&=-\sum_{y=0}^{\infty}\int_0^{A_1}\int_0^{A_2}\frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!}\\
&\qquad\qquad\times\log\frac{\frac{\int_0^{A_1}\int_0^{A_2}e^{-(x'+w'+\lambda)}(x'+w'+\lambda)^y dF_{W'}dF_{X'}}{y!}}{\frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!}}dF_W dF\\
&+\sum_{y=0}^{\infty}\int_0^{A_1}\int_0^{A_2}\frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!}\\
&\qquad\qquad\times\log\frac{\frac{\int_0^{A_1}\int_0^{A_2}e^{-(x'+w'+\lambda)}(x'+w'+\lambda)^y dF_{W'}dF_{X'}}{y!}}{\frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!}}dF_W dF_X\\
&=\int_0^{A_1}i(x,F_X)dF-I(F_X)
\end{aligned}
$$

□

B.2 Analyticity of $i(x, F_X)$

Let $\{x_n = \eta_n + j\xi_n\}_{n\geq 1}$ be a sequence of complex numbers in $\mathcal{R}_\delta$, where $\mathcal{R}_\delta$ is a portion of $\mathcal{R}_\lambda$, and

$$
\mathcal{R}_\delta \triangleq \left\{x : -\lambda < \eta_n \leq \delta_1,\ \text{and } \left|\xi_n\right| \leq \delta_2\right\}, \tag{B.8}
$$

where δ_1 and δ_2 are some positive finite numbers.

The first step is to prove that $i(x, F_X)$ is a continuous function of x, i.e., $\lim_{n\to\infty} i(x_n, F_X) = i(x, F_X)$.

$$
\begin{aligned}
&\sum_{y=0}^{\infty}\int_0^{A_2}\Bigg|\frac{e^{-(x_n+w+\lambda)}(x_n+w+\lambda)^y}{y!}\\
&\qquad\qquad\times\log\frac{\int_0^{A_1}\int_0^{A_2}e^{-(x'+w'+\lambda)}(x'+w'+\lambda)^y dF_{W'}dF_{X'}}{e^{-(x_n+w+\lambda)}(x_n+w+\lambda)^y}\Bigg|dF_W
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{y=0}^{\infty} \int_0^{A_2} \left| \frac{e^{-(x_n+w+\lambda)}(x_n+w+\lambda)^y}{y!} \log \frac{P_Y(y;F_x)y!}{e^{-(x_n+w+\lambda)}(x_n+w+\lambda)^y} \right| dF_W \\
&\le \sum_{y=0}^{\infty} \left\{ \int_0^{A_2} \left| \frac{e^{-(x_n+w+\lambda)}(x_n+w+\lambda)^y}{y!} \right| \log P_Y(y;F_n) dF_W \right. \\
&\quad \left. + \int_0^{A_2} \left| \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} \left[y\log(x+w+\lambda) + (x+w+\lambda) + \log y! \right] \right| dF_W \right\} \\
&\le \sum_{y=0}^{\infty} \frac{e^{-(\eta_n+\lambda)}\left((\eta_n+A_1+\lambda)^2+\delta_2^2\right)^{\frac{y}{2}}}{y!} \left((A_1+A_2+\lambda) + y\log\lambda + \log y!\right) \\
&\quad + \sum_{y=0}^{\infty} \frac{e^{-(\eta_n+\lambda)}\left((\eta_n+A_1+\lambda)^2+\xi_n^2\right)^{\frac{y}{2}}}{y!} \\
&\qquad\qquad \times \left(\frac{y}{2} \log\left((\eta_n+A_1+\lambda)^2+\xi_n^2\right) + \left((\eta_n+A_1+\lambda)^2+\xi_n^2\right)^{\frac{1}{2}} + \log y! \right) \\
&\le \sum_{y=0}^{\infty} \frac{\left((\delta_1+A_1+\lambda)^2+\delta_2^2\right)^{\frac{y}{2}}}{y!} \times \left(\frac{y}{2} \log\left((\delta_1+A_1+\lambda)^2+\delta_2^2\right) \right. \\
&\qquad\qquad \left. + \left((\delta_1+A_1+\lambda)^2+\delta_2^2\right)^{\frac{1}{2}} + (A_1+A_2+\lambda) + y\log\lambda + 2\log y! \right) \\
&= \sum_{y=0}^{\infty} \frac{k_\delta^y}{y!} \times \left(y\log k_\delta + k_\delta + (A_1+A_2+\lambda) + y\log\lambda + 2\log y! \right) \\
&\le e^{k_\delta} k_\delta \log k_\delta \lambda + e^{k_\delta} k_\delta + e^{k_\delta}(A_1+A_2+\lambda) + 2k_\delta^2 e^{k_\delta} \\
&< +\infty,
\end{aligned}
$$

where $((\delta_1+A_1+\lambda)^2+\delta_2^2)^{\frac{1}{2}} \triangleq k_\delta$. Through dominated convergence theorem, $\lim_{n\to\infty} i(x_n, F_X) = i(x, F_X)$.

Therefore, in view of Morera's theorem, it suffices to show that

$$\oint_\gamma i(x, F_X)dx = 0 \tag{B.9}$$

for all closed contours γ in $\mathcal{R}_\lambda$. We shall prove (B.9) by verifying the following two equations:

$$\oint_\gamma \sum_{y=0}^{\infty} \int_0^{A_2} \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} \log \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} dF_W dx = 0, \tag{B.10}$$

$$\oint_\gamma \sum_{y=0}^{\infty} \int_0^{A_2} \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} \log P_Y(y; F_X) dF_W dx = 0. \tag{B.11}$$

1. Verification of (B.10). Since

   $$\begin{aligned} &\oint_c \sum_{y=0}^{\infty} \int_0^{A_2} \left| \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} \right| \left| \log \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} \right| dF_W dx \\ &\le \oint_c \sum_{y=0}^{\infty} \int_0^{A_2} \frac{e^{-(\eta_n+w+\lambda)} \left((\eta_n+w+\lambda)^2+\xi_n^2\right)^{\frac{y}{2}}}{y!} \\ &\qquad\qquad \times \left(\Big| y \log(x+w+\lambda) \Big| + \Big| x+w+\lambda \Big| + \log y! \right) dF_W dx \\ &\le l_\gamma \left[\sum_{y=0}^{\infty} \frac{k_\delta^y}{y!} \Big(y \log k_\delta + k_\delta + \log y! \Big) \right] \\ &< +\infty, \end{aligned}$$

   where l_γ is the length of γ which is finite as γ is a closed curve. The last step is also due to the similar reason as (B.3).

   Therefore, it can be shown by invoking Fubini's theorem that

   $$\begin{aligned} &\oint_\gamma \sum_{y=0}^{\infty} \int_0^{A_2} \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} \log \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} dF_W dx \\ &= \sum_{y=0}^{\infty} \int_0^{A_2} \oint_\gamma \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} \log \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} dx dF_W. \end{aligned} \tag{B.12}$$

   Since the integrand in (B.12) is an analytic function of x on $\mathcal{R}_\lambda$, it follows from Cauchy's integral theorem that

   $$\oint_\gamma \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} \log \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} dx = 0. \tag{B.13}$$

Substituting (B.13) into (B.12) yields (B.10).

2. Verification of (B.11). Similar as previous part,

$$\begin{aligned}
&\oint_c \sum_{y=0}^{\infty} \int_0^{A_2} \left| -\frac{e^{-(x_n+w+\lambda)}(x_n+w+\lambda)^y}{y!} \log P_Y(y;F_X) \right| dF_W dx \\
&\leq \oint_c \sum_{y=0}^{\infty} \int_0^{A_2} \frac{e^{-(\eta_n+\lambda)} \left((\eta_n+A+\lambda)^2+\delta^2\right)^{\frac{y}{2}}}{y!} \\
&\qquad\qquad \times \left((A_1+A_2+\lambda) - y\log\lambda + \log y!\right) dF_W dx \\
&\leq l_\gamma \left[\sum_{y=0}^{\infty} \frac{k_\delta^y}{y!} \Big((A_1+A_2+\lambda) + y\log\lambda + \log y! \Big) \right] \\
&< +\infty.
\end{aligned}$$

Invoking Fubini's theorem,

$$\begin{aligned}
&\oint_\gamma \sum_{y=0}^{\infty} \int_0^{A_2} \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} \log P_Y(y;F_X) dF_W dx \\
&= \sum_{y=0}^{\infty} \log P_Y(y;F_X) \int_0^{A_2} \oint_\gamma \frac{e^{-(x+w+\lambda)}(x+w+\lambda)^y}{y!} dx dF_W. \qquad \text{(B.14)}
\end{aligned}$$

Since the integrand in (B.14) is an analytic function of x on $\mathcal{R}_\lambda$, it follows from from Cauchy's integral theorem that the contour integration in (B.14) is equal to zero. This completes the verification of (B.11).

Bibliography

[1] R. M. Gagliardi and S. Karp, *Optical Communications*. New York: John Wiley & Sons, 1976.

[2] S. Hranilovic, *Wireless Optical Communication Systems*, New York: Springer, 2005.

[3] E. O'Niell, *Introduction to Statistical Optics*, Reading, MA: Addison-Wesley, 1963.

[4] Z. Sodnik, B Furch, H Lutz, "Optical intersatellite communication," *IEEE J. Sel. Topics Quantum Electron*, vol. 16. No. 5, 1051-1057, 2010.

[5] R. J. McEliece, E. R. Rodemich, and A. L. Rubin, "The practical limits of photon communication," *Jet Propulsion Laboratory Deep Space Network Progress Reports*, Pasadena, CA 91103, vol. 42-55, pp. 63-67, 1979

[6] R. J. McEliece, "Practical codes for photon communication," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 4, pp. 393-398, Jul. 1981.

[7] S. Gagnon, B. Sylvestre, L. Gagnon, A. Koujelev, D. Gratton and S. Hranilovic, "Recent Developments in Satellite Laser Communications: Canadian Context," *Proc. 2012 International Conference on Space Optical Systems and Applications*, Ajaccio, Corsica, France, Oct. 9-12, 2012.

[8] H. Hemmati (ed.), *Deep space optical communication.* Chichester: Wiley InterScience, 2006.

[9] Hemmati, H., Layland, J., Lesh, J., et al, "Comparative study of optical and RF communication systems for a Mars mission," *Proc. of SPIE*, vol. 2699, San Jose, California, pp.146-164, 1996.

[10] Interagency Operations Advisory Group Optical Link Study Group, "Optical Link Study Group Final Report," Jun. 5, 2012, `https://www.ioag.org/Public%20Documents/Forms/AllItems.aspx`.

[11] S. Yamakawa, T. Hanada, and H. Kohata, "R&D status of the next generation optical communication terminals in JAXA," *IEEE International Conference on Space Optical Systems and Applications*, 2011.

[12] N.H. Roth, "Navigation and Control Design for the CanX-4/-5 Satellite Formation Flying Mission," Master of Applied Science Thesis, Institute for Aerospace Studies, University of Toronto, Canada, 2010.

[13] D.Tomren, N.Kleiner, D.Matteo, and K.Olds, "Optical intersatellite links for the Celestri satellite communications system," *Conf. on Lasers and Electro-Optics (CLEO)*, 1998.

[14] R. Fields, C. Lunde, R. Wong, J. Wicker, J. Jordan, B. Hansen, G. Muehlnikel, W. Scheel, U. Sterr, R. Kahle, and R. Meyer, "NFIRE-to-TerraSAR-X laser communication results: satellite pointing, disturbances, and other attributes consistent with successful performance," *Proc. SPIE* 7330, 73300Q, 2009.

[15] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379-423, 623-656, July, Oct. 1948.

[16] Y. Kabanov, “The capacity of a channel of the Poisson type,” *Theory Prob. Its Applic.*, vol. 23, pp. 143-147, 1978.

[17] M. H. A. Davis, “Capacity and cutoff rate for Poisson-type channels,” *IEEE Trans. Inf. Theory*, vol. IT-26, no. 6, pp. 710-715, Nov. 1980.

[18] M. R. Frey, “Information capacity of the Poisson channel,” *IEEE Trans. Inf. Theory*, vol. 37: (2), pp. 244-256, Mar. 1991.

[19] M. R. Frey, “Capacity of the Lp norm-constrained Poisson channel,” *IEEE Trans. Inf. Theory*, vol. 38: (2), pp. 445-450, Mar. 1992.

[20] 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.

[21] S. Shamai (Shitz) and A. Lapidoth, “Bounds on the capacity of a spectrally constrained Poisson channel,” *IEEE Trans. Inf. Theory*, vol. 39: (1), pp. 19-29, Jan. 1993.

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

[23] Lapidoth, Amos, Jeffrey H. Shapiro, Vinodh Venkatesan, and Ligong Wang, “The Discrete-Time Poisson Channel at Low Input Powers,” *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3260-3272, June 2011.

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

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

[26] 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.

[27] S. Shamai (Shitz) and I. Bar-David, “The capacity of average and peak-power-limited quadrature Gaussian channels,” *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 1060-1071, Jul. 1995.

[28] T. H. Chan, S. Hranilovic, and F. R. Kschischang, “Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs,” *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2073-2088, Jun. 2005.

[29] 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.

[30] J. B. Abshire, “Performance of OOK and low-order PPM modulations in optical communications when using APD-based receivers,” *IEEE Trans. Commun.*, vol. COM-32, pp. 1140-1143, Oct. 1984.

[31] C. Argon and S. W. McLaughlin, “Optical OOK-CDMA and PPM-CDMA Systems With Turbo Product Codes,” *Journal of Lightwave Tech.*, vol. 20, No. 9, pp. 1653-1663, Sep. 2002.

[32] G. M. Lee and G. W. Schroeder, "Optical PPM with Multiple Positions per Pulsewidth," *IEEE Trans. Commun.*, vol. COM-25, pp. 360-364, Mar. 1977.

[33] D. Zwillinger, "Differential PPM has a Higher Throughput than PPM for the Band-Limited and Average-Power-Limited Optical Channel," *IEEE Trans. Inf. Theory*, vol. 34, pp. 1269-1273, Sep. 1988.

[34] P. Elias, "Coding for noisy channels," *IRE Conv. Rec.*, no. 4, pp. 37-46, Mar. 1955.

[35] R. L. Dobrushin, "Asymptotic optimality of group and systematic codes for some channels," *Theory of Probability and its Applications*, vol. 8, no. 1, pp. 47-60, 1963.

[36] 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, April, 2013.

[37] J. Cao, S. Hranilovic, and J. Chen, "Capacity-achieving distributions for the discrete-time Poisson channel–Part I: General properties and numerical techniques," submitted to *IEEE Trans. Commun.*, Feb. 19, 2013 (TCOM-TPS-13-0142).

[38] E. A. Ratzer and D. MacKay, "Sparse low-density parity check codes for channels with cross-talk," in *Proc. of IEEE Inf. Theory Workshop*, Paris, France, Mar. 2003, pp. 127-130.

[39] J. Jiang and K. R. Narayanan, "Multilevel coding for channels with non-uniform inputs and rateless transmission over the BSC," in *Proc. IEEE Int. Symp. Inf. Theory*, Seattle, Washington, U.S.A., 2006, pp. 518-522.

[40] R. Friend, "Orbital Express Program Summary and Mission Overview," in *Proc. SPIE, Sensors and Systems for Space Applications II* , vol. 6958, Orlando, FL, USA, Apr. 2008.

[41] S. D'Amico and O. Montenbruck, "Proximity operations of formation-flying spacecraft using an eccentricity/inclination vector separation," *Journal of Guidance, Control, and Dynamics,* vol. 29, pp. 554-563, May-Jun. 2006.

[42] G. Krieger, A. Moreira, H. Fiedler, I. Hajnsek, M. Younis, M.Werner, and M. Zink, "The TanDEM-X mission: A satellite formation for high resolution SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3317-3341, Nov. 2007.

[43] E. Gill, S. D'Amico, and O. Montenbruck, "Autonomous formation-flying for the PRISMA mission," *Journal of Spacecraft and Rockets*, vol. 44, pp. 671-681, May-Jun. 2007.

[44] Vives, S., Lamy, P., Levacher, P., Venet, M., Boit, J., L., "In-flight Validation of the Formation Flying Technologies using the ASPIICS/PROBA-3 giant coronagaph," *Proc. SPIE* 7010-3R, 2008.

[45] M. van Mierlo, K. Yoshihara, A. Ng, L. N. Phong, and F. Chateauneuf, *Earth Observation Using Japanese/Canadian Formation Flying Nanosatellites,* pp. 165–174. Springer Berlin Heidelberg, 2010.

[46] J. K. Eyer, C. J. Damaren, R. E. Zee, and E. Cannon, "A formation flying control algorithm for the CanX-4&5 low earth orbit nanosatellite mission," *Space Technol.*, vol. 27, no. 4, pp. 147-158, 2007.

[47] M. R. LaPointe, “Formation Flying with Shepherd Satellites,” *Ohio Aerospace Institute NIAC Phase I presentation*, Oct. 2001.

[48] A. Lapidoth and S. Shamai (Shitz), “The Poisson multiple-access channel,” *IEEE Trans. Inf. Theory*, vol. 44, 1998.

[49] S. I. Bross, M. V. Burnashev, and S. Shamai (Shitz), “Error exponents for the two-user Poisson multiple-access channel,” *IEEE Trans. Inf. Theory*, vol. 47, Jul. 2001

[50] R. Gallager, *Information Theory and Reliable Communication*, New York: John Wiley and Sons, 1968

[51] J. Cao, S. Hranilovic, and J. Chen, “Capacity-achieving distributions for the discrete-time Poisson channel–Part II: Binary inputs,” submitted to *IEEE Trans. Commun.*, Feb. 19, 2013 (TCOM-TPS-13-0143).

[52] J. Cao, J. Chen, and S. Hranilovic, “Discreteness of Sum-Capacity-Achieving Distributions for Discrete-Time Poisson Multiple Access Channels with Peak Constraints”, submitted to *IEEE Commun. Letters*, April. 17, 2013 (CL2013-0872).

[53] S. B. Alexander, *Optical Communications Receiver Design*, SPIE Optical Engineering Press, Washington, 1997.

[54] C. C. Chen and C. S. Gardner, “Impact of random pointing and tracking errors on the design of coherent and incoherent optical intersatellite communication links,” *IEEE Trans. Commun.*, vol. 37, pp. 252-260, Mar. 1989.

[55] G. N. Plass and G. W. Kattawar, "Calculations of Reflected and Transmitted Radiance for Earth's Atmosphere," *Appl. Opt.*, vol. 7, pp. 1129-1135, Jun. 1968.

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

[57] J. Dauwels, "Computation of the capacity of continuous memoryless channels and the rate distortion function of memoryless continuous sources," *IEIC Technical Report (Institute of Electronics, Information and Communication Engineers)*, Japan, vol. 106, no. 59 (IT2006 7-12), 2006.

[58] K. Rose, "A mapping approach to rate-distortion computation and analysis," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1939-1952, Nov. 1994.

[59] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1995.

[60] G. Qiu, M. R. Varley, and T. J. Terrell, "Improved clustering using deterministic annealing with a gradient descent technique," *Pattern Recognition Letters*, vol. 15, pp. 607-610, 1994.

[61] K. Rose, E. Gurewitz, and G. C. Fox, "Statistical mechanics and phase transitions in clustering" *Phys. Rev. Lett.*, vol. 65, pp. 945-948, 1990.

[62] F. R. Kschischang, B. J. Frey and H-A. Loeliger, "Factor graphs and the sum-product algorithm,", *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498-519, Feb. 2001.

[63] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, pp. 619-637, Feb. 2001.

[64] EPFL Information Processing Group, LDPC degree distribution optimizer for Gaussian channel, `http://lthcwww.epfl.ch/research/ldpcopt/` .

[65] Signal Processing Microelectronics, University of Newcastle, "LOPT - Online Optimisation of LDPC and RA degree distributions," `http://sonic.newcastle.edu.au/ldpc/lopt/index.php`.

[66] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," IEEE Trans. Commun., vol. 40, no. 5, pp. 873884, 1992.

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

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

[69] J. G. Smith, "On the information capacity of peak and average power constrained Gaussian channels," Ph.D. dissertation, Dep. Elec. Eng., Univ. of California, Berkeley, CA, Dec. 1969.

[70] B. Mamandipoor, K. Moshksar, and A. Khandani, "On the sum-capacity of Gaussian MAC with peak constraint," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Boston, MA, Jul. 2012, pp. 26-30.

[71] G. D. Peterson and C. S. Gardner, "Cross-correlation interference effects in multiaccess optical communications," *IEEE Trans. Aerosp. and Electron. Syst.*, vol. AES-17, no. 2, pp. 199-207, Mar. 1981.

[72] S. Verdú, "Multiple-access channels with point-process observations: Optimum demodulation," *IEEE Trans. Inf. Theory*, vol. IT-32, Sep. 1986.

[73] R. M. Dudley, *Real Analysis and Probability*, Cambridge: Cambridge University Press, 2002.

[74] L. Hörmander, *Linear Partial Differential Operators*, Berlin: Springer Verlag, 1976.