Fundamental Limits of Multiuser Optical Wireless Communications With and Without Secrecy Constraints

A Dissertation

Presented in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

with a

Major in Electrical Engineering

in the

College of Graduate Studies

University of Idaho

by

Morteza Soltani

Major Professor: Zouheir Rezki, Ph.D.

Committee Members: Arupjyoti Bhuyan, Ph.D.; Somantika Datta, Ph.D.; Dennis Sullivan, Ph.D.

Department Administrator: Joseph Law, Ph.D.

May 2020

Authorization to Submit Dissertation

This dissertation of Morteza Soltani, submitted for the degree of Doctor of Philosophy with a Major in Electrical Engineering and titled "Fundamental Limits of Multiuser Optical Wireless Communications With and Without Secrecy Constraints," has been reviewed in final form. Permission, as indicated by the signatures and dates below is now granted to submit final copies for the College of Graduate Studies for approval.

Advisor:	Zouheir Rezki, Ph.D.	Date
Committee Members:	Arupjyoti Bhuyan, Ph.D.	Date
	Somantika Datta, Ph.D.	Date
	Dennis Sullivan, Ph.D.	Date
Department Chair:	Joseph Law, Ph.D.	Date

Abstract

Optical wireless communications (OWC) has recently gained a lot of interest among industrial and academic communities. The main inhibitor factor of this resurgence of interest is the fact that radio-frequency (RF) spectrum is already so densely occupied to handle the increasingly high demand, and hence, exploring higher frequency spectrum, including the optical range, would be a relief. Another reason behind such an interest resides in the relatively simple deployment of OWC systems. However, before a real deployment of OWC systems, there is a persistent need to establish its fundamental performance limits (e.g. capacity, secrecy capacity, and capacity region) and extract design guidelines for building efficient, reliable, and secure OWC systems. Indeed, due to different propagation channels and different transmit constraints, RF communications and OWC are fundamentally quite different. For instance, the popular intensity modulation and direct detection (IM-DD), which is a favorable scheme for OWC due to its simplicity, has some subtle differences in comparison with RF systems manifested in the nonnegativity of the transmit signal, in addition to constraints on the peak- and average-intensity of the signal. These, in turn, make the fundamental performance limits and the optimal transmission schemes for OWC based on IM-DD different from those for RF systems.

Since the fundamental performance limits of OWC play a vital role in extracting guidelines and communication protocols for designing reliable and secure systems, this dissertation addresses those limits in an OWC setting. Particularly, this dissertation presents novel contributions to the understanding of the fundamental limits of multiuser OWC with and without secrecy constraints. When a secrecy constraint is imposed, this dissertation provides analytical results on the characterization of the optimal transmission schemes for secure and reliable OWC when input-dependent Gaussian noise and Poisson noise models are considered. Additionally, an asymptotic analysis of the secrecy capacity (the fundamental performance limit for secure communications) is presented. Furthermore, a two-user optical multiple access channel model, which depicts a multiuser OWC scenario without secrecy constraints, is proposed and the optimal multiuser transmission schemes that achieve the capacity region (fundamental performance limit of this multiuser scenario) are developed. Moreover, the capacity region of the considered optical multiple access channel is explicitly characterized in a closed-form expression in the regime where the peak- and average-intensity constraints are vanishingly small. After establishing the fundamental performance limits of OWC, powerful machine learning techniques, such as deep learning, are employed for the implementation of OWC systems. In particular, a simple and cost-effective learning-based system with (near-)optimal performance is proposed and is implemented by merely taking off-the-shelf deep learning models, applying them to an OWC design problem, and tuning them based on the easily generated training data.

Acknowledgements

I would like to express my deepest and heartfelt gratitude to my advisor, Dr. Zouheir Rezki, for his continuous guidance and insightful discussions during my Ph.D. study. I will never forget the afternoons that Dr. Rezki and I discussed several research problems and the revision of the research papers. I was lucky to work closely with him and his infectious passion in complementing areas inspired this work in many different ways. I am especially grateful for his unparalleled support during this self-development journey of my Ph.D. study.

I would also like to express my appreciation to Dr. Arupjyoti Bhuyan, Prof. Somantika Datta, and Prof. Dennis Sullivan for serving in my committee and defense, and for their valuable comments and discussions. Besides, I would like to thank the ECE Department Chair, Joseph Law, and the staff, John Jacksha who always helped me.

I would like to sincerely thank my family and friends for their continuous encouragement and their moral support. This would not happen without their love.

Finally, I am grateful to the King Abdullah University of Science and Technology (KAUST) for the financial supports I have received during my Ph.D. study under the competitive research grant (CRG) OSR-2016-CRG5-2958-01.

Dedication

To my parents Hadi and Masoumeh, my siblings Mozhgan, Saeed, and Masoud, and my dearest friends Farid and Shirin without whome this Ph.D. journey would not have been completed.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Motivations

1.1.1 Reliable and Secure Wireless Communications

During the last two decades, our daily lives have become increasingly more dependent on wireless communications. As a consequence, wireless communication technologies have to be continuously enhanced to support this increasing demand. This technological evolution has led to the current high-performing wireless communication systems that are used on a daily basis.

One of the main features that distinguish wireless communications from its counterpart, wired communications, is its ability to reach multiple parties simultaneously. This is achieved through broadcasting the information. However, the broadcast nature of wireless signals imposes a critical design challenge for communicating confidential data with trusted users in the presence of unauthorized parties who can maliciously eavesdrop the ongoing communications. Communication security is a delicate issue that can, in some cases, have major ramifications if breached. For instance, consider a wireless communication system application in a bank. In such a scenario, the system must be designed in a way that ensures error-free communications (reliable communications) behind the counter and communication security elsewhere.

Figure 1.1 depicts a typical communication system and its basic building blocks. In this setup, a message W which is drawn from the message set $\mathcal{W} = \{1, 2, \ldots, |\mathcal{W}|\}$ is communicated to a receiver. To this end, first, the encoder block maps the message W into some sequence of channel symbols denoted by X^n, where n is the length of the sequence. Then the channel symbols are transmitted over the communication channel and they produce the output sequence of the channel denoted by Y^n. The output sequence is random but has a distribution that depends on the input sequence and the communication channel is defined to be this conditional distribution. Finally, the decoder block maps the channel output sequence to an estimate of the transmitted message denoted by $\widehat{W}$ and an error is occurred if the message

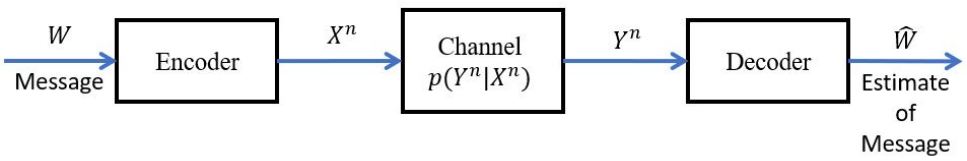

Figure 1.1: Building blocks of a communication system.

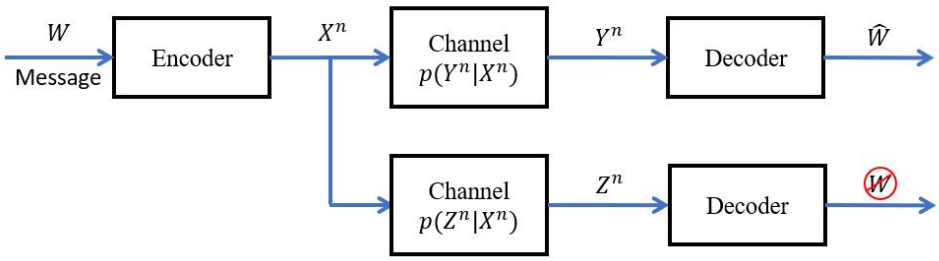

Figure 1.2: Communication of a secret message W in presence of an eavesdropper.

estimate $\widehat{W}$ is not the same as the transmitted message W.

When designing a wireless communication system, two goals have to be considered and need to be achieved: communication reliability and security. A communication system is called reliable when the transmitted messages in Fig. 1.1 can be recovered with a vanishingly small probability of error at the receiver, i.e.,

$$\Pr\left\{W \neq \widehat{W}\right\} \leq \epsilon, \tag{1.1}$$

where ϵ is an arbitrarily small positive value.

As mentioned, wireless communication systems impose a security challenge for transferring confidential data to trusted users in the presence of unauthorized parties who can maliciously eavesdrop the communications. Consider a wireless communication scenario depicted by Figure 1.2. In this setup, the secret message W is to be communicated reliably to a legitimate receiver, i.e., with a vanishingly small probability of error. Furthermore, the secret message W should be kept hidden from the eavesdropper. As such, secure communication refers to the case when a legitimate receiver can successfully decode the secret message W with a vanishingly small probability of error, while an eavesdropper cannot decode and infer the secret message W. This means that the eavesdropper's observations should not reveal any insightful information about the secret message W, i.e.,

$$I(W; Z^n) \leq \epsilon, \tag{1.2}$$

where Z^n is the channel output sequence received by the eavesdropper, ϵ is an arbitrarily small positive value, and $I(W; Z^n)$ denotes the mutual information function which measures the amount of insightful information that Z^n can reveal about the secret message W.

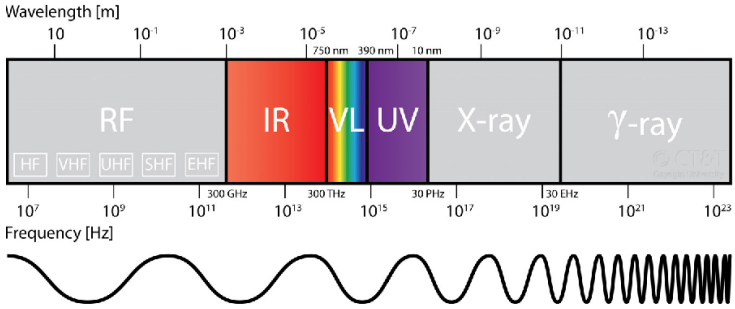

Figure 1.3: Optical frequency range.

1.1.2 Optical Wireless Communications

Coping with the increasing demand for higher data-rate wireless communications is becoming more and more challenging, especially since we are approaching the limits of what can be done with the available resources. One of these resources is bandwidth. Indeed, a communication system's capability can greatly improve if it can access a larger bandwidth. Unfortunately, this is not possible since the currently licensed spectrum is already so densely occupied. This spectrum scarcity problem has motivated researchers to explore new frequencies for wireless communications. The optical spectrum (frequencies ranging from 300 GHz to 3000 THz as shown in Figure 1.3) is one of the promising candidates due to its abundance and free license. The term optical wireless communications (OWC) refers to optical transmission in which guided visible light (VL), infrared (IR), or ultraviolet spectrum (UV) are used as propagation media. The optical wireless systems operating at IR, VL, and UL are mainly used for the terrestrial point-to-point communications as well as space and deep space communications [1].

Recently, optical wireless communications have witnessed a revival due to the invention of Li-Fi (light-fidelity). Li-Fi is a visible-light communication (VLC) technology which promises much higher rates (around Gigabit per second) than its radio-frequency (RF) counterpart (Wi-Fi). In addition to its large bandwidth, Li-Fi enjoys the property of locality, which means that it allows dense special reuse without interference, contrary to RF. Those advantages make Li-Fi an excellent technology for ensuring data coverage without relying on the RF spectrum. Li-Fi can be used for indoor applications using light fixtures (smart lighting), thus combining lighting and communication for better utilization of resources, and hence better sustainability. It can also be used for outdoor applications using light posts to provide coverage for mobile users and using traffic lights or car head/backlights to ensure connectivity between cars and infrastructure. Li-Fi can also be an excellent solution for connectivity in places where low electromagnetic interference is desired, such as hospitals and airplanes [2].

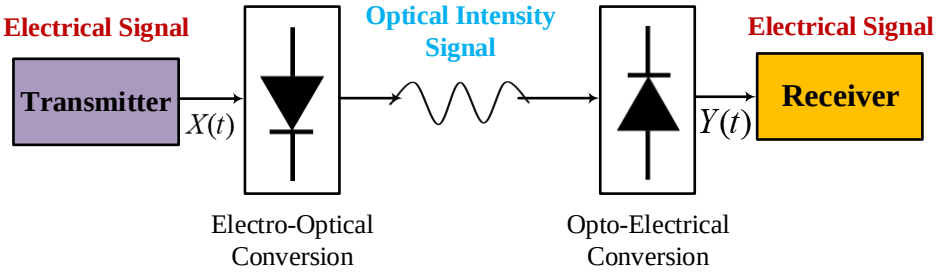

Figure 1.4: Intensity modulation and direct detection OWC.

One of the most popular communication techniques used in OWC systems is the intensity modulation and direct detection (IM-DD) due to its simplicity [2]. In this scheme, as shown in Figure 1.4, the transmitted data modulates the intensity of the emitted light from the optical transmitter (emitted light from a laser diode or a light-emitting diode). The detection of the photons takes place at the receiver by focusing the received light onto a photodetector device. The direct detection method allows the received light to impinge directly upon the photodetector device which responds to its intensity. As opposed to RF communications where the transmitted signals are in the nature of voltage and can be negative, in an OWC setting the transmitted signals are proportional to the light intensity and therefore, they are in the nature of power and are nonnegative.

Like any other form of communication, the applicability of OWC brings about the question of reliability and security. Thus, the reliability and security constraints must be considered when designing an OWC system. While several aspects of the former have been studied in the literature, this is not the case for the latter. Note that due to the differences between OWC and RF communication, results regarding RF communication security do not directly carry over to OWC.

The main thrust of this Ph.D. dissertation is thus to propose wireless communication protocols for a reliable and secure design of OWC systems.

1.2 Background

In OWC systems based on IM-DD, the photodetector device at the receiver absorbs integer number of photons and generates a real-valued output corrupted by noise. Based on the distribution of this corrupting noise, there exist several channel models for the underlying optical wireless channels. Free space optical intensity channels [2, 3, 4], optical intensity channels with input-dependent Gaussian noise [5] and Poisson optical intensity channels [6, 7, 8] are the most widely used models for OWC. Next, a brief introduction about each of these channel models is provided.

1.2.1 FREE-SPACE OPTICAL INTENSITY CHANNEL

The simplest channel model for the optical communications based on IM-DD is the free-space optical intensity channel. In this model, the corrupting noise is independent of the received optical intensity by the photodetector device and follows a zero-mean Gaussian distribution [3, 2]. This Gaussian noise accurately models the ambient light and the thermal noise existing in the optical channel and the electronic devices at the receiver, respectively. However, it neglects the effect of the nonlinearities induced by the optical devices and the photon counting process at the receiver. In this model, the addition of the channel input and the Gaussian noise and is given by [2]

$$Y = X + Z, \tag{1.3}$$

where Y denotes the channel output, X denotes the channel input and Z is the input-independent noise following a zero-mean Gaussian distribution with variance σ^2. Furthermore, since X is proportional to the light intensity, it has to satisfy nonnegativity, peak- and average-intensity constraints due to practical reasons [2]. Hence, X is constrained as

$$\begin{cases} 0 \leq X \leq \mathcal{A}, \\ \mathbb{E}[X] \leq \mathcal{E}, \end{cases} \tag{1.4}$$

where $\mathcal{A}$ and $\mathcal{E}$ are the peak- and average-intensity constraints, respectively, and $\mathbb{E}[X]$ denotes the average value of X.

1.2.2 OPTICAL INTENSITY CHANNEL WITH INPUT-DEPENDENT GAUSSIAN NOISE

A more accurate channel model than the free-space model that takes into account the additional effects of the nonlinearities of the optical devices at the receiver, is the optical intensity channel with input-dependent Gaussian noise [2, 5]. In this model, which is considered as the improved version of the free-space model, the variance of the noise depends on the received optical intensity. Despite accurately modeling the ambient light, thermal noise and the nonlinearities of the optical devices, this channel model does not capture the effect of photon arrivals at the receiver. In this setup, the channel output Y is given by [5]

$$Y = X + \sqrt{X} Z_1 + Z_0, \tag{1.5}$$

where X is the channel input satisfying the constraints (1.4), Z_0 and Z_1 follow zero-mean Gaussian distributions with variances σ_0^2 and σ_1^2, respectively, and Z_0, Z_1 and X are independent of each other.

1.2.3 Poisson Optical Channel

The most accurate channel model that can capture most of the optical channel impairments, is the Poisson optical channel model. In this model, the output is a doubly stochastic Poisson process whose rate is typically the intensity of the incident light (channel input) plus a constant "dark current". Here, the corrupting noise is called the dark current which follows a Poisson process with a constant rate [2, 6, 7]. Whether there are restrictions on the bandwidth of the input signal or not, the results regarding the Poisson channel can be divided into two categories:

- The channel input can have as large as possible bandwidth (i.e., there is no restriction on the signal bandwidth), but it has to satisfy the nonnegativity, peak- and average-intensity constraints. In this model, the channel input signal is denoted by $X(t)$. Given the channel input, the channel output $Y(t)$ is a Poisson Process with instantaneous rate $X(t) + \lambda_0$ satisfying [6]

$$\Pr\{Y(t+\tau) - Y(t) = k | X(t)\} = \frac{e^{-\Gamma}\,\Gamma^k}{k!},\ k = 1, 2, 3, \ldots, \tag{1.6}$$

  where $\Gamma = \int_t^{t+\tau} (X(u) + \lambda_0)\, du$, $\tau \geq 0$ and $\lambda_0 \geq 0$ is the dark current. This model is referred to as the continuous-time Poisson optical channel.

- In practical OWC systems, the channel input is restricted by bandwidth constraints in addition to nonnegativity, peak- and average-intensity constraints. In this case, the transmitter modulates the information bits onto continuous-time pulses of duration Δ seconds, and the receiver preprocesses the incoming continuous-time signal by integrating it over nonoverlapping intervals of length Δ. Therefore, the intensity of the input signal is fixed in each time intervals of lenght Δ, but may vary across different time intervals. In this setting, the channel input is a nonnegative sequence $\{x_k\}$, $k = 1, 2, \ldots$, where x_k corresponds to the fixed intensity of the input signal over the interval $[k\Delta, (k+1)\Delta)$. The output is a sequence $\{Y_k\}$, where Y_k denotes the number of counts registered during the interval $[k\Delta, (k+1)\Delta)$. Therefore, conditioned on the channel input, Y_k follows a Poisson distribution with mean $(x_k + \lambda_0)\Delta$, where λ_0 is the expected number of dark current counts during the interval $[k\Delta, (k+1)\Delta)$. Since Y_k depends only on x_k for $k = 1, 2, \ldots$, a memoryless setting is obtained and the time index k can be dropped. This model is called the discrete-time Poisson channel and the conditional probability mass function of the output Y given the input $X = x$ is [7]

$$p_{Y|X}(y|x) = e^{-(x+\lambda_0)\Delta}\,\frac{[(x+\lambda_0)\Delta]^y}{y!}, \quad y = 0, 1, 2, \ldots. \tag{1.7}$$

where X must satisfy the constraints in (1.4).

1.3 Problem Statement

This Ph.D. dissertation tries to highlight the potential offered by OWC technologies in terms of increasing the date-rate of existing networks, by leveraging spectrum resources in unlicensed bands. Due to its specific properties, fundamentally different from those of its RF counterpart, deployment of OWC systems requires first revisiting what is known about RF design guidelines. *Do all RF communications protocols and design guidelines extend naturally to OWC? What is the viability of OWC infrastructures regarding security requirements? How to extend security mechanisms initially developed for RF communications to OWC?*

This dissertation intends to answer the above questions, among others and plans to provide design guidelines and protocols that are sustainable for OWC systems.

1.4 Prior Work

Information Theory studies the fundamental performance limits of any form of a communication system. Information-theoretic studies helps us identify and extract the communication system design guidelines to achieve the *reliability* and *security* goals. The channel capacity and the secrecy capacity are two information-theoretic fundamental performance limits of communication systems. The channel capacity refers to the maximum *reliable* data rate, i.e., the maximum data rate at which it is guaranteed that the transmitted messages are received with a vanishingly small probability of error. The secrecy capacity denotes the maximum *reliable* data rate at which the secret messages are received by a legitimate receiver and yet the eavesdroppers cannot recover the messages.

The reliable and secure designs of OWC systems are rather difficult compared to RF systems. This is because in OWC, the input signals must be nonnegative and they are subject to peak- and average-intensity constraints (cf. equation (1.4)). On the other hand, in RF systems the transmitted signals can be negative and are generally constrained by an average power, i.e.,

$$\mathbb{E}[X^2] \leq P. \tag{1.8}$$

1.4.1 Reliability in Single-user OWC

Consider a single-user communication scenario in which a transmitter wishes to *reliably* communicate messages to a receiver. In this setting, the objective is to identify the maximum reliable data rate (capacity) and an efficient design to achieve this limit. This problem can be formulated as finding solutions to the following optimization problem [9]

$$C \triangleq \sup_{F_X(x)} I(X;Y) = \int_{\mathcal{X}} \int_{\mathcal{Y}} p(y|x) \log \frac{p(y|x)}{p(y;F_X)} \, dy \, dF_X(x), \tag{1.9}$$

where $F_X(x)$ is the cumulative distribution function of the transmitted signal X and Y is the received signal; $p(y|x)$ is the conditional distribution of the output given the input and $p(y;F_X)$ is the distribution function of the output induced by the input distribution; $\mathcal{X}$ and $\mathcal{Y}$ are the alphabet set of the transmitted and the received signals, respectively. It is noteworthy that the optimal $F_X^*(x)$, which is the solution to (1.9), helps us identify efficient system design (e.g., coding/decoding, modulation/demodulation, etc.) which achieves the capacity C.

In the context of RF communications, the transmitted signal X must satisfy (1.8) and the work in [9] showed that Gaussian distribution is the solution to the optimization problem (1.9) and found a closed-form expression of the capacity C.

However, in the context of OWC, the transmitted signals must be nonnegative and satisfy peak- and average-intensity constraints (cf. equation (1.4)). Thus, a Gaussian distribution is not admissible in an OWC setting because: 1) an input signal which is drawn from the Gaussian distribution can be negative; 2) there is no guarantee that a Gaussian distributed random variable satisfy peak- and average-intensity constraints. In fact, for free-space optical channel, the optical channel with input-dependent Gaussian noise, and the Poisson optical channels, the authors in [6, 7, 10] proved that the capacity-achieving distribution $F_X^*(x)$ is discrete with a finite number of mass points. Moreover, in general, there are no closed-form expressions for the capacity of OWC systems.

1.4.2 Security in OWC Systems

The problem of secure communication systems has been conventionally addressed by cryptographic encryption [11] without considering the imperfections introduced by the communication channel. In this scheme, the usage of *secret keys* is the main approach for having secure communication. Wyner [12], on the other hand, proved the possibility of secure communications without relying on encryption by introducing the wiretap channel. In wiretap channels, secure communication is delivered without using

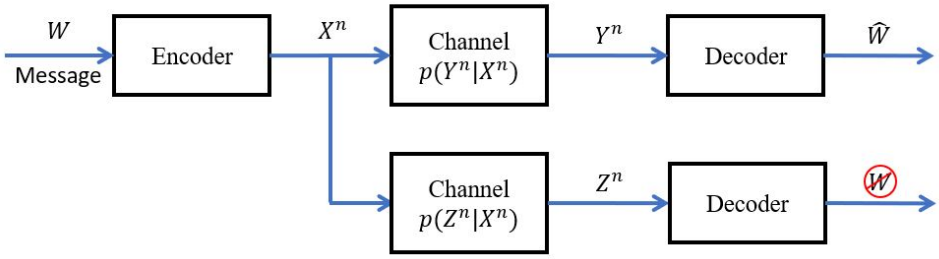

Figure 1.5: Communication of a secret message W in presence of an eavesdropper.

any secret keys and by only leveraging the randomness in the communication channels, such as noise, fading, and interference. A wiretap channel is formally defined as follows [12].

Definition 1. *As illustrated in Figure 1.5, a wiretap channel consists of a transmitter, a legitimate receiver, and an eavesdropper. In this channel, the transmitter wishes to transmit a confidential message W to a legitimate receiver and wishes to keep this message as secret as possible from an eavesdropper. An encoder at the transmitter maps the secret message W to channel input sequence X^n. Then the input symbols are transmitted over the communication channel and they produce the output sequences Y^n at the legitimate receiver and Z^n at the eavesdropper. The output sequences Y^n and Z^n are random but have distributions that depends on the input sequence. These conditional distributions of the output sequences Y^n and Z^n given the input sequence X^n are respectively denoted by $p(Y^n|X^n)$ and $p(Z^n|X^n)$. Finally, the decoder at the legitimate receiver maps the channel output sequence Y^n to an estimate of the transmitted message $\widehat{W}$ with a vanishingly small error probability $\Pr\{W \neq \widehat{W}\} \leq \epsilon$. To have secure communications, the encoder must be designed in such a way that the eavesdropper cannot decode and infer the secret message W, which means that $I(W; Z^n) \leq \epsilon$.*

Furthermore, a formal definition of a *degraded* wiretap channel is given below.

Definition 2. *A wiretap channel is called degraded if the conditional joint probability of the output sequences Y^n and Z^n given the input sequence X^n satisfies*

$$p(Z^n, Y^n|X^n) = p(Y^n|X^n)\, p(Z^n|X^n), \tag{1.10}$$

i.e., X^n and Z^n are conditionally independent given Y^n, or in other words, X^n, Y^n and Z^n form the Markov chain $X^n \to Y^n \to Z^n$.

Definition 2 implies that in a wiretap channel, if the eavesdropper observes a degraded (noisier) version of the signal obtained by the legitimate receiver, then the wiretap channel is degraded [12, 13].

Consider a degraded wiretap channel in which a transmitter wishes to *reliably* communicate secret messages to a legitimate receiver in the presence of an eavesdropper. In this setup, the objective is to identify the maximum reliable and secure data rate (secrecy capacity) and an efficient design to achieve this limit. This problem can formally be given as finding the solutions of [12, 13]

$$C_S \triangleq \sup_{F_X(x)} [I(X;Y) - I(X;Z)] = \sup_{F_X(x)} \Bigg[\int_{\mathcal{X}} \Bigg\{ \int_{\mathcal{Y}} p(y|x) \log \frac{p(y|x)}{p(y;F_X)}\, dy - \int_{\mathcal{Z}} p(z|x) \log \frac{p(z|x)}{p(z;F_X)}\, dz \Bigg\} dF_X(x) \Bigg], \quad (1.11)$$

where $F_X(x)$ is the probability distribution function of the transmitted signal X, Y and Z are the received signals at the legitimate receiver and the eavesdropper, respectively; $p(y|x)$ and $p(z|x)$ are the conditional distribution of the received signals Y and Z given the transmitted signal X, respectively; $p(y;F_X)$ and $p(z;F_X)$ is the distribution functions of Y and Z induced by the input distribution, respectively; $\mathcal{Y}$ and $\mathcal{Z}$ are the alphabet sets of the received signals Y and Z, respectively. Note that the optimal solution $F_X^*(x)$ helps us identify the efficient and secure system design (e.g., secure coding/decoding) which achieves the secrecy capacity C_S.

In the context of secure RF communications, the transmitted signal X must satisfy (1.8) and [13] showed that Gaussian distribution is the solution of the optimization problem (1.11) and found a closed-form expression of C_S.

However, in the context of secure OWC, the transmitted signals must be nonnegative and satisfy peak- and average-intensity constraints (cf. equation (1.4)), and Gaussian distributions are not admissible. In fact, for the degraded free-space optical wiretap channel and the degraded continuous-time Poisson wiretap channel, the authors in [14, 15] established that the secrecy-capacity-achieving distribution $F_X^*(x)$ is discrete with a finite number of mass points.

1.4.3 Reliability in Multiuser OWC

Multiuser communication refers to a scenario in which multiple transmitters communicate their messages with multiple receivers simultaneously. In this setting, the objective is to identify the maximum reliable set of data rates (capacity region) and an efficient system design attaining these maximum data rates.

As an example of a multiuser communication, consider a multiple access channel (MAC) in which two transmitters wishes to communicate their messages to a common receiver simultaneously with communication rates R_1 and R_2, respectively. For this scenario, the set of maximum reliable data rates is called

the capacity region and it can be given as the convex-hull of the union of all the reliable communication rates satisfying [9]

$$\begin{cases} 0 \leq R_1 \leq \sup_{F_{X_1} F_{X_2}} I(X_1; Y|X_2), \\ 0 \leq R_2 \leq \sup_{F_{X_1} F_{X_2}} I(X_2; Y|X_1), \\ 0 \leq R_1 + R_2 \leq \sup_{F_{X_1} F_{X_2}} I(X_1, X_2; Y). \end{cases} \tag{1.12}$$

where X_i, $i \in \{1, 2\}$ is the transmitted signals from the transmitter i, $i \in \{1, 2\}$, Y is the received signal, F_{X_i}, $i \in \{1, 2\}$ is cumulative distribution function of X_i, $i \in \{1, 2\}$, and $I(X_1; Y|X_2)$ is the conditional mutual information between Y and X_1 given X_2.

In the context of RF communications, the transmitted signals X_1 and X_2 are subject to a power constraint (1.8) and [9] showed that a bivariate Gaussian distribution attains the capacity region.

However, in the context of multiuser OWC, the transmitted signals are nonnegative and are subject to peak- and average-intensity constraints (cf. equation (1.4)). Thus, Gaussian distributions are not admissible. In fact, for the free-space optical MAC and the Poisson optical MAC, the authors in [14, 16, 17] established that the capacity region is attained by discrete distributions with a finite number of mass points.

1.5 Limitations of Current Solutions

1.5.1 Limitations of RF Solutions

With the increased interest in OWC based on IM-DD, it becomes natural to study the security and reliability aspects of this communication methodology. Current research studies on RF systems do not directly extend to OWC due to the physical restrictions existing in an OWC setting. For instance, using a Gaussian input has been shown to be optimal for RF systems. It is known that Gaussian distributions achieve both the capacity and the secrecy capacity of RF systems when an average power constraint is imposed. However, this is not possible in an IM-DD system since the transmit signal has to be nonnegative. Thus, the optimal input distribution that satisfies a nonnegativity constraint has to be sought. Additionally, there is a natural constraint on the peak and average input signal which is reflected as the peak and average optical intensities. These constraints make the problem of finding the optimal input distributions fundamentally different from those studied in the literature. To overcome this limitation, it is required to find input distributions that satisfy nonnegativity, peak- and average-intensity constraints and achieve the capacity of the single-user OWC channels, secrecy capacity of the optical wiretap channels, and the capacity region of multiuser OWC channels.

Furthermore, since the capacity and the secrecy capacity of the majority of OWC channel models are unknown in a closed-form expression, it is of great interest to characterize these performance limits in the regimes where the constraints (peak- and average-intensity) tend to zero (low-intensity regime) or tend to infinity (high-intensity regime). Unfortunately, the asymptotic analysis in the context of OWC is also rather more complicated and challenging than RF. Most of the available asymptotic analysis relies on the fact that Gaussian distributions are optimal under an average power constraint. However, if the transmit signal is nonnegative and is subjected to peak- and average-intensity constraints, the characterization techniques become fundamentally different, and necessitates new asymptotic analysis approaches.

1.5.2 Limitations of Existing OWC Solutions

Studying the communications performance limits of the considered channel models for OWC, i.e., free-space, input-dependent Gaussian noise, and Poisson noise models, from an information-theoretic point of view is rather difficult due to the nonnegativity, peak- and average-intensity constrained input signals. The single-user channel capacities of these channel models are shown to be achieved via discrete input distributions with a finite number of mass points under nonnegativity, peak- and average-intensity constraints [6, 7, 10, 18]. Furthermore, when the channel input is only constrained by nonnegativity and average-intensity constraints, the capacity-achieving distributions for free-space channel [19] and Poisson noise optical channel [20] are shown to be discrete but with an unbounded support set, i.e., the support set of the optimal distributions are countably infinite. However, there are no results regarding the characterization of the capacity-achieving input distributions for the input-dependent Gaussian noise model. Finally, the single-user channel capacities of the considered optical channel models are only known in closed-form in the low- or high-intensity regimes [5, 8, 21], and in general, there are no closed-form characterization of the channel capacities.

The amount of studies regarding the secure design of OWC systems with different channel models are less abundant compared to the studies on the reliable single-user system design. The existing works are limited to the free-space optical wiretap channel with peak- and average-intensity constraints [14] as well as the continuous-time Poisson wiretap channel with a peak-intensity constraint [15]. Authors in [14] studied the free-space optical wiretap channel and proved that the entire rate-equivocation region of this wiretap channel is attained by discrete input distributions with finitely many mass points, but they did not provide any asymptotic analysis for the secrecy capacity. Additionally, [15] examined the degraded continuous-time Poisson wiretap channel with a peak-intensity constraint and gave a closed-form expression for the entire boundary of the rate-equivocation region. Particularly, the authors showed

that *binary* input distributions with mass points located at the origin and the peak-intensity constraint and with a very short duty cycle attain the boundary of the rate-equivocation region. However, there are no results regarding the secrecy capacity and the rate-equivocation region of an input-dependent Gaussian noise model or a discrete-time Poisson noise model.

Information-theoretic studies have also been performed for reliable multiuser OWC systems. For instance, the work in [22] considered the free-space optical multiple access channel with nonnegativity and peak-intensity constraints, and established that the boundary of the capacity region is obtained by distributions that are discrete with a finite number of mass points. Furthermore, [23, 24] provided tight bounds on the capacity region of free-space optical multiple access channel with peak- and average-intensity constraints across several intensity regimes (low, moderate, and high). Authors in [24] characterized the capacity region of free-space optical multiple access channel with nonnegativity and average-intensity constraints in the regime where the average-intensity tends to infinity. For a continuous-time Poisson optical multiple access channel subject to peak- and average-intensity constraints, Lapidoth *et al.* established the capacity region for the two-user case in a closed-form expression. The authors showed that for achieving every point on the boundary of the capacity region, the input distributions for both users must be binary with an infinite transmission bandwidth. The discrete-time Poisson optical multiple access channel has also been considered in [16], where authors considered a two-user case and verified the optimality of discrete inputs with a finite support set for achieving the sum-capacity when nonnegativity and peak-intensity constraint are imposed. However, the authors did not verify whether or not discrete input distributions exhaust the entire capacity region. Unfortunately, there are no studies on an optical multiple access channel with an input-dependent Gaussian noise under nonnegativity, peak- and average-intensity constraints. In particular, neither the optimal input distributions exhausting the entire capacity region are known, nor does an asymptotic analysis of the capacity region exist.

Chapter 2: Contributions

Optical wireless communication is an excellent candidate as a complementary or a backup technology to RF communications for providing high data-rate connections. Nevertheless, due to its specific properties, fundamentally different from those of its RF counterpart, deployment of OWC systems requires first revisiting what is known about RF design guidelines. Toward this end, this Ph.D. dissertation studies the fundamental performance limits (e.g., capacity, secrecy capacity, and capacity region) of reliable and secure OWC systems. These fundamental performance limits play a vital role in extracting guidelines and communication protocols for designing reliable and secure OWC systems.

After establishing the fundamental performance limits of single-user and multiuser OWC scenarios, powerful machine learning techniques, such as deep learning, are employed for the implementation of OWC systems. In particular, a simple and cost-effective learning-based system with (near-)optimal performance can be implemented by merely taking off-the-shelf deep learning models, applying them to an OWC design problem, and tuning them based on the easily generated training data.

2.1 Contributions to Optical Wiretap Channel with Input-Dependent Gaussian Noise Under the Peak- and Average-Intensity Constraints

Chapter 3 studies the problem of design a secure and reliable OWC system with an input-dependent Gaussian noise channel model. This input-dependent noise model is accurate scenarios when high power optical intensity signals are considered [1, 2, 5]. Subject to nonnegativity and peak-intensity constraints on the channel input, first, a practical optical wireless communication scenario for which the considered wiretap channel is stochastically degraded is presented. It is shown that the secrecy-capacity-achieving input distribution of the wiretap channel is discrete with a finite number of mass points, one of them located at the origin. Moreover, it is established that the entire boundary of the rate-equivocation region is also obtained by discrete input distributions with a finite number of mass points. Furthermore, the optimality of discrete input distributions with finitely many mass points is established in the presence of both peak- and average-intensity constraints. Finally, the asymptotic behavior of the secrecy capacity in the low- and high-intensity regimes is analyzed. In the low-intensity regime, the secrecy capacity scales quadratically with the peak-intensity constraint. On the other hand, in the high-intensity regime, the secrecy capacity does not scale with the constraint.

In this chapter, analytical results on the characterization of the optimal input distributions that attain the fundamental performance limits (e.g. capacity, secrecy capacity, and rate-equivocation region) of an input-dependent Gaussian noise optical wiretap channel are obtained. These results significantly contribute to the current understanding of the fundamental limits of optical wireless communications with secrecy constraints, advances the knowledge of the secure design of OWC systems in an input-dependent Gaussian noise optical channel, and identify secure OWC protocols. Particularly, the chapter studies the problem of secure and reliable OWC system when nonnegativity, peak- and average-intensity constrained input signals are considered. The obtained results show that to achieve the maximum secure and reliable data rate when communicating confidential data in an OWC setting over an input-dependent Gaussian noise channel, the transmitted signals must be designed to have a discrete probability distribution with a finite number of mass points [25].

2.2 Contributions to Optical Wiretap Channel with Input-Dependent Gaussian Noise Under the Average-Intensity Constraint

This chapter considers the problem of the secure and reliable design of OWC systems over an input-dependent Gaussian noise channel when only nonnegativity and average-intensity constraints are considered. This is a practical assumption for scenarios where the input signals are not restricted by a peak-intensity constraint. In this case, it is shown that the entire boundary of the rate-equivocation is achieved by discrete input distributions with countably infinite support set, but with finitely many mass points in any bounded interval. This implies that when the transmitted optical signals are restricted by only an average-intensity constraint: 1) the secrecy capacity is achieved by a distribution which has a countably infinite support set; 2) the single-user channel capacity (the case with no secrecy constraints) is also achieved by a distribution having a countably infinite support set. Notice that these results are in contrast to the case when both peak- and average-intensity constraints are active. In the latter case, the support set of the optimal input distributions contains a finite number of mass points.

The results of this chapter significantly contribute to the current understanding of the fundamental limits of optical wireless communications with secrecy constraints when only an average-intensity constraint is active and extracts design guidelines and protocols for a secure OWC over an input-dependent Gaussian noise channel. Particularly, the obtained results show that to achieve the maximum secure and reliable data rate when communicating confidential data in an OWC setting over an input-dependent

Gaussian noise channel, the transmitted signals must be designed to have a discrete probability distribution along with an infinite number of mass points, but with finitely many mass points in any bounded interval.

2.3 Contributions to Degraded Discrete-Time Poisson Wiretap Channel

Since the Poisson noise model is the most accurate model for the underlying OWC based on IM-DD, studying the fundamental performance limits of such a model is of great importance. To this end, a discrete-time Poisson wiretap channel subject to nonnegativity, peak- and average-intensity, as well as bandwidth constraints, is considered. First, the secrecy-capacity-achieving input distribution of this wiretap channel is proved to be discrete with a finite number of mass points. Furthermore, it is shown that every point on the boundary of the rate-equivocation region of this wiretap channel is also obtained by a discrete input distribution with finitely many mass points. Additionally, the analysis is extended to the case where only an average-intensity constraint is active. In this case, it is found that the secrecy capacity, as well as the entire boundary of the rate-equivocation region, are attained by discrete distributions with a countably infinite number of mass points, but with finitely many mass points in any bounded interval. Finally, an asymptotic analysis for characterizing the behavior of the secrecy capacity in the low-intensity and high-intensity regimes is provided. It is observed that when peak- or both peak- and average-intensity constraints are active the secrecy capacity scales quadratically with the peak-intensity constraint in the low-intensity regime. However, in the high-intensity regime and when the legitimate receiver's and the eavesdropper's channel gains are identical, the secrecy capacity does not scale with the constraints and hence, it is a constant value. Moreover, when the channel gains are different, the secrecy capacity cannot scale faster than the logarithm of the square root of the constraints.

This chapter studies the most accurate, yet the most difficult channel model for OWC, i.e., discrete-time Poisson. The problem of extracting design guidelines and protocols for a reliable and secure OWC over the discrete-time Poisson wiretap channel has been an open problem for about 40 years since Davis introduced the Poisson noise channel model for OWC in 1980 [6]. This chapter successfully addresses this challenging problem and fully characterizes a reliable and secure signal design for OWC over a discrete-time Poisson channel. More specifically, this chapter establishes that to have secure and reliable OWC over the discrete-time Poisson noise channel, the transmitted signals must be designed to have discrete probability distributions with a finite number of mass points when nonnegativity, peak- and average-intensity constraints are active [26]. Furthermore, in the absence of a peak-intensity constraint, the input

signals must follow discrete probability distributions with an infinite number of mass points, but with finitely many mass points in any bounded interval.

Furthermore, this chapter addresses a conjecture that was first introduced in 1988 by Shamai [7] regarding the capacity-achieving-distributions of the discrete-time Poisson channel under an average-intensity constraint. In his paper, Shamai conjectured that the capacity-achieving-distribution of this channel under nonnegativity and average-intensity constraints is discrete with an infinite number of mass points, but he did not provide any formal proof. This chapter formally proves the conjecture and characterizes the capacity-achieving distribution to be discrete with an infinite number of mass points, but with finitely many mass points in any bounded interval.

2.4 Contributions to Optical Multiple Access Channel with an Input-Dependent Gaussian Noise

Designing reliable multiuser wireless communication systems is more challenging compared to its single-user counterpart. This is because multiple transmitters simultaneously send their messages to multiple receivers, and therefore, cause harmful interference to the ongoing communications. This chapter considers a multiuser scenario in an OWC setting, namely, an optical multiple access channel with an input-dependent Gaussian noise. In this setup, two optical transmitters wish to simultaneously and reliably communicate their messages to a common optical receiver when nonnegativity, peak- and average-intensity constraints are considered for signal transmission. This scenario applies to several optical wireless links, most notably, space optical communications as well as Li-Fi systems. Under nonnegativity, peak- and average-intensity constraints, it is shown that generating code-books of both users according to discrete distributions with finitely many mass points achieve any point on the boundary of the capacity region [27]. Furthermore, an asymptotic analysis of the capacity region is conducted in the low-intensity regime, where the capacity region is explicitly presented in a closed-form expression and it is shown that binary distributions with mass points at the origin and the peak-intensity constraint are optimal. Numerical results indicate that due to the existence of an input-dependent noise component, the geometry of the capacity region under nonnegativity, peak- and average-intensity constraints is not a pentagon as opposed to the case of the Gaussian multiple access channel with peak- and/or average-power constraints.

In particular, this chapter advances the knowledge of reliable system design in a multiuser OWC setting. It is shown that designing the transmitted signals of both users based on discrete probability distributions with a finite number of mass points results in the best reliable and simultaneous communication performance when an input-dependent Gaussian noise model is considered for the underlying

OWC.

2.5 Contributions to Learning-Based Optical Wireless Communication Systems

Studying the communications performance limits of OWC is rather difficult compared to its RF counterpart. The reason is that the transmitted signals must satisfy nonnegativity, peak- and average-intensity constraints due to the physical restrictions existing in the optical wireless channels. More importantly, traditional approaches used in constructing the signal constellations for RF channels cannot be applied directly to the optical channels due to the mentioned constraints. Therefore, one should consider designing a structured optical signal-space model that can capture all the physical restrictions in an OWC setting. This task is not straightforward and heavily depends on the considered optical channel model. Hence, seeking communications techniques (such as modulation, coding, decoding, etc.) that do not heavily depend on an existing channel model is quite appealing. Motivated by the success of learning-based autoencoders in capturing the end-to-end performance of the RF communications system, this chapter proposes a reliable design of the OWC systems in both single-user and multiuser settings based on the deep neural network structures. For each of these scenarios, a deep neural network structure is built and trained to capture the end-to-end performance of an OWC system. According to the obtained results, the learning-based OWC can perform as well as the model-based counterpart. In particular, a simple and cost-effective learning-based system with (near-)optimal performance is proposed and is implemented by merely taking off-the-shelf deep learning models, applying them to an OWC design problem, and tuning them based on the easily generated training data. According to the obtained results, the learning-based OWC can perform as reliable as the model-based counterpart [28].

The proposed learning-based structures can capture the end-to-end performance of both single-user and multiuser OWC systems and through proper training, they can lead to a system design that performs as reliably as the model-based OWC systems. Therefore, this chapter advances the knowledge of reliable OWC system design by investigating the potential of applying machine learning methods to OWC systems and proves that a simple and cost-effective OWC that is designed entirely based on deep learning algorithms can provide a reliable single-user and multiuser OWC.

2.6 Dissertation Structure

The balance of this dissertation is organized as follows.

Chapter 3 studies the optical wiretap channel with an input-dependent Gaussian noise component, in which the main distortion is caused by an additive Gaussian noise whose variance depends on the current signal strength. Subject to nonnegativity, peak- and average-intensity constraints, it is shown that optimal input distributions achieving the secrecy capacity as well as any point on the boundary of the rate-equivocation region are discrete with finitely many mass points. Furthermore, the secrecy capacity is analyzed in low- and high-intensity regimes.

Chapter 4 considers the optical wiretap channel with an input-dependent Gaussian noise when only an average-intensity constraint is active. It is established that any point on the boundary of the rate-equivocation region is attained by discrete input distributions with a countably infinite number of mass points, but with finitely many mass points in any bounded interval. This result implies that when the transmitted optical signals are restricted by nonnegativity and average-intensity constraints, the secrecy capacity and the capacity are achieved by discrete distributions with a countably infinite number of mass point, but with finitely many mass points in any bounded interval.

Chapter 5 provides an analytical characterization of the optimal input distributions achieving the secrecy capacity as well as the boundary of the rate-equivocation region of the discrete-time Poisson wiretap channel under nonnegativity, peak- and/or average-intensity constraints. It is shown that when both peak- and average-intensity constraints are active, optimal distributions are discrete with a finite support set. However, when only an average-intensity constraint is considered, optimal distributions are discrete with a countably infinite number of mass points. Additionally, the behavior of the secrecy capacity is analyzed in the low- and high-intensity regimes.

In Chapter 6, the two-user optical multiple access channel with an input-dependent Gaussian noise component is investigated. It is shown that to achieve any point on the boundary of the capacity region, input distributions must be chosen to be discrete with finitely many mass points. Moreover, in the low-intensity regime, the capacity regime is explicitly characterized in a closed-form expression.

Chapter 7 proposes the design of reliable single-user and multiuser OWC systems entirely based on deep neural network autoencoders. This chapter compares the end-to-end performance of the proposed autoencoders (learning-based OWC systems) with the state-of-the-art model-based OWC systems in terms of the block error rate (BLER) performance metric. The obtained numerical results indicate that the proposed learning-based OWC system can perform as good as the model-based counterparts in both single- and multiuser settings.

Finally, Chapter 8 presents concluding remarks and suggestions for future work.

Chapter 3: Optical Wiretap Channel with Input-Dependent Gaussian Noise Under Peak- and Average-Intensity Constraints

M. Soltani and Z. Rezki, "Optical Wiretap Channel with Input-Dependent Gaussian Noise Under Peak- and Average-Intensity Constraints," *in IEEE Transactions on Information Theory, vol. 64, no. 10, pp. 6878-6893, Oct. 2018.*

M. Soltani and Z. Rezki, "Optical Wiretap Channel with Input-Dependent Gaussian Noise Under Peak Intensity Constraint," *in Proceedings of the IEEE International Zurich Seminar on Information and Communication (IZS'2018), Zurich, Switzerland, Feb. 2018.*

3.1 Introduction

Optical wireless communication (OWC) is a promising technique for supporting high data-rate communication as a complementary or a backup technology to radio-frequency (RF) communications. It has numerous advantages in comparison to RF, including higher data-rates, more abundant unlicensed spectrum and being less demanding in terms of system infrastructure.

One of the most popular communication techniques used in OWC is the intensity modulation and direct detection (IM-DD) technique for its simplicity [2]. In this setup, the channel input modulates the intensity of the emitted light. Thus, the input signal is proportional to the light intensity and is nonnegative. The receiver is usually equipped with a photodetector (PD) which measures the intensity of the received light and generates a signal proportional to the detected intensity, corrupted by noise. The simplest existing channel model for OWC is the free-space optical (FSO) channel, where the corrupting noise at the receiver is independent of the input signal [3]. To reflect a more accurate channel model for OWC, [5] assumes the corrupting noise to be dependent on the input signal (due to the random nature of photon emission in the laser diode) and derives asymptotic upper and lower bounds on the capacity under nonnegativity, peak- and average-intensity constraints. The work in [10] also focuses on the optical intensity channels with input-dependent Gaussian noise and proves that under peak- and average-intensity constraints, discrete input distributions with finite supports are capacity-achieving.

Exchanging confidential information over a communication medium (wired, wireless or optical) in the presence of unauthorized eavesdroppers has been always a challenging problem for system designers. This problem has been conventionally addressed by cryptographic encryption [11] without considering

the imperfections introduced by the communication channel. In this model, the usage of *secret keys* is the main approach for having secure communication. Wyner [12], on the other hand, proves the possibility of secure communications without relying on encryption by introducing the stochastically degraded wiretap channel model.

The wiretap channels are studied with respect to the rate-equivocation region, which is defined as the set of rate pairs for which the transmitter can communicate confidential messages reliably with a legitimate receiver while ensuring a certain secrecy level against an eavesdropper [29]. For the class of degraded wiretap channels, it is established in [12] that there exists a single-letter characterization for the rate-equivocation region. Authors in [30] study the Gaussian wiretap channel under an average power constraint and obtain a single-letter expression for the entire rate-equivocation region. Particularly, they show that under an average power constraint, the Gaussian distribution is the optimal input distribution for attaining both the capacity and the secrecy capacity with no compromise between the communication rate and the equivocation rate at the eavesdropper. On the other hand, under a peak-power constraint, the work in [14] proves that the entire rate-equivocation region of the Gaussian wiretap channel is achieved by discrete input distributions with finite supports. More specifically, the secrecy-capacity-achieving input distribution may not be identical to the capacity-achieving counterpart in general, resulting in a tradeoff between the rate and its equivocation.

This chapter considers an *optical* wiretap channel with input-dependent Gaussian noise which consists of a transmitter, a legitimate user and an eavesdropper. We assume that the output signals at both the legitimate user's and the eavesdropper's channels are corrupted by both input-dependent and input-independent Gaussian noises. In this setup, the objective is to have a secure communication with the legitimate user over an optical channel while keeping the eavesdropper ignorant of the transmitted message as much as possible. We study the optical wiretap channel with input-dependent Gaussian noise under nonnegativity, peak- and average-intensity constraints. We first present a practical OWC scenario based on IM-DD technique for which the optical wiretap channel with input-dependent Gaussian noise is stochastically degraded. We then use the results in [12] to conclude that there exists a single-letter expression for the entire rate-equivocation region. Next, a functional optimization problem is employed to obtain necessary and sufficient conditions, also known as Karush-Kuhn-Tucker (KKT) conditions, for the optimal input distribution. Using KKT conditions, it is proved by contradiction that the secrecy capacity and the entire rate-equivocation region of this wiretap channel are obtained by discrete input distributions with a finite number of mass points. Finally, an asymptotic analysis of the secrecy capacity in the low- and high-intensity regime is presented. More specifically, it is observed that in the low-intensity regime, the secrecy capacity is achieved by a binary input distribution and it scales quadratically with the

peak-intensity constraint. In the high-intensity regime, the secrecy capacity can be upper-bounded by a constant value implying that it does not scale with the constraints. The numerical results demonstrate that similar to the case of the Gaussian wiretap channel under a peak-power constraint, here too, the secrecy capacity and the capacity are not achieved by the same distribution in general. This, in turn, implies that there is a tradeoff between the rate and its equivocation.

In the case of the optical wiretap channel with input-dependent Gaussian noise, due to the existence of input-dependent noise components, our technical proofs differ from those of [14]. More specifically, our analysis for showing the analyticity of the mutual information densities are more challenging. Additionally, our contradiction statements for proving the discreteness of the optimal input distribution are different. Besides, it is proved that the secrecy-capacity-achieving input distribution has a mass point at the origin for the case of peak-intensity constraint and the case of peak- and average-intensity constraints.

3.2 System Model

We consider a practical OWC system where IM-DD is employed for optical communication. In this setup, the channel input modulates the emitted light intensity from Light Emitting Diode (LED) at the transmitter and PDs are used for receiving the optical signal at the legitimate user's and eavesdropper's receivers. We assume that there exist line-of-sight (LoS) paths between the optical transmitter and the receivers. In such a scenario, the received power of the LoS path dominates the received power of the reflected paths. Hence, the optical wireless channel between the transmitter and the legitimate user and between the transmitter and the eavesdropper become LoS channels [31, Chapter 2]. Figure 3.1 shows the geometry of an LoS OWC link with arbitrary receiver orientation. The LoS optical wireless channels between the transmitter and the legitimate user and between the transmitter and the eavesdropper are denoted by positive reals h and g, respectively, and are given by [31, Chapter 2]

$$h = \begin{cases} \dfrac{(m+1)A_{\mathrm{B}}}{2\pi d_{\mathrm{B}}^2} \cos^m(\theta_{\mathrm{B}}) \cos(\psi_{\mathrm{B}})\, f_{\mathrm{B}}\, k_{\mathrm{B}}, & \text{if } 0 \le \psi_{\mathrm{B}} \le \Psi_{\mathrm{B}}, \\ 0, & \text{if } \psi_{\mathrm{B}} > \Psi_{\mathrm{B}}, \end{cases} \tag{3.1}$$

$$g = \begin{cases} \dfrac{(m+1)A_{\mathrm{E}}}{2\pi d_{\mathrm{E}}^2} \cos^m(\theta_{\mathrm{E}}) \cos(\psi_{\mathrm{E}})\, f_{\mathrm{E}}\, k_{\mathrm{E}}, & \text{if } 0 \le \psi_{\mathrm{E}} \le \Psi_{\mathrm{E}}, \\ 0, & \text{if } \psi_{\mathrm{E}} > \Psi_{\mathrm{E}}, \end{cases} \tag{3.2}$$

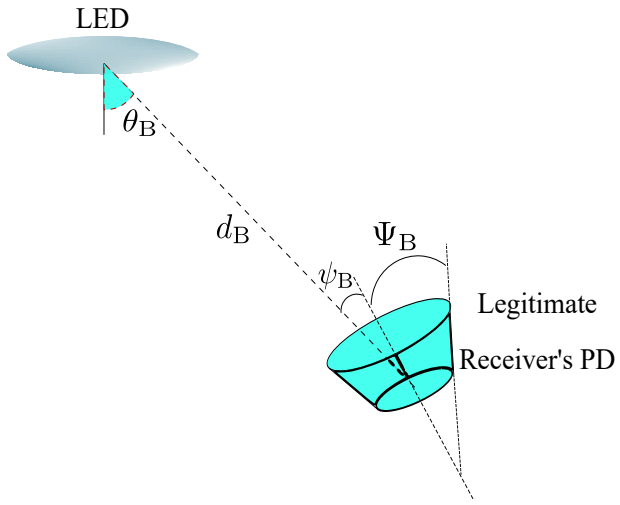

Figure 3.1: Geometry of a line of sight optical wireless link.

where m is the order of Lambertian emission, which depends on the semiangle at half illuminance of the LED $\Psi_{1/2}$ and is given by

$$m = -\frac{\ln(2)}{\ln\left(\cos\left(\Psi_{1/2}\right)\right)}. \tag{3.3}$$

The indices B and E stand for the legitimate user Bob and the eavesdropper Eve; d_{B} and d_{E} are the distances between the transmitter and the legitimate user and between the transmitter and the eavesdropper, respectively; A_{B} and A_{E} are the physical areas of the PDs at the legitimate user and the eavesdropper, respectively; $\theta_{\mathrm{B}} \in [0, \pi/2)$ and $\theta_{\mathrm{E}} \in [0, \pi/2)$ are the angles between the emitted light and the normal to the LED surface toward the legitimate receiver and the eavesdropper, respectively; ψ_{B} and ψ_{E} are the incident angle at the legitimate user and the eavesdropper, respectively; $\Psi_{\mathrm{B}} \in [0, \pi/2)$ and $\Psi_{\mathrm{E}} \in [0, \pi/2)$ are the concentrator field-of-view (FoV) of the legitimate user and the eavesdropper, respectively; f_{B}, k_{B}, f_{E} and k_{E} denote the optical filter gain and the concentrator gain of the legitimate user and the eavesdropper, respectively, and are assumed to be constant over their respective FoVs. An LoS IM-DD optical wireless link with input-dependent Gaussian noise from the transmitter to the legitimate user and from the transmitter to the eavesdropper can be respectively modeled as [31, Chapter 7]

$$\begin{cases} \widetilde{Y} = hX + \sqrt{hX}\widetilde{N}_{\mathrm{B},1} + \widetilde{N}_{\mathrm{B},1}, \\ \widetilde{Z} = gX + \sqrt{gX}\widetilde{N}_{\mathrm{E},1} + \widetilde{N}_{\mathrm{E},0}, \end{cases} \tag{3.4}$$

where X is the channel input ant it is a nonnegative random variable representing the intensity of the optical signal. Moreover, due to practical and safety restrictions, the input intensity is constrained by a

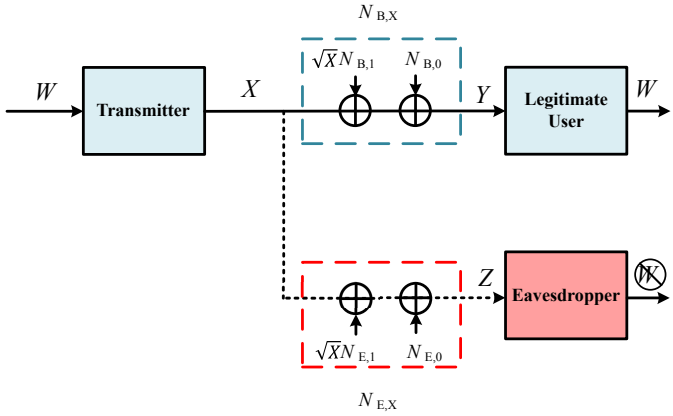

Figure 3.2: The optical wiretap channel with input-dependent Gaussian noise.

peak constraint in general, i.e., $X \leq A$ [2]. Therefore, the channel input is constrained as

$$0 \leq X \leq A. \tag{3.5}$$

$\widetilde{Y}$ and $\widetilde{Z}$ are the received optical signals at the legitimate user and the eavesdropper, respectively. h and g are the LoS channel gains between the transmitter and the legitimate user and between the transmitter and the eavesdropper, respectively, and are given by (3.1)–(3.2). $\widetilde{N}_{\mathrm{B},0} \sim \mathcal{N}(0, \widetilde{\sigma}^2_{\mathrm{B},0})$ and $\widetilde{N}_{\mathrm{B},1} \sim \mathcal{N}(0, \widetilde{\sigma}^2_{\mathrm{B},1})$ are the input-independent and input-dependent Gaussian noise components at the legitimate receiver, respectively, and are assumed to be independent. Similarly, $\widetilde{N}_{\mathrm{E},0} \sim \mathcal{N}(0, \widetilde{\sigma}^2_{\mathrm{E},0})$ and $\widetilde{N}_{\mathrm{E},1} \sim \mathcal{N}(0, \widetilde{\sigma}^2_{\mathrm{E},1})$ are the input-independent and input-dependent Gaussian noise components at the eavesdropper, respectively and are assumed to be independent. Furthermore, $\widetilde{N}_{\mathrm{B},0}$ and $\widetilde{N}_{\mathrm{E},0}$ are also assumed to be independent. We note that the input-dependent noise in (3.4) is due to the nonlinearity in the optical channel [31, Chapter 7]. Figure 3.2 depicts an optical wiretap channel that is equivalent to the optical wiretap channel described by (3.4). In this equivalent wiretap channel, each link is an optical channel with input-dependent Gaussian noise model and is given by [5]

$$\begin{cases} Y = X + \sqrt{X}N_{\mathrm{B},1} + N_{\mathrm{B},0}, \\ Z = X + \sqrt{X}N_{\mathrm{E},1} + N_{\mathrm{E},0}, \end{cases} \tag{3.6}$$

where $Y = \frac{\widetilde{Y}}{h}$, $Z = \frac{\widetilde{Z}}{g}$, $N_{\mathrm{B},0} \sim \mathcal{N}(0, \sigma^2_{\mathrm{B}})$, $N_{\mathrm{B},1} \sim \mathcal{N}(0, \sigma^2_{\mathrm{B}}\eta^2_{\mathrm{B}})$, $N_{\mathrm{E},0} \sim \mathcal{N}(0, \sigma^2_{\mathrm{E}})$ and $N_{\mathrm{E},1} \sim \mathcal{N}(0, \sigma^2_{\mathrm{E}}\eta^2_{\mathrm{E}})$ with $\sigma^2_{\mathrm{B}} = \frac{\widetilde{\sigma}^2_{\mathrm{B},0}}{h^2}$, $\eta^2_{\mathrm{B}} = \frac{\widetilde{\sigma}^2_{\mathrm{B},1}}{\widetilde{\sigma}^2_{\mathrm{B},0}}h$, $\sigma^2_{\mathrm{E}} = \frac{\widetilde{\sigma}^2_{\mathrm{E},0}}{g^2}$, and $\eta^2_{\mathrm{E}} = \frac{\widetilde{\sigma}^2_{\mathrm{E},1}}{\widetilde{\sigma}^2_{\mathrm{E},0}}g$, respectively. The variables η^2_{B} and η^2_{E} denote the ratios of the input-dependent noise variances to the input-independent noise variances of the legitimate user's and the eavesdropper's channels, respectively. Notice that changing the orientation

of the transmitter with respect to the legitimate receiver's and the eavesdropper's positions affects the channel gains h and g through ψ_{B}, θ_{B}, ψ_{E}, θ_{E}, respectively. We note that if the system parameters satisfy

$$\frac{\cos(\psi_{\mathrm{B}})}{\cos(\psi_{\mathrm{E}})}\left[\frac{\cos(\theta_{\mathrm{B}})}{\cos(\theta_{\mathrm{E}})}\right]^{m} = \frac{\widetilde{\sigma}^2_{\mathrm{B},1}}{\widetilde{\sigma}^2_{\mathrm{E},1}}\, d^2_{\mathrm{BE}}\, \rho_{\mathrm{EB}}, \tag{3.7}$$

where $d^2_{\mathrm{BE}} = \frac{d^2_{\mathrm{B}}}{d^2_{\mathrm{E}}}$ and $\rho_{\mathrm{EB}} = \frac{A_{\mathrm{E}} f_{\mathrm{E}} k_{\mathrm{E}}}{A_{\mathrm{B}} f_{\mathrm{B}} k_{\mathrm{B}}}$, then, using (3.1)–(3.2), one obtains

$$\sigma^2_{\mathrm{B}}\eta^2_{\mathrm{B}} = \sigma^2_{\mathrm{E}}\eta^2_{\mathrm{E}}, \tag{3.8}$$

which implies that the input-dependent noise components in both channels are statistically equivalent. One can show via numerical inspections that (3.8) might be satisfied for various system parameters θ_{B}, ψ_{B}, θ_{E}, ψ_{E}, $\widetilde{\sigma}^2_{\mathrm{B},1}$, $\widetilde{\sigma}^2_{\mathrm{E},1}$, d^2_{BE} and ρ_{EB}. This implies that in such cases, $N_{\mathrm{B},1}$ and $N_{\mathrm{E},1}$ are statistically equivalent and the optical wiretap channel with input-dependent Gaussian noise can be stochastically degraded.

In the sequel, without loss of generality, the wiretap channel described by (3.6) whose input is X and whose outputs are Y and Z, at the legitimate receiver and the eavesdropper, respectively, is considered. In this optical wiretap channel, if $\sigma^2_{\mathrm{B}} < \sigma^2_{\mathrm{E}}$ and $\sigma^2_{\mathrm{B}}\eta^2_{\mathrm{B}} = \sigma^2_{\mathrm{E}}\eta^2_{\mathrm{E}}$, then the random variables X, Y and Z form the Markov chain $X \to Y \to Z$ and consequently the optical wiretap channel becomes stochastically degraded. As a result, the rate-equivocation region of such an optical channel can be expressed in a single-letter form due to[12]. Furthermore, under the conditions $\sigma^2_{\mathrm{B}} \geq \sigma^2_{\mathrm{E}}$ and $\sigma^2_{\mathrm{B}}\eta^2_{\mathrm{B}} = \sigma^2_{\mathrm{E}}\eta^2_{\mathrm{E}}$, the random variables X, Y and Z form the Markov chain $X \to Z \to Y$, from which it can be easily inferred that the secrecy capacity (defined later in this section) is equal to zero.

3.2.1 The Characterization of the Rate-Equivocation Region

An $(n, 2^{nR})$ code for the peak-intensity-constrained optical wiretap channel with input-dependent Gaussian noise consists of the random variable W (message set) uniformly distributed over the set $\mathcal{W} = \{1, 2, \ldots, 2^{nR}\}$, an encoder at the transmitter $f_n : \mathcal{W} \to [0, A]^n$ satisfying the nonnegativity and peak-intensity constraints, and a decoder at the legitimate user $g_n : \mathbb{R}^n \to \mathcal{W}$. The equivocation of the confidential message W is defined as the eavesdropper's uncertainty about W and is measured by the normalized conditional entropy $\frac{1}{n} H(W|Z^n)$. The probability of error for such a code is defined as $P^n_{\mathrm{e}} = \Pr\{g_n(Y^n) \neq W\}$. A rate-equivocation pair (R, R_{eq}) is said to be achievable if there exists an

$(n, 2^{nR})$ code satisfying

$$\lim_{n\to\infty} P_{\mathrm{e}}^{n} = 0, \tag{3.9}$$

$$R_{\mathrm{eq}} \leq \lim_{n\to\infty} \frac{1}{n} H(W|Z^n). \tag{3.10}$$

The rate-equivocation region consists of all achievable rate-equivocation pairs, and is denoted by $\mathcal{E}$. A rate R is said to be perfectly secure if $R_{\mathrm{eq}} = R$, i.e., if there exists an $(n, 2^{nR})$ code satisfying $\lim_{n\to\infty} \frac{1}{n} I(W; Z^n) = 0$. The supremum of such rates is defined to be the secrecy capacity and denoted by C_{S}.

Since under the assumption of $\sigma_{\mathrm{B}}^2\eta_{\mathrm{B}}^2 = \sigma_{\mathrm{E}}^2\eta_{\mathrm{E}}^2$ and $\sigma_{\mathrm{E}}^2 > \sigma_{\mathrm{B}}^2$, the optical wiretap channel with input-dependent Gaussian noise under nonnegativity and peak-intensity constraints is stochastically degraded, its entire rate-equivocation region $\mathcal{E}$ can be expressed in a single-letter expression and the entire rate-equivocation region of the this wiretap channel is given by the union of the rate-equivocation pairs (R, R_{eq}) such that [12]

$$R \leq I(X; Y), \tag{3.11}$$

$$R_{\mathrm{eq}} \leq I(X; Y) - I(X; Z), \tag{3.12}$$

for some input distribution $F_X \in \mathcal{A}^+$, where $I(X; Y)$ and $I(X; Z)$ are the mutual information of the legitimate user's and the eavesdropper's channels, respectively, and the feasible set $\mathcal{A}^+$ is given by

$$\mathcal{A}^+ \triangleq \left\{ F_X : \int_0^A dF_X(x) = 1 \right\}. \tag{3.13}$$

3.3 Main Results

In this section, main results related to the optical wiretap channel with input-dependent Gaussian noise under nonnegativity and peak-intensity constraints are given. We first focus on the secrecy capacity and prove the discreteness of the secrecy-capacity-achieving input distribution. We then establish that the entire rate-equivocation region of this wiretap channel is also obtained by discrete input distributions with finite supports.

3.3.1 Results on the Secrecy Capacity

The secrecy capacity of the optical wiretap channel with input-dependent Gaussian noise under non-negativity and peak-intensity constraints is given by the solution of the following optimization problem

$$\max_{F_X \in \mathcal{A}^+} g_0(F_X), \tag{3.14}$$

where $g_0(F_X) = I(X;Y) - I(X;Z)$. Under the constraint (3.5), the solution of (3.14) is discrete with a finite support as stated by Theorem 1.

Theorem 1. *There exists a unique input distribution that attains the secrecy capacity of the optical wiretap channel with input-dependent Gaussian noise under nonnegativity and peak-intensity constraints. Furthermore, the support set of this optimal input distribution is a finite set.*

Proof. The proof is provided in Appendix A. ■

To prove Theorem 1, first, it is shown that the set of input distributions $\mathcal{A}^+$ that satisfies (3.13), is compact and convex. We then show that the objective function in (3.14) is continuous, strictly concave and weakly differentiable in the input distribution F_X and hence one concludes that the solution to the optimization problem (3.14) exists and is unique. We continue the proof by deriving the necessary and sufficient conditions (KKT conditions) for the optimality of the optimal input distribution F_X^* and finally by means of contradiction it is shown that this optimal input distribution is discrete with a finite number of mass points. Unlike the case of the Gaussian wiretap channel under a peak-power constraint, where the corrupting noise components are assumed to be independent from channel input [14], the existence of input-dependent noise components in the optical wiretap channel with input-dependent Gaussian noise results in several challenging problems: 1) The conditions under which this wiretap channel under a peak-intensity constraint becomes stochastically degraded is different than that of [14]; 2) The proof of the analyticity of the mutual information density functions are different from those presented in [14]; 3) The technical steps for proving the discreteness of the solution to problem (3.14) are different from those utilized in [14] and cannot be generalized to those cases.

Next, the existence of a mass point at $x = 0$ in the support set of the secrecy-capacity-achieving input distribution is established.

Proposition 1. *Let $\mathcal{S}_{F_X^*}$ be the support set of the secrecy-capacity-achieving input distribution F_X^* for the optical wiretap channel with input-dependent Gaussian noise under nonnegativity and peak-intensity constraints. Then $x = 0$ always belongs to $\mathcal{S}_{F_X^*}$.*

Proof. The proof is presented in Appendix A. ■

It is worth mentioning that the existence of a mass point at the origin has also been established in [10] for the optical channel with input-dependent Gaussian noise, but with no secrecy constraints. Furthermore, the proof of Proposition 1 also holds true for the case where $\eta_{\mathrm{B}}^2 = \eta_{\mathrm{E}}^2 = 0$, i.e., the optical wiretap channel with input-independent Gaussian noise.

3.3.2 Results on the Rate-Equivocation Region

By a time-sharing argument, it can be shown that the rate-equivocation region of the optical wiretap channel with input-dependent Gaussian noise is convex. Therefore, the region can be characterized by finding tangent lines to $\mathcal{E}$, which are given by the solutions of

$$\max_{F_X \in \mathcal{A}^+} g_\lambda(F_X), \tag{3.15}$$

where $g_\lambda(F_X) = \lambda I(X;Y) + (1-\lambda)\left[I(X;Y) - I(X;Z)\right]$, for all $\lambda \in [0,1]$. Next, it is verified that the entire rate-equivocation region of the optical wiretap channel with input-dependent Gaussian noise under the constraints (3.5) is also obtained by discrete input distributions with finite supports.

Theorem 2. *There exists a unique input distribution that achieves the boundary of the rate-equivocation region of the optical wiretap channel with input-dependent Gaussian noise under nonnegativity and peak-intensity constraints. This optimal input distribution is discrete with a finite support.*

Proof. Theorem 2 is established in Appendix A. ■

It is worth noting that for the case when $\eta_{\mathrm{B}}^2 = \eta_{\mathrm{E}}^2 = 0$ (i.e., for the optical wiretap channel with input-independent Gaussian noise), an approach similar to the one in [14] can be used to prove the discreteness of the optimal solutions (3.14) and (3.15). An interesting observation is that our contradiction argument for proving the discreteness of the optimal solutions of (3.14) and (3.15) (Equations (A.54)–(A.63)), when $\eta_{\mathrm{B}}^2, \eta_{\mathrm{E}}^2 \neq 0$ cannot be generalized to the case when $\eta_{\mathrm{B}}^2 = \eta_{\mathrm{E}}^2 = 0$. A similar observation has been made in [10], but for the case with no secrecy constraint.

3.4 Asymptotic results for a peak-intensity constraint

This section provides the asymptotic analysis on the secrecy capacity of the optical wiretap channel with input-dependent Gaussian noise under nonnegativity and peak-intensity constraints. First, the

secrecy capacity is investigated for asymptotically small values of A. Second, it is proved that for high-intensity regime, the secrecy capacity can be bounded by a constant value implying that it does not scale with the peak-intensity constraint in this regime.

3.4.1 Low-Intensity Results

For relatively small values of the peak-intensity constraint A, one can use the results shown in [32] to get

$$I(X;Y) - I(X;Z) = \frac{1}{2}[J_{\mathrm{B}}(0) - J_{\mathrm{E}}(0)] \operatorname{Var}(X) + o(A^2), \tag{3.16}$$

where $o(A^2)$ denotes a term that tends to 0 faster than A^2, $\operatorname{Var}(X)$ is the variance of the random variable X, $J_{\mathrm{B}}(0)$ and $J_{\mathrm{E}}(0)$ denote the Fisher information of the legitimate user's and wiretap channel at 0 and $J(x)$ is given by

$$J(x) = \int_y \frac{\left(\frac{d}{dx} p(y|x)\right)^2}{p(y|x)}\, dy. \tag{3.17}$$

For the channel laws (A.3) and (A.4), one can write

$$J_{\mathrm{B}}(x) = \frac{2 + \eta_{\mathrm{B}}^4 \sigma_{\mathrm{B}}^2 + 2\eta_{\mathrm{B}}^2 x}{2\sigma_{\mathrm{B}}^2 (1 + \eta_{\mathrm{B}}^2 x)^2}, \tag{3.18}$$

$$J_{\mathrm{E}}(x) = \frac{2 + \eta_{\mathrm{E}}^4 \sigma_{\mathrm{E}}^2 + 2\eta_{\mathrm{E}}^2 x}{2\sigma_{\mathrm{E}}^2 (1 + \eta_{\mathrm{E}}^2 x)^2}, \tag{3.19}$$

such that

$$J_{\mathrm{B}}(0) = \frac{2 + \eta_{\mathrm{B}}^4 \sigma_{\mathrm{B}}^2}{2\sigma_{\mathrm{B}}^2}, \tag{3.20}$$

$$J_{\mathrm{E}}(0) = \frac{2 + \eta_{\mathrm{E}}^4 \sigma_{\mathrm{E}}^2}{2\sigma_{\mathrm{E}}^2}. \tag{3.21}$$

Therefore, for small values of A, the secrecy capacity is

$$C_{\mathrm{S}} = \frac{1}{2}[J_{\mathrm{B}}(0) - J_{\mathrm{E}}(0)] \max_{F_X \in \mathcal{A}^+} \operatorname{Var}(X) + o(A^2). \tag{3.22}$$

Proposition 2. *In the regime $A \ll 1$, the secrecy capacity under the peak-intensity constraint is as follows*

$$C_{\mathrm{S}}(A) = \frac{A^2}{8}\left(\frac{1}{\sigma_{\mathrm{B}}^2} - \frac{1}{\sigma_{\mathrm{E}}^2} + \frac{1}{2}\left(\eta_{\mathrm{B}}^4 - \eta_{\mathrm{E}}^4\right)\right) + o(A^2). \tag{3.23}$$

Proof. See Appendix B. ■

Proposition 2 indicates the secrecy capacity to be a quadratic function of the peak-intensity constraint A in the low-intensity regime. Furthermore, as shown in Appendix B, under constraint (3.5), two mass points located at 0 and A with equal probabilities are optimal in this regime.

3.4.2 High-Intensity Results

In this section, an upper bound on the secrecy capacity that holds for any value of A is presented. Based on (A.40), the secrecy capacity can be simplified as

$$C_{\mathrm{S}} = h_Y(F_X^*) - h_Z(F_X^*) + \frac{1}{2}\log\left(\frac{\sigma_{\mathrm{E}}^2}{\sigma_{\mathrm{B}}^2}\right) + \frac{1}{2}\mathbb{E}_{F_X^*}\left[\log\left(\frac{1+\eta_{\mathrm{E}}^2 x}{1+\eta_{\mathrm{B}}^2 x}\right)\right]. \tag{3.24}$$

Since the optical wiretap channel with input-dependent Gaussian noise is stochastically degraded and based on the fact that $\sigma_{\mathrm{E}}^2\eta_{\mathrm{E}}^2 = \sigma_{\mathrm{B}}^2\eta_{\mathrm{B}}^2$ and $\sigma_{\mathrm{E}}^2 > \sigma_{\mathrm{B}}^2$, one can write $Z = Y + N_{\mathrm{D}}$ for some zero-mean Gaussian random variable N_D with variance $\sigma_{\mathrm{D}}^2 = \sigma_{\mathrm{E}}^2 - \sigma_{\mathrm{B}}^2$. Therefore, $h(Z) > h(Z|N_{\mathrm{D}}) = h(Y)$ and consequently $h(Z) > h(Y)$ for any nontrivial input distribution F_X^*. Furthermore, as $\eta_{\mathrm{E}}^2 < \eta_{\mathrm{B}}^2$ and $x \geq 0$, one finds $\frac{1}{2}\mathbb{E}_{F_X^*}\left[\log\left(\frac{1+\eta_{\mathrm{E}}^2 x}{1+\eta_{\mathrm{B}}^2 x}\right)\right] \leq 0$. As a result, $C_{\mathrm{S}} \leq \frac{1}{2}\log\left(\frac{\sigma_{\mathrm{E}}^2}{\sigma_{\mathrm{B}}^2}\right)$ for any value of A. This, in turn, implies that the secrecy capacity in the regime $A \to \infty$ does not scale with the peak-intensity constraint A and converges to a real and positive constant, i.e.,

$$C_{\mathrm{S}}(A) = O(1), \tag{3.25}$$

where $O(1)$ is a function such that for large enough A, the secrecy capacity is at most if k_0, for some real number $k_0 > 0$.

3.5 The case of peak- and average-intensity constraints

In this section, the discreteness of the optimal input distribution is generalized to the case when an additional average-intensity constraint is imposed on the channel input. First, the Theorems 1 and 2 are generalized to the case when both peak- and average-intensity constraints are active by establishing parallels to the proof of these theorems. Let the average intensity constraint be P. The new feasible set for the input distribution is

$$\mathcal{M}^+ \triangleq \left\{F_X : \int_0^A dF_X(x) = 1,\ \int_0^A x\, dF_X(x) \leq P\right\}. \tag{3.26}$$

We first consider the secrecy capacity

$$C_{\mathrm{S}} = \max_{F_X \in \mathcal{M}^+} [I(X;Y) - I(X;Z)]. \tag{3.27}$$

Similar to the lines provided for the proof of Theorem 1, here too, the mutual information difference $I(X;Y) - I(X;Z)$ is strictly concave and continuous in the input distribution F_X. Furthermore, $\mathcal{M}^+$ is a compact and convex set [33, Appendix A.1]. Thus the necessary and sufficient conditions in (A.38)–(A.39) take the new form

$$r_{\mathrm{eq}}(x; F_X^*) - \gamma x \leq C_{\mathrm{S}} - \gamma \, \mathbb{E}[X], \quad \forall x \in [0, A], \tag{3.28}$$

$$r_{\mathrm{eq}}(x; F_X^*) - \gamma x = C_{\mathrm{S}} - \gamma \, \mathbb{E}[X], \quad \forall x \in \mathcal{S}_{F_X^*}, \tag{3.29}$$

$$\gamma \, (\mathbb{E}[X] - P) = 0, \tag{3.30}$$

for some $\gamma \geq 0$. Assuming that the average intensity constraint is tight, we have $\gamma > 0$. For the case of $\gamma = 0$, the only imposed constraints are the nonnegativity and peak-intensity constraints and we have already proven in Theorem 1 that the optimal input distribution is discrete. Hence, (3.28)–(3.30) can be rewritten as

$$r_{\mathrm{eq}}(x; F_X^*) - \gamma x \leq C_{\mathrm{S}} - \gamma \, \mathbb{E}[X], \quad \forall x \in [0, A], \tag{3.31}$$

$$r_{\mathrm{eq}}(x; F_X^*) - \gamma x = C_{\mathrm{S}} - \gamma \, \mathbb{E}[X], \quad \forall x \in \mathcal{S}_{F_X^*}, \tag{3.32}$$

$$\mathbb{E}[X] = P. \tag{3.33}$$

Next, it is proved by contradiction that the optimal input distribution F_X^* satisfying (3.31)–(3.33) must be discrete with a finite number of mass points. To this end, let $\mathcal{S}_{F_X^*}$ have an infinite number of elements. In view of the optimality condition (3.31)–(3.33), the analyticity of $r_{\mathrm{eq}}(w; F_X)$ and w over $\mathcal{D}$ and the Identity Theorem of complex analysis, along with the Bolzano-Weierstrass Theorem, if $\mathcal{S}_{F_X^*}$ has an infinite number of mass points, one gets $r_{\mathrm{eq}}(w; F_X^*) - \gamma \, w = C_{\mathrm{S}} - \gamma \, P$ for all $w \in \mathcal{D}$, which results in

$$r_{\mathrm{eq}}(x; F_X^*) - \gamma \, x = C_{\mathrm{S}} - \gamma \, P, \quad x \in (-1/\eta_{\mathrm{B}}^2, +\infty), \tag{3.34}$$

$$\mathbb{E}[X] = P. \tag{3.35}$$

Using the bounds given in (A.57)–(A.58), one can write

$$L \leq C_{\mathrm{S}} - \gamma\, P + \gamma\, x + \frac{1}{2}\log\left(\frac{\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x)}{\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x)}\right) \leq U. \tag{3.36}$$

Defining a convergent sequence of distinct points $\{x_n\}_{n\in\mathbb{N}}$ in $\mathbb{S}$ with a limit point $x_0 = -1/\eta_{\mathrm{B}}^2$, we have: 1) x_n and $\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x_n)$ are real for all positive integers n; 2) $\lim_{n\to\infty}\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x_n) = 0$. Following the results in [10, Theorem 3] and using (3.36) one can write

$$\lim_{n\to\infty}(L - C_{\mathrm{S}}) \leq \lim_{n\to\infty}\left[\gamma\, x_n - \gamma\, P + \frac{1}{2}\log\left(\frac{\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x_n)}{\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x_n)}\right)\right] \leq \lim_{n\to\infty}(U - C_{\mathrm{S}}). \tag{3.37}$$

Since $\lim_{n\to\infty}\frac{1}{2}\log\left(\frac{\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x_n)}{\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x_n)}\right) = -\infty$ and $L - C_{\mathrm{S}}$ is a finite value, a contradiction occurs. This, in turn, implies that the support set $\mathcal{S}_{F_X^*}$ cannot have an infinite number of mass points and therefore the optimal input distribution F_X^* under the nonnegativity, peak- and average-intensity constraints, is also discrete with a finite number of mass points. Additionally, following along similar lines as in Proposition 1, one can prove that $x = 0$ belongs to the support set $\mathcal{S}_{F_X^*}$ of the secrecy-capacity-achieving input distribution with peak- and average-intensity constraints.

Finally, this contradiction argument is extended to the entire rate-equivocation region. Consider the optimization problem for determining the boundary point of the rate-equivocation region

$$\max_{F_X\in\mathcal{M}^+}\{\lambda\, I(X;Y) + (1-\lambda)\,[I(X;Y) - I(X;Z)]\}. \tag{3.38}$$

We note that if the average intensity constraint is not tight, i.e., $\mathbb{E}[X] < P$, the problem reduces to the case where only the nonnegativity and peak-intensity constraints are present, in which case the optimal input distribution is discrete with a finite support by Theorem 2. Hence, without loss of generality, one can assume that the average-intensity constraint is tight and the necessary and sufficient optimality conditions for (3.38) become

$$\lambda\, i_{\mathrm{B}}(x;F_X^*) + (1-\lambda)\, r_{\mathrm{eq}}(x;F_X^*) - \gamma\, x \leq \lambda\, I_{\mathrm{B}}(F_X^*) + (1-\lambda)[I_{\mathrm{B}}(F_X^*) - I_{\mathrm{E}}(F_X^*)] - \gamma\, P, \quad \forall x\in[0,A], \tag{3.39}$$

$$\lambda\, i_{\mathrm{B}}(x;F_X^*) + (1-\lambda)\, r_{\mathrm{eq}}(x;F_X^*) - \gamma\, x = \lambda\, I_{\mathrm{B}}(F_X^*) + (1-\lambda)[I_{\mathrm{B}}(F_X^*) - I_{\mathrm{E}}(F_X^*)] - \gamma\, P, \quad \forall x\in\mathcal{S}_{F_X^*}, \tag{3.40}$$

$$\mathbb{E}[X] = P. \tag{3.41}$$

Next, using contradiction argument, it is shown that the input distribution F_X^* satisfying (3.39)–(3.41) must be a discrete distribution with a finite support. Assume, on the contrary, that $\mathcal{S}_{F_X^*}$ has an infinite number of elements. In view of (3.39)–(3.41) and the analyticity of $i_{\mathrm{B}}(w;F_X)$, $r_{\mathrm{eq}}(w;F_X)$ and w over $\mathcal{D}$ and the Identity Theorem of complex analysis, one finds

$$\lambda\, i_{\mathrm{B}}(x;F_X^*) + (1-\lambda)\, r_{\mathrm{eq}}(x;F_X^*) - \gamma\, x = \lambda\, I_{\mathrm{B}}(F_X^*) + (1-\lambda)[I_{\mathrm{B}}(F_X^*) - I_{\mathrm{E}}(F_X^*)] - \gamma\, P,\ \forall\, x > -1/\eta_{\mathrm{B}}^2, \tag{3.42}$$

$$\mathbb{E}\,[X] = P. \tag{3.43}$$

Using the bounds presented in (A.69)–(A.70), one can verify that

$$\widetilde{L} \leq I_{\mathrm{B}}(F_X^*) - (1-\lambda)\, I_{\mathrm{E}}(F_X^*) + \frac{1}{2}\log\left(\frac{\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x)}{\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x)}\right) + \frac{\lambda}{2}\log(2\pi e\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x)) + \gamma\,(x-P) \leq \widetilde{U}. \tag{3.44}$$

Now, let $\{x_n\}_{n\in\mathbb{N}}$ be a convergent sequence of distinct points in $\mathbb{S}$ such that it is converging to a limit point $x_0 = -1/\eta_{\mathrm{B}}^2$. Based on this, we have the following results: 1) x_n and $\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x_n)$ are real for all positive integers n; 2) $\lim_{n\to\infty}\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x_n) = 0$. Following the results in [10, Theorem 3] and using (3.44), we get

$$\begin{aligned}\lim_{n\to\infty}[\widetilde{L} - I_{\mathrm{B}}(F_X^*) + (1-\lambda)\, I_{\mathrm{E}}(F_X^*)] &\leq \lim_{n\to\infty}\left[\frac{1}{2}\log\left(\frac{\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x_n)}{\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x_n)}\right) + \frac{\lambda}{2}\log(2\pi e\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x_n))\right.\\ &\qquad\left. + \gamma\,(x_n - P)\right]\\ &\leq \lim_{n\to\infty}[\widetilde{U} - I_{\mathrm{B}}(F_X^*) + (1-\lambda)\, I_{\mathrm{E}}(F_X^*)].\end{aligned} \tag{3.45}$$

Since, $\lim_{n\to\infty}\frac{1}{2}\log\left(\frac{\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x_n)}{\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x_n)}\right) = -\infty$ while $\widetilde{L} - I_{\mathrm{B}}(F_X^*) + (1-\lambda)\, I_{\mathrm{E}}(F_X^*)$ is a finite value, a contradiction occurs and thus, the support set $\mathcal{S}_{F_X^*}$ has a finite number of mass points.

3.6 ASYMPTOTIC RESULTS FOR THE SECRECY CAPACITY UNDER PEAK- AND AVERAGE-INTENSITY CONSTRAINTS

This section provides the asymptotic analysis on the secrecy capacity of the optical wiretap channel with input-dependent Gaussian noise under peak- and average-intensity constraints. First, the secrecy capacity is investigated for asymptotically small values of A and P with their ratio $\kappa \triangleq \frac{P}{A} \in (0, \frac{1}{2})$. Second, an upper bound on the secrecy capacity will be given that holds true for any value of A and κ.

3.6.1 Low-Intensity Results

For the small values of A and κ, based on the results in [32] one can write

$$C_{\mathrm{S}} = \frac{1}{2}[J_{\mathrm{B}}(0) - J_{\mathrm{E}}(0)] \max_{F_X \in \mathcal{M}^+} \mathrm{Var}\,(X) + o(A^2). \tag{3.46}$$

Proposition 3. *In the regime $A \ll 1$ and $\kappa \in (0, \frac{1}{2})$, the secrecy capacity under the peak- and average-intensity constraints is as follows*

$$C_{\mathrm{S}}(A, \kappa) = \frac{A^2}{2}\kappa(1-\kappa)\left(\frac{1}{\sigma_{\mathrm{B}}^2} - \frac{1}{\sigma_{\mathrm{E}}^2} + \frac{1}{2}(\eta_{\mathrm{B}}^4 - \eta_{\mathrm{E}}^4)\right) + o(A^2). \tag{3.47}$$

Proof. See Appendix C. ■

Similar to the case with a peak-intensity constraint, Proposition 3 reflects that the secrecy capacity under peak- and average-intensity constraints scales quadratically in A. Furthermore, as shown in Appendix C, the secrecy-capacity-achieving input distribution possesses two mass points located at 0 and A with probabilities κ and $1-\kappa$, respectively. Additionally, note that based on Appendix C, the average-intensity constraint is inactive when $\kappa \in [\frac{1}{2}, 1]$ and consequently $C_{\mathrm{S}}(A, \kappa)$ is given by (3.23).

3.6.2 High-Intensity Results

In the regime $A \to \infty$ and $P \to \infty$ with their ratio κ kept fixed, the secrecy capacity $C_{\mathrm{S}} \leq \frac{1}{2}\log\left(\frac{\sigma_{\mathrm{E}}^2}{\sigma_{\mathrm{B}}^2}\right)$ and the proof follows along similar lines as in Proposition 2. This implies that the secrecy capacity under peak- and average-intensity constraints in the high-intensity regime does not scale with the constraints and therefore converges to a real and positive constant.

3.7 Numerical Results

In this section, numerical results for the secrecy capacity and the entire rate-equivocation region of the optical wiretap channel with input-dependent Gaussian noise under nonnegativity, peak- and average-intensity constraints are provided.

Figure 3.3 provides a plot of the equivocation density for an optimal input distribution for $A = 4$, $\sigma_{\mathrm{B}}^2 = 1$, $\sigma_{\mathrm{E}}^2 = 2$, $\eta_{\mathrm{B}}^2 = 0.25$, and $\eta_{\mathrm{E}}^2 = 0.125$. We numerically found that for these parameters, the optimal input distribution is ternary with mass points located at $x = 0$, 2.025 and 4 with probability masses 0.2862, 0.3045 and 0.4093, respectively. We observe that $C_{\mathrm{S}} - r_{\mathrm{eq}}(x; F_X)$ is generally nonnegative and is

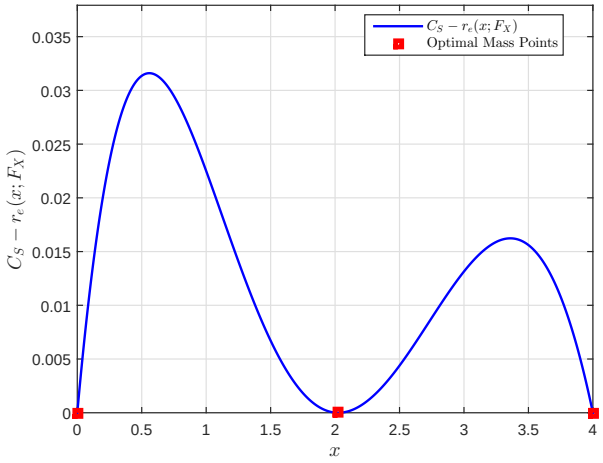

Figure 3.3: Illustration of $C_{\mathrm{S}} - r_{\mathrm{eq}}(x; F_X)$ yielded by the optimal input distribution when $\sigma_{\mathrm{B}}^2 = 1$, $\sigma_{\mathrm{E}}^2 = 2$, $\eta_{\mathrm{B}}^2 = 0.25$, $\eta_{\mathrm{E}}^2 = 0.125$ and $A = 4$.

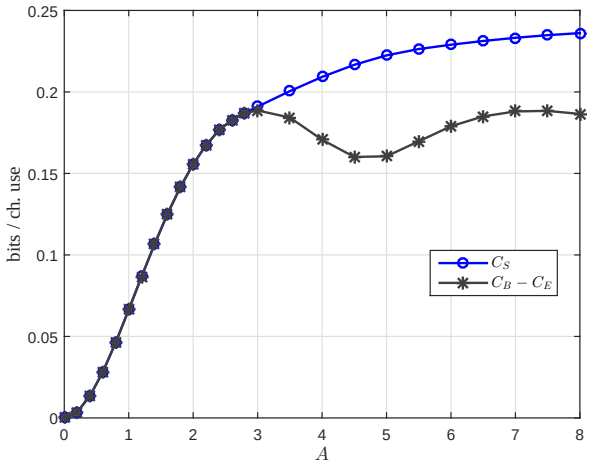

Figure 3.4: The secrecy capacity for $\sigma_{\mathrm{B}}^2 = 1$, $\sigma_{\mathrm{E}}^2 = 2$, $\eta_{\mathrm{B}}^2 = 0.25$, and $\eta_{\mathrm{E}}^2 = 0.125$ versus the peak-intensity constraint A.

equal to zero at the optimal mass points; verifying the optimality conditions in (A.38) and (A.39).

Figure 3.4 illustrates the secrecy capacity C_{S} and the difference $C_{\mathrm{B}} - C_{\mathrm{E}}$ versus the peak-intensity constraint A, where C_{B} and C_{E} are the legitimate user's and the eavesdropper's capacities, respectively. We observe that this difference is in general a lower bound for the secrecy capacity C_{S} which can be easily proven. We also observe that, for small values of A, $C_{\mathrm{B}} - C_{\mathrm{E}}$ and C_{S} are identical. However, as A increases, $C_{\mathrm{B}} - C_{\mathrm{E}}$ and C_{S} become different. Similar to the secrecy capacity results of the Gaussian wiretap channel under a peak-power constraint provided in [14], here too, $I(X;Y)$ and $I(X;Z)$ are maximized by the same discrete distribution, however, $I(X;Y) - I(X;Z)$ is maximized by a different distribution.

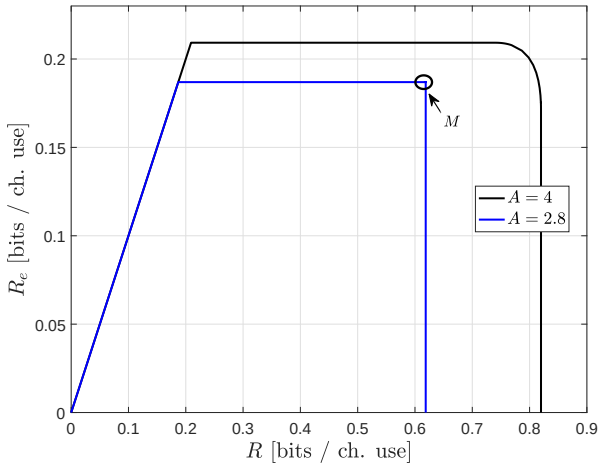

Figure 3.5: The rate-equivocation region for $\sigma_{\mathrm{B}}^2 = 1$, $\sigma_{\mathrm{E}}^2 = 2$, $\eta_{\mathrm{B}}^2 = 0.25$, and $\eta_{\mathrm{E}}^2 = 0.125$ under peak-intensity constraints $A = 2.8$ and $A = 4$. Point M refers to the case when secrecy capacity and capacity are achieved simultaneously.

As a specific example, when $A = 4$, while both $I(X;Y)$ and $I(X;Z)$ are maximized by the same *binary* distribution with mass points at $x = 0$ and 4 with probability masses 0.5088 and 0.4912, respectively, $I(X;Y) - I(X;Z)$ is maximized by a *ternary* distribution with mass points at $x = 0$, 2.025 and 4 with probability masses 0.2862, 0.3045 and 0.4093, respectively. This explains the difference between C_{S} and $C_{\mathrm{B}} - C_{\mathrm{E}}$ at $A = 4$ in this figure.

Figure 3.5 depicts the entire rate-equivocation region of the optical wiretap channel with input-dependent Gaussian noise under nonnegativity and peak-intensity constraints when $\sigma_{\mathrm{B}}^2 = 1$, $\sigma_{\mathrm{E}}^2 = 2$, $\eta_{\mathrm{B}}^2 = 0.25$, and $\eta_{\mathrm{E}}^2 = 0.125$ for two different values of A. When $A = 2.8$, it is clear from the figure that both the secrecy capacity and the capacity can be attained simultaneously (Point "M" in the figure). In particular, for $A = 2.8$, the binary input distribution with mass points located at $x = 0$ and 2.8 with probabilities 0.5183 and 0.4817, respectively, achieves both the capacity and the secrecy capacity. This implies that, when $A = 2.8$, the transmitter can communicate with the legitimate user at the capacity while achieving the maximum equivocation at the eavesdropper. On the other hand, when $A = 4$, the secrecy capacity and the capacity cannot be achieved simultaneously (notice the curved shape in the figure). More specifically, for $A = 4$, the binary input distribution with mass points located at $x = 0$ and 4 with probabilities 0.5088 and 0.4912 achieves the capacity, while a ternary distribution with mass points located at $x = 0$, 2.025, 4 with probability masses 0.2862, 0.3045 and 0.4093, respectively, achieves the secrecy capacity, i.e., the optimal input distributions for the secrecy capacity and the capacity are different. In other words, there is a tradeoff between the rate and its equivocation in the sense that, to

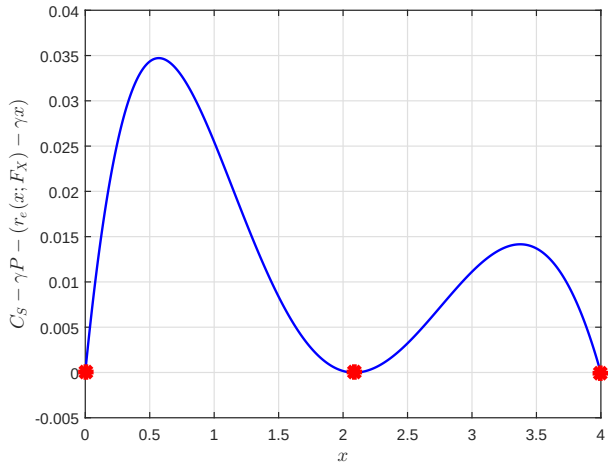

Figure 3.6: Illustration of $C_{\rm S} - r_{\rm eq}(x; F_X) + \gamma\,(x - P)$ yielded by the optimal input distribution for $\sigma_{\rm B}^2 = 1$, $\sigma_{\rm E}^2 = 2$, $\eta_{\rm B}^2 = 0.25$, $\eta_{\rm E}^2 = 0.125$, $A = 4$ and $\kappa = 0.375$. The corresponding Lagrangian multiplier is 0.0187.

increase the communication rate, one must compromise from the equivocation of this communication, and to increase the achieved equivocation, one must compromise from the communication rate.

Figure 3.6 provides illustrations for the effect of the average intensity constraint on the secrecy-capacity-achieving input distribution for $\sigma_{\rm B}^2 = 1$, $\sigma_{\rm E}^2 = 2$, $\eta_{\rm B}^2 = 0.25$, $\eta_{\rm E}^2 = 0.125$ and $A = 4$. We observe that for $\kappa > \frac{1}{2}$, the average intensity constraint is inactive. In this case, in view of Theorem 1, the optimal input distribution is a ternary distribution with mass points at $x = 0$, 2.025 and 4 with probability masses 0.2862, 0.3045 and 0.4093, respectively. We now impose an average intensity constraint with corresponding $\kappa = 0.375$. For this case, a ternary input distribution with mass points located at $x = 0$, 2.0869 and 4 with probability masses 0.4770, 0.3093 and 0.2136, respectively, is optimal and the corresponding Lagrangian multiplier is 0.0187.

Finally, Figure 3.7 plots the exact and asymptotic secrecy capacity results versus the peak-intensity constraint A for both the peak- and average-intensity constraints in the low-intensity regime. From the figure, it is observed that the asymptotic results for the secrecy capacity given in (3.23) and (3.47) are in precise agreement with the numerical results. Furthermore, it is shown that imposing the average intensity constraint in addition to the peak-intensity constraint reduces the secrecy capacity.

3.8 Conclusions

This chapter studied the optical wiretap channel with input-dependent Gaussian noise under nonnegativity, peak- and average-intensity constraints. It was shown that the secrecy capacity and the boundary of the entire rate-equivocation region are achieved by discrete input distributions with finite supports.

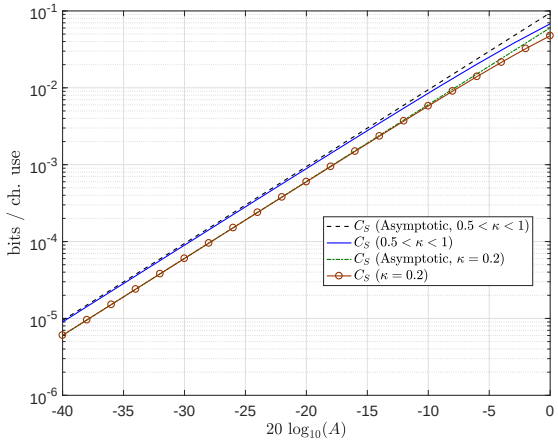

Figure 3.7: The asymptotic and exact secrecy capacity for $\sigma_{\mathrm{B}}^2 = 1$, $\sigma_{\mathrm{E}}^2 = 2$, $\eta_{\mathrm{B}}^2 = 0.25$, and $\eta_{\mathrm{E}}^2 = 0.125$ versus A for both peak- and average-intensity constraints.

Furthermore, under such constraints, the optimal input distribution always possessed a mass point at the origin.

An interesting result that this chapter reveals is that similar to the case for the Gaussian wiretap channel with a peak-power constraint, here too, it is observed that under nonnegativity and peak-intensity constraints, the secrecy capacity and the capacity could not be obtained simultaneously in general, i.e., there is a tradeoff between the rate and its equivocation in the sense that, to increase the communication rate, one must compromise from the equivocation rate, and conversely to increase the achieved equivocation rate, one must compromise from the communication rate.

We extended the discreteness result for the case when both peak- and average-intensity constraints are active. Finally, an asymptotic analysis on the secrecy capacity was presented. It was shown that in the low-intensity regime, the secrecy capacity scales quadratically with the peak-intensity constraint, while in the high-intensity regime, it is upper-bounded by a finite value implying that it does not scale with the constraints.

Chapter 4: Results on the Rate-Equivocation Region of the Degraded Signal-Dependent Noise Wiretap Channel

M. Soltani and Z. Rezki, "Results on the Rate-Equivocation Region of the Degraded Signal-Dependent Noise Wiretap Channel," *Submitted to IEEE Communications Letters, undergoing second round review.*

4.1 Introduction

In an optical wireless communications (OWC) based on the intensity modulation and direct detection (IM-DD), the information bits are communicated through light intensity [2]. In this setup, light emitting diodes (LED) and photodetectors (PD) are used for data transmission and reception, respectively. After the reception of the light intensity by a PD, an output noisy signal is generated from which information bits can be extracted. There exist several channel models for OWC, namely the free-space optical (FSO) model, the signal-dependent noise model, and the Poisson noise model [2].

Broadcast nature of optical wireless signals makes communication of confidential data with trusted users vulnerable to eavesdropping attacks. To address this issue, Wyner [12] introduced the wiretap channel (WC). In such channels, a transmitter wishes to communicate secret messages reliably with a legitimate receiver while preventing the eavesdroppers from inferring the secret messages [12, 34]. In these channels, the rate-equivocation region is considered as a fundamental performance limit which reflects the tradeoff between the secrecy and the reliability. A WC is called degraded when the observations of the eavesdropper and the secret messages given the observations of the legitimate user, are independent. For this type of channels, the rate-equivocation region is completely known [12].

The work in [14] investigated the Gaussian WC with amplitude and variance constraints and showed that distributions having a countably finite support set are secrecy-capacity-achieving. We note that the results pertaining to the Gaussian WC with amplitude and variance constraints can be directly applied to the FSO WC with peak and average optical power restrictions. Furthermore, Dytso *et al.* established that the secrecy-capacity-achieving distribution of the FSO WC with an average optical power restriction admits a countable support set [35], but did not specify whether the support set is bounded or not. The rate-equivocation regions of the degraded SDGN-WC [25], the degraded discrete-time Poisson WC [26], and the degraded continuous-time Poisson WC [15], are all exhausted by distributions having a countably finite set when input signals are restricted by peak and average optical power constraints.

This chapter studies the SDGN-WC when an average optical power constraint is active, and confirm that distributions having a countably infinite support set are optimal in the sense that they exhaust the entire rate-equivocation region. We start by noting that due to [12], the rate-equivocation region of a degraded WC is completely known. Next, a convex functional optimization problem, which is addressed in e.g. [36, 37], is considered and the optimality equations that must be satisfied by an optimal solution are derived. Using these optimality equations, first, it is shown that distributions with a countable set are optimal. This is done by providing a similar contradiction argument which appears in [25, Theorem 1]. The difference is that in this chapter, the peak optical power constraint is inactive and the bounds that were found in the proof of Theorem 1 in [25] are not valid here. Thus, to reach a contradiction, a new bound (cf. equation (D.19)) is provided. We then show that the support sets must be unbounded which is done via providing another contradiction argument. In particular, our contradiction argument hinges on the fact that if the support set is bounded, then the cost function growing linearly in x is lower bounded by the rate-equivocation density growing quadratically in x, and thus reaching a contradiction. From these result, it is concluded that distributions with a countably infinite support set are optimal. Finally, our analysis show that optimal distributions in the FSO WC with an average optical power constraint also admit a countably infinite support set[1]. The countability of the sets can be shown using [35, Theorem 3, Section IV], but a rigorous conclusion regarding the boundedness/unboundedness of the support sets was not provided.

4.2 SDGN-WC with an Average Optical Power Constraint

4.2.1 Channel Model

In the SDGN-WC, an IM-DD system based on pulse amplitude modulation (PAM) scheme is considered due to its popularity and simplicity of implementation. In this setup setup along with a high optical power setting, the received optical signals at the legitimate receiver and the eavesdropper can be respectively given by [2, 5]

$$\begin{cases} Y_1 = X + N_{\mathrm{B}}(X), \\ Y_2 = X + N_{\mathrm{E}}(X). \end{cases} \tag{4.1}$$

In (3.6), the transmitted signal X is nonnegative, Y_1 is the channel output of the legitimate receiver, and Y_2 is the channel output of the eavesdropper; $N_{\mathrm{B}}(X)$ is distributed according to a Gaussian distribution with mean zero and variance $\sigma^2_{\mathrm{B}}(X) \triangleq \sigma^2_{\mathrm{B},0} + \sigma^2_{\mathrm{B},1} X$, where $\sigma^2_{\mathrm{B},0}$ and $\sigma^2_{\mathrm{B},1}$ are positive constants; likewise,

[1] This refers to the case when $\sigma^2_{\mathrm{B},1} = \sigma^2_{\mathrm{E},1} = 0$ in (4.1).

$N_{\mathrm{E}}(X)$ is a zero-mean Gaussian noise at the eavesdropper with variance $\sigma^2_{\mathrm{E}}(X) \triangleq \sigma^2_{\mathrm{E},0} + \sigma^2_{\mathrm{E},1}X$, where $\sigma^2_{\mathrm{E},0}$ and $\sigma^2_{\mathrm{E},1}$ are positive constants. Moreover, the transmitted PAM intensity signal is restricted by an average optical power constraint due to the optical power consumption considerations [2]. Thus, we have [7, 4]

$$\mathbb{E}[X] \leq \mathcal{E}, \tag{4.2}$$

with $\mathcal{E} > 0$ being the average optical power constraint. Observe that if [25]

$$\begin{cases} \sigma^2_{\mathrm{B},1} = \sigma^2_{\mathrm{E},1} \\ \sigma^2_{\mathrm{E},0} > \sigma^2_{\mathrm{B},0}, \end{cases} \tag{4.3}$$

then one can construct a random variable $\widetilde{Y}_2 = X + N_B(X) + N_D$, where N_D is a zero-mean Gaussian distributed random variable with variance $\sigma^2_{\mathrm{E},0} - \sigma^2_{\mathrm{B},0}$ and it is independent from $N_B(X)$. Thus, $\widetilde{Y}_2 = Y_1 + N_D$. Now, observe that $Y_2|X$ and $\widetilde{Y}_2|X$ have same distributions, but $\widetilde{Y}_2$ depends only on Y_1 and not on X, i.e., we have the Markov chain $X \to Y_1 \to \widetilde{Y}_2$. Since the rate-equivocation region of a wiretap channel depends only on the marginals [34], the considered wiretap channel becomes stochastically degraded, and its rate-equivocation region can be single-letterized [12].

4.2.2 Wiretap Codes for the SDGN-WC with an Average Optical Power Constraint

An $(n, 2^{nR})$ wiretap code for the SDGN-WC with an average optical power constraint is given by a set $\mathcal{N} = \{1, 2, \ldots, 2^{nR}\}$, an encoding function e_n and a decoding function d_n. The elements of $\mathcal{N}$ are described by a uniformly distributed random variable N over $\mathcal{N}$. The encoder performs the mapping $e_n : \mathcal{N} \to \mathbb{R}^n_+$ subject to the constraint (4.2), and the decoder performs the mapping $d_n : \mathbb{R}^n \to \mathcal{N}$. The equivocation of N is measured by $\frac{1}{n} H(N|Y_2^n)$, where $H(\cdot|\cdot)$ is the conditional entropy. The error probability of the code is $P_{\mathrm{e}}^n = \Pr\{d_n(Y_1^n) \neq N\}$. If a rate-equivocation pair (R, R_{eq}) satisfies

$$\lim_{n\to\infty} P_{\mathrm{e}}^n = 0, \tag{4.4}$$

$$R_{\mathrm{eq}} \leq \lim_{n\to\infty} \frac{1}{n} H(N|Y_2^n), \tag{4.5}$$

it is said to be achievable and the rate-equivocation region is the union of all achievable (R, R_{eq}). Furthermore, the supremum of the rates satisfying $R_{\mathrm{eq}} = R$ is called the secrecy capacity.

Since the SDGN-WC is stochastically degraded due to (4.3), the rate-equivocation region is the union

of $(R, R_{\text{eq}}) \in \mathbb{R}_+^2$ given by [12]

$$\begin{cases} R \leq I(X;Y_1), \\ R_{\text{eq}} \leq I(X;Y_1) - I(X;Y_2), \end{cases} \tag{4.6}$$

for $F_X \in \mathcal{P}^+$, where $I(\cdot;\cdot)$ is the mutual information between two random variables and the set $\mathcal{P}^+$ is denoted as

$$\mathcal{P}^+ \triangleq \left\{ F_X : \int_0^{+\infty} dF_X(x) = 1, \int_0^{+\infty} x \, dF_X(x) \leq \mathcal{E} \right\}. \tag{4.7}$$

4.3 Main Results

The rate-equivocation region of the degraded SDGN-WC is convex [34, Theorem 1]. Thus, the boundary of this region can be expressed by the following optimization problem

$$\sup_{F_X \in \mathcal{P}^+} f_\mu(F_X) \stackrel{\triangle}{=} \sup_{F_X \in \mathcal{P}^+} [\, \mu \, I(X;Y_1) + (1-\mu)\big\{ I(X;Y_1) - I(X;Y_2) \big\}], \ \forall \, \mu \in [0,1]. \tag{4.8}$$

Under the constraints (4.2) and for each $\mu \in [0,1]$, a unique solution to (4.8) exists and the support set of the optimal solution is countably infinite. This is formally stated below.

Theorem 3. *The rate-equivocation region of the degraded SDGN-WC with nonnegativity and average optical power constraints is exhausted by distributions having a countably infinite support set.*

Proof. For convenience, the proof is relegated to Appendix D. ■

Theorem 3 is established as follows. Firstly, the set $\mathcal{P}^+$ is shown to be convex and compact. Secondly, it is shown that the objective functional is continuous, weakly differentiable and strictly concave in F_X, and thus, a unique solution to (4.8) exists. Thirdly, the optimality equations for a solution F_X^* is derived. Fourthly, it is established that the intersection of the support set of F_X^* with any bounded interval is countably finite. This is done by providing a contradiction argument, i.e., it is assumed that this intersection has an infinite cardinality. Then, invoking Identity and Bolzano-Weierstrass Theorems, it is found that a constant and finite value is upper bounded by $-\infty$ which is clearly a contradiction. Lastly, by resorting to another contradiction argument, the support set is shown to be an unbounded set. This is done by assuming to the contrary that the support set is bounded. Based on this assumption, it is found that the cost function, which grows linearly in x, must be lower bounded by the rate-equivocation density, which grows quadratically in x, and thus reaching a contradiction and the result follows. It is worth mentioning that a similar analysis can be observed in the work by Fahs *et al.* [19], but for the channels with input-independent additive noise and without secrecy constraints.

Theorem 3 has the following consequences: 1) when $\mu = 0$ (the point corresponding to the *secrecy capacity* of the degraded SDGN-WC when input signals are restricted by an average optical power), the optimal distribution has a countably infinite support set; 2) when $\mu = 1$ (the point corresponding to the *capacity* of this channel under the mentioned constraint), the optimal solution also admits a countably infinite support set.

We note that since the optimal distributions admit a countably infinite support, numerical computation of the region is not feasible.

4.4 Conclusions

This chapter studied the degraded SDGN-WC with nonnegativity and average-intensity constraints. It is established that under these constraints, distributions having a countably infinite number of mass points, but finitely many mass points in any bounded interval, are optimal in the sense that they exhaust the entire rate-equivocation region. From this result, it was inferred that the secrecy-capacity-achieving and the capacity-achieving distributions also admit a countably infinite support set.

The obtained results imply that numerical computations of the boundary of the rate-equivocation region as well as the secrecy capacity of the considered wiretap channel under an average-intensity constraint is not feasible. Therefore, to evaluate the secrecy performance of an OWC in such a setting, it is of great importance to provide inner and outer bounds on the rate-equivocation region based on discrete distributions with a finite number of mass points. These inner and outer bounds help to characterize near-optimal secure transmission schemes for OWC systems operating over the optical wiretap channel with an input-dependent Gaussian noise when only an average-intensity constraint is active.

Chapter 5: The Degraded Discrete-Time Poisson Wiretap Channel

M. Soltani and Z. Rezki, "The Discrete-Time Poisson Optical Wiretap Channel with Peak Intensity Constraints," *in Proceedings of the 2019 IEEE International Symposium on Information Theory (ISIT'2019), Paris, France, Jun. 2019.*

M. Soltani and Z. Rezki, "The Degraded Discrete-Time Poisson Optical Wiretap Channel," *Submitted to the IEEE Transactions on Information Theory.*

5.1 Introduction

Intensity modulation and direct detection (IM-DD) is the simplest and the most commonly used technique for optical wireless communications. In this scheme, the channel input modulates the intensity of the emitted light. Thus, the input signal is proportional to the light intensity and is nonnegative. The receiver is usually equipped with a photodetector which absorbs integer number of photons and generates a real valued output corrupted by noise. Based on the distribution of the corrupting noise there exist several models for the underlying optical wireless communication channels. Free space optical (FSO) channels [2, 3], optical channels with input-dependent Gaussian noise [2, 5], and Poisson optical channels [2, 7, 8, 18] are the most widely used models for optical wireless communications. Among these models, the most accurate one that can capture most of the optical channel impairments is the Poisson model. The studies conducting research on Poisson optical channels are mainly categorized in two mainstreams. The first category considers the continuous-time Poisson model where the input signals can admit arbitrarily waveforms and there are no bandwidth constraints on the transmission. The second category concerns the discrete-time Poisson channel and deals with the cases where stringent transmission bandwidths are assumed.

For the discrete-time Poisson channel, Shamai [7] studied the single-user channel capacity and showed that the capacity-achieving distribution under nonnegativity, peak- and average-intensity constraints is discrete with a finite number of mass points. In [8, 38], authors provided asymptotic analysis of the channel capacity in the regimes where the peak- and/or average-intensity constraints tend to zero (low-intensity regime) or to infinity (high-intensity regime). The work in [38] focused on characterizing the channel capacity in the low-intensity regime of an average-intensity constrained or an average- and peak-intensity constrained inputs and found upper and lower bounds which in some cases, the bounds coincide. Authors

in [8] investigated the high-intensity behavior of the channel capacity for a peak- and average-intensity constrained input and presented tight bounds, thus fully characterizing the channel capacity in the high-intensity regime. While the capacity of the discrete-time Poisson channel is unknown, the capacity of the continuous-time Poisson channel where the signaling bandwidth is not restricted is known in closed-form. For the peak-intensity constrained or peak- and average-intensity constrained inputs the capacity of the continuous-time Poisson channel is achieved by a binary distribution with mass points located at the origin and at the peak-intensity constraint [6], however, the channel capacity of the average-intensity constrained input is *infinite* and the capacity-achieving input is unknown.

The broadcast nature of optical wireless signals imposes a security challenge especially in the presence of unauthorized eavesdroppers. This problem has been conventionally addressed by cryptographic encryption [11] without considering the imperfections introduced by the communication channels. Wyner [12], on the other hand, proved the possibility of secure communications without relying on encryption by introducing the notion of a degraded wiretap channel. This result was later generalized by Csiszar and Korner by dropping the degradedness assumption of the wiretap channel [34].

The wiretap channels are studied with respect to the rate-equivocation region, which is defined as the set of all rate pairs for which the transmitter can communicate confidential messages reliably with a legitimate receiver at a certain secrecy level against an eavesdropper [13]. A wiretap channel is called degraded when the observations of the eavesdropper and the secret messages given the observations of the legitimate user are independent. For this type of channels, Wyner established that there exists a single-letter characterization for the rate-equivocation region [12].

Authors in [14] studied the degraded Gaussian wiretap channel under amplitude and variance constraints, and prove that the entire rate-equivocation region of this wiretap channel is attained by discrete input distributions with finitely many mass points. Furthermore, the authors observed that the secrecy-capacity-achieving input distribution may not be identical to the capacity-achieving counterpart in general, resulting in a tradeoff between the rate and its equivocation. It is worth mentioning that the results pertaining to the Gaussian wiretap channel with amplitude and variance constraints can be directly applied to characterize the optimal distributions exhausting the entire rate-equivocation region of the FSO wiretap channel with peak- and average-intensity constraints. Furthermore, Dytso *et al.* established that the secrecy-capacity-achieving distribution of the FSO wiretap channel with nonnegativity and average-intensity constraints is discrete with finitely many mass points in any bounded interval, but did not specify whether the support set of the optimal distribution is bounded or not [35].

The work in [25] considers the degraded optical wiretap channel with input-dependent Gaussian noise under peak- and average-intensity constraints and verified the optimality of distributions with a

finitely many mass points for attaining the entire boundary of the rate-equivocation region. Besides, the authors provided asymptotic behavior of the secrecy capacity in the low- and high-intensity regimes. For this wiretap channel, authors observed that, in general, there is a tradeoff between the rate and its equivocation. Finally, [15] examined the degraded continuous-time Poisson wiretap channel (PWC) with a peak-intensity constraint and gave a closed-form expression for the secrecy capacity. Particularly, the authors showed that *binary* input distributions with mass points located at the origin and the peak-intensity constraint and with a very short duty cycle exhaust the entire rate-equivocation region.

In this chapter, a degraded *discrete*-time PWC (DT–PWC) is considered. The DT–PWC consists of a transmitter, a legitimate user and an eavesdropper. In this setup, the input signals are restricted to have finite bandwidths. This fact distinguishes the DT–PWC from its continuous-time counterpart, where input signals can have infinite bandwidths. Using an IM-DD system, the photodetectors at the legitimate user and the eavesdropper counts the number of received photons and output signals that follow Poisson distribution. Here, the objective is to have secure communication with the legitimate user over a discrete-time Poisson channel while keeping the eavesdropper ignorant of the transmitted messages as much as possible.

We start by the secrecy capacity of the degraded DT–PWC and employ the functional optimization problems addressed in, for example [7, 14, 25, 36], to derive the necessary and sufficient optimality equations, also known as Karush-Kuhn-Tucker (KKT) conditions, that must be satisfied by an optimal solution. Using these equations, it is confirmed that a unique distribution with a countably finite number of mass points achieves the secrecy capacity of the degraded DT–PWC when only peak-intensity or both peak- and average-intensity constraints are active. This is done by providing a contradiction argument. We start by assuming, on the contrary, that the support set of the optimal solutions contains an infinite number of elements. Then recalling the Identity and Bolzano-Weierstrass Theorems from complex analysis one can conclude that: 1) when the legitimate user's and the eavesdropper's channel gains are not identical, a nonnegative constant must be lower bounded by a logarithmically increasing function in x where $x \geq 0$, which is a contradiction; 2) when the channel gains are identical, the nonnegative constant must be upper bounded by $-\infty$ and a contradiction occurs. Following along similar lines of the above mentioned analysis, the optimality of distributions with a finite number of mass points is extended to the entire boundary of the rate-equivocation.

Additionally, the secrecy capacity of the DT–PWC with nonnegativity and average-intensity constraints is investigated, and it is verified that a unique distribution with the following structural properties is secrecy-capacity-achieving: 1) the support set of the optimal solution contains a finitely many mass points in any bounded interval; 2) the support set of the optimal solution is an unbounded set. These two

properties imply that the optimal distribution is discrete with countably infinite number of mass points, but with finitely many mass points in any bounded interval. The first property is shown by means of contradiction. We assume, on the contrary, that for some bounded interval, the intersection of the support set of the optimal solution and the bounded interval has an infinite number of mass points. Then, using the KKT conditions and invoking the Bolzano-Weierstrass and Identity Theorems from complex analysis, one finds that a nonnegative constant is upper bounded by $-\infty$ which results in a contradiction. The second property is also shown through a contradiction approach. We assume that the optimal support set is bounded and the following cases are considered: 1) when legitimate user's and the eavesdropper's channel gains are not identical, our contradiction hinges on the fact that a linearly increasing function in x must be lower bounded by another function which grows as fast as $x \log x$. This is not possible for large values of x and hence a contradiction occurs; 2) when the channel gains are identical, one finds that the Lagrangian multiplier must be lower bounded by a constant and thus, using the Envelope Theorem [39], it is observed that the secrecy capacity would at least grow linearly in the average-intensity constraint. However, in Appendix E, it is shown that the secrecy capacity is always upper bounded by a constant for all values of the average-intensity constraint. Therefore, the desired contradiction is reached and the result follows. Moreover, following along similar lines, it is established that every point on the boundary of the rate-equivocation region is also attained by a unique distribution with countably infinite number of mass points, but finitely many mass points in any bounded interval. This, in turn, implies that the capacity of the discrete-time Poisson channel with average-intensity constraint is also achieved by a discrete distribution with countably infinite number of mass points and settles down Shamai's conjecture in [7].

Furthermore, considering the peak- or both peak- and average-intensity constraints we shed light on the asymptotic behavior of the secrecy capacity in the low- or high-intensity regimes. In the low-intensity regime, a full characterization of the secrecy capacity and the secrecy-capacity-achieving distribution is given. It is observed that the secrecy capacity scales quadratically with the peak-intensity constraint and the secrecy-capacity-achieving input distribution is binary with mass points located at the origin and the peak-intensity constraint. We characterize the secrecy capacity in the low-intensity regime by deriving lower and upper bounds on the secrecy capacity and showing that these bounds coincide. We note that a valid upper bound on the secrecy capacity of the DT–PWC is the secrecy capacity of the continuous-time PWC across all intensity regimes. This is because in the continuous-time version, input signals are not restricted to have a finite transmission bandwidth and can have arbitrary waveforms with an infinite transmission bandwidth. Thus, under the same constraints, i.e., peak- and/or average-intensity constraints, the secrecy capacity of the continuous-time PWC is greater than that of the DT–PWC. Also,

a legitimate lower bound on the secrecy capacity of the DT–PWC is the difference between the capacities of the legitimate user's and the eavesdropper's channels. For the high-intensity regime and when the legitimate receiver's and the eavesdropper's channel gains are identical, it is proved that the secrecy capacity does not scale with the constraints and hence, it is a constant value. In addition, when the channel gains are different, the secrecy capacity can be upper bounded by the capacity of a discrete-time Poisson channel and thus, it cannot scale faster than the logarithm of the square root of the peak- and/or average-intensity constraints.

Finally, through numerical inspections, it is found that when peak-intensity or both peak- and average-intensity constraints are active, the secrecy capacity and the capacity of the DT–PWC are not achieved by the same distributions in general. Therefore, there is a tradeoff between the rate and its equivocation. This is also true for the continuous-time PWC when peak-intensity or both peak- and average-intensity constraints are active [15]. However, note that for the continuous-time PWC when the eavesdropper's observations are just the thinned version of those of the legitimate receiver's, the secrecy capacity and the capacity are attained with identical binary input distributions as pointed out in [15] which implies that there is no tradeoff between the rate and its equivocation. Lastly, since with only average-intensity constraint the optimal distributions admit a countably infinite number of mass points, numerical computation of the secrecy-capacity and the boundary of the rate-equivocation region is not feasible.

The rest of this chapter is structured as follows. The discrete-time degraded PWC is formally defined in Section 5.2. The main results of the chapter regarding the characterization of the optimal distributions attaining the secrecy capacity as well as exhausting the entire rate-equivocation region are presented in Section 5.3. Proofs of the main results are provided in Appendix E. Asymptotic analysis of the secrecy capacity in the low- and high-intensity regimes are demonstrated in Section 5.3. Numerical results are shown in Section 5.4, and finally, conclusions are drawn in Section 5.5.

5.2 The Degraded Discrete-Time Poisson Wiretap Channel

We consider a practical optical wireless communication system where IM-DD is employed. In this setup, the channel input modulates the emitted light intensity from the light emitting diode (LED) at the transmitter and photodetectors are used for receiving the optical signal at the legitimate user's and eavesdropper's receivers. We assume that there exist line-of-sight (LoS) paths between the optical transmitter and the receivers. In such a scenario, the received power of the LoS path dominates the received power of the reflected paths. Hence, the optical wireless channel between the transmitter and the legitimate user and between the transmitter and the eavesdropper become LoS channels with positive

and constant channel gains [31, Chapter 2].

In the considered wiretap channel, confidential data are transmitted by sending pulse amplitude modulated (PAM) intensity signals which are constant in discrete time slots of Δ seconds. The receiver is modeled as a photon counter which generates an integer representing the number of received photons. Specifically, in each time slot of Δ seconds an input intensity X is corrupted by the LoS channel gains α_B and α_E and the combined impact of background radiation as well as the photodetectors' dark currents λ_B and λ_E at the legitimate user's and the eavesdropper's receivers, respectively. The channel outputs at the legitimate receiver and the eavesdropper are denoted by Y and Z, respectively, and are random variables related to the number of received photon in Δ seconds. These channel outputs conditioned on the input signal obey the Poisson distributions with mean $(\alpha_B X+\lambda_B)\Delta$ and $(\alpha_E X+\lambda_E)\Delta$, respectively, i.e., [7, equation 16]

$$p_{Y|X}(y|x) = e^{-(\alpha_B x+\lambda_B)\Delta}\,\frac{\left[(\alpha_B x+\lambda_B)\Delta\right]^y}{y!},\ y\in\mathbb{N}, \tag{5.1}$$

$$p_{Z|X}(z|x) = e^{-(\alpha_E x+\lambda_E)\Delta}\,\frac{\left[(\alpha_E x+\lambda_E)\Delta\right]^z}{z!},\ z\in\mathbb{N}, \tag{5.2}$$

where $\mathbb{N}$ is the set of all nonnegative integers. This model is referred to as the DT–PWC where a bandwidth constraint is imposed on the input signals by constraining the signals to be rectangular PAM. We note that this model is in contrast to the continuous-time PWC where the input signals can admit arbitrary waveforms with very large transmission bandwidth [15, 7].

In the DT–PWC, the channel input X is a nonnegative random variable representing the intensity of the optical signal. Since intensity is constrained due to practical and safety restrictions by peak- and average-intensity constraints, the input must satisfy [2]

$$0\le X\le \mathcal{A}, \tag{5.3}$$

$$\mathbb{E}[X]\le \mathcal{E}. \tag{5.4}$$

In this chapter, the *degraded* DT–PWC is studied. Therefore, the case where the following conditions hold is considered:

$$\alpha_B \ge \alpha_E, \tag{5.5}$$

$$\frac{\lambda_B}{\alpha_B}\le\frac{\lambda_E}{\alpha_E}, \tag{5.6}$$

which implies that the random variables X, Y, and Z form the Markov chain $X\to Y\to Z$ and con-

sequently, the DT–PWC becomes stochastically degraded [12, 15, 40]. In the sequel, without loss of generality, it is assumed that at least of the inequalities (5.5) or (5.6) is strict. This is because if both are tight, then the legitimate receiver's and eavesdropper's channels become identical and the secrecy capacity (defined later in this section) is equal to zero.

5.2.1 The Rate-Equivocation Characterization of the DT–PWC

An $(n, 2^{nR})$ code for the DT–PWC consists of the random variable W (message set) uniformly distributed over $\mathcal{W} = \{1, 2, \cdots, 2^{nR}\}$, an encoder at the transmitter $f_n : \mathcal{W} \to \mathbb{R}_+^n$ satisfying the constraints (5.3)–(5.4), and a decoder at the legitimate user $g_n : \mathbb{N}^n \to \mathcal{W}$. Equivocation of a code is measured by the normalized conditional entropy $\frac{1}{n} H(W|Z^n)$. The probability of error for such a code is defined as $P_e^n = \Pr[g_n(Y^n) \neq W]$. A rate-equivocation pair (R, R_e) is said to be achievable if there exists an $(n, 2^{nR})$ code satisfying

$$\lim_{n\to\infty} P_e^n = 0, \tag{5.7}$$

$$R_e \leq \frac{1}{\Delta} \lim_{n\to\infty} \frac{1}{n} H(W|Z^n), \tag{5.8}$$

where $H(W|Z^n)$ is the conditional entropy of W given the observations Z^n. The rate-equivocation region consists of all achievable rate-equivocation pairs. A rate R is said to be perfectly secure if $R_e = R$, that is, if there exists an $(n, 2^{nR})$ code satisfying $\frac{1}{\Delta} \lim_{n\to\infty} \frac{1}{n} I(W; Z^n) = 0$, where $I(W; Z^n)$ is the mutual information between the random variables W and Z^n. The supremum of such rates is defined to be the secrecy capacity and is denoted by C_S.

Since under the assumptions (5.5)–(5.6), the DT–PWC is degraded, its entire rate-equivocation region, denoted by $\mathcal{R}$, can be expressed in a single-letter expression and it is given by the union of all rate-equivocation pairs (R, R_e) such that [12]

$$\mathcal{R} = \begin{cases} 0 \leq R \leq \frac{1}{\Delta} I(X; Y), \\ 0 \leq R_e \leq \frac{1}{\Delta} [I(X; Y) - I(X; Z)], \end{cases} \tag{5.9}$$

for some input distribution $F_X \in \mathcal{F}^+$ where the feasible set $\mathcal{F}^+$ is given by one of the following sets

$$\Omega^+_{\mathcal{A},\,\mathcal{E}} \triangleq \left\{ F_X : \int_0^{\mathcal{A}} dF_X(x) = 1,\, \int_0^{\mathcal{A}} x\, dF_X(x) \leq \mathcal{E} \right\}, \tag{5.10}$$

$$\Omega^+_{\mathcal{A}} \triangleq \left\{ F_X : \int_0^{\mathcal{A}} dF_X(x) = 1 \right\}, \tag{5.11}$$

$$\Omega^+_{\mathcal{E}} \triangleq \left\{ F_X : \int_0^{\infty} dF_X(x) = 1,\, \int_0^{\infty} x\, dF_X(x) \leq \mathcal{E} \right\}. \tag{5.12}$$

5.3 Main Results

This section presents the main results related to the DT–PWC. We first consider that both of the constraints (5.3) and (5.4) are active which happens when $\mathcal{E} < \mathcal{A}$. In this case, it is proved that discrete distributions with finitely many mass points achieve the secrecy capacity and exhaust the entire rate-equivocation region of the DT–PWC. As a byproduct of this analysis, one can also establish the optimality of discrete distributions with a finite number of mass points when only a peak-intensity constraint is active, i.e., the case when $\mathcal{E} \geq \mathcal{A}$. Finally, the case when only an average-intensity constraint is active (this happens when $\mathcal{A} \to \infty$ while $\mathcal{E}$ is fixed) is considered and the optimal distributions attaining the secrecy capacity and exhausting the entire rate-equivocation region are characterized. It is shown that the support set of the optimal solutions has countably infinite many mass points, but only finitely many mass points in any bounded interval.

5.3.1 Results on the Secrecy Capacity

For the degraded DT–PWC, the secrecy capacity is given by a single-letter expression as [13, Chap. 3]

$$C_S = \sup_{F_X \in \mathcal{F}^+} f_0(F_X) \triangleq \frac{1}{\Delta} \sup_{F_X \in \mathcal{F}^+} \left[I(X;Y) - I(X;Z) \right], \tag{5.13}$$

where the feasible set $\mathcal{F}^+$ is given by one of the sets in (5.10)–(5.12).

We start by characterizing the secrecy-capacity-achieving distribution when $\mathcal{F}^+ = \Omega^+_{\mathcal{A},\,\mathcal{E}}$ in (5.13). Under the constraints (5.3)–(5.4), the solution to the optimization problem in (5.13) exists, is unique and is discrete with finitely many mass points. This is formally presented by the following theorem.

Theorem 4. *There exists a unique input distribution that attains the secrecy capacity of the DT–PWC with nonnegativity, peak- and average-intensity constraints. Furthermore, the support set of this optimal input distribution is a finite set.*

Proof. For convenience, the proof is relegated to Appendix E. ■

The proof of Theorem 4 is sketched as follows. Firstly, the set of input distributions $\Omega^{+}_{\mathcal{A},\,\mathcal{E}}$ is shown to be sequentially compact in the Lévy metric sense and convex. Secondly, it is shown that the objective functional is continuous, weakly differentiable and strictly concave in F_X. Thus, a unique solution to (5.13) exists. Thirdly, the necessary and sufficient KKT conditions that must be satisfied by an optimal solution F^{*}_{X} are derived. Fourthly, it is established that the support set of F^{*}_{X} contains finitely many mass points. This is done by providing a contradiction argument. We start by assuming, on the contrary, that the support set contains an infinite number of elements. Next, we invoke the Identity and Bolzano-Weierstrass Theorems from complex analysis and we conclude that: 1) when the legitimate user's and the eavesdropper's channel gains are not identical, the Lagrangian multiplier (which is a nonnegative constant) must be lower bounded by a logarithmically increasing function in x which is a contradiction; 2) when the channel gains are identical, the Lagrangian multiplier is upper bounded by $-\infty$ which again is a contradiction. Following along similar lines of the proof of Theorem 4, the optimality of distributions with a finite number of mass points is extended to the entire boundary of the rate-equivocation.

It is worth mentioning that in the continuous-time PWC studied in [15] the secrecy-capacity-achieving input distribution is always binary with mass points located at the origin and the value of the peak-intensity constraint [15, Theorem 1]. Furthermore, to achieve the secrecy capacity, input signals must have very short duty cycle (i.e., $\Delta \to 0$ or equivalently, a very large transmission bandwidth is required). However, in the DT–PWC the number of mass points of the optimal distribution depends on the value of Δ and the peak- or both peak- and average-intensity constraints, and in general, it is greater than two.

Next, a corollary which concerns the characterization of the optimal distribution attaining the secrecy capacity of the DT–PWC with nonnegativity and peak-intensity constraints is presented.

Corollary 1. *The secrecy capacity of the DT–PWC with nonnegativity and peak-intensity constraints, i.e., the case when $\mathcal{F}^{+} = \Omega^{+}_{\mathcal{A}}$ in* (5.13)*, is achieved by a unique and discrete input distribution with a finite number of mass points.*

Proof. The proof follows along similar lines of those mentioned in the proof of Theorem 4. ■

Next, consider the case where $\mathcal{F}^{+} = \Omega^{+}_{\mathcal{E}}$ in (5.13). For this case, it is shown that a discrete distribution with countably infinite number of mass points, but with finitely many mass points in any bounded interval achieve the secrecy capacity when nonnegativity and average-intensity constraints (no peak-intensity constraint) are active.

Theorem 5. *There exists a unique input distribution which attains the secrecy capacity of the DT–PWC with nonnegativity and average-intensity constraints. The optimal distribution is discrete with countably infinite number of mass points, but only finitely many mass points in any bounded interval.*

Proof. Theorem 5 is established in Appendix E. ■

To prove Theorem 5, we first prove that the set of input distributions $\Omega_{\mathcal{E}}^{+}$ is compact and convex. We then invoke similar arguments to those presented in the proof of Theorem 4 to show that the objective function in (5.13) is continuous, strictly concave and weakly differentiable in the input distribution F_X. Therefore, we conclude that the solution to the optimization problem (5.13) exists and is unique. We continue the proof by showing that first, the intersection of the support set of the optimal input distribution denoted by $\mathcal{S}_{F_X^*}$ with any bounded interval B contains a finite number of mass points, i.e., $|\mathcal{S}_{F_X^*} \cap B| < \infty$, where $|B|$ denotes the cardinality of the set B. Next, we show that $\mathcal{S}_{F_X^*}$ must be an unbounded set. These structural properties imply that the optimal distribution is discrete with countably infinite number of mass points, but with finitely many mass points in any bounded interval. The first property is shown by means of contradiction. We assume that $|\mathcal{S}_{F_X^*} \cap B| = \infty$. Then, using the KKT conditions and invoking the Bolzano-Weierstrass and Identity Theorems from complex analysis, we find that the Lagrangian multiplier is upper bounded by $-\infty$ which is a contradiction. The second property is also shown through contradiction. Assuming that the optimal support set is bounded, we consider the following cases: 1) if the legitimate user's and the eavesdropper's channel gains are not identical, our contradiction hinges on the fact that a linearly increasing function in x must be lower bounded by another function which grows as fast as $x \log x$ which is a contradiction for large values of x; 2) if the channel gains are identical, we find that the Lagrangian multiplier would be lower bounded by a constant and using the Envelope Theorem [39], we observe that the secrecy capacity must at least grow linearly in the average-intensity constraint. However, in Appendix E we establish that the secrecy capacity is always upper bounded by a constant for all values of the average-intensity. Therefore, the desired contradiction occurs.

Finally, we establish the existence of a mass point at $x = 0$ in the support set of the secrecy-capacity-achieving input distributions under all the possible choices for $\mathcal{F}^+$ given by (5.10)–(5.12).

Proposition 4. *Let $\mathcal{S}_{F_X^*}$ be the support set of the secrecy-capacity-achieving input distribution F_X^* for the DT–PWC under one of the constraints in* (5.10)–(5.12)*. Then $x = 0$ always belong to $\mathcal{S}_{F_X^*}$.*

Proof. The proof is by contradiction and follows along similar lines as in [25, Proposition 1] with the difference that the conditional channel laws follow Poisson distribution. For completeness, the proof is relegated to Appendix E. ■

It is worth mentioning that the existence of a mass point at the origin has also been established in [41, Corollary 2] for the discrete-time memoryless Poisson channel, but with no secrecy constraints. Furthermore, a modified version of the proof of Proposition 4 can be used to alternatively prove the existence of a mass point at the origin for the discrete-time memoryless Poisson channel.

5.3.2 Rate-Equivocation Region

By a time-sharing argument, it can be shown that the rate-equivocation region of the DT–PWC is convex. Therefore, the region can be characterized by finding tangent lines to $\mathcal{R}$ which are given by the solutions of

$$\sup_{F_X \in \mathcal{F}^+} f_\mu(F_X) \triangleq \sup_{F_X \in \mathcal{F}^+} \frac{\mu}{\Delta} I(X;Y) + \sup_{F_X \in \mathcal{F}^+} \frac{1-\mu}{\Delta} \left[I(X;Y) - I(X;Z) \right], \quad \forall\, \mu \in [0,1], \tag{5.14}$$

where the feasible set $\mathcal{F}^+$ is one of the sets given by (5.10)–(5.12). We start by proving that the entire boundary of the rate-equivocation region of the DT–PWC with nonnegativity, peak- and average-intensity constraints (i.e., $\mathcal{F}^+ = \Omega^+_{\mathcal{A},\,\mathcal{E}}$) is obtained by discrete input distributions with a finite number of mass points.

Theorem 6. *Every point on the boundary of the rate-equivocation region of the DT–PWC with nonnegativity, peak- and average-intensity constraints, is achieved by a unique input distribution which is discrete with a finite number of mass points.*

Proof. For brevity, Theorem 6 is established in Appendix E. ■

The proof of Theorem 6 follows along similar lines as the one in the proof of Theorem 4 with the difference in the contradiction argument. Here, our contradiction is based on the fact that (regardless of having $\alpha_B = \alpha_E$ or not) the Lagrangian multiplier is lower bounded by a function that grows logarithmically in x.

Next, we present a corollary which states that the entire boundary of the rate-equivocation region of the DT–PWC under nonnegativity and peak-intensity constraints is attained by discrete distributions with finitely many mass points.

Corollary 2. *Every point on the boundary of the rate-equivocation region of the DT–PWC with nonnegativity and peak-intensity constraints is achieved by a unique and discrete input distribution with a finite number of mass points.*

Proof. The proof follows by invoking similar arguments to those in the proof of Theorem 6. ■

Finally, we consider the case where $\mathcal{F}^+ = \Omega_{\mathcal{E}}^+$ in (5.13) and characterize the optimal distributions exhausting the entire rate-equivocation region when nonnegativity and average-intensity constraints are active.

Theorem 7. *Every point on the boundary of the rate-equivocation region of the DT–PWC with nonnegativity and average-intensity constraints is achieved by a unique and discrete input distribution with countably infinite number of mass points, but finitely many mass points in any bounded interval.*

Proof. The proof is presented in Appendix E. ■

The proof of Theorem 7 follows along similar lines as the ones in the proof of Theorem 5 with a difference in the unboundedness proof of the optimal support set. Here, we do not consider different cases on the channel gains and the desired contradiction occurs by showing that a linearly increasing function in x would be lower bounded by another function growing as fast as $x \log x$.

A direct consequence of Theorem 7 is that when $\mu = 1$ in (5.14) (the point corresponding to the capacity of the discrete time Poisson channel with nonnegativity and average-intensity constraints), the optimal distribution is discrete with a countably infinite number of mass points, but finitely many mass points in any bounded interval. This result settles down Shamai's conjecture in [7] using different and simpler arguments than those appeared in [20, Theorem 15].

5.3.3 Asymptotic Analysis of the Secrecy Capacity

This section provides the asymptotic analysis on the secrecy capacity of the DT–PWC. Firstly, the secrecy capacity is investigated for asymptotically small values of $\mathcal{A}$ and $\mathcal{E}$ with their ratio held fixed at $p \triangleq \frac{\mathcal{E}}{\mathcal{A}}$. Secondly, the behavior of the secrecy capacity is analyzed in the regime where the constraints tend to infinity.

5.3.3.1 Low-Intensity Results

The following theorem gives a closed-form expression of the secrecy capacity in the low-intensity regime.

Theorem 8. *In the regime where the peak-intensity constraint $\mathcal{A} \to 0$ or both peak- and average-intensity constraints $\mathcal{A} \to 0,\ \mathcal{E} \to 0$ while their ratio is fixed at $p \triangleq \frac{\mathcal{E}}{\mathcal{A}}$, the secrecy capacity is given by*

$$C_S = \begin{cases} \frac{\mathcal{A}^2}{8}\left(\frac{\alpha_B^2}{\lambda_B} - \frac{\alpha_E^2}{\lambda_E}\right), & \text{if } \frac{1}{2} \le p \le 1, \\ \frac{\mathcal{A}^2}{2}\, p\,(1-p)\left(\frac{\alpha_B^2}{\lambda_B} - \frac{\alpha_E^2}{\lambda_E}\right), & \text{if } 0 < p < \frac{1}{2}. \end{cases} \tag{5.15}$$

Proof. The proof is based on deriving tight lower and upper bounds that coincide in the low-intensity regime. These lower and upper bounds are given by the Lemma 1 and Lemma 2, respectively. ■

We note that when $\frac{1}{2} \leq p \leq 1$ the average-intensity constraint is inactive, and only the peak-intensity constraint is active. Furthermore, we observe that in the low-intensity regime when peak- or both peak- and average-intensity constraints are active the secrecy capacity scales quadratically with the peak-intensity constraint. Additionally, in this regime the secrecy capacity is independent of the pulse duration Δ and thus, there is no tradeoff between the secrecy capacity and the transmission bandwidth.

It is worth mentioning that in the low-intensity regime, the secrecy-capacity-achieving input distribution is $F_X^*(x) = \frac{1}{2}u(x) + \frac{1}{2}u(x - \mathcal{A})$ when the peak-intensity constraint is active, and it is $F_X^*(x) = (1-p)\,u(x) + p\,u(x - \mathcal{A})$ when both peak- and average-intensity constraints are active, where $u(\cdot)$ is the unit step function.

To derive (5.15), we provide lower and upper bounds on the secrecy capacity and show that these bounds coincide in the low-intensity regime. To that end, we consider the secrecy capacity of the continuous-time PWC and we note that it is a valid upper bound on the secrecy capacity of the DT–PWC across all the intensity regimes. This is because in the continuous-time version, input signals are not restricted to be PAM and can have arbitrary waveforms with an infinite transmission bandwidth. Furthermore, it can be easily shown that the difference between the capacities of the legitimate user's and the eavesdropper's channels is a valid lower bound on the secrecy capacity.

Based on these arguments, we present two lemmas that provide closed-form expressions for the lower and the upper bounds on the secrecy capacity in the low-intensity regime.

Lemma 1. *The secrecy capacity of the DT–PWC in the low-intensity regime when peak- or both peak- and average-intensity constraints are active is lower bounded by*

$$C_S \geq C_B - C_E \geq \begin{cases} \frac{\mathcal{A}^2}{8}\left(\frac{\alpha_B^2}{\lambda_B} - \frac{\alpha_E^2}{\lambda_E}\right), & \text{if } \frac{1}{2} \leq p \leq 1, \\ \frac{\mathcal{A}^2}{2}\,p\,(1-p)\left(\frac{\alpha_B^2}{\lambda_B} - \frac{\alpha_E^2}{\lambda_E}\right), & \text{if } 0 < p < \frac{1}{2}. \end{cases} \tag{5.16}$$

where C_B is the capacity of the legitimate receiver's channel and C_E is the capacity of the eavesdropper's channel.

Proof. The proof is presented in Appendix E. ■

Lemma 2. *The secrecy capacity of the DT–PWC in the low-intensity regime when peak- or both peak-*

and average-intensity constraints are active is upper bounded by

$$C_S \leq C_S^{CT} = \begin{cases} \frac{\mathcal{A}^2}{8}\left(\frac{\alpha_B^2}{\lambda_B} - \frac{\alpha_E^2}{\lambda_E}\right), & \text{if } \frac{1}{2} \leq p \leq 1, \\ \frac{\mathcal{A}^2}{2}\, p\,(1-p)\left(\frac{\alpha_B^2}{\lambda_B} - \frac{\alpha_E^2}{\lambda_E}\right), & \text{if } 0 < p < \frac{1}{2}. \end{cases} \tag{5.17}$$

where C_S^{CT} is the secrecy capacity of the degraded continuous-time PWC with either peak- or both peak- and average-intensity constraints.

Proof. The proof is relegated to Appendix E. ■

We observe that the lower and upper bounds on the secrecy capacity of the DT–PWC coincide, when peak- or both peak- and average-intensity constraints are active. Thus, we can fully characterize the secrecy capacity of the DT–PWC in this regime.

5.3.4 High-Intensity Results

This section presents the behavior of the secrecy capacity of the DT–PWC when the constraints tend to infinity. We consider two scenarios: 1) the condition in (5.5) is tight and the condition in (5.6) is strict; 2) the condition in (5.5) is strict and the condition in (5.6) is either strict or tight.

We observe that the secrecy capacity of scenario 1 can be upper bounded by a constant across all intensity regimes which implies that in the high-intensity regime, the secrecy capacity of the DT–PWC does not scale with the constraints. However, the secrecy capacity of scenario 2 in the high-intensity regime cannot scale faster than the logarithm of the square root of the constraints. Before we present the main results regarding the asymptotic behavior of the secrecy capacity in the high-intensity regime, we state a lemma which we frequently use in our analysis throughout this section.

Lemma 3. *For a degraded DT–PWC (i.e., when the conditions in* (5.5)–(5.6) *hold true), the mutual information difference $f_0(F_X) = I(X;Y) - I(X;Z)$ can be upper bounded as*

$$f_0(F_X) = I(X;Y) - I(X;\widetilde{Y}) + I(X;\widetilde{Y}) - I(X;Z) \leq I(X;Y) - I(X;\widetilde{Y}) + I(X;\widetilde{Z}). \tag{5.18}$$

where $\widetilde{Y} \triangleq Y + N_D$, with N_D being a Poisson distributed random variable with mean $\lambda_D\Delta$ independent of X and Y, where $\lambda_D \triangleq \frac{\alpha_B}{\alpha_E}\lambda_E - \lambda_B$. Moreover, $\widetilde{Z}|X$ is a Poisson random variable with mean $\left[(\alpha_B - \alpha_E)X + \left(\frac{\alpha_B}{\alpha_E} - 1\right)\lambda_E\right]\Delta$ independent of $Z|X$ and such that $\widetilde{Y}|X = Z|X + \widetilde{Z}|X$.

Proof. The proof is given by [15, Lemma 1, Lemma 7]. ■

5.3.5 Scenario 1: $\alpha_B = \alpha_E$ and $\frac{\lambda_B}{\alpha_B} < \frac{\lambda_E}{\alpha_E}$

For this case observe that $\widetilde{Z} \equiv 0$ and we can write

$$C_S = \frac{1}{\Delta} f_0(F_X^*) = \frac{1}{\Delta}\left[H_Y(F_X^*) - H_{\widetilde{Y}}(F_X^*) + H_{\widetilde{Y}|X}(F_X^*) - H_{Y|X}(F_X^*)\right], \tag{5.19}$$

where $F_X^* \in \mathcal{F}^+$ with $\mathcal{F}^+$ being one of the feasible sets defined in (5.10)–(5.12), and $H_Y(F_X^*)$ and $H_{\widetilde{Y}}(F_X^*)$ are the entropies of the discrete random variables Y and $\widetilde{Y}$, respectively, induced by the optimal input distribution F_X^*. Furthermore, $H_{Y|X}(F_X^*)$ and $H_{\widetilde{Y}|X}(F_X^*)$ are the conditional entropies of $Y|X$ and $\widetilde{Y}|X$, respectively, induced by F_X^*. Now, we are ready to present the upper bound on the secrecy capacity of the DT–PWC in the high-intensity regime.

Proposition 5. *The secrecy capacity of the DT–PWC with either of the considered constraints in* (5.10)–(5.12) *can be upper bounded by*

$$C_S \leq \frac{\frac{\lambda_D^2}{2} + \frac{\lambda_D}{\Delta}}{\lambda_B}, \tag{5.20}$$

Proof. For convenience the proof is relegated to Appendix E. ■

First, notice that the upper bound in (5.20) holds for all the values of the peak- and/or average-intensity constraints. This implies that the secrecy capacity of the DT–PWC in the regimes where either of the constraints $\mathcal{A} \to \infty$ or $\mathcal{E} \to \infty$ does not scale with the constraints and approaches a positive constant, i.e.,

$$C_S = O(1). \tag{5.21}$$

Observe that in the unconstrained bandwidth regime ($\Delta \to 0$), the upper bound in (5.20) diverges to infinity.

5.3.6 Scenario 2: $\alpha_B > \alpha_E$ and $\frac{\lambda_B}{\alpha_B} \leq \frac{\lambda_E}{\alpha_E}$

In this case, first note that using the upper bound in (5.20), the secrecy capacity of the DT–PWC is upper bounded by

$$C_S \leq \frac{\frac{\lambda_D^2}{2} + \frac{\lambda_D}{\Delta}}{\lambda_B} + \frac{1}{\Delta} I(X^*; \widetilde{Z}^*), \tag{5.22}$$

where $I(X^*; \widetilde{Z}^*)$ is mutual information between X and $\widetilde{Z}$ induced by the secrecy-capacity-achieving distribution $F_X^* \in \mathcal{F}^+$. Now, observe that $I(X^*; \widetilde{Z}^*)$ can be upper bounded by the capacity of a discrete-time Poisson channel whose input is X and output is $\widetilde{Z}$. Therefore, in the high-intensity regime and following along similar lines of [8, Theorem 3, Theorem 4, Theorem 7], one can upper bound $I(X^*; \widetilde{Z}^*)$

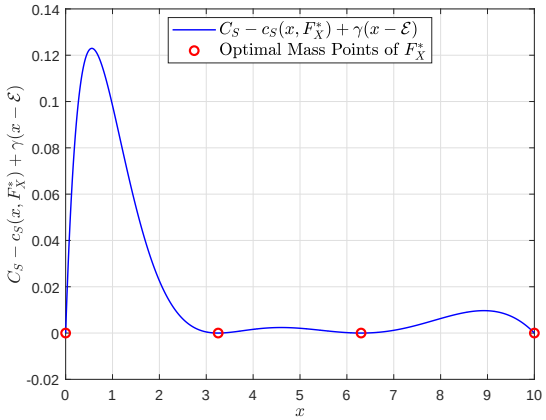

Figure 5.1: Illustration of $C_S - c_S(x; F_X^*) + \gamma(x - \mathcal{E})$ yielded by the optimal input distribution when $A = 10$, $\mathcal{E} = \frac{\mathcal{A}}{4}$, $\alpha_B = 2$, $\lambda_B = 1$, $\alpha_E = 1$, $\lambda_E = 2$, and $\Delta = 0.5$ seconds.

by

$$I(X^*; \widetilde{Z}^*) \leq \frac{1}{2} \log\left[(\alpha_B - \alpha_E)\mathcal{A}\Delta\right], \quad \text{for constraints (5.10)–(5.11)},$$
$$I(X^*; \widetilde{Z}^*) \leq \frac{1}{2} \log\left[(\alpha_B - \alpha_E)\mathcal{E}\Delta\right], \quad \text{for constraint (5.12)}. \tag{5.23}$$

Therefore, C_S is upper bounded by

$$C_S \leq \frac{\frac{\lambda_D^2}{2} + \frac{\lambda_D}{\Delta}}{\lambda_B} + \frac{1}{2\Delta} \log\left[(\alpha_B - \alpha_E)\mathcal{A}\Delta\right], \quad \text{for constraints (5.10)–(5.11)}, \tag{5.24}$$
$$C_S \leq \frac{\frac{\lambda_D^2}{2} + \frac{\lambda_D}{\Delta}}{\lambda_B} + \frac{1}{2\Delta} \log\left[(\alpha_B - \alpha_E)\mathcal{E}\Delta\right], \quad \text{for constraint (5.12)}. \tag{5.25}$$

As can be seen the secrecy capacity of the DT–PWC cannot scale faster than the logarithm of the square root of constraints, i.e.,

$$C_S = O\left(\log\left(\sqrt{\mathcal{A}}\right)\right), \quad \text{for constraints (5.10)–(5.11)}, \tag{5.26}$$
$$C_S = O\left(\log\left(\sqrt{\mathcal{E}}\right)\right), \quad \text{for constraint (5.12)}. \tag{5.27}$$

5.4 Numerical Results

In this section, we provide numerical results for the secrecy capacity and the entire rate-equivocation region of the DT–PWC.

Figure 5.1 provides a plot of the KKT conditions given by (E.28)–(E.29) for an optimal input dis-

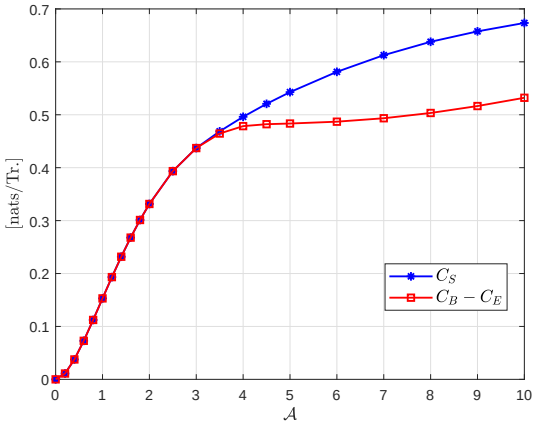

Figure 5.2: The secrecy capacity when $\mathcal{E} = \frac{\mathcal{A}}{4}$, $\alpha_B = 2$, $\lambda_B = 1$, $\alpha_E = 1$, $\lambda_E = 2$, and $\Delta = 0.5$ seconds versus the peak-intensity constraint $\mathcal{A}$.

tribution when $A = 10$, $\mathcal{E} = \frac{\mathcal{A}}{4}$, $\alpha_B = 2$, $\lambda_B = 1$, $\alpha_E = 1$, $\lambda_E = 2$, and $\Delta = 0.5$ seconds. We numerically found that for these parameters, the optimal input distribution has four mass points located at $x = 0, 3.2541, 6.3032$, and 10 with probability masses 0.4799, 0.3630, 0.0683, and 0.0888, respectively. Furthermore, the corresponding Lagrange multiplier is $\gamma = 0.0513$. We observe that $C_S - c_S(x; F_X^*) + \gamma(x - \mathcal{E})$ is generally nonnegative and is equal to zero at the optimal mass points; verifying the optimality conditions in (E.28)–(E.29).

Figure 5.2 illustrates the secrecy capacity C_S and the difference $C_B - C_E$ versus the peak-intensity constraint $\mathcal{A}$, where C_B and C_E are the legitimate user's and the eavesdropper's channel capacities, respectively. First, we observe that the secrecy capacity is an increasing function in $\mathcal{A}$. Furthermore, we see that this difference is a lower bound on the secrecy capacity C_S. We also observe that, for small values of $\mathcal{A}$, $C_B - C_E$ and C_S are identical. However, as $\mathcal{A}$ increases $C_B - C_E$ and C_S become different. Similar to the secrecy capacity results of the FSO wiretap channel and optical wiretap channel with input-dependent Gaussian noise under a peak- and average-intensity constraints provided in [14, 25], here too, $I(X;Y)$ and $I(X;Z)$ are maximized by the same discrete distribution, however, $I(X;Y) - I(X;Z)$ is maximized by a different distribution. As a specific example, when $\mathcal{A} = 4$, while both $I(X;Y)$ and $I(X;Z)$ are maximized by the same *binary* distribution with mass points at $x = 0$ and 4 with probability masses 0.75 and 0.25, respectively, $I(X;Y) - I(X;Z)$ is maximized by a *ternary* distribution with mass points at $x = 0$, 2.6848, and 4 with probability masses 0.6884, 0.1872, and 0.1244, respectively. This explains the difference between C_S and $C_B - C_E$ at $\mathcal{A} = 4$ in this figure.

In Figure 5.3, we plot the effect of pulse duration Δ on the secrecy capacity of the DT–PWC with nonnegativity, peak- and average-intensity constraints. From the figure, we observe that, in the low-

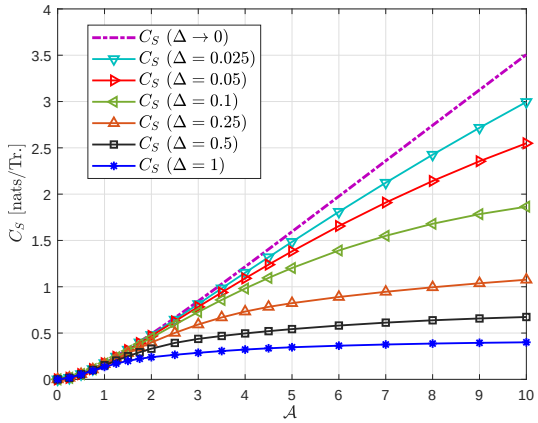

Figure 5.3: The secrecy capacity of the DT–PWC when $\mathcal{E} = \frac{\mathcal{A}}{4}$, $\alpha_B = 2$, $\lambda_B = 1$, $\alpha_E = 1$, and $\lambda_E = 2$ versus the peak-intensity constraint $\mathcal{A}$ for different values of pulse duration Δ.

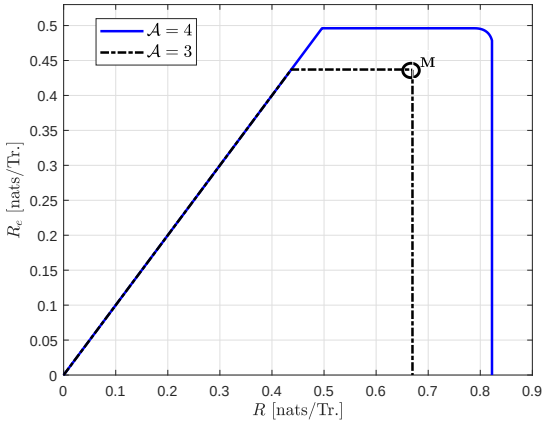

Figure 5.4: The rate-equivocation region when $\mathcal{E} = \frac{\mathcal{A}}{4}$, $\alpha_B = 2$, $\lambda_B = 1$, $\alpha_E = 1$, $\lambda_E = 2$, and $\Delta = 0.5$ for peak-intensity constraints $\mathcal{A} = 3$ and $\mathcal{A} = 4$. Point M refers to the case when secrecy capacity and capacity are achieved simultaneously.

intensity regime, the effect of decreasing Δ on the secrecy capacity is not significant. However, in the moderate- to high-intensity regime, Δ becomes significantly influential and the decrease in Δ results in a higher secrecy capacity. Furthermore, we see that the secrecy capacity of the continuous-time PWC (when $\Delta \to 0$) is always an upper bound on the secrecy capacity of the DT–PWC.

Figure 5.4 depicts the entire rate-equivocation region of the DT–PWC with nonnegativity, peak- and average-intensity constraints when $\mathcal{E} = \frac{\mathcal{A}}{4}$, $\alpha_B = 2$, $\lambda_B = 1$, $\alpha_E = 1$, $\lambda_E = 2$, and $\Delta = 0.5$ for two different values of $\mathcal{A}$. When $\mathcal{A} = 3$, it is clear from the figure that both the secrecy capacity and the capacity can be attained simultaneously (Point "M" in the figure). In particular, for $\mathcal{A} = 3$, the binary input distribution with mass points located at $x = 0$ and 3 with probabilities 0.75 and 0.25, respectively,

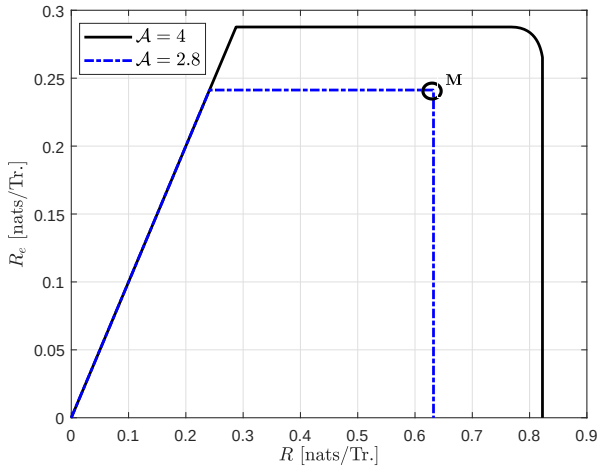

Figure 5.5: The rate-equivocation region when $\mathcal{E} = \frac{\mathcal{A}}{4}$, $\alpha_B = 2$, $\lambda_B = 1$, $\alpha_E = 1$, $\lambda_E = 0.5$, and $\Delta = 0.5$ for peak-intensity constraints $\mathcal{A} = 2.8$ and $\mathcal{A} = 4$. Point M refers to the case when secrecy capacity and capacity are achieved simultaneously.

achieves both the capacity and the secrecy capacity. This implies that, when $\mathcal{A} = 3$, the transmitter can communicate with the legitimate user at the capacity while achieving the maximum equivocation at the eavesdropper. On the other hand, when $\mathcal{A} = 4$ the secrecy capacity and the capacity cannot be achieved simultaneously (notice the curved shape in the figure). More specifically, for $\mathcal{A} = 4$ the binary input distribution with mass points at $x = 0$ and 4 with probability masses 0.75 and 0.25, respectively, achieves the capacity, while a ternary distribution with mass points located at $x = 0$, 2.6848, and 4 with probability masses 0.6884, 0.1872, and 0.1244, respectively, achieves the secrecy capacity. This implies that the optimal input distributions for the secrecy capacity and the capacity are different. In other words, there is a tradeoff between the rate and its equivocation in the sense that, to increase the communication rate, one must compromise on the equivocation of this communication, and to increase the achieved equivocation, one must compromise on the communication rate.

Figure 5.5 illustrates the entire rate-equivocation region of the DT–PWC with nonnegativity, peak- and average-intensity constraints for the case when $\alpha_B > \alpha_E$ and $\frac{\lambda_B}{\alpha_B} = \frac{\lambda_E}{\alpha_E}$. In this case, the eavesdropper's observations are just the thinned version of those of the legitimate receiver's and [15] shows that for the continuous-time PWC, $C_S = C_B - C_E$, i.e., there is no tradeoff between the rate and its equivocation. This is in contrast to the case of the DT–PWC as shown in this figure. We observe that even in this extreme case, in general, there is a tradeoff between the rate and its equivocation.

Finally, in Figure 5.6, we plot the exact and asymptotic secrecy capacity results in the low-intensity regime versus the peak-intensity constraint $\mathcal{A}$ when peak- or both the peak- and average-intensity con-

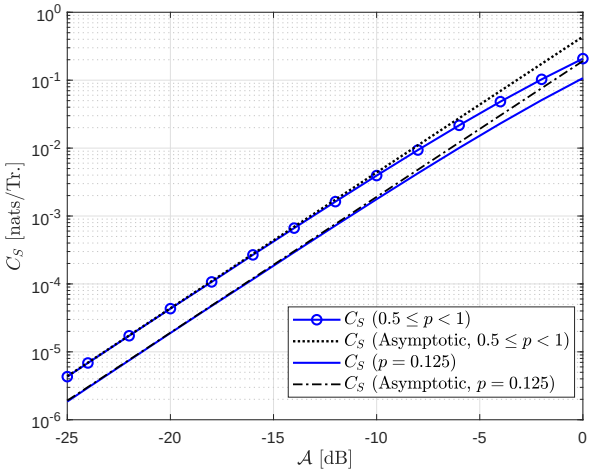

Figure 5.6: The asymptotic and exact secrecy capacity for $\alpha_B = 2$, $\lambda_B = 1$, $\alpha_E = 1$, and $\lambda_E = 2$ versus $\mathcal{A}$ for both peak- and average-intensity constraints.

straints are active. From the figure, we observe that the asymptotic results for the secrecy capacity given in (5.15) are in precise agreement with the numerical result.

5.5 Conclusions

We studied the DT–PWC where a combination of peak- and average-intensity constraints were considered. We formally characterized the secrecy-capacity-achieving input distribution to be unique and discrete with a finite number of mass points when peak-intensity or both peak- and average-intensity constraints were active. Also, we established that the entire rate-equivocation region of the DT–PWC under peak- or both peak- and average-intensity constraints is exhausted by discrete distributions with finitely many mass points. However, when only an average-intensity constraint is imposed we showed that the secrecy capacity as well as the entire boundary of the rate-equivocation region are attained by discrete distributions with countably infinite number of mass points, but finitely many mass points in any bounded interval. In some cases, we also have been able to characterize the optimal distributions for the continuous-time PWC (i.e., the case where pulse duration is very small).

Besides, we characterized the behavior of the secrecy capacity in the low- and high-intensity regimes. In the low-intensity regime, we fully characterized the secrecy capacity and the secrecy-capacity-achieving input distribution when peak- or both peak- and average-intensity constraints are active. It was shown that in this regime the secrecy capacity scales quadratically with the peak-intensity constraint and the optimal input distribution is binary. In the high-intensity regime we proved that when the legitimate receiver's and the eavesdropper's channel gains are identical, the secrecy capacity is upper bounded by

a positive constant, thus, it does not scale with the constraints. Moreover, when the channel gains are different, the secrecy capacity is upper bounded by the capacity of a discrete-time Poisson channel and the secrecy capacity cannot scale faster than the logarithm of the square root of the constraints.

Finally, our numerical results indicated that under nonnegativity and peak- and average-intensity constraints, the secrecy capacity and the capacity of the DT–PWC channel cannot be obtained simultaneously in general, i.e., there is a tradeoff between the rate and its equivocation.

Chapter 6: The Capacity Region of the Input-Dependent Gaussian Noise Optical Multiple Access Channel with Peak- and Average-Intensity Constraints

M. Soltani, Z. Rezki and A. Chaaban, "Sum-Capacity-Achieving Distributions in the Input-Dependent Gaussian Noise Optical Multiple Access Channel with Peak and Average Intensity Constraints," *in Proceedings of the 2019 IEEE Canadian Workshop on Information Theory (CWIT'2019), Hamilton, ON, Canada, 2019, pp. 1-6.*

M. Soltani, Z. Rezki and A. Chaaban, "The Capacity Region of the Input-Dependent Gaussian Noise Optical Multiple Access Channel with Peak- and Average-Intensity Constraints," *Submitted to IEEE Transactions on Wireless Communications*

6.1 Introduction

Optical wireless communication (OWC) is a promising technique for supporting high data-rate communication as a complementary or a backup technology to radio-frequency (RF) communications. It has numerous advantages in comparison to RF, including higher data-rates, more abundant unlicensed spectrum and being less demanding in terms of system infrastructure. Furthermore, OWC is utilized in deep-space optical communications and visible light communications [2, 1]. One of the most popular communication techniques used in OWC is the intensity modulation and direct detection (IM-DD) technique for its simplicity [2]. In this setup, the channel input modulates the intensity of the emitted light. Thus, the input signal is proportional to the light intensity and is nonnegative. The receiver is equipped with a photodetector (PD) which measures the intensity of the received light and generates an output signal proportional to the detected intensity, corrupted by noise.

6.1.1 Channel Models for OWC

Based on the distribution of the corrupting noise at the output of the receiver's PD, there are several channel models for the underlying OWC. Free space optical (FSO) channels [2, 3], optical channels with input-dependent Gaussian noise (IDGN) [2, 5] and Poisson optical channels [2, 7, 18, 8] are the most widely used channel models for optical wireless communications.

The most accurate channel model for OWC based on IM-DD that can capture most of the optical channel impairments, i.e., nonlinearities of optical devices, the photon counting process at the receiver's PD, the ambient light in the environment, and the thermal noise in the electronic devices of the receiver, is the Poisson optical channel model. In this model, the output is a doubly stochastic Poisson process whose rate is typically the intensity of the incident light plus a constant called "dark current" [2, 7, 18, 8, 6, 1]. Whether there are restrictions on the bandwidth of the input signal or not, the Poisson optical channel can be divided into two categories: 1) continuous-time Poisson (CT-P) optical channel model [18, 6], in which the channel input can have as large as possible bandwidth, i.e., there is no restriction on the transmission bandwidth; 2) discrete-time Poisson (DT-P) optical channel model [7, 8], which refers to the cases when there exists a finite transmission bandwidth constraint.

When the mean of the Poisson process in the Poisson channel model is very large, which corresponds to the cases when the intensity of the incident light plus the dark current is very large, the optical channel can be well-approximated by an additive Gaussian noise whose variance depends on the signal intensity [2],[5, Appendix C],[1, 42]. This signal-dependent noise accurately models the ambient light, thermal noise, and the nonlinearities of the optical devices, but fails to capture the photon arrivals at the receiver's PD.

When the ambient light in the environment and the thermal noise at the electronic devices of the receiver are dominant, the optical wireless channel can be approximated by an input-independent Gaussian noise. This model is known as the FSO channel. In such a model, the corrupting noise is independent of the received optical intensity and follows a zero-mean Gaussian distribution [2, 3, 4, 1, 21]. As such, the channel output is simply the addition of the channel input and the input-independent Gaussian noise.

6.1.2 Single-User OWC

Studying the communications performance limits (such as the channel capacity) of the aforementioned channel models from an information-theoretic point of view is rather difficult. This is because the channel input must satisfy nonnegativity, peak- and average-intensity constraints due to eye safety and practical considerations [2]. Considering the single-user capacity of the aforementioned channel models, the works in [6, 7, 10, 18] showed that the capacity-achieving input distributions are discrete with a finite number of mass points when the channel input is constrained by nonnegativity, peak- and average-intensity constraints. This is on the contrary to the case of the Gaussian channels with average-power constrained inputs, where Gaussian input distribution, which is a continuous distribution, is capacity-achieving [9]. Furthermore, when the channel input is only constrained by nonnegativity and average-intensity con-

straints, the capacity-achieving distributions for FSO channel [19] and DT-P optical channel [20] were shown to be discrete but with an unbounded support set, i.e., the support set of the optimal distributions are countably infinite. Finally, although the capacity of the CT-P optical channel is known in closed-form due to [6, 18], the capacity of the other channel models (FSO, IDGN, and DT-P) are only characterized in closed-form for the low- and high-intensity regimes [5, 8, 21], and in general, the capacity of these channels is still unknown.

6.1.3 Mulit-User OWC

6.1.3.1 Optical Intensity Multiple Access Channel

Information-theoretic studies have also been performed for the multiuser OWC. For instance, the work in [22] considered the Gaussian multiple access channel with amplitude and variance constraints, and established that the boundary of the capacity region is obtained by distributions that are discrete with a finite number of mass points. These results are directly applicable to the free-space optical multiple access channel (FSO-MAC) with nonnegativity and peak- and average-intensity constraints. Thus, the capacity region of the FSO-MAC under nonnegativity, peak- and average-intensity constraints is exhausted by discrete input distributions with a finite support set. Furthermore, [23, 24] provided tight bounds on the capacity region of FSO-MAC with peak- and/or average-intensity constraints across several intensity regimes (low, moderate, and high). Specifically, in the regime where both peak- and average-intensity constraints tend to zero with their ratio held fixed, [23] shows that an ON-OFF keying scheme combined with successive interference cancellation at the receiver is capacity-achieving. Authors in [24] characterized the capacity region of FSO-MAC with nonnegativity and average-intensity constraints in the regime where the average-intensity tends to infinity. In particular, it was shown that an exponential distribution is asymptotically optimal and attains the capacity region of the FSO-MAC with an average-intensity constraint. For a continuous-time Poisson OMAC subject to peak- and average-intensity constraints, Lapidoth *et al.* established the capacity region of the Poisson MAC for the two-user case in a closed-form expression. The authors showed that for achieving every point on the boundary of the capacity region, the input distributions for both users must be binary with an infinite transmission bandwidth. The discrete-time Poisson OMAC has also been considered in [16], where authors studied the two-user case and verified the optimality of discrete inputs with a finite support set for achieving the *sum-capacity* when nonnegativity and peak-intensity constraint are considered. However, the authors did not verify whether or not discrete input distributions exhaust the entire capacity region of discrete-time Poisson OMAC with peak-intensity constraint.

6.1.3.2 Optical Intensity Broadcast Channel

Compared to the research that has been conducted for OMAC with a variety of intensity constraints, results regarding the optical broadcast channel (OBC) are less abundant. The existing results pertaining to the OBC are limited to free-space OBC (FSOBC) [43, 44, 45] and continuous-time Poisson OBC (CTP-OBC) [40, 46]. In regards to FSOBC, authors in [43, 44] analyze the performance of an orthogonal code-division multiple-access in an FSOBC. This orthogonalization allows serving multiple users in the FSOBC without interference. Thus, the channel from the transmitter to each user reduces to an FSO channel and the capacity results of [10] can be applied. On the other hand, [9] shows that for a stochastically degraded broadcast channels, superposition coding (SC) is optimal and orthogonalizing users is not efficient. Inspired by this fact, the work in [45] studies the N-user FSOBC under peak- and average-intensity constraints, and finds fairly tight inner and outer bounds for the capacity region. In particular, it shows that in the low-intensity regime, ON-OFF keying along with time-division multiple-access is capacity-achieving. However, in general, it does not specify and characterize the capacity-achieving input distributions. In [40], authors study the CTP-OBC under peak- and average-intensity constraints and establish the conditions under which the CTP-OBC is degraded; hence the capacity region is achieved using superposition coding. Furthermore, binary input distributions along with timesharing achieve all the points on the boundary of the capacity region. In [46], the authors show that superposition coding is optimal for other classes of CTP-OBC, such as less noisy and more capable broadcast channels when peak- and average-intensity constraints are active. In particular, the authors proved that binary distributions are optimal in the sense that they achieve the capacity region of the CTP-OBC.

6.1.3.3 Other Multiuser Optical Intensity Channels

Unfortunately, there exist a few limited studies on other types of multiuser optical intensity channels, such as interference channel, relay channel, etc. The only existing researches concerning these types of multiuser channels are the works on the FSO interference channel (FSO-IC) with an average-intensity constraint [47] and on the CT-P interference channel (CTP-IC) with a peak-intensity constraint [48]. For the FSO-IC with an average-intensity constraint, inner and outer bounds on the capacity region have been derived [47]. The bounds were shown to coincide asymptotically at the strong interference regime when average-intensity constraint tends to infinity, thus characterizing the strong interference capacity region. Authors in [48] provided the conditions for the strong interference regime and the corresponding capacity region for the CTP-IC. In particular, they showed that binary distribution achieves the strong interference capacity region of the CTP-IC.

In this chapter, a discrete-time IDGN-OMAC which reflects an OWC scenario in which two optical transmitters wish to communicate their messages to a common optical receiver, is studied. In this setup, the input signals are restricted by nonnegativity, peak- and average-intensity constraints. Using an IM-DD system, the photodetector at the receiver counts the number of received photons and outputs a signal that is corrupted by an additive Gaussian noise whose variance depends on the input signals. The objective is to have reliable communications with the receiver such that it can decode the messages of both users reliably. To this end, it is first noted that the capacity region of an OMAC can be characterized by solving a weighted sum-rate maximization problem [22, Lemma 4].

Next, the optimal input distributions that exhaust the capacity region of the IDGN-OMAC with peak- and average-intensity constraints is characterized by solving a convex optimization framework addressed in, e.g., [16, 36, 22]. This is done by deriving the necessary and sufficient Karush-Kuhn-Tucker (KKT) conditions that the optimal distributions must satisfy. Using these optimality conditions, it is shown that the optimal distributions must be discrete and admit a finite number of mass points. This step is established via a proof by contradiction method. In particular, the contradiction argument hinges on the fact that if the support set of the optimal input distributions has an infinite number of elements, then the cost function which grows linearly in x should grow faster than the rate-region density which grows quadratically in x.

Finally, a closed-form expression for the capacity region of the IDGN-OMAC with peak- and average-intensity constraints in the low-intensity regime is presented, i.e., the regime where both peak- and average-intensity constraints tend to zero with their ratio held fixed. To this end, the single-user capacity of the input-dependent Gaussian noise channel with peak- and average-intensity constraints for each user is considered. The single-user channel capacity of the first user and the second user are respectively denoted by C_1 and C_2. It is noted that capacity region of the IDGN-OMAC with peak- and average-intensity constraints, is contained in the region $[0, C_1] \times [0, C_2]$. Afterwards, according to [5, Theorem 10], a closed-form expression for the single-user channel capacity C_1 and C_2 in the low-intensity regime is given. Finally, it is shown that the region $[0, C_1] \times [0, C_2]$ is contained in the capacity region of IDGN-OMAC with peak- and average-intensity constraints in the low-intensity regime. This implies that in the low-intensity regime, the capacity region of the IDGN-OMAC with peak- and average-intensity constraints is indeed the rectangular region $[0, C_1] \times [0, C_2]$. It is worth mentioning that in the low-intensity regime, the optimal input distributions exhausting the entire capacity region are binary with mass points at the origin and the peak-intensity constraint.

6.2 Input-Dependent Gaussian Noise OMAC

In a two-user OMAC based on the IM-DD technique and pulse amplitude modulation scheme, the channel inputs of both users modulate the emitted light intensity from the light-emitting diode (LED) at the transmitters and a photodetector is used for receiving the optical signal at the receiver. Thus, the channel input $X_i,\ i \in \{1,2\}$ is a nonnegative random variable representing the intensity of the optical signal for user i. Since intensity is constrained due to practical and safety restrictions by a peak- and average-constraints in general, the input has to satisfy [2, Chapter 7], [7]

$$0 \leq X_i \leq \mathcal{A}_i,\ i \in \{1,2\}, \tag{6.1}$$

$$\mathbb{E}\left[X_i\right] \leq \mathcal{E}_i,\ i \in \{1,2\}. \tag{6.2}$$

In this setup, conditional on the channel inputs X_i's, the output signal at the receiver Y is Gaussian distributed with mean $X_1 + X_2$ and variance $\sigma^2(x_1,x_2) \triangleq \sigma_0^2 + \sigma_1^2(x_1+x_2)$, where σ_0^2 and σ_1^2 are positive constants. Hence, the conditional channel law is given by [5]

$$p_{Y|X_1,X_2}(y|x_1,x_2) = \frac{1}{\sqrt{2\pi\sigma^2(x_1,x_2)}} \exp\left[-\frac{(y-x_1-x_2)^2}{2\sigma^2(x_1,x_2)}\right],\ y \in \mathbb{R}. \tag{6.3}$$

Based on (6.3), the channel output Y can be written as

$$Y = X_1 + X_2 + \sqrt{X_1+X_2}\, Z_1 + Z_0, \tag{6.4}$$

where Z_0 is a zero-mean Gaussian noise component with variance σ_0^2 and Z_1 is independent of Z_0 and is distributed according to a zero-mean Gaussian distribution with variance σ_1^2. We note that when $\sigma_1^2 = 0$, the IDGN-OMAC becomes and OMAC with only input-independent Gaussian noise which corresponds to an FSO-MAC.

A general coding scheme for the IDGN-OMAC can be described as follows. Transmitter $i,\ i \in \{1,2\}$ wishes to communicate a message W_i chosen uniformly from the message set $\mathcal{W}_i = \{1,\ldots,|\mathcal{W}_i|\}$ to the receiver, where $|\mathcal{W}_i|$ is the cardinality of the set $\mathcal{W}_i$. This message is encoded into a codeword of length $n \in \mathbb{N}$ denoted by $\mathbf{X}_i \in \mathbb{R}_+^n$, and then transmitted, one symbol at a time. The codewords at transmitter i constitute a codebook that must satisfy the constraints (6.1)–(6.2). The receiver collects the received symbols over n transmission in $\mathbf{Y} \in \mathbb{R}^n$, and then uses a decoder to decode $\widehat{W}_i \in \mathcal{W}_i$, for all $i \in \{1,2\}$, from $\mathbf{Y}$. The transmission rate from transmitter i to the receiver is then defined as $R_i = \frac{\log(|\mathcal{W}_i|)}{n}$ in nats

per transmission.

We are interested in the set of achievable rate pairs (R_1, R_2) such that the error probability denoted by $\Pr\{W_i \neq \widehat{W}_i,\ i \in \{1,2\}\}$ can be made arbitrarily small by increasing the code length n. The set of all achievable rate pairs is the capacity region denoted by $\mathcal{C}$. The capacity region of a two-user discrete-time memoryless MAC is known due to [49] and is given by the closure of the convex-hull of all rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying

$$R_1 \leq I(X_1; Y|X_2), \tag{6.5a}$$

$$R_2 \leq I(X_2; Y|X_1), \tag{6.5b}$$

$$R_1 + R_2 \leq I(X_1, X_2; Y), \tag{6.5c}$$

for some input distributions $F_{X_1,X_2}(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2)$ over $\mathcal{F}_1^+ \times \mathcal{F}_2^+$, where $I(X_1; Y|X_2)$ is the conditional mutual information between X_1 and Y given X_2, $I(X_1, X_2; Y)$ is the mutual information between (X_1, X_2) and Y, $F_X(\cdot)$ is the cumulative distribution function of a random variable X, and the feasible sets $\mathcal{F}_i^+, i \in \{1,2\}$ is given by one of the following sets

$$\Omega^+_{\mathcal{A}_i,\,\mathcal{E}_i} \triangleq \left\{ F_{X_i} : \int_0^{\mathcal{A}_i} dF_{X_i}(x) = 1,\ \int_0^{\mathcal{A}_i} x\, dF_{X_i}(x) \leq \mathcal{E}_i,\ i \in \{1,2\} \right\}, \tag{6.6}$$

$$\Omega^+_{\mathcal{A}_i} \triangleq \left\{ F_{X_i} : \int_0^{\mathcal{A}_i} dF_{X_i}(x) = 1,\ i \in \{1,2\} \right\}, \tag{6.7}$$

This capacity region is achievable by jointly decoding (W_1, W_2) at the receiver, or by successive decoding combined with time-sharing [49].

6.3 IDGN-OMAC Capacity Region Characterization

This section presents the main results of the paper related to the characterization of the input distributions that exhaust the capacity region of the IDGN-OMAC with nonnegativity, peak- and average-intensity constraints.

As mentioned in Section 6.2, the capacity region is given by the closure of the convex-hull of the all the rate pairs satisfying (6.5). Thus, any point on the boundary of this capacity region corresponds to a solution for the optimization problem $\sup_{(R_1,R_2)\in\mathcal{C}} R_1 + \mu R_2$ for some $\mu > 0$. In other words, maximizing the weighted sum-rate over all the achievable rate pairs yields all the points on the boundary of the capacity region by letting μ vary in $(0, \infty)$. Using the structure of the capacity region, authors in [22,

Lemma 4] established that $\sup_{(R_1,R_2)\in\mathcal{C}} R_1+\mu R_2$ can be alternatively given by

$$\sup_{(R_1,R_2)\in\mathcal{C}} R_1+\mu R_2=\begin{cases}\sup_{F_{X_i}\in\mathcal{F}_i^+,\, i\in\{1,2\}} I(X_1;Y|X_2)+\mu I(X_2;Y), & 0<\mu<1\\ \sup_{F_{X_i}\in\mathcal{F}_i^+,\, i\in\{1,2\}} I(X_1,X_2;Y), & \mu=1\\ \sup_{F_{X_i}\in\mathcal{F}_i^+,\, i\in\{1,2\}} I(X_1;Y)+\mu I(X_2;Y|X_1), & \mu>1,\end{cases} \tag{6.8}$$

where $\mathcal{F}_i^+$ is one of the feasible sets in (6.6)–(6.7).

In what follows, we prove that the optimal input distributions that are solutions to the optimization problem in (6.8) for the possible choices of $\mathcal{F}_i^+$ mentioned in (6.6)–(6.7) are always discrete with a finite number of mass points, i.e., their support set is a countable finite set. These results are formally stated by the following theorems.

Theorem 9. *In a two-user IDGN-OMAC with nonnegativity, peak- and average-intensity constraints, i.e., $\mathcal{F}_i^+=\Omega^+_{\mathcal{A}_i,\mathcal{E}_i}$, $i\in\{1,2\}$, discrete input distributions with a countably finite support set exhaust the capacity region.*

Proof. The proof is presented in Appendix F. ■

We establish Theorem 9 in a few steps. Here, we only provide a proof sketch and the details of the proof are relegated to Appendix F. The first part of the proof is showing that the supremum in (6.8) is achievable. This implies that the supremum is actually a maximum and there exists at least one element $F_{X_i}\in\Omega^+_{\mathcal{A}_i,\mathcal{E}_i}, i\in\{1,2\}$ that achieves the maximum. To this end, we need to show that: 1) the set $\Omega^+_{\mathcal{A}_i,\mathcal{E}_i}$ is compact and convex; 2) the objective functional in (6.8) for all $\mu>0$ is continuous in F_{X_i}. The second part of the proof is focused on showing that the objective functional in (6.8) is weakly differentiable and concave in F_{X_i} for all $\mu>0$. Taking the weak derivative of the objective functional with respect to F_{X_i} and using the concavity, in the last part of the proof we derive the necessary and sufficient Karuch-Kuhn-Tucker (KKT) optimality conditions that an optimal distribution $F^*_{X_i}$ must satisfy. We continue the proof by showing that the optimal solution $F^*_{X_i}\in\Omega^+_{\mathcal{A}_i,\mathcal{E}_i}$ must be discrete with a countably finite set. This is done by proof via a contradiction approach, i.e., we assume to the contrary that the support set of the optimal solution $F^*_{X_i}$ contains an infinite number of elements; then we extend the corresponding rate-region densities (defined later in Appendix F) to the complex plane and observe that these densities are analytic over some open connected sets in the complex plane; afterward, leveraging the Identity Theorem from complex analysis and the Bolzano-Weierstrass Theorem, we find that a linearly growing function in x is lower bounded by another function which grows quadratically in x, and thus reaching the desired contradiction implying that the support set of $F^*_{X_i}$ must be countably finite.

It is noteworthy that our analysis for reaching a contradiction is flexible in the sense that following along similar lines of the provided analysis, one can show that the capacity region of the FSO-MAC with nonnegativity, peak- and average-intensity constraints is also exhausted by discrete distributions with a finite number of mass points. In particular, our contradiction argument holds for the case when $\sigma_1^2 = 0$, and thus, the result follows.

In the IDGN-OMAC, when the ratio of the average-intensity constraint to the peak-intensity constraint is ≥ 1, the average-intensity constraint is inactive and only the peak-intensity constraint is imposed [5]. Therefore, it is of interest to also characterize the optimal input distributions attaining the capacity region when only the peak-intensity constraint is considered. Remarkably, dropping the average-intensity constraint does not change the result of Theorem 9. Invoking the KKT conditions and a slight modification of the proof of Theorem 9 shows that any point on the capacity region of an IDGN-OMAC with nonnegativity and peak-intensity constraint can be achieved by discrete input distributions with a finite number of mass points. This is formally stated by the following corollary.

Corollary 3. *In a two-user IDGN-OMAC with nonnegativity and peak-intensity constraints, i.e., $\mathcal{F}_i^+ = \Omega_{\mathcal{A}_i}^+$, $i \in \{1,2\}$, discrete input distributions with finitely many mass points achieve all the points on the boundary of the capacity region.*

Next, we characterize the capacity region of the IDGN-OMAC in the low-intensity regime, i.e., in the regime where both peak- and average-intensity constraints tend to zero while their ratio is held fixed at $\alpha_i \triangleq \frac{\mathcal{E}_i}{\mathcal{A}_i}, i \in \{1,2\}$.

Theorem 10. *In the regime where $\mathcal{E}_i \to 0$ and $\mathcal{A}_i \to 0$ with their ratio held fixed at α_i with $i \in \{1,2\}$, the capacity region of the IDGN-OMAC is given by*

$$\mathcal{C} = \begin{cases} R_1 \leq \dfrac{\mathcal{A}_1^2}{2}\alpha_1(1-\alpha_1)\left(\dfrac{1}{\sigma_0^2} + \dfrac{\sigma_1^4}{2\sigma_0^4}\right), \\ R_2 \leq \dfrac{\mathcal{A}_2^2}{2}\alpha_2(1-\alpha_2)\left(\dfrac{1}{\sigma_0^2} + \dfrac{\sigma_1^4}{2\sigma_0^4}\right), \end{cases} \tag{6.9}$$

for $\alpha_i \in (0, \frac{1}{2}], i \in \{1,2\}$, and

$$\mathcal{C} = \begin{cases} R_1 \leq \dfrac{\mathcal{A}_1^2}{8}\left(\dfrac{1}{\sigma_0^2} + \dfrac{\sigma_1^4}{2\sigma_0^4}\right), \\ R_2 \leq \dfrac{\mathcal{A}_2^2}{8}\left(\dfrac{1}{\sigma_0^2} + \dfrac{\sigma_1^4}{2\sigma_0^4}\right), \end{cases} \tag{6.10}$$

for $\alpha_i \in [\frac{1}{2}, 1]$.

Proof. Theorem 10 is proven in Appendix F. ■

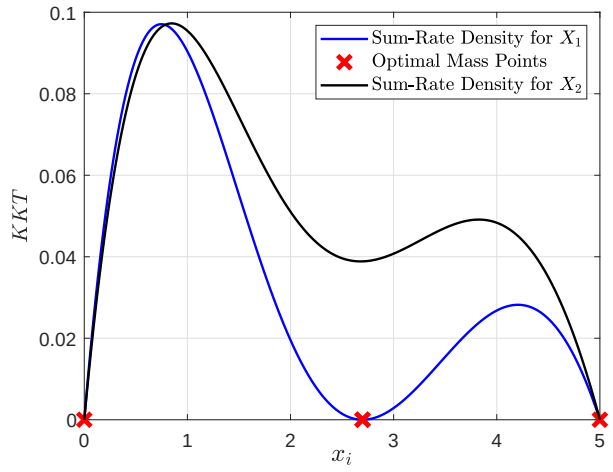

Figure 6.1: Illustration of KKT conditions satisfied by the optimal input distributions $F^*_{X_1}$ and $F^*_{X_2}$ when $\sigma_0^2 = 1$, $\sigma_1^2 = 0.25$, $\mathcal{A}_1 = \mathcal{A}_2 = 5$, and $\mathcal{E}_1 = \mathcal{E}_2 = 1$.

The proof of Theorem 10 is outlined as follows. First, we note that the capacity region of the IDGN-OMAC satisfies $\mathcal{C} \subset [0, C_1] \times [0, C_2]$, where C_1 is the single-user capacity of the first user and C_2 is the single-user capacity of the second user. Then, according to [5, Theorem 10], we have a closed-form expression for the single user capacity C_i, $i \in \{1, 2\}$ in the low-intensity regime as

$$C_i = \begin{cases} \dfrac{\mathcal{A}_i^2}{2}\alpha_i(1-\alpha_i)\left(\dfrac{1}{\sigma_0^2} + \dfrac{\sigma_1^4}{2\sigma_0^4}\right), & \alpha_i \in (0, \frac{1}{2}), \\ \dfrac{\mathcal{A}_i^2}{8}\left(\dfrac{1}{\sigma_0^2} + \dfrac{\sigma_1^4}{2\sigma_0^4}\right), & \alpha_i \in [\frac{1}{2}, 1]. \end{cases} \tag{6.11}$$

where C_i is attained by a binary input distribution with mass points located at $\{0, \mathcal{A}_i\}$ with corresponding probability masses $\{1-\alpha_i, \alpha_i\}$ when $\alpha_i \in (0, \frac{1}{2})$ and $\{\frac{1}{2}, \frac{1}{2}\}$ when $\alpha_i \in [\frac{1}{2}, 1]$. We continue the proof by showing that $[0, C_1] \times [0, C_2] \subset \mathcal{C}$ which will imply that $\mathcal{C} = [0, C_1] \times [0, C_2]$. To that end, we establish that the point (C_1, C_2) is achievable by showing that in the low-intensity regime, the sum-capacity (the maximum achievable sum-rate) is strictly greater than $C_1 + C_2$. This results in the achievability of the point (C_1, C_2), and consequently the result follows.

6.4 Numerical Results

This section provides numerical inspections for the capacity region of the IDGN-OMAC with peak- and average-intensity constraints along with the characterization of the optimal input distributions that correspond to some points on the boundary of the capacity region.

Figure 6.1 illustrates the rate-region density with respect to the optimal input distributions $F^*_{X_1}$ and

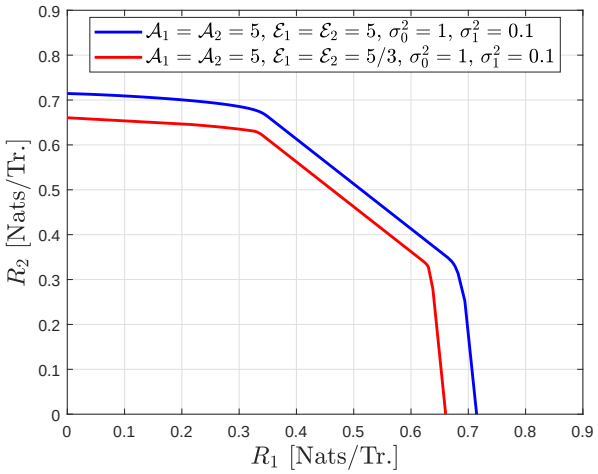

Figure 6.2: The capacity region of IDGN-OMAC with with nonnegativity, peak- and average-intensity constraints for $\sigma_0^2 = 1$, $\sigma_1^2 = 0.25$, $\mathcal{A}_1 = \mathcal{A}_2 = 5$, and two sets of values for the average-intensity constraints $\mathcal{E}_1$ and $\mathcal{E}_2$.

$F_{X_2}^*$ for $\mu = 1$, $\sigma_0^2 = 1$, $\sigma_1^2 = 0.25$, $\mathcal{A}_1 = \mathcal{A}_2 = 5$ and $\mathcal{E}_1 = \mathcal{E}_2 = 1$. We numerically found that for these parameters, the optimal input distribution for X_1 is ternary with mass points located at $\{0, 2.6986, 5\}$ with corresponding probability masses $\{0.7159, 0.1826, 0.1015\}$ and with the corresponding Lagrangian multiplier $\lambda_1 = 0.1627$. Moreover, the optimal input distribution for X_2 is binary with mass points at $\{0, 5\}$ with corresponding probability masses $\{0.8, 0.2\}$ and the Lagrangian multiplier $\lambda_2 = 0.155$. We observe that $\Xi(F_{X_i}^*) - \xi(x_i; F_{X_i}^*) + \lambda\,(x_i - \mathcal{E}_i),\ i \in \{1, 2\}$ is generally nonnegative and is equal to zero at the optimal mass points; verifying the optimality conditions in (F.22)–(F.24).

In Figure 6.2, we plot the entire boundary of the capacity region of the IDGN-OMAC with peak- and average-intensity constraints for $\sigma_0^2 = 1$, $\sigma_1^2 = 0.25$, $\mathcal{A}_1 = \mathcal{A}_2 = 5$ and two different sets of values for the average-intensity constraints $\mathcal{E}_1$ and $\mathcal{E}_2$. We note that every point on the boundary of the capacity region is achieved by discrete input distributions $F_{X_i}^*$, $i \in \{1, 2\}$ with finitely many mass points. For the case when $\mathcal{E}_1 = \mathcal{E}_2 = 5$, the average-intensity constraint is inactive and only the peak-intensity is imposed. However, when $\mathcal{E}_1 = \mathcal{E}_2 = 5/3$, both the peak- and average-intensity constraints are active. We observe that the capacity region of the IDGN-OMAC with peak- and average-intensity constraints is contained in the capacity region of the IDGN-OMAC with only peak-intensity constraints. We would like to draw the reader's attention to the geometry of the capacity region in the figure. As can be seen, due to the existence of an input-dependent noise component, the capacity region of the IDGN-OMAC is not a pentagon as opposed to the capacity region of the Gaussian multiple access channel with peak- and/or average power constraint [9, 22].

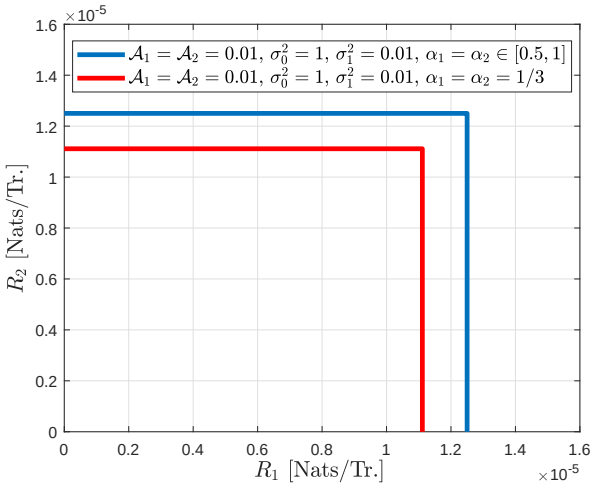

Figure 6.3: The capacity region of IDGN-OMAC with with nonnegativity, peak- and average-intensity constraints in the low-intensity regime for $\sigma_0^2 = 1$, $\sigma_1^2 = 0.01$, $\mathcal{A}_1 = \mathcal{A}_2 = 0.01$, and two sets of values for the average-intensity to peak-intensity ratios α_1 and α_2.

Finally, Figure 6.3 depicts the capacity region of the IDGN-OMAC with nonnegativity, peak- and average-intensity constraints in the low-intensity regime. As mentioned in Section 6.3, the capacity region is a rectangle formed by $[0, C_1] \times [0, C_2]$, where C_i, $i \in \{1, 2\}$ is the single-user capacity of user i. We note that every point on the boundary of the region is achieve by binary input distributions with mass points at $\{0, \mathcal{A}_i\}$, $i \in \{1, 2\}$ with corresponding probability masses $\{2/3, 1/3\}$ for the case $\alpha_i = 1/3$ and $\{0.5, 0.5\}$ for the case $\alpha_i \in [0.5, 1]$. We note that when $\alpha_i = 1/3$, both peak- and average-intensity constraints are active. However, when $\alpha_i \in [0.5, 1]$, only the peak-intensity constraint is active.

6.5 Conclusions

In this chapter, an IDGN-OMAC with nonnegativity, peak- and average-intensity constraints was considered. The optimal input distributions achieving any point on the boundary of the capacity region were fully characterized. It was shown that these optimal distributions must be discrete and posses a finite number of mass points.

Furthermore, the asymptotic behavior of the capacity region in the regime where both peak- and average-intensity constraints approach zero while their ratio is held fixed at a constant was provided. In this regime, the capacity region was fully characterized in a closed-form expression and it was shown that binary distributions with mass points at the origin and the peak-intensity constraint are optimal. We observed that in the low-intensity regime the capacity region has a rectangular shape which is formed by the single-user channel capacities of the users.

The numerical results indicated that due to the existence of an input-dependent noise component, the geometry of the capacity region of an IDGN-OMAC with nonnegativity, peak- and average-intensity constraints is not a pentagon, which is in contrast to the case of Gaussian multiple access channel with peak- and/or average-power constraints.

Chapter 7: Autoencoder-Based Optical Wireless Communications Systems

M. Soltani, W. Fatnassi, A. Aboutaleb, Z. Rezki, A. Bhuyan, and P. Titus, "Autoencoder-Based Optical Communications," *in Proceedings of the 2018 IEEE GLOBECOM Workshops on Machine Learning, Abu Zhabi, UAE, Dec. 2018.*

7.1 Introduction

Optical wireless communications (OWC) is a promising technique for supporting high data-rate communications as a complementary or a backup technology to radio-frequency (RF) communications. It has numerous advantages in comparison to RF, including higher data-rates, more abundant unlicensed spectrum and being less demanding in terms of system infrastructure.

One of the most popular communication techniques used in OWC is the intensity modulation and direct detection (IM-DD) technique for its simplicity [2]. In this setup, information modulate the intensity of the emitted light from the laser diode at the transmitter. Thus, the transmitted signal is proportional to the light intensity and is nonnegative. The receiver is usually equipped with a photodetector which measures the intensity of the received light and generates a signal proportional to the detected intensity, corrupted by noise.

Studying the communications performance limits (such as the channel capacity) of this simple implementation is rather difficult. The reason is that the transmitted signal must satisfy nonnegativity, peak and average intensity constraints due to the physical restrictions existing in the optical wireless channels [3, 21]. More importantly, traditional approaches used in constructing the signal constellations for RF channels cannot be applied directly to the optical channels due to the mentioned constraints. Therefore, one should consider designing structured optical signal-space model that can capture all the physical restrictions in the optical channels [4]. This task is not straightforward and heavily depends on the considered optical channel model. Hence, seeking for communications techniques (such as modulation, coding, decoding, etc.) that does not heavily depend on an existing channel model is quite appealing.

Recently, machine learning (ML) and deep learning (DL) approaches have been proposed for problems related to the physical layer of the communications network, such as modulation classifications [50, 51], coding and decoding [51, 52, 53], detection of the transmitted symbols [53, 54], channel estimation and equalization [53, 54]. These learning-based schemes are based on deep neural networks (DNNs) and do not heavily depend on the communications channel models. Among these techniques, autoencoders are

of special interest as they can capture the end-to-end performance of the entire communications system building blocks (such as encoding, transmission, reception, detection, equalization and decoding). In [51], O'Shae *et. al.* consider single- and multiuser communications over an additive white Gaussian noise (AWGN) RF channel and show that the performance of the learning-based communications systems (communications system based on autoencoders) can be competitive with respect to model-based algorithms, such as Hamming coding with maximum likelihood detector. Additionally, the authors in [53] demonstrate the feasibility of using autoencoders for practical over-the-air RF communications.

Motivated by the success of DL-based autoencoders in capturing the end-to-end performance of RF communications system, single- and multiuser OWC scenarios based on the autoencoders are proposed. For each of these scenarios, a complete OWC system solely composed of DNNs is designed and trained, and its end-to-end performance are compared with the model-based optical communications systems in terms of the block error rate (BLER) performance metric. Both single- and multiuser OWC systems are considered. In the single-user case, the BLER performance of the trained autoencoder is compared with that of an OWC system employing ON-OFF Keying (OOK) modulations (a modulation scheme often used in OWC systems [2, 1, 55]) along with Hamming coding scheme and hard- and soft-decision decoders. According to our obtained results, the learning-based OWC is able to perform as reliable as the model-based counterpart. In the multiuser case, an optical multiple access channel (MAC) based on autoencoders is studied and its BLER performance is compared with a multiple access system employing OOK modulations along with either joint decoding or time-sharing schemes. The numerical results demonstrate that the learning-based optical MAC can outperform the model-based MAC.

The rest of this chapter is organized as follows. Section 7.2 presents a single-user OWC system based on autoencoders. Section 7.3 provides the autoencoder-based implementation of an optical MAC (a multiuser scenario). Section 7.4 compares the end-to-end performance of the autoencoder-based single- and multiuser OWC systems with the model-based counterparts. Finally, section 7.5 concludes this chapter.

7.2 Single-user OWC Based on Autoencoders

Consider an autoencoder-based single-user OWC system as shown in Figure 7.1, in which the transmitter sends the message $s \in \mathcal{M},\ \mathcal{M} = \{1, \ldots, M\}$, to the receiver over an optical channel subject to nonnegativity and peak intensity constraints. The message s is represented as a one-hot vector $\mathbf{1}(s) \in \mathbb{R}^M$.

Then, the NN transmitter encodes the message s according to the mapping $\mathbf{g} : \mathcal{M} \to \mathbb{R}^n$ to generate the transmitted vector $\mathbf{x} = \mathbf{g}(s)$. Furthermore, to ensure the nonnegativity and peak intensity constraints

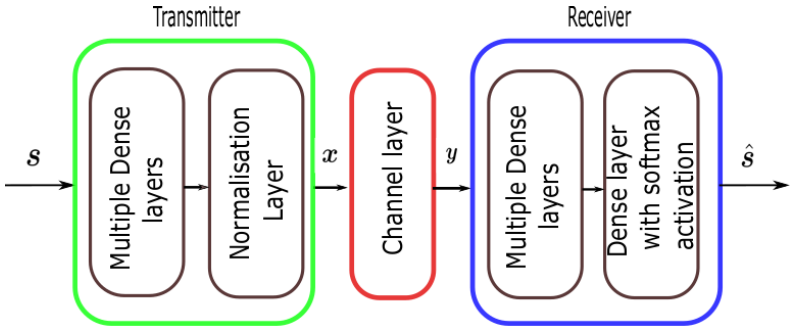

Figure 7.1: An autoencoder-based single-user OWC system.

Layer	Output dimensions
Input	M
Dense + linear	M
Dense + linear	n
Normalization	n
Channel	n
Dense + linear	M
Dense + linear	M

Table 7.1: Layout of the autoencoder used in Figure 7.1.

on the transmitted signal, the normalization layer restricts the elements of the encoded vector $\mathbf{x} \in \mathbb{R}^n$ as $0 \leq \mathbf{x}(i) \leq A$, $i = 1, \ldots, n$, using a weighted sigmoid activation function, i.e., $A \times \text{sigmoid}(\cdot)$, where A is the peak intensity constraint. The communication rate of this OWC system is $R = k/n$ bits/channel use, where $k = \log_2 M$ number of bits are transmitted through n channel use (alternatively, this is denoted by the pair (n, k)). We represent the channel layer by an AWGN with a fixed variance $\sigma^2 = (1/R\rho)$, where ρ is the signal to noise ratio. This channel model is widely used in OWC systems and is considered to be an accurate model in scenarios where the ambient light and the thermal noise are the dominant sources for noise [2, 3, 4]. Finally, the NN receiver decodes the received vector $\mathbf{y} \in \mathbb{R}^n$ and generates the estimate of the transmitted message $\hat{s}$, based on the mapping $\mathbf{h} : \mathbb{R}^n \to \mathcal{M}$. The multiple dense layers at the transmitter (two dense layers at the transmitter) and the receiver (two dense layer at the receiver) use a linear activation function. The last layer at the receiver uses a softmax activation function whose output $\mathbf{p} \in (0, 1)^M$ is a probability vector over all possible messages. Then, the decoded message $\hat{s}$ corresponds to the index of the element of $\mathbf{p}$ with the highest probability.

The structure of the neural networks used in each layer of the considered autoencoder is given in Table 7.1. We train the autoencoder at a fixed value of ρ to optimize the overall BLER performance which is defined as $\Pr\{\hat{s} \neq s\}$. In Section 7.4, we compare the BLER performance of the learning-based OWC system with a model-based OWC system that employs OOK modulations along with hard- and

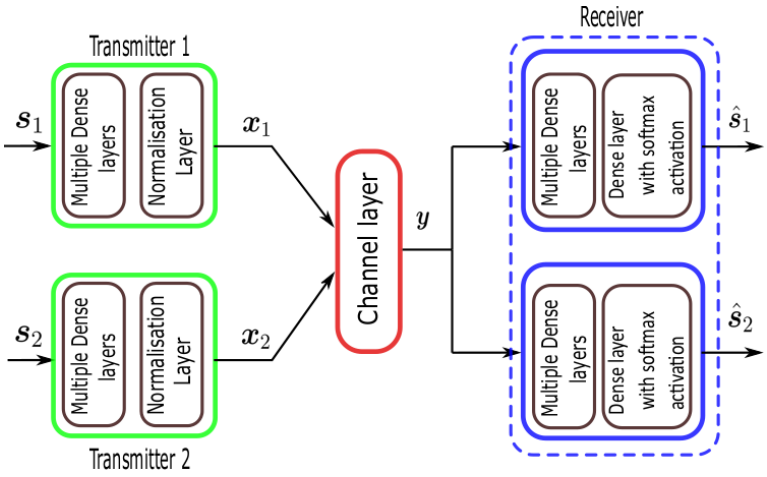

Figure 7.2: Implementation of an optical MAC based on Autoencoders.

soft-decision decoders.

7.3 Multiuser OWC Based on Autoencoders

It is also possible to express a multiuser OWC system, e.g., multiple access channel (MAC), based on autoencoders. Figure 7.2 depicts the schematic of an optical MAC channel constructed by deep neural networks. In this MAC, transmitters 1 and 2 wish to communicate messages $s_1 \in \mathcal{M}$ and $s_2 \in \mathcal{M}$, respectively, to the common receiver over an optical channel subject to nonnegativity and peak intensity constraints. To this end, both of the NN transmitters first encode their messages s_1 and s_2 to vectors $\mathbf{x}_1 \in \mathbb{R}^n$ and $\mathbf{x}_2 \in \mathbb{R}^n$, respectively. Afterwards, the existing normalization layers at each of the transmitters impose the constraints $0 \leq \mathbf{x}_1(i) \leq A_1$ and $0 \leq \mathbf{x}_2(i) \leq A_2$, with $i = 1, \ldots, n$, on the transmitted symbols using weighted sigmoid functions. On the receiver's side the input $\mathbf{y}$ to the NNs is given by

$$\mathbf{y} = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{w}, \tag{7.1}$$

where $\mathbf{w} \sim (\mathbf{0}, \sigma_w^2 \mathbf{I}_n)$ is the zero-mean AWGN component and $\sigma_w^2 = 1/R\rho$ with R as the sum-rate of both transmitters. These coupled autoencoders can be trained to minimize the following loss function

$$L = \max(L_1, L_2), \tag{7.2}$$

where $L_c = -\sum_{i=1}^{K} \mathbf{t}_c(i) \log \mathbf{p}_c(i)$, $c = \{1, 2\}$ are the individual cross-entropy loss functions for the first and second transmitter, respectively; K is the length of the output vector at the last layers of the receiver,

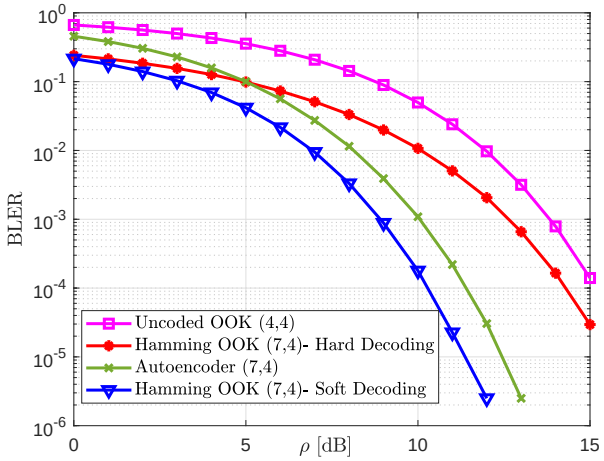

Figure 7.3: The BLER performance of the autoencoder and an OWC system employing OOK modulations with and without Hamming coding for the peak intensity constraint $A = 2$.

$\mathbf{t}_c(i) \in \{0, 1\}$ is the ith target label for the transmitter c, and $0 \leq \mathbf{p}_c(i) \leq 1$ is the ith output of softmax activation functions at the last layers of the receiver.

We note that this min-max problem ensures that $\max\left(\Pr\{s_1 \neq \hat{s}_1\}, \Pr\{s_2 \neq \hat{s}_2\}\right)$ is minimized. This in turn, implies that the receiver is able to decode both messages with a small probability of error. We have observed that the considered min-max problem results in a better BLER performance than the considered minimization of the combined loss functions presented in [51, Sec. III]. Furthermore, from an information-theoretic perspective, this min-max problem is indeed considered as the performance metric for evaluating the reliability of a multiple access setting [9, Ch. 15]. In Section 7.4, we compare the BLER of the learning-based optical MAC with two multiple access systems. The first system employs OOK modulations along with joint decoding and the second system employs OOK modulations along with time-sharing.

7.4 Simulation Results

This section demonstrates the BLER performance of the proposed autoencoder-based single- and multiuser OWC systems and gives a detailed comparison between the obtained results based on the learning-based approach and that of the model-based systems. In the simulations, the structure of all the autoencoders follows the layout given in Table 7.1 and we have used the Stochastic Gradient Descent Algorithm for optimizing the performance of the autoencoders.

In Figure 7.3, we compare the BLER performance of the autoencoder-based OWC system against the

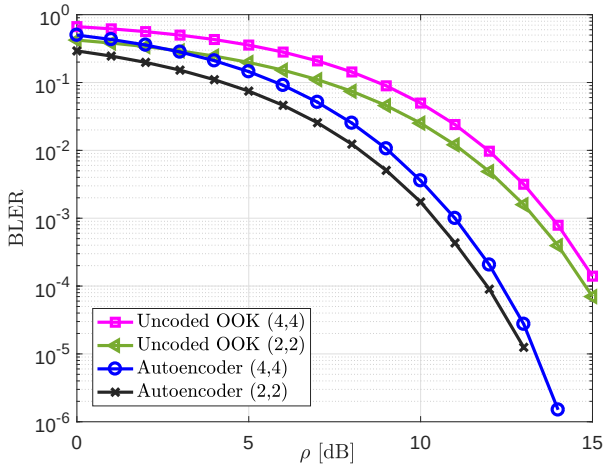

Figure 7.4: The BLER performance of the autoencoder and an OWC system employing uncoded OOK modulations for the peak intensity constraint $A = 2$.

BLER performance of an OWC system employing OOK modulations and a Hamming code with either hard- or soft-decision decoding schemes with a fixed peak intensity constraint $A = 2$ for both systems. We also provide the BLER of the uncoded OOK modulations with maximum likelihood decoder. The results indicate that the autoencoder has learned, without any prior knowledge, encoding and decoding functions that achieve better BLER performance than the hard-decision decoder for $\rho > 5$ dB. Furthermore, the BLER performance of the autoencoder-based OWC system is only 1 dB inferior to that of the soft-decision decoder when ρ exceeds 7 dB. Additionally, we observe that the BLER performance of the autoencoder is better than the BLER performance of the OWC system employing uncoded OOK modulations with maximum likelihood decoder. In our simulations, we have trained the autoencoder at a fixed value of $\rho = 10$ dB using Adam optimizer with the learning rate of 0.001.

In Figure 7.4, we provide a similar BLER comparison for the $(2, 2)$ and $(4, 4)$ OWC systems. We observe that the autoencoder outperforms the OWC system employing OOK modulations for both $(2, 2)$ and $(4, 4)$ cases. Based on this, one can infer that the autoencoder has learned some joint coding and modulation schemes such that a coding gain is achieved.

In Figure 7.5, we plot the learned representations $\mathbf{x}$ of all messages as real constellation points along with their relative frequency of occurrences for different values of (n, k). Surprisingly, in both $(4, 4)$ and $(7, 4)$ autoencoder systems, we observe that the autoencoder learned an OOK modulations with constellation points located at 0 and $A = 2$. For the $(4, 4)$ autoencoder system, both points occur with the same relative frequency (i.e., standard OOK modulations) for representing $M = 2^k$ messages across n channel uses. However, for the $(7, 4)$ autoencoder system, the point at 0 has a higher relative frequency

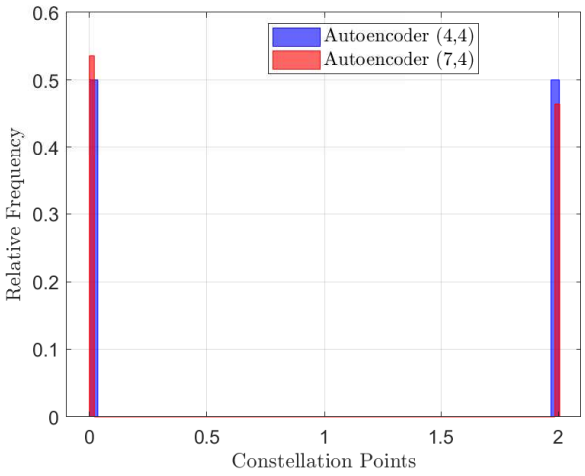

Figure 7.5: Constellation points along with their relative frequency of occurrence generated by the autoencoder for the peak intensity constraint $A = 2$.

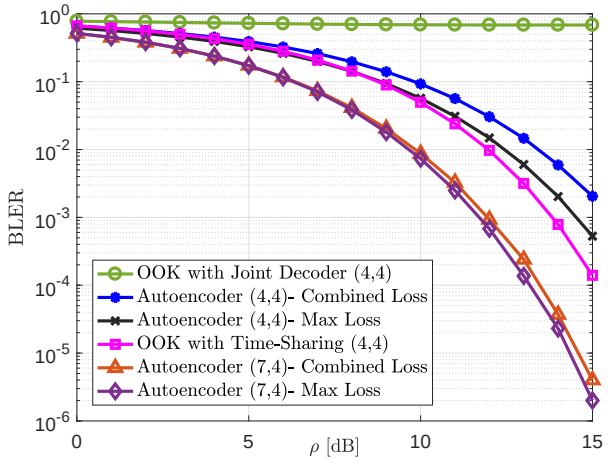

Figure 7.6: BLER versus ρ for the autoencoder-based optical MAC and optical MAC with OOK modulations with joint decoding and time-sharing settings and with peak intensity constraints $A_1 = A_2 = 2$.

of occurrence than the point at $A = 2$ which differs from the standard OOK modulations.

In Figure 7.6, we compare the BLER performance of the optical MAC based on autoencoders against the BLER performance of the optical MAC with OOK modulations along with either joint decoding or time-sharing schemes. First, we observe that, the $(4, 4)$ autoencoder system outperforms the joint decoding system over the full range of ρ. The reason is that with joint decoding, when OOK modulations is used, the receiver always fails to decode the received messages. In particular, when both transmitters send the symbols 0 or $A_1 = A_2$, the receiver fails to distinguish the transmitted symbols and therefore, it cannot do any better than a random guess. However, as mentioned earlier, autoencoders learn through

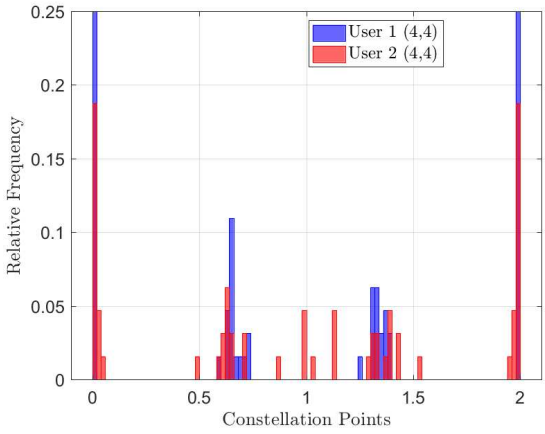

Figure 7.7: Learned constellation points for the $(4, 4)$ autoencoder-based optical multiple access system.

training an efficient coding and representation of the messages which enables them to decode the messages correctly. Next, we see that the performance of the $(4, 4)$ autoencoder-based multiple access system is the same as that of the MAC with time-sharing setting until $\rho = 10$ dB and is only 0.5 dB inferior at $\rho = 15$ dB. Finally, we observe that the autoencoder-based MAC (both $(4, 4)$ and $(7, 4)$ autoencoder systems) optimized by our proposed min-max approach outperforms the autoencoder-based MAC optimized by the minimization of the combined weighted loss functions proposed in [51, Sec. III], where in each mini-batch the weights are updated.

In the simulations for the MAC scenario, we trained the autoencoders at a fixed value of $\rho = 15$ dB using Adam optimizer with the learning rate of 0.0005. It is worth mentioning that in this MAC scenario, R refers to the sum-rate of both users and a symmetric MAC is considered, where each transmitter communicates with the rate k/n bits/channel use and therefore, $R = 2k/n$ bits/channel use.

Finally, in Figures 7.7 and 7.8, we illustrate the learned constellation points of each of the users in the autoencoder-based optical MAC for different communications rates. It is interesting to observe that while in the single-user case, the learned constellation points for the $(4, 4)$ autoencoder system are located at 0 and $A = 2$ with equal relative frequency, in the MAC setting, the constellation points of the users, shown in Figure 7.7, are scattered in the interval $[0, 2]$ with different relative frequencies. A similar observation can be made for the $(7, 4)$ autoencoder system depicted in Figure 7.8. These results indicate that the autoencoders successfully learned efficient coding, modulation and decoding schemes in a multiple access scenario.

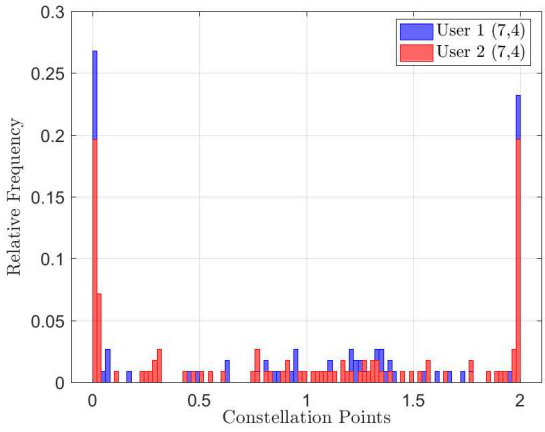

Figure 7.8: Learned constellation points for the $(7, 4)$ autoencoder-based optical multiple access system.

7.5 Conclusions

Since the transmitted signals in an OWC setting must satisfy nonnegativity, peak- and average-intensity constraints, traditional approaches used in constructing the signal constellations for RF channels cannot be applied directly to the optical channels. Therefore, one should consider designing a structured optical signal-space model that can capture all the physical restrictions in an OWC setting. This task is not straightforward and heavily depends on the considered optical channel model. Hence, seeking communications techniques (such as modulation, coding, decoding, etc.) that do not heavily depend on an existing channel model is quite appealing. Motivated by the success of learning-based autoencoders in capturing the end-to-end performance of the RF communications system, this chapter proposed the design of the OWC systems in single-user and multiuser scenarios based on the autoencoders. In particular, a simple and cost-effective autoencoder system with (near-)optimal performance is proposed and is implemented by merely taking off-the-shelf deep learning models, applying them to an OWC design problem, and tuning them based on the easily generated training data. According to the obtained results, using autoencoders for both single- and multiuser OWC scenarios can lead to a reliable OWC systems. The comparison of the end-to-end BLER performance of the designed and trained autoencoders in both single- and multiuser OWC scenarios against several baseline model-based OWC systems indicated that the autoencoders are able to learn efficient encoding, modulation and decoding functions, and in some cases can outperform the baseline model-based systems in terms of the BLER performance. Therefore, one can conclude that autoencoders can be a promising solution for OWC system where a precise channel model and efficient communications techniques, such as coding, modulations and decoding are not available.

Chapter 8: Conclusions and Future Work

8.1 Conclusions

This dissertation studies the fundamental performance limits of the multiuser OWC system with and without secrecy constrains. It presents techniques of secure and reliable system design for OWC for the input-dependent Gaussian noise and the Poisson noise models. Major contributions are presented in the following areas: 1) characterization of the optimal input distributions attaining the secrecy capacity and the entire boundary of the rate-equivocation region of an input-dependent Gaussian noise optical wiretap channel when peak- and average-intensity constraints are active; 2) asymptotic analysis of the secrecy capacity of an input-dependent Gaussian noise optical wiretap channel in the low- and high-intensity regimes when peak- and average-intensity constraints are active; 3) characterization of the optimal input distributions achieving the entire boundary of the rate-equivocation region of the input-dependent Gaussian noise optical wiretap channel when only an average-intensity constraint is active; 4) characterization of the optimal secure transmission schemes for the discrete-time Poisson optical wiretap channel with peak- and/or average-intensity constraints; 5) asymptotic analysis of the secrecy capacity of the discrete-time Poisson wiretap channel; 6) characterization of the optimal input distributions exhausting the entire capacity region of the optical multiple access channel with an input-dependent Gaussian noise and deriving closed-form expression of the capacity region in the low-intensity regime; 7) proposal of deep neural network autoencoders for designing reliable single-user and multiuser OWC systems.

The secrecy-capacity-achieving input distribution of the input-dependent Gaussian noise optical wiretap channel is shown to be discrete with a finite number of mass points when peak- and average-intensity constraints are active. Moreover, the entire rate-equivocation region of the considered wiretap channel is also obtained by discrete input distributions with a finite support set. Finally, the asymptotic behavior of the secrecy capacity in the low- and high-intensity regimes is analyzed. In the low-intensity regime, the secrecy capacity scales quadratically with the peak-intensity constraint and it is achieved by a binary distribution. On the other hand, in the high-intensity regime, the secrecy capacity does not scale with the constraint and hence, it is constant.

When only nonnegativity and average-intensity constraints are considered, the entire boundary of the rate-equivocation of the input-dependent Gaussian noise optical wiretap channel is achieved by discrete input distributions with countably infinite support set, but with finitely many mass points in any bounded interval. This implies that when the transmitted optical signals are restricted by only an average-intensity constraint: 1) the secrecy capacity is achieved by a distribution which has a countably infinite support

set; 2) the channel capacity is also achieved by a distribution having a countably infinite support set.

Since the Poisson noise model is the most accurate model for the underlying OWC based on IM-DD, studying the fundamental performance limits of such a model is of great importance. To this end, a discrete-time Poisson wiretap channel subject to nonnegativity, peak- and/or average-intensity, as well as bandwidth constraints, is considered. It is shown that every point on the boundary of the rate-equivocation region of this wiretap channel is obtained by a discrete input distribution with finitely many mass points. Additionally, the analysis is extended to the case where only an average-intensity constraint is active. In this case, it is found that the boundary of the rate-equivocation region is achieved by discrete distributions with a countably infinite number of mass points, but with finitely many mass points in any bounded interval. Finally, asymptotic analysis for characterizing the behavior of the secrecy capacity in the low- and high-intensity regimes is provided. It is observed in the low-intensity regime, the secrecy capacity scales quadratically with the peak-intensity constraint. However, in the high-intensity regime and when the legitimate receiver's and the eavesdropper's channel gains are identical, the secrecy capacity is constant. Moreover, when the channel gains are different, the secrecy capacity cannot scale faster than the logarithm of the square root of the peak-intensity constraint.

The capacity region of a two-user input-dependent Gaussian noise optical multiple access channel is considered. It is established that under nonnegativity, peak- and average-intensity constraints, generating code-books of both users according to discrete distributions with finitely many mass points achieve any point on the boundary of the capacity region. Furthermore, an asymptotic analysis of the capacity region is conducted in the low-intensity regime, where the capacity region is explicitly presented in a closed-form expression and it is shown that binary distributions with mass points at the origin and the peak-intensity constraint are optimal. Numerical results indicate that due to the existence of an input-dependent noise component, the geometry of the capacity region under nonnegativity, peak- and average-intensity constraints is not a pentagon as opposed to the case of the Gaussian multiple access channel with peak- and/or average-power constraints.

Since the transmitted signals in OWC must satisfy nonnegativity, peak- and average-intensity constraints due to the physical restrictions existing in the optical wireless channels, traditional approaches used in constructing the signal constellations for RF channels cannot be applied directly to the optical channels. Therefore, one should consider designing a structured optical signal-space model that can capture all the physical restrictions in an OWC setting. This task is not straightforward and heavily depends on the considered optical channel model. Hence, seeking communications techniques (such as modulation, coding, decoding, etc.) that do not heavily depend on an existing channel model is quite appealing. Motivated by the success of learning-based autoencoders in capturing the end-to-end performance of the

RF communications system, this dissertation proposes the design of the OWC systems in single-user and multiuser scenarios based on the autoencoders. In particular, a simple and cost-effective learning-based system with (near-)optimal performance is proposed and is implemented by merely taking off-the-shelf deep learning models, applying them to an OWC design problem, and tuning them based on the easily generated training data. According to the obtained results, the learning-based OWC can perform as well as the model-based counterpart.

8.2 Future Work

In Chapter 3, although the secrecy capacity is shown to be constant in the high-intensity regime, the value of the constant is not fully determined. Therefore, it would be of interest to fully characterize the secrecy capacity in this regime and to determine the constant. The reason is this constant is the maximum possible secrecy capacity that can be achieved in an input-dependent Gaussian noise setting and therefore, finding it helps to evaluate the potential of secure OWC over the optical wiretap channel with an input-dependent Gaussian noise.

In Chapter 4, the optimal input distributions are proved to be discrete with an infinite number of mass points, but with finitely many mass points in any bounded interval. This implies that numerical computations of the boundary of the rate-equivocation region of the wiretap channel with an input-dependent Gaussian noise under an average-intensity constraint is not feasible. Therefore, to evaluate the secrecy performance of an OWC in such a setting, it is of great importance to provide inner and outer bounds on the rate-equivocation region based on discrete distributions with a finite number of mass points. These inner and outer bounds help to characterize near-optimal secure transmission schemes for OWC systems operating over the optical wiretap channel with an input-dependent Gaussian noise when only an average-intensity constraint is active.

In Chapter 5, the behavior of the secrecy capacity in the high-intensity regime is analyzed only through providing loose upper bounds, and full characterization of the secrecy capacity that results in a closed-form expression of the secrecy capacity is missing. Therefore, in order to fully understand the potential of secure OWC over the discrete-time Poisson wiretap channel in the high-intensity regime, tight inner and upper bounds need to be sought.

In Chapter 6, the boundary of the capacity region is shown to be achieved by discrete input distributions with finitely many mass points when both peak- and average-intensity constraints are active. However, it is not known whether the support set of the optimal distributions remains finite for the case when only an average-intensity constraint is imposed. Furthermore, the capacity region is only charac-

terized in a closed-form expression in the low-intensity regime and it is shown that binary distributions achieve the boundary. It would be of great interest to find a closed-form expression of the boundary of the capacity region in the high-intensity regime and to characterize the optimal distributions in this regime.

The proposed learning-based autoencoders in Chapter 7 can capture the end-to-end performance of both single-user and multiuser OWC systems and can perform as good as the model-based OWC system. In this chapter, the considered channel model for the underlying OWC is the free-space optical channel which is not an accurate channel model for most of the OWC scenarios. Therefore, other channel models, such as the input-dependent Gaussian noise model, or the Poisson optical model should be considered when designing these learning-based autoencoders. Furthermore, the only considered multiuser scenario in this chapter is the multiple access channel. Therefore, to further investigate the potential of designing learning-based autoencoders for OWC systems, it is necessary to design, build, and train the autoencoders for other multiuser settings, such as a broadcast channel, an interference channel, etc., and compare their performance with the model-based multiuser OWC scenarios.

References

[1] H. Hemmati, *Deep Space Optical Communications*, ser. JPL Deep-Space Communications and Navigation Series. Wiley, 2006.

[2] S. Arnon, J. Barry, G. Karagiannidis, R. Schober, and M. Uysal, *Advanced optical wireless communication systems*, 1st ed. New York, NY, USA: Cambridge University Press, 2012.

[3] A. Lapidoth, S. M. Moser, and M. A. Wigger, "On the capacity of free-space optical intensity channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4449–4461, Oct. 2009.

[4] S. Hranilovic and F. R. Kschischang, "Optical intensity-modulated direct detection channels: signal space and lattice codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 6, pp. 1385–1399, 2003.

[5] S. M. Moser, "Capacity results of an optical intensity channel with input-dependent Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 207–223, Jan. 2012.

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

[7] S. Shamai, "Capacity of a pulse amplitude modulated direct detection photon channel," *IEE Proceedings I - Commun., Speech and Vision*, vol. 137, no. 6, pp. 424–430, Dec. 1990.

[8] 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.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York, NY, USA: Wiley-Interscience, 1991.

[10] 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.

[11] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, no. 4, pp. 656–715, 1949.

[12] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Jan. 1975.

[13] M. Bloch and J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering.* Cambridge University Press, 2011.

[14] O. Ozel, E. Ekrem, and S. Ulukus, "Gaussian wiretap channel with amplitude and variance constraints," *IEEE Trans. Inf. Theory*, vol. 61, no. 10, pp. 5553–5563, Oct. 2015.

[15] A. Laourine and A. B. Wagner, "The degraded Poisson wiretap channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 12, pp. 7073–7085, Dec 2012.

[16] J. Cao, J. Chen, and S. Hranilovic, "Discreteness of sum-capacity-achieving distributions for discrete-time poisson multiple access channels with peak constraints," *IEEE Commun. Lett.*, vol. 17, no. 8, pp. 1644–1647, Aug. 2013.

[17] A. Lapidoth and S. Shamai, "The poisson multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 488–501, Mar. 1998.

[18] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel. i," *IEEE Trans. Inf. Theory*, vol. 34, no. 6, pp. 1449–1461, Nov. 1988.

[19] J. Fahs and I. Abou-Faycal, "On properties of the support of capacity-achieving distributions for additive noise channel models with input cost constraints," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 1178–1198, Feb. 2018.

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

[21] A. Chaaban, J. M. Morvan, and M. S. Alouini, "Free-space optical communications: Capacity bounds, approximations, and a new sphere-packing perspective," *IEEE Trans. Commun.*, vol. 64, no. 3, pp. 1176–1191, Mar. 2016.

[22] B. Mamandipoor, K. Moshksar, and A. K. Khandani, "Capacity-achieving distributions in Gaussian multiple access channel with peak power constraint," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6080–6092, Oct. 2014.

[23] A. Chaaban, O. M. S. Al-Ebraheemy, T. Y. Al-Naffouri, and M. S. Alouini, "Capacity bounds for the Gaussian IM-DD optical multiple-access channel," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3328–3340, May 2017.

[24] J. Zhou and W. Zhang, "Bounds on the capacity region of the optical intensity multiple access channel," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7629–7641, Nov. 2019.

[25] M. Soltani and Z. Rezki, "Optical wiretap channel with input-dependent Gaussian noise under peak- and average-intensity constraints," *IEEE Trans. Inf. Theory*, vol. 64, no. 10, pp. 6878–6893, Oct 2018.

[26] M. Soltani and Z. Rezki, "Discrete-time Poisson optical wiretap channel with peak intensity constraint," in *Proc. IEEE Int. Symp. on Inf. Theory*, Jul. 2019, pp. 136–140.

[27] M. Soltani, Z. Rezki, and A. Chaaban, "Sum-capacity-achieving distributions in the input-dependent Gaussian noise optical multiple access channel with peak and average intensity constraints," *2019 16th Canadian Workshop on Information Theory (CWIT)*, pp. 1–6, 2019.

[28] M. Soltani, W. Fatnassi, A. Aboutaleb, Z. Rezki, A. Bhuyan, and P. Titus, "Autoencoder-based optical wireless communications systems," *2018 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, 2018.

[29] Y. Liang, H. V. Poor, and S. Shamai, "Information theoretic security," *Foundations and Trends in Commun. and Inf. Theory*, vol. 5, no. 4-5, pp. 355–580, 2009.

[30] S. Leung-Yan-Cheong and M. Hellman, "The Gaussian wiretap channel," *IEEE Trans. Inf. Theory*, vol. 24, no. 4, pp. 451–456, Jul. 1978.

[31] Z. Wang, Q. Wang, W. Huang, and Z. Xu, *Visible light communications: modulation and signal processing*, 1st ed. Piscataway, NJ, USA: Wiley and Sons, Incorporated, John, Nov. 2017.

[32] V. V. Prelov and E. C. van der Meulen, "An asymptotic expression for the information and capacity of a multidimensional channel with weak input signals," *IEEE Trans. Inf. Theory*, vol. 39, no. 5, pp. 1728–1735, Sep. 1993.

[33] C. Luo., *Communication for Wideband Fading Channels: On Theory and Practice.* PhD Thesis, Massachusetts Institute of Technology, Feb. 2006.

[34] I. Csiszar and J. Korner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, 1978.

[35] A. Dytso, M. Egan, S. M. Perlaza, H. V. Poor, and S. S. Shitz, "Optimal inputs for some classes of degraded wiretap channels," in *Proc. IEEE Information Theory Workshop*, Nov. 2018.

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

[37] I. C. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290–1301, May 2001.

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

[39] D. G. Luenberger, *Optimization by Vector Space Methods*, 1st ed. USA: John Wiley and Sons, Inc., 1997.

[40] A. Lapidoth, I. E. Telatar, and R. Urbanke, "On wide-band broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3250–3258, Dec. 2003.

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

[42] M. Safari, "Efficient optical wireless communication in the presence of signal-dependent noise," *2015 IEEE International Conference on Communication Workshop (ICCW)*, pp. 1387–1391, Jun. 2015.

[43] B. M. Ghaffari, M. D. Matinfar, and J. A. Salehi, "Wireless optical cdma lan: digital design concepts," *IEEE Trans. Wireless Commun.*, vol. 56, no. 12, pp. 2145–2155, Dec. 2008.

[44] U. N. Griner and S. Arnon, "Multiuser diffuse indoor wireless infrared communication using equalized synchronous cdma," *IEEE Trans. Commun.*, vol. 54, no. 9, pp. 1654–1662, Sep. 2006.

[45] A. Chaaban, Z. Rezki, and M. S. Alouini, "On the capacity of the intensity-modulation direct-detection optical broadcast channel," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3114–3130, May 2016.

[46] H. Kim, B. Nachman, and A. E. Gamal, "Superposition coding is almost always optimal for the poisson broadcast channel," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1782–1794, Apr. 2016.

[47] Z. Zhang and A. Chaaban, "On the capacity of the two-user im/dd interference channel," in *Canadian Wrkshp on Inf. Theory (CWIT)*, Jun. 2019, pp. 1–6.

[48] L. Lai, Y. Liang, and S. S. Shitz, "On the capacity bounds for poisson interference channels," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 223–238, Jan. 2015.

[49] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[50] A. Fehske, J. Gaeddert, and J. H. Reed, "A new approach to signal classification using spectral correlation and neural networks," *First IEEE Int. Symp. on New Frontiers in Dynamic Spectrum Access Networks*, pp. 144–150, Nov. 2005.

[51] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[52] H. Kim, Y. Jiang, R. Rana, S. Kannan, S. Oh, and P. Viswanath, "Communication algorithms via deep learning," *The International Zurich Seminar on Inf. and Commun. (IZS 2018)*, pp. 48–50, 2018.

[53] S. Dorner, S. Cammerer, J. Hoydis, and S. t. Brink, "Deep learning based communication over the air," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 132–143, Feb. 2018.

[54] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in ofdm systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Sep. 2018.

[55] H. Hemmati, A. Biswas, and I. B. Djordjevic, "Deep-space optical communications: Future perspectives and applications," *Proceedings of the IEEE*, vol. 99, no. 11, pp. 2020–2039, Nov. 2011.

[56] R. M. Dudley, *Real Analysis and Probability.* Cambridge University Press, vol. 74, 2002.

[57] A. Khisti, A. Tchamkerten, and G. W. Wornell, "Secure broadcasting over fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2453–2469, Jun. 2008.

[58] J. Fahs and I. Abou-Faycal, "On the finiteness of the capacity of continuous channels," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 166–173, Jan. 2016.

[59] A. Elmoslimany and T. M. Duman, "On the discreteness of capacity-achieving distributions for fading and signal-dependent noise channels with amplitude-limited inputs," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 1163–1177, Feb. 2018.

[60] C. Luo., *Communication for Wideband Fading Channels: On Theory and Practice.* PhD Thesis, Massachusetts Institute of Technology, Feb. 2006.

Appendix A: Proof of the Main Results in Chapter 3

In this section, we first provide the required preliminaries for the development of the main results. We then give the detailed proofs of the theorems stated in Section 3.3.

A.1 Preliminaries and Notation

Since both channels are AWGN with input-dependent noise, the output densities for Y and Z exist for any input distribution F_X, and are given by

$$p_Y(y;F_X) = \int_0^A p(y|x)\, dF_X(x),\ y \in \mathbb{R}, \tag{A.1}$$

$$p_Z(z;F_X) = \int_0^A p(z|x)\, dF_X(x),\ z \in \mathbb{R}, \tag{A.2}$$

where $p(y|x)$ and $p(z|x)$ are given by [5]

$$p(y|x) = \frac{1}{\sqrt{2\pi\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x)}} \exp\left(-\frac{(y-x)^2}{2\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x)}\right), \tag{A.3}$$

$$p(z|x) = \frac{1}{\sqrt{2\pi\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x)}} \exp\left(-\frac{(z-x)^2}{2\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x)}\right). \tag{A.4}$$

We define the rate-equivocation density $r_{\mathrm{eq}}(x;F_X)$ as

$$r_{\mathrm{eq}}(x;F_X) = i_{\mathrm{B}}(x;F_X) - i_{\mathrm{E}}(x;F_X), \tag{A.5}$$

where $i_{\mathrm{B}}(x;F_X)$ and $i_{\mathrm{E}}(x;F_X)$ are the mutual information densities for the legitimate user's and eavesdropper's channel, respectively, and are given by

$$i_{\mathrm{B}}(x;F_X) = -\int_{\mathbb{R}} p(y|x) \log\left(p_Y(y;F_X)\right)\, dy - \frac{1}{2}\log(2\pi e\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x)), \tag{A.6}$$

$$i_{\mathrm{E}}(x;F_X) = -\int_{\mathbb{R}} p(z|x) \log\left(p_Z(z;F_X)\right)\, dz - \frac{1}{2}\log(2\pi e\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x)). \tag{A.7}$$

The mutual information and the mutual information density are related through

$$I(X;Y) = \int_0^A i_{\mathrm{B}}(x;F_X)\, dF_X(x), \tag{A.8}$$

$$I(X;Z) = \int_0^A i_{\mathrm{E}}(x;F_X)\, dF_X(x). \tag{A.9}$$

Since the channel input X satisfies (3.5), it can be shown that the conditional densities in (A.3) and (A.4) can be bounded as [10, Lemma 3]

$$\exp(-\alpha - \beta' y^2) \leq p(y|x) \leq \exp(\alpha - \beta y^2), \tag{A.10}$$

$$\exp(-\mu - \xi' z^2) \leq p(z|x) \leq \exp(\mu - \xi z^2), \tag{A.11}$$

for all $x \in [0, A], y, z \in \mathbb{R}$, where $\alpha, \beta, \beta', \mu, \xi$ and ξ' are positive constants. Hence, for all $F_X \in \mathcal{A}^+$

$$\exp(-\alpha - \beta' y^2) \leq p_Y(y;F_X) \leq \exp(\alpha - \beta y^2), \tag{A.12}$$

$$\exp(-\mu - \xi' z^2) \leq p_Z(z;F_X) \leq \exp(\mu - \xi z^2). \tag{A.13}$$

Thus, we can write

$$|\log\left(p_Y(y;F_X)\right)| \leq \alpha + \beta' y^2, \tag{A.14}$$

$$|\log\left(p_Z(z;F_X)\right)| \leq \mu + \xi' z^2. \tag{A.15}$$

Next, we prove Theorem 1 using the preliminaries provided in this section.

A.2 Proof of Theorem 1

A.2.1 The feasible set $\mathcal{A}^+$ is compact and convex

The proof follows along similar lines as in [33, Appendix A.1].

A.2.2 $g_0(F_X)$ is continuous in F_X

In order to show that $g_0(F_X)$ is a continuous function in F_X, it is sufficient to show that $I(X;Y)$ is continuous in F_X. The continuity of $I(X;Z)$ in F_X can be shown by following along similar lines as those in the proof of the continuity of $I(X;Y)$ in F_X. To this end, let us consider a sequence $\{F_X^{(n)}\}_{n\in\mathbb{N}}$

in $\mathcal{A}^+$ such that $F_X^{(n)} \to F_X$ for some $F_X \in \mathcal{A}^+$. It is evident that $p(y|x)$ is a continuous and bounded function in x and y, thus,

$$\begin{aligned}\lim_{n\to\infty} p_Y(y; F_X^{(n)}) &= \lim_{n\to\infty} \int_0^A p(y|x)\, dF_X^{(n)}(x) \\ &= \int_0^A p(y|x)\, dF_X(x), \end{aligned} \tag{A.16}$$

where (A.16) follows by the Helly-Bray Theorem [56]. Then,

$$\lim_{n\to\infty} p(y|x) \log\left(p_Y(y; F_X^{(n)})\right) = p(y|x) \log\left(p_Y(y; F_X)\right). \tag{A.17}$$

Moreover, by observing (A.10) and (A.14), we conclude that

$$\left|p(y|x) \log\left(p_Y(y; F_X^{(n)})\right)\right| \leq \exp(\alpha - \beta y^2)[\alpha + \beta' y^2]. \tag{A.18}$$

Since the right hand side of (A.18) is absolutely integrable, we have

$$\int_{-\infty}^{+\infty} \left|p(y|x) \log\left(p_Y(y; F_X^{(n)})\right)\right| \, dy < \infty. \tag{A.19}$$

Thus, by applying the Dominated Convergence Theorem, we get

$$\begin{aligned}\lim_{n\to\infty} -\int_{\mathbb{R}} p(y|x) \log\left(p_Y(y; F_X^{(n)})\right) \, dy &= -\int_{\mathbb{R}} \lim_{n\to\infty} p(y|x) \log\left(p_Y(y; F_X^{(n)})\right) \, dy \\ &= -\int_{\mathbb{R}} p(y|x) \log(p_Y(y; F_X))\, dy. \end{aligned} \tag{A.20}$$

Additionally, $\frac{1}{2}\log(2\pi e\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x))$ is a bounded and continuous function for all $x \in [0, A]$. Therefore, we conclude that $i_{\mathrm{B}}(x; F_X)$ is a bounded and continuous function in F_X. Finally, applying the Helly-Bray Theorem results in

$$\lim_{n\to\infty} \int_0^A i_{\mathrm{B}}(x; F_X)\, dF_X^{(n)}(x) = \int_0^A i_{\mathrm{B}}(x; F_X)\, dF_X(x), \tag{A.21}$$

which implies that $I(X;Y)$ is continuous in F_X. Similar steps lead to the fact that $I(X;Z)$ is also a continuous function in F_X. This further implies that the objective function $g_0(F_X)$ is continuous in F_X.

A.2.3 $g_0(F_X)$ IS STRICTLY CONCAVE IN F_X

To show that $g_0(F_X)$ is a strictly concave function in F_X, we first note that $g_0(F_X) = I(X;Y|Z)$ when random variables X, Y and Z form the Markov chain $X \to Y \to Z$. Next, we present a lemma

that establishes that $I(X;Y|Z)$ is a strictly concave function in F_X.

Lemma 4. *If the random variables X, Y and Z form the Markov chain $X \to Y \to Z$, then the conditional mutual information $I(X;Y|Z)$ is a strictly concave function in input distribution F_X. Furthermore, the output distributions are unique, i.e., if F_{X_1} and F_{X_2} are both secrecy-capacity-achieving, then $p_Y(y;F_{X_1}) = p_Y(y;F_{X_2})$ and $p_Z(z;F_{X_1}) = p_Z(z;F_{X_2})$.*

Proof. We start the proof by noting that for random variables X, Y and Z that form the Markov chain $X \to Y \to Z$, $I(X;Y|Z)$ is a concave function in F_X [57, Appendix A]. Now, let X_1 and X_2 be two channel inputs generated by F_{X_1} and F_{X_2}, respectively, and Q be a binary-valued random variable such that

$$p(y,z,x|q) = \begin{cases} p(y,z|x)\, p_{X_1}(x), & q = 1, \\ p(y,z|x)\, p_{X_2}(x), & q = 2, \end{cases} \tag{A.22}$$

where $p_{X_1}(x)$ and $p_{X_2}(x)$ be the probability density functions (PDF) of the random variables X_1 and X_2. Based on (A.22), we have the following Markov chain

$$Q \to X \to Y \to Z. \tag{A.23}$$

Following along the same lines as [57, Appendix A], one can show that

$$I(X;Y|Z,Q) - I(X;Y|Z) = -I(Q;Y|Z). \tag{A.24}$$

Since $I(Q;Y|Z) \geq 0$, $I(X;Y|Z,Q) \leq I(X;Y|Z)$. This implies that $I(X;Y|Z)$ is a concave function in F_X. Now, we prove that with the Markov chain $Q \to X \to Y \to Z$, $I(X;Y|Z)$ is strictly concave in F_X, i.e., $I(Q;Y|Z) > 0$. Assume, to the contrary, that $I(Q;Y|Z) = 0$. This implies that random variables Q, Y and Z also form the Markov chain

$$Q \to Z \to Y. \tag{A.25}$$

Furthermore, from the Markov chain (A.23), we have

$$Q \to X \to Z. \tag{A.26}$$

Combining Markov chains (A.25) and (A.26) results in a new Markov chain given by

$$Q \to X \to Z \to Y. \tag{A.27}$$

Now, based on (A.23) and (A.27), we obtain the following

$$\begin{aligned} p(y,z,x)\big|_{\text{Markov chain (A.23)}} &= p(y,z,x)\big|_{\text{Markov chain (A.27)}} \\ p_X(x)\,p(y|x)\,p(z|y) &= p_X(x)\,p(z|x)\,p(y|z) \\ \frac{p(y|x)}{p(z|x)} &= \frac{p(y|z)}{p(z|y)}. \end{aligned} \tag{A.28}$$

We note that (A.28) holds for any $y, z \in \mathbb{R}$ and $x \in \mathcal{S}_{F_X}$, where $\mathcal{S}_{F_X}$ is the support set of F_X. As a result, for fixed values of y and z the right hand side (RHS) of (A.29) is fixed, while the left hand side (LHS) is a function of x. Since $Y|X \sim \mathcal{N}(x, \sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x))$ and $Z|X \sim \mathcal{N}(x, \sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x))$, (A.28) reduces to

$$\sqrt{\frac{\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x)}{\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x)}}\exp\left(\frac{(z-x)^2}{2\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x)} - \frac{(y-x)^2}{2\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x)}\right) = \frac{p(y|z)}{p(z|y)}. \tag{A.29}$$

To reach a contradiction, let us choose $y = z = 0$. For the contradiction, it is sufficient to show that the LHS of (A.29) is not a constant function in x. To this end, let us denote the LHS of (A.29) for $y = z = 0$ as $f(x)$. We show that $\frac{d[\log(f(x))]}{dx} < 0$ for all $x \in \mathcal{S}_{F_X}$[1]. The derivate of $\log(f(x))$ is given by

$$\begin{aligned} \frac{d[\log(f(x))]}{dx} =& \frac{\sigma_{\mathrm{E}}^2\eta_{\mathrm{E}}^2}{2}\left[\frac{(\sigma_{\mathrm{B}}^2-\sigma_{\mathrm{E}}^2)}{(\sigma_{\mathrm{E}}^2+\sigma_{\mathrm{E}}^2\eta_{\mathrm{E}}^2 x)(\sigma_{\mathrm{B}}^2+\sigma_{\mathrm{B}}^2\eta_{\mathrm{B}}^2 x)}\right] + x\left[\frac{(\sigma_{\mathrm{B}}^2-\sigma_{\mathrm{E}}^2)}{(\sigma_{\mathrm{E}}^2+\sigma_{\mathrm{E}}^2\eta_{\mathrm{E}}^2 x)(\sigma_{\mathrm{B}}^2+\sigma_{\mathrm{B}}^2\eta_{\mathrm{B}}^2 x)}\right] \\ &+ \frac{x^2\,\sigma_{\mathrm{E}}^2\eta_{\mathrm{E}}^2}{2}\left[\frac{(\sigma_{\mathrm{E}}^4-\sigma_{\mathrm{B}}^4)+2\sigma_{\mathrm{E}}^2\eta_{\mathrm{E}}^2(\sigma_{\mathrm{E}}^2-\sigma_{\mathrm{B}}^2)x}{(\sigma_{\mathrm{E}}^2+\sigma_{\mathrm{E}}^2\eta_{\mathrm{E}}^2 x)^2(\sigma_{\mathrm{B}}^2+\sigma_{\mathrm{B}}^2\eta_{\mathrm{B}}^2 x)^2}\right] \\ =& \frac{\sigma_{\mathrm{E}}^2\eta_{\mathrm{E}}^2}{2}\left[\frac{(\sigma_{\mathrm{B}}^2-\sigma_{\mathrm{E}}^2)}{(\sigma_{\mathrm{E}}^2+\sigma_{\mathrm{E}}^2\eta_{\mathrm{E}}^2 x)(\sigma_{\mathrm{B}}^2+\sigma_{\mathrm{B}}^2\eta_{\mathrm{B}}^2 x)}\right] + \frac{x^2\sigma_{\mathrm{E}}^2\eta_{\mathrm{E}}^2(\sigma_{\mathrm{B}}^4-\sigma_{\mathrm{E}}^4)+2x\sigma_{\mathrm{B}}^2\sigma_{\mathrm{E}}^2(\sigma_{\mathrm{B}}^2-\sigma_{\mathrm{E}}^2)}{2(\sigma_{\mathrm{E}}^2+\sigma_{\mathrm{E}}^2\eta_{\mathrm{E}}^2 x)(\sigma_{\mathrm{B}}^2+\sigma_{\mathrm{B}}^2\eta_{\mathrm{B}}^2 x)}. \end{aligned} \tag{A.30}$$

Now, we note that since $\sigma_{\mathrm{E}}^2 > \sigma_{\mathrm{B}}^2$ and x is nonnegative (as x must satisfy the nonnegativity constraint), (A.30) is strictly negative for all $x \in \mathcal{S}_{F_X}$ and consequently, $\frac{df(x)}{dx} < 0$ for all $x \in \mathcal{S}_{F_X}$. This implies that for $y = z = 0$, $f(x)$ is not a constant function of x, which is a contradiction. This, in turn, implies that $I(Q;Y|Z) > 0$ and as a result, $I(X;Y|Z)$ is strictly concave in F_X. Furthermore, the output distributions are unique, i.e., if F_{X_1} and F_{X_2} are both secrecy-capacity-achieving, then $p_Y(y;F_{X_1}) = p_Y(y;F_{X_2})$ and $p_Z(z;F_{X_1}) = p_Z(z;F_{X_2})$. ■

[1]We note that for $y = z = 0$, $f(x) > 0$ for all $x \in \mathcal{S}_{F_X}$ and as a result, the sign of the derivative of $\frac{df(x)}{dx}$ is the same as that of $\frac{d[\log(f(x))]}{dx}$.

A.2.4 $g_0(F_X)$ IS WEAKLY DIFFERENTIABLE

Defining $F_{X_\theta} = (1-\theta)\, F_{X_0} + \theta F_X,\ \forall F_X \in \mathcal{A}^+,\ \theta \in [0,1]$, we have to show that the following limit exists

$$\lim_{\theta \to 0} \frac{g_0((1-\theta)\, F_{X_0} + \theta F_X) - g_0(F_{X_0})}{\theta}. \tag{A.31}$$

Substituting (A.6) and (A.7) into (A.31), we get

$$\begin{aligned}
&\lim_{\theta \to 0} \left[\frac{\int_0^A i_{\mathrm{B}}(x; F_{X_\theta})\, dF_{X_\theta} - \int_0^A i_{\mathrm{E}}(x; F_{X_\theta})\, dF_{X_\theta}}{\theta} + \frac{\int_0^A i_{\mathrm{E}}(x; F_{X_0})\, dF_{X_0} - \int_0^A i_{\mathrm{B}}(x; F_{X_0})\, dF_{X_0}}{\theta} \right] \\
&= \lim_{\theta \to 0} \left[\frac{(1-\theta) \int_0^A i_{\mathrm{B}}(x; F_{X_\theta})\, dF_{X_0} + \theta \int_0^A i_{\mathrm{B}}(x; F_{X_\theta})\, dF_X}{\theta} \right. \\
&\quad \left. - \frac{(1-\theta) \int_0^A i_{\mathrm{E}}(x; F_{X_\theta})\, dF_{X_0} + \theta \int_0^A i_{\mathrm{E}}(x; F_{X_\theta})\, dF_X}{\theta} - \frac{\int_0^A i_{\mathrm{B}}(x; F_{X_0})\, dF_{X_0} - \int_0^A i_{\mathrm{E}}(x; F_{X_0})\, dF_{X_0}}{\theta} \right] \\
&= \lim_{\theta \to 0} \left[\frac{\int_0^A \left[i_{\mathrm{B}}(x; F_{X_\theta}) - i_{\mathrm{B}}(x; F_{X_0}) \right]\, dF_{X_0}}{\theta} + \int_0^A i_{\mathrm{B}}(x; F_{X_\theta})\, dF_X - \int_0^A i_{\mathrm{B}}(x; F_{X_\theta})\, dF_{X_0} \right. \\
&\quad \left. - \frac{\int_0^A \left[i_{\mathrm{E}}(x; F_{X_\theta}) - i_{\mathrm{E}}(x; F_{X_0}) \right]\, dF_{X_0}}{\theta} - \int_0^A i_{\mathrm{E}}(x; F_{X_\theta})\, dF_X + \int_0^A i_{\mathrm{E}}(x; F_{X_\theta})\, dF_{X_0} \right].
\end{aligned} \tag{A.32}$$

Next, we show that

$$\lim_{\theta \to 0} \frac{\int_0^A \left[i_{\mathrm{B}}(x; F_{X_\theta}) - i_{\mathrm{B}}(x; F_{X_0}) \right]\, dF_{X_0}}{\theta} = 0, \tag{A.33}$$

$$\lim_{\theta \to 0} \frac{\int_0^A \left[i_{\mathrm{E}}(x; F_{X_\theta}) - i_{\mathrm{E}}(x; F_{X_0}) \right]\, dF_{X_0}}{\theta} = 0. \tag{A.34}$$

To this end, we first prove (A.33). The proof of (A.34) follows along a similar line as that of (A.33). We start the proof by substituting $i_{\mathrm{B}}(x; F_X) = \int_{\mathbb{R}} p(y|x) \log\left(\frac{p(y|x)}{p_Y(y; F_X)}\right)\, dy$ into the left hand side (A.33) to obtain

$$\lim_{\theta \to 0} \frac{-\int_0^A \int_{\mathbb{R}} p(y|x) \log\left(\dfrac{p_Y(y; F_{X_\theta})}{p_Y(y; F_{X_0})}\right)\, dy\, dF_{X_0}(x)}{\theta}. \tag{A.35}$$

By noting that $p_Y(y;F_{X_\theta}) = (1-\theta)\,p_Y(y;F_{X_0}) + \theta\,p_Y(y;F_X)$ and substituting this into (A.35), we get

$$\begin{aligned}
&\lim_{\theta\to 0} \frac{-\int_0^A \int_{\mathbb{R}} p(y|x) \log\left(\dfrac{p_Y(y;F_{X_\theta})}{p_Y(y;F_{X_0})}\right) dy\, dF_{X_0}(x)}{\theta} \\
&= \lim_{\theta\to 0} \frac{-\int_0^A \int_{\mathbb{R}} p(y|x) \log\left(1+\theta\left[\dfrac{p_Y(y;F_X)}{p_Y(y;F_{X_0})}-1\right]\right) dy\, dF_{X_0}(x)}{\theta} \\
&\overset{(a)}{=} \lim_{\theta\to 0} \frac{-\int_0^A \int_{\mathbb{R}} p(y|x)\,\theta\left[\dfrac{p_Y(y;F_X)}{p_Y(y;F_{X_0})}-1\right] dy\, dF_{X_0}(x)}{\theta} \\
&= -\int_{\mathbb{R}}\int_0^A \frac{p_Y(y;F_X)}{p_Y(y;F_{X_0})}\, p(y|x)\, dF_{X_0}(x)\, dy + \int_{\mathbb{R}}\int_0^A p_Y(y|x)\, dF_{X_0}(x)\, dy \\
&= \int_{\mathbb{R}} p_Y(y;F_{X_0})\, dy - \int_{\mathbb{R}} \frac{p_Y(y;F_X)}{p_Y(y;F_{X_0})} p_Y(y;F_{X_0})\, dy = 1-1=0,
\end{aligned} \tag{A.36}$$

where (a) follows from the fact that when $\theta \to 0$, $\log(1+\theta) \to \theta$ and the limit exists. By substituting (A.33) and (A.34) into (A.32) and noting that $F_{X_\theta} \to F_{X_0}$ as $\theta \to 0$, (A.31) becomes

$$\begin{aligned}
\lim_{\theta\to 0} \frac{g_0((1-\theta)\,F_{X_0}+\theta F_X) - g_0(F_{X_0})}{\theta} &= \int_0^A \left[i_{\mathrm{B}}(x;F_{X_0}) - i_{\mathrm{E}}(x;F_{X_0})\right] dF_X \\
&\quad - \int_0^A \left[i_{\mathrm{B}}(x;F_{X_0}) - i_{\mathrm{E}}(x;F_{X_0})\right] dF_{X_0} \\
&= \int_0^A r_{\mathrm{eq}}(x;F_{X_0})\, dF_X - g_0(F_{X_0}),
\end{aligned} \tag{A.37}$$

which implies that the objective function $g_0(F_X)$ is weakly differentiable. Since the feasible set $\mathcal{A}^+$ is compact and convex and the objective function $g_0(F_X)$ is continuous, strictly concave and weakly differentiable, steps analogous to [10, Theorem 2], [36, Corollary 1] yield the following necessary and sufficient conditions for the optimality of the distribution F_X^*

$$r_{\mathrm{eq}}(x;F_X^*) \le C_{\mathrm{S}},\ \forall x \in [0,A], \tag{A.38}$$

$$r_{\mathrm{eq}}(x;F_X^*) = C_{\mathrm{S}},\ \forall x \in \mathcal{S}_{F_X^*}, \tag{A.39}$$

where $\mathcal{S}_{F_X^*}$ is the support set of F_X^* and the secrecy capacity C_{S} is expressed as

$$C_{\mathrm{S}} = I_{\mathrm{B}}(F_X^*) - I_{\mathrm{E}}(F_X^*) = h_Y(F_X^*) - h_Z(F_X^*) + \frac{1}{2}\mathbb{E}_{F_X^*}\left[\log\left(\frac{\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x)}{\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x)}\right)\right], \tag{A.40}$$

where $I_{\mathrm{B}}(F_X^*)$ and $I_{\mathrm{E}}(F_X^*)$ are the mutual information for Bob and Eve, respectively, generated by the optimal input distribution F_X^*. Similarly, $h_Y(F_X^*)$ and $h_Z(F_X^*)$ are the differential entropies of Y and Z,

respectively, generated by the input distribution F_X^*. Moreover, $\mathbb{E}_{F_X^*}$ denotes the expectation operator with respect to the optimal distribution F_X^*. We now prove by contradiction that the secrecy-capacity-achieving input distribution F_X^* has a finite number of mass points. To reach a contradiction, we use the KKT conditions in (A.38) and (A.39). To this end, we first show that both $i_{\mathrm{B}}(x;F_X)$ and $i_{\mathrm{E}}(x;F_X)$ have analytic extensions over some open connected set in the complex plane $\mathbb{C}$ that includes the nonnegative real line $\mathbb{R}_0^+$.

A.2.5 The rate-equivocation density $r_{eq}(x;F_X)$ has an analytic extension to some open connected set in the complex plane $\mathbb{C}$

To prove the analyticity of $r_{\mathrm{eq}}(x;F_X)$ over some open connected set in the complex plane $\mathbb{C}$, it is sufficient to prove that $i_{\mathrm{B}}(x;F_X)$ has an analytic extension to the open connected set. Invoking similar steps as those in the proof of the analyticity of $i_{\mathrm{B}}(x;F_X)$, one can show that $i_{\mathrm{E}}(x;F_X)$ has also an analytic extension to the open connected set. We start by denoting $i_{\mathrm{B}}(w;F_X)$ as the extension of mutual information density of the legitimate user's channel to the complex plane. Now, we have

$$i_{\mathrm{B}}(w;F_X) = -\int_{\mathbb{R}} p(y|w)\log\left(p_Y(y;F_X)\right)\,dy - \frac{1}{2}\log(2\pi e\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 w)), \tag{A.41}$$

where w is the complex variable. Note that $\log(2\pi e\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 w))$ is analytic over $\mathcal{D}_1 \stackrel{\triangle}{=} \left\{w : \Re(w) > \frac{-1}{\eta_{\mathrm{B}}^2}\right\}$, where $\Re(\cdot)$ is the real part of a complex variable. Similarly, $\log(2\pi e\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 w))$ is analytic over $\mathcal{D}_2 \stackrel{\triangle}{=} \left\{w : \Re(w) > \frac{-1}{\eta_{\mathrm{E}}^2}\right\}$. Since $\eta_{\mathrm{B}}^2\sigma_{\mathrm{B}}^2 = \eta_{\mathrm{E}}^2\sigma_{\mathrm{E}}^2$ and $\sigma_{\mathrm{E}}^2 > \sigma_{\mathrm{B}}^2$, we have $\frac{-1}{\eta_{\mathrm{B}}^2} > \frac{-1}{\eta_{\mathrm{E}}^2}$. Defining $\mathcal{D}$ as $\mathcal{D} \stackrel{\triangle}{=} \mathcal{D}_1$, one can have both of the logarithm functions to be analytic over $\mathcal{D}$. We note that $\mathcal{D}$ is an open connected set in the complex plane $\mathbb{C}$. Next, we show that the continuation of $-\int_{\mathbb{R}} p(y|w)\log\left(p_Y(y;F_X)\right)\,dy$ to the complex plane is continuous over $\mathcal{D}$. To this end, let $\{w_n\}_{n\in\mathbb{N}}$ be a sequence of of complex numbers in $\mathcal{D}$ converging to $w \in \mathcal{D}$, where $w_n = a_n + jb_n$. Since w_n converges, there exist a positive real $\delta > 0$ such that $|w_n| < \delta$ and for some $n > N$. This further implies that $|b_n| < \delta$ for $n > N$. Now, let $\sigma^2_{\mathrm{B},X,\,r}(w_n)$ and $\sigma^2_{\mathrm{B},X,\,i}(w_n)$ be the real and imaginary parts of $\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 w_n)$, respectively, i.e.,

$$\sigma^2_{\mathrm{B},X,\,r}(w_n) = \Re(\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 w_n)), \tag{A.42}$$

$$\sigma^2_{\mathrm{B},X,\,i}(w_n) = \Im(\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 w_n)), \tag{A.43}$$

where $\Im(\cdot)$ is the imaginary part of a complex variable. We have

$$\begin{aligned}|\sigma^2_{\mathrm{B}}(1+\eta^2_{\mathrm{B}}w_n)|^2 &= (\sigma^2_{B,X,\,r}(w_n))^2 + (\sigma^2_{\mathrm{B},X,\,i}(w_n))^2\\ &\geq (\sigma^2_{\mathrm{B},X,\,r}(w_n))^2. \end{aligned} \tag{A.44}$$

Since $w_n \in \mathcal{D}$, we have $\Re(w_n) > -1/\eta^2_{\mathrm{B}}$ and as a result, $\sigma^2_{\mathrm{B},X,\,r}(w_n)$ is a positive real value. Now, we can write

$$\begin{aligned}|p(y|w_n)| =& \left| \frac{1}{\sqrt{\ 2\pi\sigma^2_{\mathrm{B}}(1+\eta^2_{\mathrm{B}}w_n)}} \exp\left(-\frac{(y-w_n)^2}{2\sigma^2_{\mathrm{B}}(1+\eta^2_{\mathrm{B}}w_n)}\right)\right|\\ \leq& \frac{1}{\sqrt{2\pi\sigma^2_{\mathrm{B},X,\,r}(w_n)}} \left| \exp\left(-\frac{(y-a_n-jb_n)^2}{2(\sigma^2_{\mathrm{B},X,\,r}(w_n)+j\sigma^2_{\mathrm{B},X,\,i}(w_n))}\right)\right|\\ =& \frac{1}{\sqrt{2\pi\sigma^2_{\mathrm{B},X,\,r}(w_n)}} \exp\left(-\frac{\sigma^2_{\mathrm{B},X,\,r}(w_n)\left[(y-a_n)^2-b_n^2\right]}{2|\sigma^2_{\mathrm{B}}(1+\eta^2_{\mathrm{B}}w_n)|^2} + \frac{2\,b_n\,\sigma^2_{\mathrm{B},X,\,i}(w_n)\ (y-a_n)}{2|\sigma^2_{\mathrm{B}}(1+\eta^2_{\mathrm{B}}w_n)|^2}\right)\\ =& \frac{1}{\sqrt{2\pi\sigma^2_{\mathrm{B},X,r}(w_n)}} \exp\left(\frac{b_n^2}{2\sigma^2_{\mathrm{B},X,r}(w_n)}\right) \times \exp\left(-\frac{\sigma^2_{\mathrm{B},X,r}(w_n)\,(y-c_n)^2}{2|\sigma^2_{\mathrm{B}}(1+\eta^2_{\mathrm{B}}w_n)|^2}\right)\\ \leq& \frac{1}{\sqrt{2\pi\sigma^2_{\mathrm{B},X,\,r}(w_n)}} \exp\left(\frac{\delta^2}{2\sigma^2_{\mathrm{B},X,\,r}(w_n)}\right) \times \exp\left(-\frac{\sigma^2_{\mathrm{B},X,\,r}(w_n)\,(y-c_n)^2}{2|\sigma^2_{\mathrm{B}}(1+\eta^2_{\mathrm{B}}w_n)|^2}\right)\\ \leq& M(\delta)\exp\left(-\frac{(y-c_n)^2}{d_n^2}\right), \end{aligned} \tag{A.45}$$

where $M(\delta)$ is a bounded function of δ, $c_n \triangleq a_n + b_n\frac{\sigma^2_{\mathrm{B},X,\,i}(w_n)}{\sigma^2_{\mathrm{B},X,\,r}(w_n)}$ and $d_n^2 \triangleq \frac{2|\sigma^2_{\mathrm{B}}(1+\eta^2_{\mathrm{B}}w_n)|^2}{\sigma^2_{\mathrm{B},X,\,r}(w_n)}$. Using (A.45) and (A.14), we get

$$|p(y|w_n)\log(p_Y(y;F_X))| \leq M(\delta)\exp\left(-\frac{(y-c_n)^2}{d_n^2}\right)[\alpha+\beta' y^2]. \tag{A.46}$$

Now, let us define $h(y) \triangleq M(\delta)\exp\left(-\frac{(y-c_n)^2}{d_n^2}\right)[\alpha+\beta' y^2]$ for $y \in \mathbb{R}$. It is a straightforward task to verify that $\int_{\mathbb{R}} h(y)\,dy < \infty$. Hence, by Dominated Convergence Theorem, $i_{\mathrm{B}}(w;F_X)$ is continuous over $\mathcal{D}$.

To show that the function $i_{\mathrm{B}}(w;F_X)$ is analytic over $\mathcal{D}$, it is sufficient to show that if $\oint_C i_{\mathrm{B}}(w;F_X)\,dw = 0$, for any closed contour C in $\mathcal{D}$, then Morera's Theorem applies and it results in the analyticity of $i_{\mathrm{B}}(w;F_X)$ over $\mathcal{D}$. This contour integral is given by

$$\oint_C i_{\mathrm{B}}(w;F_X)\,dw = \oint_C \int_{\mathbb{R}} p(y|w)\log\left(p_Y(y;F_X)\right)\,dy\,dw - \oint_C \frac{1}{2}\log(2\pi e\sigma^2_{\mathrm{B}}(1+\eta^2_{\mathrm{B}}w))\,dw. \tag{A.47}$$

Since $h(y)$ is finite, we define Γ_w as

$$\Gamma_w = \max_{w\in\mathcal{D}} \int_{\mathbb{R}} |p(y|w_n) \log\left(p_Y(y;F_X)\right)|\, dy, \tag{A.48}$$

and we can write

$$\begin{aligned}\left|\oint_C i_{\mathrm{B}}(w;F_X)\, dw\right| &= \left|\oint_C \int_{\mathbb{R}} p(y|w) \log\left(p_Y(y;F_X)\right)\, dy\, dw\right| \\ &\le \oint_C \int_{\mathbb{R}} |p(y|w) \log\left(p_Y(y;F_X)\right)|\, dy\, dw \\ &\le \Gamma_w \ell_C < \infty, \end{aligned} \tag{A.49}$$

where ℓ_C is the length of C which is finite as C is a closed curve. Therefore, by applying Fubini Theorem [56], one can change the order of integration in (A.47) and get

$$\oint_C i_{\mathrm{B}}(w;F_X)\, dw = \int_{\mathbb{R}} \log\left(p_Y(y;F_X)\right)\, dy \oint_C p(y|w)\, dw - \oint_C \frac{1}{2}\log(2\pi e\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 w))\, dw. \tag{A.50}$$

It is clear that the complex functions $p(y|w)$ and $\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 w)$ are analytic over $\mathcal{D}$. This implies that

$$\oint_C p(y|w)\, dw = 0, \tag{A.51}$$

$$\oint_C \frac{1}{2}\log(2\pi e\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 w))\, dw = 0, \tag{A.52}$$

which results in $\oint_C i_{\mathrm{B}}(w;F_X)\, dw = 0$ and thus by Morera's Theorem, $i_{\mathrm{B}}(w;F_X)$ is analytic over $\mathcal{D}$. Similarly, it can be shown that $i_{\mathrm{E}}(w;F_X)$ is also analytic over $\mathcal{D}$ and therefore, the equivocation density $r_{\mathrm{eq}}(w;F_X) = i_{\mathrm{B}}(w;F_X) - i_{\mathrm{E}}(w;F_X)$ is analytic over $\mathcal{D}$.

A.2.6 The secrecy-capacity-achieving input distribution is discrete with a finite number of mass points

To prove the discreteness of the optimal input distribution F_X^*, we use a contradiction approach. To this end, let us assume that $\mathcal{S}_{F_X^*}$ has an infinite number of elements. In view of the optimality condition (A.39), the analyticity of $r_{\mathrm{eq}}(w;F_X)$ over $\mathcal{D}$ and the Identity Theorem of complex analysis along with the Bolzano-Weierstrass Theorem, if $S_{F_X^*}$ has an infinite number of mass points, we get $r_{\mathrm{eq}}(w;F_X^*) = C_{\mathrm{S}}$ for all $w \in \mathcal{D}$. Since $(-1/\eta_{\mathrm{B}}^2, +\infty) \subset \mathcal{D}$, any real variable $x \in (-1/\eta_{\mathrm{B}}^2, +\infty)$ also belongs

to $\mathcal{D}$. This, in turn, implies that if $r_{\rm eq}(w; F_X) = C_{\rm S}$, $\forall\, w \in \mathcal{D}$, then

$$r_{\rm eq}(x; F_X^*) = C_{\rm S}, \quad \forall\, x \in (-1/\eta_{\rm B}^2,\, +\infty). \tag{A.53}$$

Next, we show that (A.53) results in a contradiction. By observing the bounds given in (A.10)–(A.15), one can easily show that

$$\begin{aligned}\int_{\mathbb{R}} \exp(-\alpha - \beta' y^2)[-\alpha - \beta' y^2]\, dy &\leq \int_{\mathbb{R}} p(y|x) \log(p_Y(y; F_X^*))\, dy \\ &\leq \int_{\mathbb{R}} \exp(\alpha - \beta y^2)[\alpha + \beta' y^2]\, dy,\end{aligned} \tag{A.54}$$

for all $x \in (-1/\eta_{\rm B}^2,\, A) \subset (-1/\eta_{\rm B}^2,\, +\infty)$. Similarly,

$$\begin{aligned}\int_{\mathbb{R}} \exp(-\mu - \xi' z^2)[-\mu - \xi' z^2]\, dz &\leq \int_{\mathbb{R}} p(z|x) \log(p_Z(z; F_X^*))\, dz \\ &\leq \int_{\mathbb{R}} \exp(\mu - \xi z^2)[\mu + \xi' y^2]\, dz,\end{aligned} \tag{A.55}$$

for all $x \in (-1/\eta_{\rm B}^2,\, A)$. Therefore, we can write

$$L \leq -\int_{\mathbb{R}} p(y|x) \log(p_Y(y; F_X^*))\, dy + \int_{\mathbb{R}} p(z|x) \log(p_Z(z; F_X^*))\, dz \leq U, \tag{A.56}$$

where the lower bound L and the upper bound U are given respectively as

$$L = \int_{\mathbb{R}} [-\mu - \xi' z^2] \exp(-\mu - \xi' z^2)\, dz + \int_{\mathbb{R}} [-\alpha - \beta' y^2]\, \exp(\alpha - \beta y^2)\, dy, \tag{A.57}$$

$$U = \int_{\mathbb{R}} [\mu + \xi' z^2] \exp(\mu - \xi z^2)\, dz + \int_{\mathbb{R}} [\alpha + \beta' y^2] \exp(-\alpha - \beta' y^2)\, dy. \tag{A.58}$$

Next, we establish that for finite positive real values of β, β', ξ, ξ', μ and α, L and U are finite values. To prove the finiteness of L and U, it is sufficient to prove that L is finite as the finiteness of U follows along a similar line as that of L. We start by expanding L as $L = L_1 + L_2$, where L_1 and L_2 are given respectively as

$$L_1 = \int_{\mathbb{R}} [-\mu - \xi' z^2]\, \exp(-\mu - \xi' z^2)\, dz, \tag{A.59}$$

$$L_2 = \int_{\mathbb{R}} [-\alpha - \beta' y^2]\, \exp(\alpha - \beta y^2)\, dy. \tag{A.60}$$

Since the proof of the finiteness of L_2 is quite similar to that of L_1, we only show the finiteness of L_1.

By direct integration, one can show that $L_1 = -\exp(-\mu)\,(\mu + \frac{1}{2})\sqrt{\frac{\pi}{\xi'}}$, which is a finite value for finite positive reals μ and ξ'. This implies that L is a finite value for finite positive reals μ, α, ξ', β and β'. Next, by substituting (A.6) and (A.7) into (A.53) and using the bounds in (A.57)–(A.58), we can write

$$L \leq C_{\mathrm{S}} + \frac{1}{2}\log\left(\frac{\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x)}{\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x)}\right) \leq U. \tag{A.61}$$

Now, we define the sequence $\{x_n\}_{n\in\mathbb{N}}$ of distinct points in the interval $\mathbb{S} \stackrel{\triangle}{=} (-1/\eta_{\mathrm{B}}^2, A)$ such that it is convergent to a limit point $x_0 = -1/\eta_{\mathrm{B}}^2$. We note that the limit point does not necessarily belong to $(-1/\eta_{\mathrm{B}}^2, A)$. Based on this, we have the following points:

- x_n and $\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x_n)$ are real values for all positive integers n.
- The limit of $\lim_{n\to\infty}\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x_n)$ exists and is equal 0. This is established as follows:

$$\lim_{n\to\infty}\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x_n) = \sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 \lim_{n\to\infty} x_n) \stackrel{(a)}{=} 0, \tag{A.62}$$

  where (a) follows from the fact that $\lim_{n\to\infty} x_n = x_0 = -1/\eta_{\mathrm{B}}^2$ and $(1+\eta_{\mathrm{B}}^2 x_0) = 0$.

Following the results in [10, Theorem 3] and using (A.61), we can write

$$\lim_{n\to\infty}(L - C_{\mathrm{S}}) \leq \lim_{n\to\infty}\frac{1}{2}\log\left(\frac{\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x_n)}{\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x_n)}\right) \leq \lim_{n\to\infty}(U - C_{\mathrm{S}}). \tag{A.63}$$

Now, we note that $\lim_{n\to\infty}\frac{1}{2}\log\left(\frac{\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x_n)}{\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x_n)}\right) = -\infty$ (as $\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x_0)$ is a positive finite value due to the fact that $\eta_{\mathrm{B}}^2 > \eta_{\mathrm{E}}^2$) and the $\lim_{n\to\infty}(L - C_{\mathrm{S}}) = L - C_{\mathrm{S}}$ is a finite value[2], thus a contradiction occurs. This, in turn, implies that the support set $\mathcal{S}_{F_X^*}$ cannot have an infinite number of elements and therefore the optimal input distribution F_X^* is discrete with a finite number of mass points.

A.3 Proof of Proposition 1

Suppose, to the contrary, that $x = 0$ does not belong to the support set of the optimal input distribution $\mathcal{S}_{F_X^*}$. Let $0 < x_1 \leq x_2 \leq \ldots \leq x_N$ be the mass points in the set $\mathcal{S}_{F_X^*}$. Consider two optical wiretap channels with input-dependent Gaussian noises depicted in Figure A.1. Wiretap channel 1 is the original optical wiretap channel, and wiretap channel 2 is obtained from wiretap channel 1 by appending a pre-coder and a post-coder before and after the inner optical channel in the legitimate user's

[2]It is shown in Section 3.4.B that $C_{\mathrm{S}} \leq \frac{1}{2}\log\left(\frac{\sigma_{\mathrm{E}}^2}{\sigma_{\mathrm{E}}^2}\right)$. Furthermore, as $C_{\mathrm{S}} \geq 0$, we have $0 \leq C_{\mathrm{S}} \leq \frac{1}{2}\log\left(\frac{\sigma_{\mathrm{E}}^2}{\sigma_{\mathrm{E}}^2}\right)$. This implies that C_{S} is a finite value.

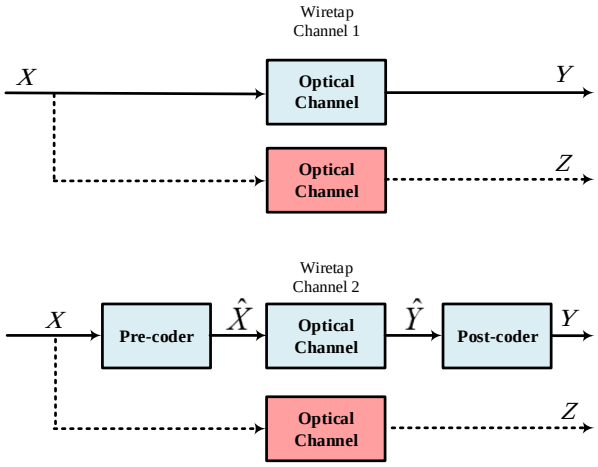

Figure A.1: Two optical wiretap channels with input-dependent Gaussian noise.

link. Specifically, $\hat{X} = X - x_1$ and $Y = \hat{Y} + \hat{N}_{\mathrm{B}}$, where $\hat{N}_{\mathrm{B}}$ is an additive Gaussian noise with mean x_1 and variance $x_1\sigma_{\mathrm{B}}^2\eta_{\mathrm{B}}^2$ and is independent from $\hat{Y}$. For any $x \geq x_1$, the conditional probability density functions $p(y|x)$ and $p(z|x)$ are the same in both wiretap channels. Thus, the joint probability density functions of $p(y, x)$ and $p(z, x)$ in the two wiretap channels are also the same, if the input distribution is F_X^*. As a result, C_{S} is identical in both wiretap channels.

In the second wiretap channel, as $X, \hat{X}, \hat{Y}, Y$ and Z form the Markov chain $X \to \hat{X} \to \hat{Y} \to Y \to Z$, we have $I(\hat{X}; \hat{Y}|Z) \geq I(X; Y|Z)$ by the data processing inequality. This indicates that $I(\hat{X}; \hat{Y}) - I(\hat{X}; Z) \geq I(X; Y) - I(X; Z)$. Now, let $F_{\hat{X}}^*$ be the distribution function of $\hat{X}$ when the distribution function of X is F_X^*. Clearly, $F_{\hat{X}}^*$ satisfies the nonnegativity and peak-intensity constraints. Hence, $F_{\hat{X}}^*$ is also secrecy-capacity-achieving for wiretap channel 1. Based on Lemma 4, the secrecy-capacity-achieving output distribution is unique, as a result, $p_Y(y; F_X^*) = p_Y(y; F_{\hat{X}}^*)$. Therefore, for wiretap channel 2, given the input distribution function of X is F_X^*, the probability density functions for Y and $\hat{Y}$ are the same, which is not possible since $\mathbb{E}[Y] = \mathbb{E}[\hat{Y}] + x_1$. Hence, we reach a contradiction and the proposition follows.

A.4 Proof of Theorem 2

This section presents the proof of Theorem 2 by extending the analysis in the previous section to the entire rate-equivocation region. This extension entails generalizing the contradiction argument in the proof of Theorem 1 to the case when an additional mutual information term is present in the objective function. We start by noting that the objective function $g_\lambda(F_X)$ in (3.15) is strictly concave, and the feasible set $\mathcal{A}^+$ is compact and convex, therefore, the optimization problem in (3.15) has a unique

maximizer. We denote the optimal input distribution for (3.15) as F_X^* which depends on the value λ.

Now, we obtain the KKT optimality conditions for the optimal input distribution of the optimization problem in (3.15). Since the objective function g_λ is weakly differentiable, we have

$$\begin{aligned}&\lim_{\theta\to 0}\frac{g_\lambda((1-\theta)\,F_{X_0}+\theta F_X)-g_\lambda(F_{X_0})}{\theta}\\&=\int_0^A[\lambda\,i_{\mathrm{B}}(x;F_{X_0})+(1-\lambda)\,r_{\mathrm{eq}}(x;F_X)]\,dF_X(x)-g_\lambda(F_{X_0}).\end{aligned}\tag{A.64}$$

Following similar steps mentioned in the proof of Theorem 1, the KKT optimality conditions for the optimality of F_X^* are obtained as follows

$$\lambda\,i_{\mathrm{B}}(x;F_X^*)+(1-\lambda)\,r_{\mathrm{eq}}(x;F_X^*)\le\lambda I_{\mathrm{B}}(F_X^*)+(1-\lambda)(I_{\mathrm{B}}(F_X^*)-I_{\mathrm{E}}(F_X^*)),\quad\forall x\in[0,A],\tag{A.65}$$

$$\lambda\,i_{\mathrm{B}}(x;F_X^*)+(1-\lambda)\,r_{\mathrm{eq}}(x;F_X^*)=\lambda\,I_{\mathrm{B}}(F_X^*)+(1-\lambda)(I_{\mathrm{B}}(F_X^*)-I_{\mathrm{E}}(F_X^*)),\quad\forall x\in\mathcal{S}_{F_X^*}.\tag{A.66}$$

Now, we show that the optimal input distribution F_X^* has a finite support. To this end, we use similar steps mentioned in the proof of Theorem 1 and prove the discreteness of F_X^* by a contradiction approach and using the optimality conditions in (A.65)–(A.66).

Let us assume that $\mathcal{S}_{F_X^*}$ has an infinite number of elements. Under such an assumption, (A.66), the analyticity of $i_{\mathrm{B}}(w;F_X^*)$ and $i_{\mathrm{E}}(w;F_X^*)$ over $\mathcal{D}$ in the complex plane and the Identity Theorem of complex analysis imply that

$$\lambda\,i_{\mathrm{B}}(x;F_X^*)+(1-\lambda)\,r_{\mathrm{eq}}(x;F_X^*)=\lambda\,I_{\mathrm{B}}(F_X^*)+(1-\lambda)(I_{\mathrm{B}}(F_X^*)-I_{\mathrm{E}}(F_X^*)),\quad\forall\,x\in(-1/\eta_{\mathrm{B}}^2,+\infty).\tag{A.67}$$

Next, we show that (A.67) results in a contradiction. To this end, by using (A.54) and (A.55) and the fact that $(1-\lambda)$ is nonnegative for all $\lambda\in[0,1]$, we can bound (A.67) as

$$\widetilde{L}\le I_{\mathrm{B}}(F_X^*)-(1-\lambda)\,I_{\mathrm{E}}(F_X^*)+\frac{1}{2}\log\left(\frac{\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2x)}{\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2x)}\right)+\frac{\lambda}{2}\log(2\pi e\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2x))\le\widetilde{U},\tag{A.68}$$

where $\widetilde{L}$ and $\widetilde{U}$ are given by

$$\widetilde{L}=\,(1-\lambda)\int_{\mathbb{R}}[-\mu-\xi'z^2]\,\exp(-\mu-\xi'z^2)\,dz+\int_{\mathbb{R}}[-\alpha-\beta'y^2]\,\exp(\alpha-\beta y^2)\,dy,\tag{A.69}$$

$$\widetilde{U}=\,(1-\lambda)\int_{\mathbb{R}}[\mu+\xi'z^2]\,\exp(\mu-\xi z^2)\,dz+\int_{\mathbb{R}}[\alpha+\beta'y^2]\,\exp(-\alpha-\beta'y^2)\,dy.\tag{A.70}$$

Invoking similar arguments for the proving the finiteness of L and U given in (A.57)–(A.58), one can show

that the lower bound $\widetilde{L}$ and the upper bound $\widetilde{U}$ are also finite values. Now, let $\{x_n\}_{n\in\mathbb{N}}$ be a convergent sequence of distinct points in $\mathbb{S}$ with a limit point $x_0 = -1/\eta_{\mathrm{B}}^2$. It is clear that 1) x_n and $\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x_n)$ are real for all positive integers n; 2) $\lim_{n\to\infty} \sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x_n) = 0$. Following the results in [10, Theorem 3] and using (A.68), we get

$$\begin{aligned}\lim_{n\to\infty}\left[\widetilde{L} - I_{\mathrm{B}}(F_X^*) + (1-\lambda)\, I_{\mathrm{E}}(F_X^*)\right] &\leq \lim_{n\to\infty}\left[\frac{1}{2}\log\left(\frac{\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x_n)}{\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x_n)}\right) + \frac{\lambda}{2}\log(2\pi e\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x_n))\right] \\ &\leq \lim_{n\to\infty}\left[\widetilde{U} - I_{\mathrm{B}}(F_X^*) + (1-\lambda)\, I_{\mathrm{E}}(F_X^*)\right], \end{aligned} \tag{A.71}$$

We note that $\lim_{n\to\infty} \frac{1}{2}\log\left(\frac{\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{B}}^2 x_n)}{\sigma_{\mathrm{E}}^2(1+\eta_{\mathrm{E}}^2 x_n)}\right) = -\infty$, while $\frac{\lambda}{2}\log(2\pi e\sigma_{\mathrm{B}}^2(1+\eta_{\mathrm{E}}^2 x_0))$ and $\widetilde{L} - I_{\mathrm{B}}(F_X^*) + (1-\lambda)\, I_{\mathrm{E}}(F_X^*)$ are finite values. Hence, we reach a contradiction; implying that the optimal input distribution F_X^* has a finite support.

Appendix B: Secrecy Capacity in the Low-Intensity Regime Under a Peak-Intensity Constraint

In the low-intensity regime, the secrecy capacity can be written as

$$C_{\mathrm{S}} = \frac{1}{2}[J_{\mathrm{B}}(0) - J_{\mathrm{E}}(0)] \max_{F_X \in \mathcal{A}^+} \mathrm{Var}\,(X) + o(A^2). \tag{B.1}$$

Therefore, the optimal input distribution that attains the secrecy capacity under nonnegativity and peak-intensity constraints in the low-intensity regime, also maximizes the variance of the input random variable $\mathrm{Var}\,(X)$. One can show that $\mathrm{Var}\,(X) = \mathbb{E}[X^2] - (\mathbb{E}\,[X])^2$ is a strictly concave function in F_X. Since the set $\mathcal{A}^+$ is compact and convex and the functional $\mathrm{Var}\,(X)$ is continuous and strictly concave in F_X, the maximizer of the optimization problem in (B.1) exists and is unique. Moreover, the condition for the optimality of F_X^* is as follows [9, Chapter 12]

$$\frac{\delta\,\mathrm{Var}\,(X)}{\delta\, f_X^*(x)} = x^2 - 2\,x \int_0^A t\, f_X^*(t)\, dt = 0, \tag{B.2}$$

where $f_X^*(x)$ is the optimal probability density function (PDF) of random variable X and $\frac{\delta\,\mathrm{Var}\,(X)}{\delta\, f_X^*(x)}$ is the functional derivative of $\mathrm{Var}\,(X)$ with respect to $f_X^*(x)$. Now, let us assume that the optimal input distribution that satisfies (B.2) is $f_X^*(x) = p_0\,\delta(x-x_0)+p_1\,\delta(x-x_1)$, where $\delta(\cdot)$ is the dirac delta function. Substituting this distribution into (B.2) results in

$$x_0^2 - 2x_0^2\,p_0 - 2x_0x_1\,p_1 = 0, \tag{B.3}$$

$$x_1^2 - 2x_1^2\,p_1 - 2x_1x_0\,p_0 = 0. \tag{B.4}$$

One can verify that the optimal mass points are located at $\{x_0 = 0,\, x_1 = A\}$ and their corresponding probabilities are $\{p_0 = p_1 = 0.5\}$. Hence, $\max_{F_X \in \mathcal{A}^+} \mathrm{Var}\,(X) = \frac{A^2}{4}$.

Appendix C: Secrecy Capacity in the Low-Intensity Regime Under Peak- and Average-Intensity Constraints

In the low-intensity regime, the secrecy capacity can be written as

$$C_{\mathrm{S}} = \frac{1}{2}[J_{\mathrm{B}}(0) - J_{\mathrm{E}}(0)] \max_{F_X \in \mathcal{M}^+} \mathrm{Var}\,(X) + o(A^2). \tag{C.1}$$

Therefore, the optimal input distribution that attains the secrecy capacity under both the peak- and average-intensity constraints for low-intensity regime, also maximizes the variance of the input random variable $\mathrm{Var}\,(X)$. Since the set $\mathcal{M}^+$ is compact and convex and the functional $\mathrm{Var}\,(X)$ is continuous and strictly concave in F_X, the maximizer of the optimization problem in (C.1) exists and is unique. Moreover, the optimization problem in (C.1) is equivalent to the following

$$h(f_X, \ell) = \max_{F_X \in \mathcal{M}^+} \mathrm{Var}\,(X) = \max_{F_X \in \mathcal{A}^+} \mathrm{Var}\,(X) - \ell(\mathbb{E}\,[X] - P), \tag{C.2}$$

where ℓ is the Lagrangian multiplier and positive. Therefore, the optimality conditions for F_X^* can be given by [9, Chapter 12]

$$\frac{\delta\, h(f_X^*, \ell)}{\delta\, f_X^*(x)} = 0, \tag{C.3}$$

$$\frac{\partial h(f_X^*, \ell)}{\partial \ell} = 0. \tag{C.4}$$

Next, let us consider that the optimal input distribution that satisfies (C.3)–(C.4) is $f_X^*(x) = p_0\,\delta(x - x_0) + p_1\,\delta(x - x_1)$, where $\delta(\cdot)$ is the dirac delta function. Substituting this distribution into (C.3)–(C.4) results in

$$x_0^2 - 2x_0^2\,p_0 - 2x_0x_1\,p_1 - \ell\,x_0 = 0, \tag{C.5}$$

$$x_1^2 - 2x_1^2\,p_1 - 2x_1x_0\,p_0 - \ell\,x_1 = 0, \tag{C.6}$$

$$x_0\,p_0 + x_1\,p_1 = P. \tag{C.7}$$

One can show that for $\ell = A(1 - 2\kappa)$ the optimal mass points are located at $\{x_0 = 0,\, x_1 = A\}$ and their corresponding probabilities are $\{p_0 = 1-\kappa, p_1 = \kappa\}$. Thus, $\max_{F_X \in \mathcal{A}^+} \mathrm{Var}\,(X) = \kappa(1-\kappa)A^2$. Since $\ell > 0$

(average intensity constraint is active), then $\kappa \in (0, \frac{1}{2})$. When $\kappa \in [\frac{1}{2}, 1]$ the average intensity constraint is not active and the result follows from Appendix B. That is, in the low-intensity regime, $\kappa \in (0, \frac{1}{2})$. A similar observation has been made in [5], but for the case with no secrecy constraint.

Appendix D: Proof of Theorem 3

D.1 Preliminaries

As mentioned, the conditional probability laws of Y_1 and Y_2 given X are described as

$$p_{Y_1|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma_{\mathrm{B}}^2(x)}} \exp\left[-\frac{(y-x)^2}{2\sigma_{\mathrm{B}}^2(x)}\right], \tag{D.1}$$

$$p_{Y_2|X}(t|x) = \frac{1}{\sqrt{2\pi\sigma_{\mathrm{E}}^2(x)}} \exp\left[-\frac{(t-x)^2}{2\sigma_{\mathrm{E}}^2(x)}\right], \tag{D.2}$$

and we have

$$p_{Y_1}(y; F_X) = \int_0^{+\infty} p_{Y_1|X}(y|x)\, dF_X(x),\ y \in \mathbb{R}, \tag{D.3}$$

$$p_{Y_2}(t; F_X) = \int_0^{+\infty} p_{Y_2|X}(t|x)\, dF_X(x),\ t \in \mathbb{R}. \tag{D.4}$$

We define the secrecy-rate density as $c_{\mathrm{S}}(x; F_X) \triangleq i_{\mathrm{B}}(x; F_X) - i_{\mathrm{E}}(x; F_X)$, where

$$i_{\mathrm{B}}(x; F_X) \triangleq \int_{\mathbb{R}} p_{Y_1|X}(y|x) \log \frac{1}{p_{Y_1}(y; F_X)}\, dy - \frac{1}{2}\log(2\pi e\sigma_{\mathrm{B}}^2(x)), \tag{D.5}$$

$$i_{\mathrm{E}}(x; F_X) \triangleq \int_{\mathbb{R}} p_{Y_2|X}(t|x) \log \frac{1}{p_{Y_2}(t; F_X)}\, dt - \frac{1}{2}\log(2\pi e\sigma_{\mathrm{E}}^2(x)). \tag{D.6}$$

Observe that $I(X;Y_1) = \mathbb{E}_X[i_{\mathrm{B}}(x; F_X)]$ and $I(X;Y_2) = \mathbb{E}_X[i_{\mathrm{E}}(x; F_X)]$.

D.2 Proof of the Theorem

Theorem 3 is proved as follows.

D.2.1 The set $\mathcal{P}^+$ is convex and compact

Invoking similar argument that appeared in [58, Theorem 3], one can show that the set $\mathcal{P}^+$ satisfies the desired property.

D.2.2 $f_\mu(F_X)$ is continuous in F_X

We need to show that $I(X;Y_1)$ and $I(X;Y_2)$ are both continuous in F_X. To this end, we only show that $I(X;Y_1)$ is continuous in F_X, as one can similarly show that $I(X;Y_2)$ is continuous. We start by

noting that $I(X;Y_1) = h(Y_1) - h(Y_1|X)$, where $h(\cdot)$ is the differential entropy and $h(\cdot|\cdot)$ is the conditional differential entropy. Now, observe that $p_{Y_1}(y;F_X) \leq k_0 \int_0^{+\infty} \phi(y,x)\, dF_X(x)$, where $k_0 \triangleq \frac{1}{\sqrt{2\pi\sigma^2_{\mathrm{B},0}}}$ and $\phi(y,x) \triangleq \exp\left(-\frac{(y-x)^2}{2\sigma^2_{\mathrm{B}}(x)}\right) \in [0,1]$. Therefore, we have

$$|p_{Y_1}(y;F_X)\log p_{Y_1}(y;F_X)| \leq |k_0 \log k_0|\psi(y) + k_0|\psi(y)\log\psi(y)|, \tag{D.7}$$

where $\psi(y) \triangleq \int_0^{\infty} \phi(y,x)\, dF_X(x) \in [0,1]$. Note that $|\psi(y)\log\psi(y)| \leq e^{-1}$ for $\psi(y) \in [0, e^{-1}]$ and $|\psi(y)\log\psi(y)| \leq \psi(y)$ for $\psi(y) \in [e^{-1}, 1]$. Furthermore, $\phi(y,x)$ is increasing in x for $y > x > 0$ and decreasing in x for $0 < y < x$. Next, we show that (D.7) can be bounded above by an integrable function $\beta(y)$, i.e., $\int_{\mathbb{R}} \beta(y)\, dy < \infty$. To that end, observe that based on (D.7) and the arguments appear after it, one can write

$$\beta(y) \triangleq \begin{cases} k_0'\psi(y), & |y| > y_0 \\ k_0'', & |y| \leq y_0 \end{cases} \tag{D.8}$$

where y_0, k_0', k_0'' are positive constants. Note that when $y < -y_0$, we have $\phi(y,x) \leq \exp\left(-\frac{y^2}{2\sigma^2_B(x)}\right) \in [0,1]$ and

$$\begin{aligned}\psi(y) = \int_0^{\infty} \phi(y,x)\, dF_X(x) &\leq \int_0^{|y|^{\frac{3}{2}}} \exp\left(-\frac{y^2}{2\sigma^2_{\mathrm{B}}(x)}\right) dF_X(x) + \int_{|y|^{\frac{3}{2}}}^{\infty} \exp\left(-\frac{y^2}{2\sigma^2_{\mathrm{B}}(x)}\right) dF_X(x) \\ &\leq \exp\left(-\frac{y^2}{2\sigma^2_{\mathrm{B}}(|y|^{\frac{3}{2}})}\right) + \frac{\mathcal{E}}{|y|^{\frac{3}{2}}},\end{aligned} \tag{D.9}$$

where the first term in (D.9) follows since $\sigma^2_{\mathrm{B}}(x)$ is linearly increasing in x, and the second term is due to Markov's inequality. Also, when $y > y_0$ we get

$$\begin{aligned}\psi(y) &= \int_0^{y/2} \phi(y,x)\, dF_X(x) + \int_{3y/2}^{\infty} \phi(y,x)\, dF_X(x) + \int_{y/2}^{3y/2} \phi(y,x)\, dF_X(x) \\ &\leq 2\exp\left(-\frac{y^2}{8\sigma^2_{\mathrm{B}}(3y/2)}\right) + \frac{32\,\sigma^4_B(3y/2)}{y^4},\end{aligned} \tag{D.10}$$

where the first term (D.10) follows because $\phi(y,x)$ is increasing in x for $0 < x < y/2$ and decreasing in x for $0 < 3y/2 < x$, and the second terms is due to the inequality $e^{-u} \leq u^{-2}$, where $u \triangleq \frac{2\sigma^2_B(3y/2)}{(y-x)^2} > 0$, and noting that u is decreasing in x. From (D.9)–(D.10), we can conclude that $\beta(y)$ is integrable.

As such, the reasoning in [37, Appendix I.B] establishes that $h(Y_1)$ is continuous in F_X.

Additionally, since $h(Y_1|X=x)=\frac{1}{2}\log(2\pi e\sigma^2_{\mathrm{B}}(x))$, we observe that

$$-\infty<\frac{1}{2}\log\,2\pi e\sigma^2_{\mathrm{B},0}\leq h(Y_1|X=x)\leq k_1+k_2\sqrt{x}, \tag{D.11}$$

where k_1 and k_2 are some positive constants. Now, we have $\int_b^{+\infty} h(Y_1|X=x)\,dF_X(x)\leq \mathcal{E}/\sqrt{b}$ for all $F_X(x)\in\mathcal{M}^+$ and b large enough. Thus, by invoking similar arguments that appeared in [37, Appendix I.B], one can show that $h(Y_1|X)$ is continuous in F_X.

Steps 1 and 2 imply that the supremization in (4.8) is a maximization problem.

D.2.3 Strict concavity of $f_\mu(F_X^*)$

This step can be proved by a similar contradiction argument which is appeared in [25, Appendix A]. This step implies that the answer to (4.8), denoted by F_X^*, is unique.

D.2.4 $f_\mu(F_X)$ is weakly differentiable in $\mathcal{P}^+$

Following along similar lines of [25], one can show the weak differentiability of the functional $f_\mu(F_X)$ in $F_X\in\mathcal{P}^+$ and the weak derivative at $F_X^o\in\mathcal{P}^+$, denoted by $f'_\mu(F_X,F_X^o)$, is

$$\begin{aligned} f'_\mu(F_X,F_X^o)&\triangleq\lim_{\theta\to 0}\frac{f_\mu((1-\theta)F_X^o+\theta F_X)-f_\mu(F_X^o)}{\theta}\\ &=\int_0^{+\infty}[\mu i_B(x;F_X^o)+(1-\mu)c_{\mathrm{S}}(x;F_X^o)]dF_X(x)-f_\mu(F_X^o),\ \theta\in[0,1]. \end{aligned}\tag{D.12}$$

After establishing the steps 1 through 4 and following along similar lines of [37, Appendix II.B], the optimal solution F_X^* must satisfy the following necessary and sufficient optimality equations

$$\mu i_{\mathrm{B}}(x;F_X^*)+(1-\mu)c_{\mathrm{S}}(x;F_X^*)-\gamma x\leq f_\mu(F_X^*)-\gamma\mathcal{E},\qquad \forall\, x\in[0,+\infty), \tag{D.13}$$

$$\mu i_{\mathrm{B}}(x;F_X^*)+(1-\mu)c_{\mathrm{S}}(x;F_X^*)-\gamma x=f_\mu(F_X^*)-\gamma\mathcal{E},\qquad \forall\, x\in\mathcal{S}_{F_X^*}, \tag{D.14}$$

where $\gamma\geq 0$ is the Lagrangian multiplier and $\mathcal{S}_{F_X^*}$ is the optimal solution's support set.

We now prove by contradiction that $\mathcal{S}_{F_X^*}$ must be in such a way that the cardinality of $\mathcal{S}_{F_X^*}\cap B$, where $B\subset[0,+\infty)$ is any bounded interval, is finite. For reaching a contradiction, we use the equations in (D.13)–(D.14). We first show that both $i_{\mathrm{B}}(x;F_X)$ and $i_{\mathrm{E}}(x;F_X)$ are analytic in the complex plane.

D.2.5 The analyticity of $c_{\mathrm{S}}(x; F_X)$

Following along similar lines of [25], we have the analyticity of $i_{\mathrm{B}}(z; F_X)$ over $\mathcal{C}_1 \stackrel{\triangle}{=} \{z : \Re(z) > -\sigma^2_{\mathrm{B},0}/\sigma^2_{\mathrm{B},1}\}$, and the analyticity of $i_{\mathrm{E}}(z; F_X)$ over $\mathcal{C}_2 \stackrel{\triangle}{=} \{z : \Re(z) > -\sigma^2_{\mathrm{E},0}/\sigma^2_{\mathrm{E},1}\}$, where z is the complex variable and $\Re(\cdot)$ is the real part operator. Thus, By virtue of the conditions (4.3), $c_{\mathrm{S}}(z; F_X)$ is analytic in $\mathcal{C}_1 \cap \mathcal{C}_2 = \mathcal{C}_1$.

D.2.6 The intersection of $\mathcal{S}_{F_X^*}$ with any bounded interval has a finite cardinality

Assume, to the contrary, that there exists a bounded interval B such that $\mathcal{S}_{F_X^*} \cap B$ has an infinite cardinality. Using (D.14), the analyticity of $c_{\mathrm{S}}(z; F_X^*)$ and $i_{\mathrm{B}}(z; F_X^*)$, Identity and Bolzano-Weierstrass Theorems, we have that if $|\mathcal{S}_{F_X^*} \cap B| = \infty$, where $|A|$ is the cardinality of set A, then $\mu i_{\mathrm{B}}(z; F_X^*) + (1-\mu)c_{\mathrm{S}}(z; F_X^*) - \gamma z = f_\mu(F_X^*) - \gamma\mathcal{E}$ for all $z \in \mathcal{C}_1$. Since $(-\sigma^2_{\mathrm{B},0}/\sigma^2_{\mathrm{B},1}, +\infty) \subset \mathcal{C}_1$, any real variable $x \in (-\sigma^2_{\mathrm{B},0}/\sigma^2_{\mathrm{B},1}, +\infty)$ also belongs to $\mathcal{C}_1$. This implies

$$\mu i_{\mathrm{B}}(x; F_X^*) + (1-\mu)c_{\mathrm{S}}(x : F_X^*) - \gamma x = f_\mu(F_X^*) - \gamma\mathcal{E}, \quad \forall\, x \in (-\sigma^2_{\mathrm{B},0}/\sigma^2_{\mathrm{B},1}, +\infty). \tag{D.15}$$

Substituting (D.5) and (D.6) into (D.15) and rearranging the terms, one obtains

$$\begin{aligned}\frac{1}{2}\log\frac{\sigma^2_{\mathrm{B}}(x)}{\sigma^2_{\mathrm{E}}(x)} + \frac{\mu}{2}\log(2\pi e\sigma^2_{\mathrm{E}}(x)) + \gamma(x-\mathcal{E}) + f_\mu(F_X^*) &= \int_{\mathbb{R}} p_{Y_1|X}(y|x)\log\frac{1}{p_{Y_1}(y; F_X^*)}\,dy \\ &\quad - (1-\mu)\int_{\mathbb{R}} p_{Y_2|X}(t|x)\times\log\frac{1}{p_{Y_2}(t; F_X^*)}\,dt \\ &\stackrel{\triangle}{=} L(x), \quad \forall\, x > -\sigma^2_{\mathrm{B},0}/\sigma^2_{\mathrm{B},1}. \end{aligned}\tag{D.16}$$

Observe that $p_{Y_1}(y; F_X^*) \le k_0$ for all $y \in \mathbb{R}$. Thus $-\log p_{Y_1}(y; F_X^*) \ge \frac{1}{2}\log(2\pi\sigma^2_{\mathrm{B},0})$. Furthermore, let A be a constant such that $\Pr(X \le A) \ge \frac{1}{2}$. Therefore,

$$p_{Y_2}(t; F_X^*) \ge \int_0^A p_{Y_2|X}(t|x)\,dF_X^*(x) \stackrel{(a)}{\ge} g(t), \tag{D.17}$$

where (a) follows from [59], with $g(t)$ given by

$$g(t) \stackrel{\triangle}{=} \begin{cases} \frac{1}{\sqrt{8\pi\sigma^2_{\mathrm{E}}(A)}} e^{-(t-A)^2/2\sigma^2_{\mathrm{E},0}}, & t \le A/2, \\ \frac{1}{\sqrt{8\pi\sigma^2_{\mathrm{E}}(A)}} e^{-t^2/2\sigma^2_{\mathrm{E},0}}, & t > A/2. \end{cases} \tag{D.18}$$

Hence, $L(x)$ can be lower bounded as

$$L(x) \geq \frac{1}{2}\log(2\pi\sigma^2_{\mathrm{B},0}) + (1-\mu)\left[\log(k_3) - \int_{\mathbb{R}} p_{Y_2|X}(t|x) \times \frac{t^2}{2\sigma^2_{\mathrm{E},0}}\,dt + \frac{A}{2\sigma^2_{\mathrm{E},0}}\int_{-\infty}^{A/2}(2t-A)p_{Y_2|X}(t|x)\,dt\right]$$
$$\geq \frac{1}{2}\log(2\pi\sigma^2_{\mathrm{B},0}) + (1-\mu)\left[\log(k_4) - \frac{x^2+\sigma^2_{\mathrm{E}}(x)}{2\sigma^2_{\mathrm{E},0}} + \frac{A}{\sigma^2_{\mathrm{E},0}}\left(x - \sigma^2_{\mathrm{E}}(x)\frac{p_{Y_2|X}(A/2|x)}{P_{Y_2|X}(A/2|x)}\right)\right], \tag{D.19}$$

where $k_3 = \frac{1}{\sqrt{8\pi\sigma^2_{\mathrm{E}}(A)}}$ and $k_4 = k_3 e^{-A^2/2\sigma^2_{\mathrm{E},0}}$ are constants, $P_{Y_2|X}(\cdot|x)$ is the cumulative distribution function of $Y_2|X$, and the last term in (D.19) is obtained using integration by parts. Now, observe that in (D.19), when $x \to -\frac{\sigma^2_{\mathrm{B},0}}{\sigma^2_{\mathrm{B},1}}^{+}$, $L(x)$ is lower bounded by a finite and constant value, say k_5, due to (4.3). Thus, taking the limit from the sides of (D.16) as $x \to -\frac{\sigma^2_{\mathrm{B},0}}{\sigma^2_{\mathrm{B},1}}^{+}$ results in

$$k_5 \leq \lim_{x\to -\frac{\sigma^2_{\mathrm{B},0}}{\sigma^2_{\mathrm{B},1}}^{+}} \frac{1}{2}\log\frac{\sigma^2_{\mathrm{B}}(x)}{\sigma^2_{\mathrm{E}}(x)} + \frac{\mu}{2}\log\left[2\pi e(\sigma^2_{\mathrm{E},0} - \sigma^2_{\mathrm{B},0})\right] + \gamma\left(-\frac{\sigma^2_{\mathrm{B},0}}{\sigma^2_{\mathrm{B},1}} - \mathcal{E}\right) + f_\mu(F^*_X). \tag{D.20}$$

Observe that in light of (4.3), we have

$$\lim_{x\to -\frac{\sigma^2_{\mathrm{B},0}}{\sigma^2_{\mathrm{B},1}}^{+}} \frac{1}{2}\log\frac{\sigma^2_{\mathrm{B}}(x)}{\sigma^2_{\mathrm{E}}(x)} \overset{(b)}{=} \frac{1}{2}\log\left[\lim_{x\to -\frac{\sigma^2_{\mathrm{B},0}}{\sigma^2_{\mathrm{B},1}}^{+}} \frac{\sigma^2_{\mathrm{B}}(x)}{\sigma^2_{\mathrm{E}}(x)}\right] = \frac{1}{2}\log\frac{0^+}{\sigma^2_{\mathrm{E},0} - \sigma^2_{\mathrm{B},0}} = -\infty, \tag{D.21}$$

where (b) is justified because of the continuity of the logarithm. Combining (D.20) and (D.21) results in the desired contradiction. Therefore, $\mathcal{S}_{F^*_X} \cap B$ must have a finite cardinality.

D.2.7 The unboundedness of $\mathcal{S}_{F^*_X}$

To prove this, we again resort to a contradiction approach. Assume, to the contrary, that $\mathcal{S}_{F^*_X}$ is bounded, i.e., $\mathcal{S}_{F^*_X} \subseteq [0, h]$, where $h < +\infty$. In the previous section, we proved that the intersection of $\mathcal{S}_{F^*_X}$ with any bounded interval has a finite cardinality. Since, we are assuming that $\mathcal{S}_{F^*_X}$ is bounded, thus, it has a finite cardinality. This implies that $F^*_X(x) = \sum_{i=1}^N p_i u(x - x_i)$, where $N < +\infty$, $0 \leq x_1 < x_2 < \cdots < x_N \leq h$ are the mass points with corresponding probabilities $\{p_1, \ldots, p_N\}$. Furthermore, we can write

$$p_{Y_2}(t; F^*_X) = \sum_{i=1}^N p_i p_{Y_2|X}(t|x_i) > p_N p_{Y_2|X}(t|x_N). \tag{D.22}$$

Therefore, $\log p_{Y_2}(t;F_X^*) > \log p_N + \log p_{Y_2|X}(t|x_N)$, and $-i_{\mathrm{E}}(x;F_X^*)$ can be lower bounded as

$$-\,i_{\mathrm{E}}(x;F_X^*) \geq -\int_{\mathbb{R}} p_{Y_2|X}(t|x)\frac{(t-x_N)^2}{2\sigma_{\mathrm{E}}^2(x_N)} + \frac{1}{2}\log\frac{e\,\sigma_{\mathrm{E}}^2(x)}{\sigma_{\mathrm{E}}^2(x_N)} + \log p_N = -\frac{x^2}{2\sigma_{\mathrm{E}}^2(x_N)} + o(x^2), \tag{D.23}$$

where $o(x^2)$ is a function which satisfies $\lim_{x\to+\infty}\frac{o(x^2)}{x^2}=0$. Furthermore, since $x\in[0,x_N]$, we can write $p_{Y_1}(y;F_X^*)\leq G(y)$ [59], where $G(y)$ is given by

$$G(y) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma_{\mathrm{B},0}^2}}e^{-\frac{y^2}{2\sigma_{\mathrm{B}}^2(x_N)}}, & y<0,\\ \frac{1}{\sqrt{2\pi\sigma_{\mathrm{B},0}^2}}e^{-\frac{(y-x_N)^2}{2\sigma_{\mathrm{B}}^2(x_N)}}, & y>x_N,\\ \frac{1}{\sqrt{2\pi\sigma_{\mathrm{B},0}^2}}, & 0\leq y\leq x_N. \end{cases} \tag{D.24}$$

As a result, $i_{\mathrm{B}}(x;F_X^*)$ can be lower bounded as

$$\begin{aligned} i_{\mathrm{B}}(x;F_X^*) &\geq \frac{1}{2}\log\frac{\sigma_{\mathrm{B},0}^2}{e\,\sigma_{\mathrm{B}}^2(x)} + \int_{\mathbb{R}} p_{Y_1|X}(y|x)\,\frac{y^2}{2\sigma_{\mathrm{B}}^2(x_N)}\,dy - \underbrace{\int_0^{x_N} p_{Y_1|X}(y|x)\,\frac{y^2}{2\sigma_{\mathrm{B}}^2(x_N)}\,dy}_{\geq -\frac{x_N^2}{2\sigma_{\mathrm{B}}^2(x_N)}\ \text{as}\ y\leq x_N,\ \int_0^{x_N} p_{Y_1|X}(y|x)\,dy\leq 1}\\ &\quad - \int_{x_N}^{+\infty}\frac{x_N\,y\,p_{Y_1|X}(y|x)}{\sigma_{\mathrm{B}}^2(x_N)}\,dy + \frac{x_N^2}{2\sigma_{\mathrm{B}}^2(x_N)}\int_{x_N}^{+\infty} p_{Y_1|X}(y|x)\,dy\\ &\overset{(c)}{\geq} \frac{1}{2}\log\frac{\sigma_{\mathrm{B},0}^2}{e\,\sigma_{\mathrm{B}}^2(x)} + \frac{x^2+\sigma_{\mathrm{B}}^2(x)}{2\sigma_{\mathrm{B}}^2(x_N)}\,dy - \frac{x_N^2}{2\sigma_{\mathrm{B}}^2(x_N)} - \frac{x_N}{\sigma_{\mathrm{B}}^2(x_N)}\left[x-\sigma_{\mathrm{B}}^2(x)\,\Xi(x)\right], \end{aligned} \tag{D.25}$$

where (c) follows because $p_{Y_1|X}(y|x)\geq 0$, and $\Xi(x)\triangleq\frac{p_{Y_1|X}(+\infty|x)-p_{Y_1|X}(x_N|x)}{1-P_{Y_1|X}(x_N|x)}$. Note that $p_{Y_1|X}(+\infty|x)\geq 0$, and $p_{Y_1|X}(x_N|x)\leq k_0$. Furthermore, $P_{Y_1|X}(x_N|x)\approx 0$ for large values of x and $1-P_{Y_1|X}(x_N|x)\approx 1$. Hence, for large values of x, (D.25) can be further lower bounded as

$$i_{\mathrm{B}}(x;F_X^*) \geq \frac{x^2}{2\sigma_{\mathrm{B}}^2(x_N)} + o(x^2). \tag{D.26}$$

Combining (D.13), (D.23), and (D.26), we find

$$\begin{aligned} f_\mu(F_X^*) + \gamma(x-\mathcal{E}) &\geq \mu\,i_{\mathrm{B}}(x;F_X^*) + (1-\mu)\,c_{\mathrm{S}}(x;F_X^*)\\ &\geq \mu\frac{x^2}{2\sigma_{\mathrm{B}}^2(x_N)} + (1-\mu)\left[\frac{x^2}{2\sigma_{\mathrm{B}}^2(x_N)} - \frac{x^2}{2\sigma_{\mathrm{E}}^2(x_N)}\right] + o(x^2). \end{aligned} \tag{D.27}$$

Notice that in light of the degradedness conditions in (4.3), the left-hand-side of (D.27) grows linearly in x, but the right-hand-side of it grows quadratically in x for all $\mu\in[0,1]$. Thus, we reach a contradiction implying that $\mathcal{S}_{F_X^*}$ must be an unbounded set. From Steps 6 and 7, we infer that F_X^* must have a countably

infinite support set. Finally, we note that the contradiction in (D.27) also holds for $\sigma^2_{\mathrm{B},1} = \sigma^2_{\mathrm{E},1} = 0$, i.e., the support set of optimal solutions for the FSO WC with an average optical power is also unbounded. This conclusion cannot be reached using [35, Theorem 3, Section IV]. Combining this result along with the countability of the support (discreteness of the distributions) shown in [35, Theorem 3, Sec. (iv)], we conclude that in FSO WC with an average optical power constraint, optimal inputs admit a countably infinite support set.

Appendix E: Proof of the Main Results in Chapter 5

In this section, we first provide the required preliminaries for the development of the main results. We then give the detailed proofs of theorems and the proposition mentioned in Sec. 5.3.

E.1 Preliminaries

Since both legitimate user's and eavesdropper's channels are discrete-time Poisson channels, the output densities for Y and Z exist for any input distribution F_X, and are given by

$$P_Y(y;F_X) = \int_0^{\mathcal{A}} p(y|x)\, dF_X(x),\ y \in \mathbb{N}, \tag{E.1}$$

$$P_Z(z;F_X) = \int_0^{\mathcal{A}} p(z|x)\, dF_X(x),\ z \in \mathbb{N}, \tag{E.2}$$

where $p(y|x)$ and $p(z|x)$ are given by (5.1)–(5.2). We define the secrecy rate density $c_s(x;F_X)$ as

$$c_S(x;F_X) \stackrel{\triangle}{=} \frac{1}{\Delta}\left[i_B(x;F_X) - i_E(x;F_X)\right], \tag{E.3}$$

where $i_B(x;F_X)$ and $i_E(x;F_X)$ are the mutual information densities for the legitimate user's and the eavesdropper's channels, respectively, and are as follows

$$i_B(x;F_X) = \sum_{y=0}^{+\infty} p(y|x)\, \log \frac{p(y|x)}{P_Y(y;F_X)}, \tag{E.4}$$

$$i_E(x;F_X) = \sum_{z=0}^{+\infty} p(z|x)\, \log \frac{p(z|x)}{P_Z(z;F_X)}. \tag{E.5}$$

Plugging (5.1)–(5.2) and (E.1)–(E.2) into (E.4)–(E.5) and after some algebra, we get

$$i_B(x;F_X) = [(\alpha_B x + \lambda_B)\Delta]\log[(\alpha_B x + \lambda_B)\Delta] - \alpha_B x\Delta - \sum_{y=0}^{+\infty} p(y|x)\log g_B(y;F_X), \tag{E.6}$$

$$i_E(x;F_X) = [(\alpha_E x + \lambda_E)\Delta]\log[(\alpha_E x + \lambda_E)\Delta] - \alpha_E x\Delta - \sum_{z=0}^{+\infty} p(z|x)\log g_E(z;F_X), \tag{E.7}$$

where $g_B(y;F_X)$ and $g_E(z;F_X)$ are respectively defined as

$$g_B(y;F_X) \triangleq \int_0^{\mathcal{A}} e^{-\alpha_B x\Delta}\left[(\alpha_B x+\lambda_B)\Delta\right]^y dF_X(x), \tag{E.8}$$

$$g_E(z;F_X) \triangleq \int_0^{\mathcal{A}} e^{-\alpha_E x\Delta}\left[(\alpha_E x+\lambda_E)\Delta\right]^z dF_X(x). \tag{E.9}$$

Furthermore, we have the following identities

$$I(X;Y) = \int_0^{\mathcal{A}} i_B(x;F_X)\, dF_X(x) \triangleq I_B(F_X), \tag{E.10}$$

$$I(X;Z) = \int_0^{\mathcal{A}} i_E(x;F_X)\, dF_X(x) \triangleq I_E(F_X), \tag{E.11}$$

$$f_0(F_X) = \int_0^{\mathcal{A}} c_S(x;F_X)\, dF_X(x). \tag{E.12}$$

Next, we prove Theorem 4 using the preliminaries provided in this section.

E.2 Proof of Theorem 4

We start by proving that the set of input distributions $\Omega^+_{\mathcal{A},\,\mathcal{E}}$ is compact and convex. We then show that the objective functions $f_0(F_X)$ in (5.13) is continuous, strictly concave and weakly differentiable in the input distribution F_X and hence, we conclude that the optimization problems in (5.13) has a unique solutions. We continue the proof by deriving the necessary and sufficient conditions (KKT conditions) for the optimality of the optimal input distribution F_X^*. Finally, by means of contradiction we show that the optimal input distributions are discrete with a finite number of mass points. The proof is then streamlined into a few lemmas which we state below.

Lemma 5. *The feasible set $\Omega^+_{\mathcal{A},\,\mathcal{E}}$ is convex and sequentially compact in the Levy metric sense.*

Proof. The proof follows along similar lines as [7, Lemma 1] ■

Lemma 6. *The functional $f_0 : \Omega^+_{\mathcal{A},\,\mathcal{E}} \to \mathbb{R},\ f_0(F_X) = I_B(F_X) - I_E(F_X)$ is continuous in F_X.*

Proof. The proof follows along similar lines as presented in [7, Lemma 3]. ■

From Lemma 5 and Lemma 6, $f_0(F_X)$ is continuous in F_X over $\Omega^+_{\mathcal{A},\,\mathcal{E}}$ which itself is a compact set, then by the Extreme Value Theorem, $f_0(F_X)$ is bounded above and attains its supremum. That is, the supremum in (5.13) is actually a maximum which is achievable by at least one input distribution F_X.

Lemma 7. *The functional $f_0(F_X)$ is strictly concave in F_X.*

Proof. The proof is by contradiction and follows along similar lines as in [25, Appendix A] with the difference that the conditional channel laws follow Poisson distribution.

We start the proof by noting that for random variables X, Y and Z that form the Markov chain $X \to Y \to Z$, $I(X;Y|Z) = I(X;Y) - I(X;Z)$ is a concave functional in F_X [57, Appendix A]. Now, let X_1 and X_2 be two channel inputs generated by F_{X_1} and F_{X_2}, respectively, and Q be a binary-valued random variable such that

$$p(y,z,x|q) = \begin{cases} p(y,z|x)\, p_{X_1}(x), & q = 1, \\ p(y,z|x)\, p_{X_2}(x), & q = 2, \end{cases} \tag{E.13}$$

where $p_{X_1}(x)$ and $p_{X_2}(x)$ are the probability density functions of the random variables X_1 and X_2. Based on (E.13), we have the following Markov chain

$$Q \to X \to Y \to Z. \tag{E.14}$$

Following along the same lines as [57, Appendix A], one can show that

$$I(X;Y|Z,Q) - I(X;Y|Z) = -I(Q;Y|Z). \tag{E.15}$$

Since $I(Q;Y|Z) \geq 0$, $I(X;Y|Z,Q) \leq I(X;Y|Z)$. This implies that $I(X;Y|Z)$ is a concave function in F_X. Now, we prove that with the Markov chain $Q \to X \to Y \to Z$, $I(X;Y|Z)$ is strictly concave in F_X, i.e., $I(Q;Y|Z) > 0$. Assume, to the contrary, that there exists an F_X such that $I(Q;Y|Z) = 0$. This implies that random variables Q, Y and Z also form the Markov chain

$$Q \to Z \to Y. \tag{E.16}$$

Furthermore, from the Markov chain (E.14), we have

$$Q \to X \to Z. \tag{E.17}$$

Combining Markov chains (E.16) and (E.17) results in a new Markov chain given by

$$Q \to X \to Z \to Y. \tag{E.18}$$

Now, based on (E.14) and (E.18), we obtain the following

$$\begin{aligned} p(y,z,x)\big|_{\text{Markov chain (E.14)}} &= p(y,z,x)\big|_{\text{Markov chain (E.18)}} \\ p_X(x)\,p(y|x)\,p(z|y) &= p_X(x)\,p(z|x)\,p(y|z) \\ \frac{p(y|x)}{p(z|x)} &= \frac{p(y|z)}{p(z|y)}. \end{aligned} \tag{E.19}$$

We note that (E.19) holds for any $y, z \in \mathbb{N}$ and $x \in \mathcal{S}_{F_X}$, where $\mathcal{S}_{F_X}$ is the support set of F_X. As a result, for fixed values of y and z the RHS of (E.19) is fixed, while the left hand side (LHS) is a function of x. Since $Y|X$ and $Z|X$ are Poisson distributed with mean $(\alpha_B x + \lambda_B)\Delta$ and $(\alpha_E x + \lambda_E)\Delta$, respectively, (E.19) reduces to

$$\frac{e^{-(\alpha_B x+\lambda_B)\Delta}[(\alpha_B x+\lambda_B)\Delta]^{\,y}/y!}{e^{-(\alpha_E x+\lambda_E)\Delta}[(\alpha_E x+\lambda_E)\Delta]^{\,z}/z!} = \frac{p(y|z)}{p(z|y)}. \tag{E.20}$$

To reach a contradiction, let us choose $y = z = 1$. Now, it is sufficient to show that the LHS of (E.20) is not a constant function in x. To this end, let $h(x)$ denote LHS (E.20) for $y = z = 1$. In this case, we have $h(x) = e^{[(\alpha_E-\alpha_B)x+(\lambda_E-\lambda_B)]\Delta}\frac{\alpha_B x+\lambda_B}{\alpha_E x+\lambda_E}$. It is clear that $h(x)$ is not a constant function in x, for $x \in \mathcal{S}_{F_X}$. This is because at leas one of the inequalities in (5.5) or (5.6) is strict. Therefore, we reach a contradiction. This, in turn, implies that $I(Q;Y|Z) > 0$ and as a result, $I(X;Y|Z)$ is strictly concave in F_X. Furthermore, the output distributions are unique, i.e., if F_{X_1} and F_{X_2} are both secrecy-capacity-achieving, then $p_Y(y;F_{X_1}) = p_Y(y;F_{X_2})$ and $p_Z(z;F_{X_1}) = p_Z(z;F_{X_2})$. ■

Lemma 7 implies that the answer to the optimization problem in (5.13) for $\mathcal{F}^+ = \Omega^+_{\mathcal{A},\,\mathcal{E}}$, denoted by F_X^*, is unique.

Lemma 8. *The functional $f_0(F_X)$ is weakly differentiable in $\Omega^+_{\mathcal{A},\,\mathcal{E}}$ and its weak derivative at the point F_X^o, denoted by $f_0'(F_X^o)$ is given by*

$$f_0'(F_X, F_X^o) \triangleq \lim_{t\to 0}\frac{f_0((1-t)F_X^o + tF_X) - f_0(F_X^o)}{t} = \int_0^{\mathcal{A}} c_S(x;F_X^o)\,dF_X(x) - f(F_X^o), \tag{E.21}$$

where $t \in [0,1]$.

Proof. The proof is based on the definition of the weak derivative and follows along similar lines as the one in [25]. ■

From Lemma 5, Lemma 7, and Lemma 8, we have a strictly concave and weak-differentiable function $f_0(F_X)$ over $\Omega^+_{\mathcal{A},\,\mathcal{E}}$ which is a convex set, then the necessary and sufficient conditions for an input

distribution F_X^* to be optimal is

$$f_0'(F_X, F_X^*) \leq 0, \quad \forall\ F_X,\ F_X^* \in \Omega^+_{\mathcal{A},\,\mathcal{E}}. \tag{E.22}$$

Now, we define the mapping

$$g(F_X) = \int_0^{\mathcal{A}} x\, dF_X(x) - \mathcal{E}, \tag{E.23}$$

from $\Omega^+_{\mathcal{A},\,\mathcal{E}}$ to $\mathbb{R}$. This mapping is linear in F_X and hence convex. Furthermore, the weak-derivative of $g(F_X)$ at the point F_X^o is given by

$$g'(F_X, F_X^o) = g(F_X) - g(F_X^o). \tag{E.24}$$

Using the Lagrangian Theorem, and noting that $f_0(F_X) - \gamma g(F_X)$ (where $\gamma \geq 0$ is the Lagrangian coefficient) is weakly differentiable and strictly concave in F_X, the necessary and sufficient conditions for $F_X^* \in \Omega^+_{\mathcal{A},\,\mathcal{E}}$ to be optimal is

$$f_0'(F_X, F_X^*) - \gamma g'(F_X, F_X^*) \leq 0, \quad \forall\ F_X,\ F_X^* \in \Omega^+_{\mathcal{A},\,\mathcal{E}}, \tag{E.25}$$

that is

$$\int_0^{\mathcal{A}} \left[c_S(x; F_X^*) - \gamma x\right]\, dF_X(x) \leq C_S - \gamma\mathcal{E}, \tag{E.26}$$

where the secrecy capacity is $C_S = I_B(F_X^*) - I_E(F_X^*) = f_0(F_X^*)$. Next, we present a theorem which states the KKT conditions for the optimality of $F_X^* \in \Omega^+_{\mathcal{A},\,\mathcal{E}}$.

Theorem 11. *Let $\mathcal{S}_{F_X^*} \subset [0, \mathcal{A}]$ be the support set of F_X^*, then*

$$\int_0^{\mathcal{A}} \left[c_S(x; F_X^*) - \gamma x\right]\, dF_X(x) \leq C_S - \gamma\mathcal{E}, \tag{E.27}$$

for all $F_X \in \Omega^+_{\mathcal{A},\,\mathcal{E}}$ if and only if

$$c_S(x; F_X^*) - \gamma x \leq C_S - \gamma\mathcal{E}, \quad \forall\ x \in [0, \mathcal{A}], \tag{E.28}$$

$$c_S(x; F_X^*) - \gamma x = C_S - \gamma\mathcal{E}, \quad \forall\ x \in \mathcal{S}_{F_X^*}. \tag{E.29}$$

Proof. The implication from (E.28) to (E.27) is immediate. For the converse, assume (E.28) is false. Then there exists an $\hat{x}$ such that

$$c_S(\hat{x}; F_X^*) > C_S + \gamma(\hat{x} - \mathcal{E}). \tag{E.30}$$

If $F_X(x) = u(x - \hat{x})$, where $u(\cdot)$ is the unit step function, then

$$\int_0^{\mathcal{A}} [c_S(x; F_X^*) - \gamma x] \, dF_X(x) = c_S(\hat{x}; F_X^*) - \gamma \hat{x} > C_S - \gamma \mathcal{E}, \tag{E.31}$$

which contradicts (E.27). Now, assume that (E.28) is true, but (E.29) is false, i.e., there exists $\hat{x} \in \mathcal{S}_{F_X^*}$ such that

$$c_S(\hat{x}; F_X^*) < C_S + \gamma(\hat{x} - \mathcal{E}). \tag{E.32}$$

Since all the functions in the above equation are continuous in x, the inequality is satisfied strictly on a neighborhood $\mathcal{S}'$ of $\hat{x}$. Now, by definition of a support set, the set $\mathcal{S}'$ necessarily satisfies $\int_{\mathcal{S}'} dF_X^*(x) = \epsilon \in [0, 1]$. Hence,

$$\begin{aligned} C_S - \gamma \mathcal{E} = f_0(F_X^*) - \gamma \mathcal{E} &= \int_0^{\mathcal{A}} [c_S(x; F_X^*) - \gamma x] \, dF_X^*(x) \\ &= \int_{\mathcal{S}'} [c_S(x; F_X^*) - \gamma x] dF_X^*(x) + \int_{\mathcal{S}_{F_X^*} - \mathcal{S}'} [c_S(x; F_X^*) - \gamma x] dF_X^*(x) \\ &< \epsilon(C_S - \gamma \mathcal{E}) + (1 - \epsilon)(C_S - \gamma \mathcal{E}) < (C_S - \gamma \mathcal{E}), \end{aligned} \tag{E.33}$$

which is a contradiction, and hence the result follows. ■

We now prove by contradiction that the secrecy-capacity-achieving input distribution F_X^* has a finite number of mass points. To reach a contradiction, we use the KKT conditions in (E.28)–(E.29). To this end, the following lemma establishes that both $i_B(x; F_X)$ and $i_E(x; F_X)$ have analytic extensions over some open connected set in the complex plane $\mathbb{C}$.

Lemma 9. *The secrecy rate density $c_S(x; F_X) - \gamma x$ has an analytic extension to the open connected set $\mathcal{O} \stackrel{\triangle}{=} \{w \in \mathbb{C} : \Re(w) > -\frac{\lambda_B}{\alpha_B}\}$, where $\Re(w)$ is the real part of the complex variable w.*

Proof. The mutual information densities $i_B(w; F_X)$ and $i_E(w; F_X)$ have analytic extension to the open connected sets $\mathcal{O}_B \stackrel{\triangle}{=} \{w \in \mathbb{C} : \Re(w) > -\frac{\lambda_B}{\alpha_B}\}$ and $\mathcal{O}_E \stackrel{\triangle}{=} \{w \in \mathbb{C} : \Re(w) > -\frac{\lambda_E}{\alpha_E}\}$, respectively, according to [7]. Therefore, the secrecy rate density $c_S(w; F_X) - \gamma w$ has an analytic extension to the open connected set $\mathcal{O} = \mathcal{O}_B \cap \mathcal{O}_E$. Since $\frac{\lambda_E}{\alpha_E} \geq \frac{\lambda_B}{\alpha_B}$ (based on (5.6)), we have $\mathcal{O} = \mathcal{O}_B$. This completes the proof of Lemma 9. ■

Now, we are ready to prove the discreteness and finiteness of the support set of F_X^* using a contradiction argument. We start by assuming that $\mathcal{S}_{F_X^*}$ has an infinite number of elements. In view of the optimality condition (E.29), the analyticity of $c_S(w; F_X) - \gamma w$ over $\mathcal{O}$ and the Identity Theorem from

complex analysis along with Bolzano-Weierstrass Theorem, if $\mathcal{S}_{F_X^*}$ has an infinite number of mass points, we deduce that $r_e(w; F_X^*) - \gamma w = C_S - \gamma\mathcal{E}$ for all $w \in \mathcal{O}$. Since $(-\frac{\lambda_B}{\alpha_B}, +\infty) \subset \mathcal{O}$, we conclude that

$$c_S(x; F_X^*) - \gamma x = C_S - \gamma\mathcal{E}, \quad \forall\, x > -\frac{\lambda_B}{\alpha_B}. \tag{E.34}$$

Next, we show that (E.34) results in a contradiction. Observe that (E.34) implies that $c_S(x; F_X^*) - \gamma x$ is a constant function in x for all $x \in (-\frac{\lambda_B}{\alpha_B}, +\infty)$. Therefore, to reach a contradiction, we show that $c_S(x; F_X^*) - \gamma x$ is not a constant function over this interval. To that end, we take the derivative of both sides of (E.34) with respect to x and we find

$$\frac{dc_S(x; F_X^*)}{dx} = \gamma, \quad \forall\, x > -\frac{\lambda_B}{\alpha_B}. \tag{E.35}$$

Substituting (E.6)–(E.7) into (E.3) and taking the derivative with respect to x, we can write

$$\begin{aligned}\frac{dc_S(x; F_X^*)}{dx} =& \alpha_B \log[(\alpha_B x + \lambda_B)\Delta] + \alpha_B \sum_{y=0}^{+\infty} p(y|x) \log \frac{g_B(y; F_X^*)}{g_B(y+1; F_X^*)} \\ &- \alpha_E \log[(\alpha_E x + \lambda_E)\Delta] - \alpha_E \sum_{z=0}^{+\infty} p(z|x) \log \frac{g_E(z; F_X^*)}{g_E(z+1; F_X^*)}, \forall\, x > -\frac{\lambda_B}{\alpha_B}. \end{aligned} \tag{E.36}$$

It can be easily shown that

$$\lambda_B \Delta \leq \frac{g_B(y+1; F_X^*)}{g_B(y; F_X^*)} \leq (\alpha_B \mathcal{A} + \lambda_B)\Delta, \tag{E.37}$$

$$\lambda_E \Delta \leq \frac{g_E(z+1; F_X^*)}{g_E(z; F_X^*)} \leq (\alpha_E \mathcal{A} + \lambda_E)\Delta, \tag{E.38}$$

Using the bounds in (E.37)–(E.38), one obtains

$$\begin{aligned}\frac{dc_S(x; F_X^*)}{dx} \geq& (\alpha_B - \alpha_E) \log[(\alpha_B x + \lambda_B)\Delta] + \alpha_E \log \frac{\alpha_B x + \lambda_B}{\alpha_E x + \lambda_E} - \alpha_B \log[(\alpha_B \mathcal{A} + \lambda_B)\Delta] \\ &+ \alpha_E \log(\lambda_E \Delta) \\ =& (\alpha_B - \alpha_E) \log \frac{\alpha_B x + \lambda_B}{\alpha_B \mathcal{A} + \lambda_B} + \alpha_E \log \frac{\alpha_B x + \lambda_B}{\alpha_E x + \lambda_E} + \alpha_E \log \frac{\lambda_E}{\alpha_B \mathcal{A} + \lambda_B}, \forall\, x > -\frac{\lambda_B}{\alpha_B}. \end{aligned} \tag{E.39}$$

Finally, we consider two cases and for each case we provide a contradiction argument.

- Case 1: $\alpha_B > \alpha_E$

  In this case, we note that for sufficiently large values of x, the right-hand-side (RHS) of (E.39) scales logarithmic in x, i.e., $\frac{dc_S(x; F_X^*)}{dx} = \Omega(\log x)$ which means that there exist constants $c > 0$ and

$x_0 > -\frac{\lambda_B}{\alpha_B}$ such that $\frac{dc_S(x;F_X^*)}{dx} \geq c \log x$ for all $x > x_0$. However, this results in a contradiction since based on (E.35), $\frac{dc_S(x;F_X^*)}{dx}$ must be a constant function in x for all $x > -\frac{\lambda_B}{\alpha_B}$.

- Case 2: $\alpha_B = \alpha_E$

  For this case, using the bounds in (E.37)–(E.38), we first upper bound $\frac{dc_S(x;F_X^*)}{dx}$ as follows

$$\begin{aligned}\frac{dc_S(x;F_X^*)}{dx} &\leq (\alpha_B - \alpha_E) \log \frac{\alpha_B x + \lambda_B}{\alpha_E \mathcal{A} + \lambda_E} + \alpha_E \log \frac{\alpha_B x + \lambda_B}{\alpha_E x + \lambda_E} + \alpha_B \log \frac{\alpha_E \mathcal{A} + \lambda_E}{\lambda_B} \\ &= \alpha_B \log \frac{x + \frac{\lambda_B}{\alpha_B}}{x + \frac{\lambda_E}{\alpha_E}} + \alpha_B \log \frac{\alpha_E \mathcal{A} + \lambda_E}{\lambda_B}, \quad \forall\ x > -\frac{\lambda_B}{\alpha_B}. \end{aligned} \tag{E.40}$$

  Recall that at least one of the inequalities in (5.5)–(5.6) is strict (due to the degradedness assumption). Therefore, in this case, (5.6) is strict. Now, to reach a contradiction, it suffices to compute the limit of the RHS of (E.40) as $x \to -\frac{\lambda_B}{\alpha_B}^+$. For this purpose and in regard of (E.35), we have

$$\gamma \leq \lim_{x \to -\frac{\lambda_B}{\alpha_B}^+} \alpha_B \log \frac{x + \frac{\lambda_B}{\alpha_B}}{x + \frac{\lambda_E}{\alpha_E}} + \alpha_B \log \frac{\alpha_E \mathcal{A} + \lambda_E}{\lambda_B}. \tag{E.41}$$

  Observe that since $\frac{\lambda_E}{\alpha_E} > \frac{\lambda_B}{\alpha_B}$, the limit $\lim_{x \to -\frac{\lambda_B}{\alpha_B}^+} \log \frac{x + \frac{\lambda_B}{\alpha_B}}{x + \frac{\lambda_E}{\alpha_E}} = -\infty$ and therefore, we get $\gamma \leq -\infty$ which is a contradiction because γ is a nonnegative constant.

Hence, for each case we reach a contradiction which implies that the support set $\mathcal{S}_{F_X^*}$ must have finitely many mass points in the interval $[0, \mathcal{A}]$. This completes the proof of Theorem 4.

We note that since our bounds in (E.39) and (E.40) do not depend on Δ, the result also holds true for the case where $\Delta \to 0$. That is, the optimal input distribution for the degraded continuous-time PWC with nonnegativity, peak- and average-intensity constraints is also discrete with a finite number of mass points in the interval $[0, \mathcal{A}]$.

E.3 Proof of Theorem 5

This section presents the proof of Theorem 5 by extending the analysis in the previous section to the case where only an average-intensity constraint is active. We start the proof by noting that the feasible set $\Omega_{\mathcal{E}}^+$ is convex and sequentially compact in the Lévy metric sense [37, Appendix I.A]. Furthermore, the functional $f_0 : \Omega_{\mathcal{E}}^+ \to \mathbb{R}, f_0(F_X) = I_B(F_X) - I_E(F_X)$ is continuous in F_X. This is because each one of the mutual information terms $I_B(F_X)$ and $I_E(F_X)$ are continuous in F_X based on [20, Lemma 17]. Therefore, we conclude that the supremum in (5.13) for $\mathcal{F}^+ = \Omega_{\mathcal{E}}^+$ is achieved by at least one element $F_X \in \Omega_{\mathcal{E}}^+$. Furthermore, the functional $f_0(F_X)$ is strictly concave, and weakly differentiable by following

along similar lines of Lemma 7 and Lemma 8. Hence, the maximum is achieved by a unique distribution. Finally, invoking similar arguments that appear in the statement of Theorem 11, we find the following necessary and sufficient KKT conditions for the optimality of the input distribution F_X^* as

$$c_S(x; F_X^*) - \gamma x \leq C_S - \gamma \mathcal{E}, \quad \forall\, x \in [0, +\infty), \tag{E.42}$$

$$c_S(x; F_X^*) - \gamma x = C_S - \gamma \mathcal{E}, \quad \forall\, x \in \mathcal{S}_{F_X^*}. \tag{E.43}$$

Next, we prove that the secrecy-capacity-achieving input distribution F_X^* has the following structural properties: 1) the intersection of $\mathcal{S}_{F_X^*}$ with any bounded interval B contains a finite number of mass points, i.e., $|\mathcal{S}_{F_X^*} \cap B| < \infty$; 2) the support set of the optimal distribution is an unbounded set. These two properties imply that $\mathcal{S}_{F_X^*}$ is a countably infinite set. The first property is shown by means of contradiction. We assume, on the contrary, that for some bounded interval B, $\mathcal{S}_{F_X^*} \cap B$ contains an infinite number of elements. Then, using the KKT conditions in (E.42)–(E.43), the analyticity of the secrecy rate density $c_S(x; F_X^*)$ over $\mathcal{O}$, and invoking the Bolzano-Weierstrass and Identity Theorems, we find that $\gamma \leq -\infty$ which is not possible, and hence results in a contradiction. The second property is also shown through a contradiction approach. We consider two cases for the channel gains α_B and α_E and for each case, we provide a contradiction arguments. These cases are as follows: 1) when $\alpha_B > \alpha_E$, our contradiction hinges on the fact that if $\mathcal{S}_{F_X^*}$ is a bounded set, then the cost function which grows linearly in x must be lower bounded by the secrecy rate density which grows as fast as $x \log x$. This is not possible for large values of x and hence a contradiction occurs; 2) when the channel gains are identical, we find that the Lagrangian multiplier must be lower bounded by a constant and thus, using the Envelope Theorem [39] we observe that the secrecy capacity must at least grow linearly in the average-intensity constraint. However, in this section, it is shown that the secrecy capacity is always upper bounded by a constant for all values of the average-intensity. Therefore, the desired contradiction is reached and the result follows.

E.3.1 The support set of the optimal solution has finitely many mass points in any bounded interval

Let B be a bounded interval and assume, to the contrary, that $\mathcal{S}_{F_X^*} \cap B$ has an infinite number of elements. Now based on the optimality equation (E.43), the analyticity of $c_S(x; F_X^*)$ over $\mathcal{O}$, and the

Bolzano-Weierstrass and Identity Theorems from complex analysis, one can find

$$c_S(x; F_X^*) - \gamma x = C_S - \gamma \mathcal{E}, \quad \forall\, x > -\frac{\lambda_B}{\alpha_B}. \tag{E.44}$$

Next, we show that this results in a contradiction. To this end, we note that

$$\begin{aligned} g_E(z+1; F_X^*) &= \int_0^{+\infty} e^{-\alpha_E x \Delta} \left[(\alpha_E x + \lambda_E)\Delta\right]^{z+1} dF_X(x) = e^{\lambda_E \Delta}(z+1)! \underbrace{\int_0^{+\infty} p(z|x)\, dF_X^*(x)}_{\leq 1 \text{ as } p(z|x) \leq 1} \\ &\leq e^{\lambda_E \Delta}(z+1)!. \end{aligned} \tag{E.45}$$

Furthermore, observe that

$$g_E(z; F_X^*) \geq (\lambda_E \Delta)^z\, \mathbb{E}_{F_X^*}[e^{-\alpha_E X \Delta}] \overset{(i)}{\geq} (\lambda_E \Delta)^z\, e^{-\alpha_E \Delta\, \mathbb{E}_{F_X^*}[X]} = (\lambda_E \Delta)^z\, e^{-\alpha_E \mathcal{E} \Delta}, \tag{E.46}$$

where (i) is due to the Jensen's Inequality as $e^{-\alpha_B x \Delta}$ is a convex function in x. Plugging the bounds in (E.45)–(E.46) into (E.36), we get

$$\begin{aligned} \frac{dc_S(x; F_X^*)}{dx} \leq\, & (\alpha_B - \alpha_E) \log[(\alpha_B x + \lambda_B)\Delta] + \alpha_E \log \frac{\alpha_B x + \lambda_B}{\alpha_E x + \lambda_E} \\ & + \alpha_B \underbrace{\sum_{y=0}^{+\infty} p(y|x) \log \frac{g_B(y; F_X^*)}{g_B(y+1; F_X^*)}}_{\triangleq \Xi_B(x)} + \alpha_E \underbrace{\sum_{z=0}^{+\infty} p(z|x) \log \frac{e^{\lambda_E \Delta}(z+1)!}{e^{-\alpha_E \mathcal{E} \Delta}(\lambda_E \Delta)^z}}_{\triangleq \Xi_E(x)}. \end{aligned} \tag{E.47}$$

Next, we provide upper bounds on $\Xi_B(x)$ and $\Xi_E(x)$ as follows

$$\Xi_B(x) \overset{(ii)}{\leq} \sum_{y=0}^{+\infty} p(y|x) \log \frac{1}{\lambda_B \Delta} = -\log(\lambda_B \Delta) \tag{E.48}$$

$$\begin{aligned} \Xi_E(x) &= \mathbb{E}_{Z|X}[\log(Z+1)! - Z\, \log(\lambda_E \Delta)] + (\alpha_E \mathcal{E} + \lambda_E)\Delta \\ &= \mathbb{E}_{Z|X}[\log\, Z!] + \mathbb{E}_{Z|X}[\log(Z+1)] - [(\alpha_E x + \lambda_E)\Delta] \log(\lambda_E \Delta) + (\alpha_E \mathcal{E} + \lambda_E)\Delta \\ &\overset{(iii)}{\leq} \mathbb{E}_{Z|X}[\log\, Z!] + \log(\mathbb{E}_{Z|X}[Z] + 1) - [(\alpha_E x + \lambda_E)\Delta] \log(\lambda_E \Delta) + (\alpha_E \mathcal{E} + \lambda_E)\Delta \\ &\overset{(iv)}{\leq} \frac{1}{2} \log[2\pi e(\mathbb{E}_{Z|X}[Z] + \frac{1}{12})] - \mathbb{E}_{Z|X}[Z] + \mathbb{E}_{Z|X}[Z] \log(\mathbb{E}_{Z|X}[Z]) + \log(\mathbb{E}_{Z|X}[Z] + 1) \\ &\quad - [(\alpha_E x + \lambda_E)\Delta] \log(\lambda_E \Delta) + (\alpha_E \mathcal{E} + \lambda_E)\Delta \\ &\leq [(\alpha_E x + \lambda_E)\Delta] \log[(\alpha_E x + \lambda_E)\Delta] - [(\alpha_E x + \lambda_E)\Delta](1 + \log(\lambda_E \Delta)) \\ &\quad + \frac{3}{2} \log[(\alpha_E x + \lambda_E)\Delta + 1] + (\alpha_E \mathcal{E} + \lambda_E)\Delta + \frac{1}{2} \log(2\pi e), \end{aligned} \tag{E.49}$$

where (ii) follows from (E.37), (iii) is due to the Jensen's Inequality as $\log x$ is a concave function, and (iv) follows from an upper bound on the entropy of the Poisson random variable [8, Lemma 10]. Combining (E.47)–(E.49), we get

$$\begin{aligned}\frac{dc_S(x;F_X^*)}{dx} \leq\,& (\alpha_B-\alpha_E)\log[(\alpha_B x+\lambda_B)\Delta] + \alpha_E \log\frac{x+\frac{\lambda_B}{\alpha_B}}{x+\frac{\lambda_E}{\alpha_E}} + \alpha_E\log\frac{\alpha_B}{\alpha_E}\\ &+\alpha_E\Big([(\alpha_E x+\lambda_E)\Delta]\log[(\alpha_E x+\lambda_E)\Delta] - [(\alpha_E x+\lambda_E)\Delta](1+\log(\lambda_E\Delta))\\ &+\frac{3}{2}\log[(\alpha_E x+\lambda_E)\Delta+1] + (\alpha_E\mathcal{E}+\lambda_E)\Delta + \frac{1}{2}\log(2\pi e)\Big)\\ &-\alpha_B\log(\lambda_B\Delta), \quad \forall\, x > -\frac{\lambda_B}{\alpha_B}. \end{aligned} \tag{E.50}$$

In order to see a contradiction it suffices to compute the limit of the RHS of (E.50) as $x\to -\frac{\lambda_B}{\alpha_B}^+$. For this purpose and in regard of (E.35) and (E.50), we have

$$\begin{aligned}\gamma \leq\,& \lim_{x\to-\frac{\lambda_B}{\alpha_B}^+}(\alpha_B-\alpha_E)\log\frac{\alpha_B x+\lambda_B}{\lambda_B} + \lim_{x\to-\frac{\lambda_B}{\alpha_B}^+}\alpha_E\log\frac{x+\frac{\lambda_B}{\alpha_B}}{x+\frac{\lambda_E}{\alpha_E}} + \alpha_E\log\frac{\alpha_B}{\alpha_E}\\ &+\alpha_E\underbrace{\lim_{x\to-\frac{\lambda_B}{\alpha_B}^+}[[(\alpha_E x+\lambda_E)\Delta]\log[(\alpha_E x+\lambda_E)\Delta] - [(\alpha_E x+\lambda_E)\Delta](1+\log(\lambda_E\Delta))]}_{\text{finite value for } \frac{\lambda_E}{\alpha_E}\geq\frac{\lambda_B}{\alpha_B}}\\ &+\alpha_E\underbrace{\lim_{x\to-\frac{\lambda_B}{\alpha_B}^+}\left[\frac{3}{2}\log[(\alpha_E x+\lambda_E)\Delta+1] + (\alpha_E\mathcal{E}+\lambda_E)\Delta + \frac{1}{2}\log(2\pi e)\right]}_{\text{finite value for } \frac{\lambda_E}{\alpha_E}\geq\frac{\lambda_B}{\alpha_B}}\\ &-\alpha_E\log(\lambda_B\Delta). \end{aligned} \tag{E.51}$$

Thus, we obtain that $\gamma\leq-\infty$ which is a contradiction as γ is a nonnegative constant. Therefore, the $\mathcal{S}_{F_X^*}\cap B$ has a finite cardinality. This implies that the optimal input distribution F_X^* possess a countably finite number of mass points in any bounded interval. It is noteworthy that the upper bound in (E.51) depends on Δ. Therefore, in this case we cannot conclude that the secrecy-capacity-achieving distribution of the continuous-time PWC with nonnegativity and average-intensity constraints admits a finite number of mass points in any bounded interval.

E.3.2 The support set of the optimal distribution $\mathcal{S}_{F_X^*}$ is unbounded

To prove this, we again resort to a contradiction approach. Assume, to the contrary, that $\mathcal{S}_{F_X^*}$ is a bounded set, i.e., $\mathcal{S}_{F_X^*}\subseteq[0,h]$, where h is some finite positive constant. In the previous section, we proved that the intersection of $\mathcal{S}_{F_X^*}$ with any bounded interval has a finite cardinality. Since, we are assuming

that $\mathcal{S}_{F_X^*}$ is bounded, thus, it has a finite cardinality. This implies that $F_X^*(x) = \sum_{i=1}^N p_i u(x - x_i)$, where $N < +\infty$, $0 \leq x_1 < x_2 < \cdots < x_N \leq h$ are the mass points with corresponding probabilities $\{p_1, \ldots, p_N\}$. Furthermore, we can write

$$g_E(z; F_X^*) = \int_0^h e^{-\alpha_E x \Delta} \left[(\alpha_E x + \lambda_E)\Delta\right]^z dF_X^*(x) = \sum_{i=1}^N p_i e^{-\alpha_E x_i \Delta} \left[(\alpha_E x_i + \lambda_E)\Delta\right]^z$$
$$> p_N e^{-\alpha_E x_N \Delta} \left[(\alpha_E x_N + \lambda_E)\Delta\right]^z. \tag{E.52}$$

$$g_B(y; F_X^*) = \int_0^h e^{-\alpha_B x \Delta} \left[(\alpha_B x + \lambda_B)\Delta\right]^y dF_X^*(x) = \sum_{i=1}^N p_i e^{-\alpha_B x_i \Delta} \left[(\alpha_B x_i + \lambda_B)\Delta\right]^y$$
$$\leq \left[(\alpha_B x_N + \lambda_B)\Delta\right]^y \tag{E.53}$$

Therefore, $\log g_E(z; F_X^*) > \log p_N - \alpha_E x_N \Delta + z \log[(\alpha_E x_N + \lambda_E)\Delta]$ and $\log g_B(y; F_X^*) \leq y \log[(\alpha_B x_N + \lambda_B)\Delta]$. In light of the optimality equation (E.42) and using these bounds we obtain

$$\begin{aligned}
C_S + \gamma(x - \mathcal{E}) \geq\, & (\alpha_B x + \lambda_B) \log[(\alpha_B x + \lambda_B)\Delta] - (\alpha_E x + \lambda_E) \log[(\alpha_E x + \lambda_E)\Delta] \\
& + (\alpha_E - \alpha_B)x + \frac{1}{\Delta} \sum_{z=0}^{+\infty} p(z|x) \log g_E(z; F_X^*) - \frac{1}{\Delta} \sum_{y=0}^{+\infty} p(y|x) \log g_B(y; F_X^*) \\
> \,& (\alpha_B x + \lambda_B) \log[(\alpha_B x + \lambda_B)\Delta] - (\alpha_E x + \lambda_E) \log[(\alpha_E x + \lambda_E)\Delta] \\
& + (\alpha_E - \alpha_B)x + \frac{\log p_N}{\Delta} - \alpha_E x_N + (\alpha_E x + \lambda_E) \log[(\alpha_E x_N + \lambda_E)\Delta] \\
& - (\alpha_B x + \lambda_B) \log[(\alpha_B x_N + \lambda_B)\Delta] \\
= \,& (\alpha_B - \alpha_E)x \log[(\alpha_B x + \lambda_B)\Delta] + \alpha_E x \log \frac{\alpha_B x + \lambda_B}{\alpha_E x + \lambda_E} \\
& + x \left[(\alpha_E - \alpha_B) + (\alpha_E - \alpha_B) \log[(\alpha_E x_N + \lambda_E)\Delta] + \alpha_B \log \frac{\alpha_E x_N + \lambda_E}{\alpha_B x_N + \lambda_B}\right] \\
& + \lambda_B \log \frac{\alpha_B x + \lambda_B}{\alpha_B x_N + \lambda_B} - \lambda_E \log \frac{\alpha_E x + \lambda_E}{\alpha_E x_N + \lambda_E} + \frac{\log p_N}{\Delta} - \alpha_E x_N, \quad \forall\, x \geq 0.
\end{aligned} \tag{E.54}$$

Now, we consider the following cases and for each case we provide a contradiction argument.

- Case 1: $\alpha_B > \alpha_E$

  Observe that in this case, the RHS of (E.54) scales like $x \log x$ for sufficiently large values of x, i.e., $C_S + \gamma(x - \mathcal{E}) = \Omega(x \log x)$. However, this is clearly a contradiction because $C_S + \gamma(x - \mathcal{E})$ grows linearly in x. Thus, the optimal support set $\mathcal{S}_{F_X^*}$ must be an unbounded set.

- Case 2: $\alpha_B = \alpha_E$

In this case, (E.54) can be simplified further as

$$C_S + \gamma(x - \mathcal{E}) \geq \alpha_B x \left[\log \frac{x_N + \frac{\lambda_E}{\alpha_E}}{x_N + \frac{\lambda_B}{\alpha_B}} + \log \frac{x + \frac{\lambda_B}{\alpha_B}}{x + \frac{\lambda_E}{\alpha_E}} \right] + \lambda_B \log \frac{x + \frac{\lambda_B}{\alpha_B}}{x_N + \frac{\lambda_B}{\alpha_B}} - \lambda_E \log \frac{x + \frac{\lambda_E}{\alpha_E}}{x_N + \frac{\lambda_E}{\alpha_E}} + \frac{\log p_N}{\Delta} - \alpha_E x_N, \ \forall\, x \geq 0. \tag{E.55}$$

Observe that the RHS of (E.55) grows linearly in x for large values of x. Thus, dividing the sides of (E.55) into $x > 0$ and taking the limit as $x \to \infty$, we find

$$\gamma \geq \alpha_B \log \frac{x_N + \frac{\lambda_E}{\alpha_E}}{x_N + \frac{\lambda_B}{\alpha_B}}. \tag{E.56}$$

We note that since $\alpha_B = \alpha_E$, the inequality in (5.6) is strict, i.e., $\frac{\lambda_E}{\alpha_E} > \frac{\lambda_B}{\alpha_B}$ and therefore, $\alpha_B \log \frac{x_N + \frac{\lambda_E}{\alpha_E}}{x_N + \frac{\lambda_B}{\alpha_B}} > 0$. Next, we show that this lower bound on the Lagrangian multiplier γ results in a contradiction. To that end, we first note that the Lagrangian multiplier γ and the location of the last mass point in the support set of the optimal distribution depend on the value of the average-intensity constraint. Thus, in (E.56) one must replace γ by $\gamma(\mathcal{E})$ and x_N by $x_N(\mathcal{E})$. Now, we recall the Envelope Theorem [39] which shows that the Lagrangian multiplier γ and the secrecy capacity (the optimal value of the objective functional) are related as follows

$$\frac{dC_S(\mathcal{E})}{d\mathcal{E}} = \gamma(\mathcal{E}). \tag{E.57}$$

In light of this relationship and the lower bound in (E.56), the following lower bound can be found

$$C_S(\mathcal{E}) = \int_0^{\mathcal{E}} \gamma(t)\, dt \geq \int_0^{\mathcal{E}} \alpha_B \log \frac{x_N(t) + \frac{\lambda_E}{\alpha_E}}{x_N(t) + \frac{\lambda_B}{\alpha_B}}\, dt = \int_0^{\mathcal{E}} \alpha_B \log \left[1 + \frac{\frac{\lambda_E}{\alpha_E} - \frac{\lambda_B}{\alpha_B}}{x_N(t) + \frac{\lambda_B}{\alpha_B}} \right] dt. \tag{E.58}$$

Now, based on the contradiction assumption we have $x_N(\mathcal{E}) < h$ with h being a finite positive constant. Therefore, (E.58) can be further lower bounded as

$$C_S(\mathcal{E}) \geq \int_0^{\mathcal{E}} \alpha_B \log \left[1 + \frac{\frac{\lambda_E}{\alpha_E} - \frac{\lambda_B}{\alpha_B}}{h + \frac{\lambda_B}{\alpha_B}} \right] dt = \alpha_B \log \left[1 + \frac{\frac{\lambda_E}{\alpha_E} - \frac{\lambda_B}{\alpha_B}}{h + \frac{\lambda_B}{\alpha_B}} \right] \mathcal{E}, \tag{E.59}$$

which must hold for all $\mathcal{E} \geq 0$. Since $h > 0$ and $\frac{\lambda_E}{\alpha_E} > \frac{\lambda_B}{\alpha_B}$, the logarithm term is always positive implying that $C_S(\mathcal{E})$ must at least grow linearly in $\mathcal{E}$. However, later in this section, it is shown that the secrecy capacity of the DT–PWC with nonnegativity and average-intensity constraints when $\alpha_B = \alpha_E$ is upper bounded by a constant for all $\mathcal{E} \geq 0$. Therefore, the implication in (E.59) results

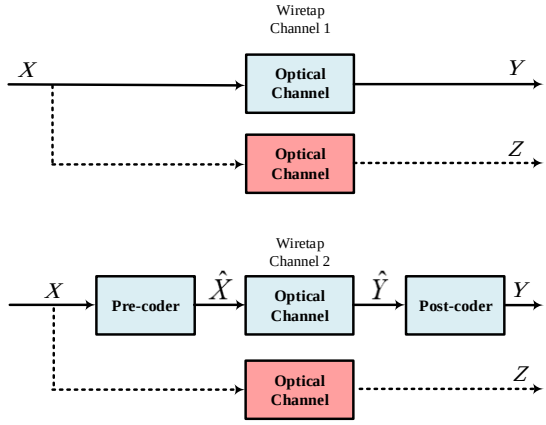

Figure E.1: Two discrete-time Poisson wiretap channels.

in a contradiction. This implies that the optimal support set $\mathcal{S}_{F_X^*}$ must be an unbounded set.

Showing that $\mathcal{S}_{F_X^*}$ is an unbounded set for these considered cases completes the proof of Theorem 5.

E.4 Proof of Proposition 4

Suppose, to the contrary, that $x = 0$ does not belong to the support set of the optimal input distribution $\mathcal{S}_{F_X^*}$. Let $0 < x_1 \leq x_2 \leq \ldots \leq x_N \leq \infty$ be the mass points in the set $\mathcal{S}_{F_X^*}$. Consider two DT–PWC depicted in Figure E.1. Wiretap channel 1 is the original optical wiretap channel, and wiretap channel 2 is obtained from wiretap channel 1 by appending a pre-coder and a post-coder before and after the inner optical channel in the legitimate user's link. Specifically, $\hat{X} = X - x_1$ and $Y|X = \hat{Y}|X + \hat{N}_B$, where $\hat{N}_B$ is a Poisson random variable with mean $\alpha_B x_1 \Delta$ and is independent from $\hat{Y}|X$. For any $x \geq x_1$, the conditional probability density functions $p(y|x)$ and $p(z|x)$ are the same in both wiretap channels. Thus, the joint probability density functions of $p(y, x)$ and $p(z, x)$ in the two wiretap channels are also the same, if the input distribution is F_X^*. As a result, C_S is identical in both wiretap channels.

In the second wiretap channel, as $X, \hat{X}, \hat{Y}, Y$ and Z form the Markov chain $X \to \hat{X} \to \hat{Y} \to Y \to Z$, we have $I(\hat{X}; \hat{Y}|Z) \geq I(X; Y|Z)$ by the data processing inequality. This indicates that $I(\hat{X}; \hat{Y}) - I(\hat{X}; Z) \geq I(X; Y) - I(X; Z)$. Now, let $F_{\hat{X}}^*$ be the distribution function of $\hat{X}$ when the distribution function of X is F_X^*. Clearly, $F_{\hat{X}}^*$ satisfies either of the constraints in (5.10)–(5.12) that are active. Hence, $F_{\hat{X}}^*$ is also secrecy-capacity-achieving for wiretap channel 1. Based on Lemma 7, the secrecy-capacity-achieving output distribution is unique, as a result, $p_Y(y; F_X^*) = p_Y(y; F_{\hat{X}}^*)$. Therefore, for wiretap channel 2, given the input distribution function of X is F_X^*, the probability density functions for $Y|X$ and $\hat{Y}|X$ are the same, which is not possible since $\mathbb{E}[Y|X] = \mathbb{E}[\hat{Y}|X] + \alpha_B x_1 \Delta$. Hence, we reach a contradiction and the

proposition follows.

E.5 Proof of Theorem 6

We start the proof by noting that the feasible set $\Omega^{+}_{\mathcal{A},\,\mathcal{E}}$ is compact and convex, and the objective function $f_\mu(F_X)$ in (5.14) is continuous in F_X, strictly concave, and weakly differentiable. Therefore, the optimization problem in (5.14) has a *unique* maximizer. We denote the optimal input distribution for (5.14) by F_X^* which depends on the value μ.

Next, we obtain the KKT conditions for the optimal input distribution of the optimization problem in (5.14). Following along similar lines of the proof of Theorem 4 and noting that the objective function $f_\mu(F_X)$ is weakly differentiable with a weak derivative given as

$$f'_\mu(F_X, F_X^*) = \int_0^{\mathcal{A}} \Big[\frac{\mu}{\Delta}\, i_B(x; F_X^*) + (1-\mu)\, c_S(x; F_X^*)\Big]\, dF_X(x) - f_\mu(F_X^*), \tag{E.60}$$

the KKT conditions for the optimality of F_X^* are obtained as follows

$$\frac{\mu}{\Delta} i_B(x; F_X^*) + (1-\mu)\, c_S(x; F_X^*) - \gamma x \leq \frac{\mu}{\Delta} I_B(F_X^*) + \frac{1-\mu}{\Delta}\left[I_B(F_X^*) - I_E(F_X^*)\right] - \gamma\mathcal{E}, \quad \forall\, x \in [0, \mathcal{A}], \tag{E.61}$$

$$\frac{\mu}{\Delta} i_B(x; F_X^*) + (1-\mu)\, c_S(x; F_X^*) - \gamma x = \frac{\mu}{\Delta} I_B(F_X^*) \; + \frac{1-\mu}{\Delta}\left[I_B(F_X^*) - I_E(F_X^*)\right] - \gamma\mathcal{E}, \quad \forall\, x \in \mathcal{S}_{F_X^*}. \tag{E.62}$$

Next, we show that the optimal input distribution F_X^* has a finite support. To this end, assume to the contrary, that $\mathcal{S}_{F_X^*}$ has an infinite number of elements. Under such an assumption, (E.62), the analyticity of $i_B(w; F_X^*)$ and $i_E(w; F_X^*)$ over $\mathcal{O}$ in the complex plane and the Bolzano-Weierstrass and Identity Theorems of complex analysis, one obtains

$$\frac{\mu}{\Delta} i_B(x; F_X^*) + (1-\mu)\, c_S(x; F_X^*) - \gamma x = \frac{\mu}{\Delta} I_B(F_X^*) + \frac{1-\mu}{\Delta}\left[I_B(F_X^*) - I_E(F_X^*)\right] - \gamma\mathcal{E}, \quad \forall\, x > -\frac{\lambda_B}{\alpha_B}. \tag{E.63}$$

We continue the proof by showing that (E.63) results in a contradiction. To do so, we first observe that RHS of (E.63) does not depend on x and hence, it is a constant function in x. Taking the derivative of both sides of (E.63) with respect to x, we get

$$\frac{\mu}{\Delta} \frac{d i_B(x; F_X^*)}{dx} + (1-\mu) \frac{d c_S(x; F_X^*)}{dx} = \gamma, \quad \forall\, x > -\frac{\lambda_B}{\alpha_B}, \tag{E.64}$$

or equivalently

$$\gamma = \mu \left[\alpha_B \log[(\alpha_B x + \lambda_B)\Delta] + \alpha_B \sum_{y=0}^{+\infty} p(y|x) \log \frac{g_B(y; F_X^*)}{g_B(y+1; F_X^*)} \right] + (1-\mu) \Bigg[(\alpha_B - \alpha_E) \times \log[(\alpha_B x + \lambda_B)\Delta] + \alpha_E \log \frac{\alpha_B x + \lambda_B}{\alpha_E x + \lambda_E} + \alpha_B \sum_{y=0}^{+\infty} p(y|x) \log \frac{g_B(y; F_X^*)}{g_B(y+1; F_X^*)} - \alpha_E \sum_{z=0}^{+\infty} p(z|x) \log \frac{g_E(z; F_X^*)}{g_E(z+1; F_X^*)} \Bigg], \quad \forall\ x > -\frac{\lambda_B}{\alpha_B}. \tag{E.65}$$

Using the bounds in (E.37)–(E.38), the RHS of (E.64) can be lower bounded as

$$\gamma \geq \mu\, \alpha_B \log \frac{\alpha_B x + \lambda_B}{\alpha_B \mathcal{A} + \lambda_B} + (1-\mu) \Bigg[(\alpha_B - \alpha_E) \log \frac{\alpha_B x + \lambda_B}{\alpha_B \mathcal{A} + \lambda_B} + \alpha_E \log \frac{\alpha_B x + \lambda_B}{\alpha_E x + \lambda_E} + \alpha_E \log \frac{\lambda_E}{\alpha_B \mathcal{A} + \lambda_B} \Bigg], \quad \forall\ x > -\frac{\lambda_B}{\alpha_B}. \tag{E.66}$$

Observe that the RHS of (E.66) scales logarithmically, i.e., $\Omega(\log x)$ for large values of x. This is clearly a contradiction because the constant value γ cannot be greater than a logarithmically increasing function. This implies that $\mathcal{S}_{F_X^*}$ cannot have infinite elements in the interval $[0, \mathcal{A}]$. Hence, F_X^* is discrete with a finite number of mass points. Additionally, we note that for $\mu = 0$, F_X^* must be discrete with a finite support according to Theorem 4, and for $\mu = 1$ (the point corresponding to the capacity of the discrete-time Poisson channel with peak- and average-intensity constraints), F_X^* is also discrete with a finite number of mass points; reproving the results presented in [7]. Consequently, the entire rate-equivocation region of the DT–PWC with peak- and average-intensity constraints is exhausted by discrete input distributions with finitely many mass points. This completes the proof of Theorem 6.

Finally, observe that the lower bound in (E.66) does not depends on Δ. Thus, the result also holds true for the case where $\Delta \to 0$. That is, every point on the boundary of the rate-equivocation region of the degraded continuous-time PWC with nonnegativity, peak- and average-intensity constraints is also achieved by a unique and discrete input distribution with a finite number of mass point in the interval $[0, \mathcal{A}]$.

E.6 Proof of Theorem 7

We start the proof by noting that the feasible set $\Omega_{\mathcal{E}}^{+}$ is compact and convex, and the objective function $f_\mu(F_X)$ in (5.14) is continuous in F_X, strictly concave, and weakly differentiable. Therefore, the optimization problem in (5.14) has a *unique* maximizer. We denote the optimal input distribution for

(5.14) by F_X^* which depends on μ.

The KKT conditions for the optimal input distribution F_X^* of the optimization problem in (5.14) is given by

$$\frac{\mu}{\Delta} i_B(x;F_X^*) + (1-\mu)\, c_S(x;F_X^*) - \gamma x \leq \frac{\mu}{\Delta} I_B(F_X^*) + \frac{1-\mu}{\Delta}\left[I_B(F_X^*) - I_E(F_X^*)\right] - \gamma\mathcal{E},\ \forall\, x \geq 0, \quad \text{(E.67)}$$

$$\frac{\mu}{\Delta} i_B(x;F_X^*) + (1-\mu)\, c_S(x;F_X^*) - \gamma x = \frac{\mu}{\Delta} I_B(F_X^*) + \frac{1-\mu}{\Delta}\left[I_B(F_X^*) - I_E(F_X^*)\right] - \gamma\mathcal{E},\ \forall\, x \in \mathcal{S}_{F_X^*} \quad \text{(E.68)}$$

We show that the optimal input distribution F_X^* has the following structural properties: 1) the intersection of the optimal support set with any bounded interval contains finitely many mass points; 2) The optimal support set itself is an unbounded set. Theses properties are proved via similar contradiction approaches that appear in the proof of Theorem 5.

E.6.1 The intersection of the optimal support set with any bounded interval contains a finite number of elements

Let B be a bounded interval and assume, to the contrary, that $\mathcal{S}_{F_X^*} \cap B$ has an infinite number of elements. Now based on the optimality equation (E.68), the analyticity of $i_B(x;F_X^*)$ and $c_S(x;F_X^*)$ over $\mathcal{O}$, the Bolzano-Weierstrass and Identity Theorems from complex analysis, we get

$$\frac{\mu}{\Delta} i_B(x;F_X^*) + (1-\mu)\, c_S(x;F_X^*) - \gamma x = \frac{\mu}{\Delta} I_B(F_X^*) + \frac{1-\mu}{\Delta}\left[I_B(F_X^*) - I_E(F_X^*)\right] - \gamma\mathcal{E}, \quad \forall\, x > -\frac{\lambda_B}{\alpha_B}, \quad \text{(E.69)}$$

and we show that (E.69) results in a contradiction. By taking the derivative of both sides of (E.69) with respect to x we find

$$\frac{\mu}{\Delta}\frac{di_B(x;F_X^*)}{dx} + (1-\mu)\frac{dc_S(x;F_X^*)}{dx} = \gamma, \quad \forall\, x > -\frac{\lambda_B}{\alpha_B}. \quad \text{(E.70)}$$

Using the bounds in (E.47)–(E.49) the RHS of (E.70) can be upper bounded as

$$\begin{aligned}
\gamma \leq\, & \mu\,\alpha_B \log\frac{\alpha_B x + \lambda_B}{\lambda_B} + (1-\mu)\big[(\alpha_B - \alpha_E)\log[(\alpha_B x + \lambda_B)\Delta] + \alpha_E \log\frac{\alpha_B}{\alpha_E} \\
& + \alpha_E \log\frac{x + \frac{\lambda_B}{\alpha_B}}{x + \frac{\lambda_E}{\alpha_E}} + \alpha_E\big([(\alpha_E x + \lambda_E)\Delta]\log[(\alpha_E x + \lambda_E)\Delta] - [(\alpha_E x + \lambda_E)\Delta] \\
& \times (1 + \log(\lambda_E\Delta)) + \frac{3}{2}\log[(\alpha_E x + \lambda_E)\Delta + 1] + (\alpha_E\mathcal{E} + \lambda_E)\Delta + \frac{1}{2}\log(2\pi e)\big) \\
& - \alpha_B \log(\lambda_B\Delta)\big], \quad \forall\, x > -\frac{\lambda_B}{\alpha_B}. \quad \text{(E.71)}
\end{aligned}$$

Taking the limit from both sides of (E.71) as $x \to -\frac{\lambda_B}{\alpha_B}^+$, we obtain $\gamma \leq -\infty$. This is a contradiction and we conclude that $\mathcal{S}_{F_X^*} \cap B$ must contain finitely many mass points. Notice that this holds true for all $\mu \in [0,1]$ implying that the support set of the capacity-achieving input distribution for the discrete-time Poisson channel with nonnegativity and average-intensity constraints has a finite number of mass points in any bounded interval. Notice that the upper bound in (E.71) depends on Δ for all $\mu \in [0,1)$ and surprisingly does not depend on Δ for $\mu = 1$. Therefore, in this case we conclude that the *capacity-achieving* distribution of the continuous-time PWC with nonnegativity and average-intensity constraints admits a finite number of mass points in any bounded interval. Nevertheless, the capacity of the continuous-time version under an average-intensity constraint is infinite [6].

E.6.2 The support set of the optimal distribution $\mathcal{S}_{F_X^*}$ for all $\mu \in [0,1]$ is unbounded

Assume, to the contrary, that $\mathcal{S}_{F_X^*}$ is a bounded set, i.e., $\mathcal{S}_{F_X^*} \subseteq [0,h]$ where h is some finite positive constant. In the previous section, we proved that the intersection of $\mathcal{S}_{F_X^*}$ with any bounded interval has a finite number of elements for all $\mu \in [0,1]$. Since we are assuming that $\mathcal{S}_{F_X^*}$ is bounded, thus, it must contain finitely many mass points. This implies that $F_X^*(x) = \sum_{i=1}^{N} p_i u(x - x_i)$, where $N < +\infty$, $0 \leq x_1 < x_2 < \cdots < x_N \leq h$ are the mass points with corresponding probabilities $\{p_1, \ldots, p_N\}$. Following along similar lines of the proof of Theorem 5 and in view of the optimality condition (E.67), one can write

$$\begin{aligned}
\Psi(\mu, \Delta, F_X^*) + \gamma(x - \mathcal{E}) &\geq (1-\mu)\, c_S(x; F_X^*) + \frac{\mu}{\Delta} i_B(x; F_X^*) \\
&> (1-\mu)\Bigg[(\alpha_B - \alpha_E) x \log[(\alpha_B x + \lambda_B)\Delta] + \alpha_E x \log \frac{\alpha_B x + \lambda_B}{\alpha_E x + \lambda_E} \\
&\quad + x\left[(\alpha_E - \alpha_B)(1 + \log[(\alpha_E x_N + \lambda_E)\Delta]) + \alpha_B \log \frac{\alpha_E x_N + \lambda_E}{\alpha_B x_N + \lambda_B}\right] \\
&\quad + \lambda_B \log \frac{\alpha_B x + \lambda_B}{\alpha_B x_N + \lambda_B} - \lambda_E \log \frac{\alpha_E x + \lambda_E}{\alpha_E x_N + \lambda_E} + \frac{\log p_N}{\Delta} - \alpha_E x_N \Bigg] \\
&\quad + \mu\left[(\alpha_B x + \lambda_B) \log \frac{\alpha_B x + \lambda_B}{\alpha_B x_N + \lambda_B} - \alpha_B x\right], \ \forall\, x \geq 0,
\end{aligned} \tag{E.72}$$

where $\Psi(\mu, \Delta, F_X^*) \triangleq \frac{\mu}{\Delta} I_B(F_X^*) + \frac{1-\mu}{\Delta}[I_B(F_X^*) - I_E(F_X^*)]$. Observe that the RHS of (E.72) scales like $x \log x$ for large values of x and for all $\mu \in (0,1]$, but the LHS of (E.72) is a linear function in x. Therefore, for all $\mu \in (0,1]$ we reach a contradiction and we have that $\mathcal{S}_{F_X^*}$ must be an unbounded set. Furthermore, we have already established in Theorem 5 that when $\mu = 0$ (the point corresponding to the secrecy capacity) $\mathcal{S}_{F_X^*}$ is also unbounded. Consequently, we conclude that $\mathcal{S}_{F_X^*}$ is an unbounded set for

all $\mu \in [0,1]$. This completes the proof of Theorem 7.

Lastly, observe that the bound in (E.72) does not depend on Δ for $\mu = 1$ implying that the support set of the capacity-achieving distribution for the continuous-time PWC with nonnegativity and average-intensity constraint is an unbounded set. Combining this with the results in the previous section, we find that the capacity-achieving distribution for the continuous-time version is discrete and has a countably infinite number of mass points, but finitely many mass points in any bounded interval.

E.7 Lower Bound on the Secrecy Capacity of the DT–PWC in the Low-Intensity Regime

We start the proof by noting that $C_B \geq \frac{1}{\Delta} I(X^b;Y)$ where X^b is the channel input with a binary distribution. We choose the input distribution to be either $F_X(x) = \frac{1}{2}u(x) + \frac{1}{2}u(x-\mathcal{A})$ when only the peak-intensity constraint is active or to be $F_X(x) = (1-p)u(x) + pu(x-\mathcal{A}),\ 0 < p < \frac{1}{2}$, when both peak- and average-intensity constraints are active and they both go to zero with their ratio held fixed at p. Now, we follow along similar lines of [38, Proposition 2] to find the closed-form expression of the mutual information $I(X^b;Y)$ in the low-intensity regime when both peak- and average-intensity constraints are active, i.e., $0 < p < \frac{1}{2}$. We note that

$$\begin{aligned}
I(X^b;Y) &= H(Y) - H(Y|X) = -\sum_{y=0}^{+\infty} [(1-p)p(y|0) + p\,p(y|\mathcal{A})] \log\left[(1-p)p(y|0) + p\,p(y|\mathcal{A})\right] \\
&\quad + (1-p)\sum_{y=0}^{+\infty} p(y|0)\log(p(y|0)) + p\sum_{y=0}^{+\infty} p(y|\mathcal{A})\log(p(y|\mathcal{A})) \\
&= -p\sum_{y=0}^{+\infty} p(y|\mathcal{A})\left(\log\frac{p(y|0)}{p(y|\mathcal{A})} + \log\left((1-p) + p\frac{p(y|\mathcal{A})}{p(y|0)}\right)\right) - (1-p)\sum_{y=0}^{+\infty} p(y|0) \\
&\quad \times \log\left((1-p) + p\frac{p(y|\mathcal{A})}{p(y|0)}\right) \\
&= \underbrace{p\sum_{y=0}^{+\infty} p(y|\mathcal{A})\log\frac{p(y|\mathcal{A})}{p(y|0)}}_{\triangleq T_1(\mathcal{A})} - \sum_{y=0}^{+\infty}((1-p)p(y|0) + pp(y|\mathcal{A}))\underbrace{\log\left((1-p) + p\frac{p(y|\mathcal{A})}{p(y|0)}\right)}_{\triangleq T_2(\mathcal{A},y)}.
\end{aligned} \tag{E.73}$$

Note that $T_1(\mathcal{A}) = -p\,\alpha_B\mathcal{A}\Delta + p\,(\lambda_B + \alpha_B\mathcal{A})\Delta\log\left(1 + \frac{\alpha_B\mathcal{A}}{\lambda_B}\right)$. Now, consider the Taylor expansion of $T_1(\mathcal{A})$ around $\mathcal{A} = 0$ to get

$$T_1(\mathcal{A}) = -p\Delta\,\alpha_B\mathcal{A} + p\Delta\left(\alpha_B\mathcal{A} + \frac{\alpha_B^2}{2\lambda_B}\mathcal{A}^2 + o(\mathcal{A}^2)\right) = p\Delta\left(\frac{\mathcal{A}^2}{2\lambda_B}\alpha_B^2 + o(\mathcal{A}^2)\right), \tag{E.74}$$

where $o(\mathcal{A}^2)$ contains all the term that tend to zero faster than $\mathcal{A}^2$, i.e., $\lim_{\mathcal{A}\to 0}\frac{o(\mathcal{A}^2)}{\mathcal{A}^2}=0$. Furthermore, observe that $T_2(\mathcal{A},y)=\log\left((1-p)+p\,e^{-\alpha_B\mathcal{A}\Delta}\left(1+\frac{\alpha_B\mathcal{A}}{\lambda_B}\right)^y\right)$, and the Taylor expansion of $T_2(\mathcal{A},y)$ around $\mathcal{A}=0$ gives

$$T_2(\mathcal{A},y)=p\,\alpha_B\left(\frac{y}{\lambda_B}-\Delta\right)\mathcal{A}+\left[p\left(\frac{\alpha_B^2\Delta^2}{2}-\frac{\alpha_B^2 y}{2\lambda_B^2}+\frac{\alpha_B^2 y^2}{2\lambda_B^2}-\frac{\alpha_B^2 y\Delta}{\lambda_B}\right)-p^2\alpha_B^2\frac{(\Delta-\frac{y}{\lambda_B})^2}{2}\right]\mathcal{A}^2+o(\mathcal{A}^2 y). \tag{E.75}$$

Plugging (E.75) into (E.73), the second term in (E.73) denoted by $T_3(\mathcal{A})$ becomes

$$\begin{aligned}T_3(\mathcal{A})\stackrel{\triangle}{=}&-\sum_{y=0}^{+\infty}\left((1-p)p(y|0)+pp(y|\mathcal{A})\right)T_2(\mathcal{A},y)=-(1-p)p\left(\frac{\lambda_B\Delta}{\lambda_B}-\Delta\right)\mathcal{A}-(1-p)p\\&\times\left(\frac{\alpha_B^2\Delta^2}{2}-\frac{\alpha_B^2\Delta}{2\lambda_B}+\frac{\alpha_B^2\Delta^2}{2}+\frac{\alpha_B^2\Delta}{2\lambda_B}-\alpha_B^2\Delta^2\right)\mathcal{A}^2+(1-p)p^2\alpha_B^2\frac{\mathcal{A}^2}{2\lambda_B}\Delta\\&-p^2\alpha_B^2\frac{\mathcal{A}^2}{\lambda_B}\Delta-p^2\mathcal{A}^2\Big(\frac{\alpha_B^2\Delta^2}{2}-\frac{\alpha_B^3\mathcal{A}\Delta}{2\lambda_B^2}-\frac{\alpha_B^2\Delta}{2\lambda_B}+\frac{\alpha_B^4\mathcal{A}^2\Delta^2}{2\lambda_B^2}+\frac{\alpha_B^3\mathcal{A}\Delta^2}{\lambda_B^2}\\&+\frac{\alpha_B^2\Delta^2}{2}+\frac{\alpha_B^3\mathcal{A}\Delta}{2\lambda_B^2}+\frac{\alpha_B^2\Delta}{2\lambda_B}-\frac{\alpha_B^3\mathcal{A}\Delta^2}{\lambda_B}-\alpha_B^2\Delta^2\Big)+p^3\alpha_B^2\frac{\mathcal{A}^2}{2}\Big(\Delta^2-\frac{2\alpha_B\mathcal{A}\Delta}{\lambda_B}-2\Delta^2\\&+\frac{\alpha_B^2\mathcal{A}^2\Delta^2}{\lambda_B^2}+\Delta^2+\frac{2\alpha_B\mathcal{A}\Delta^2}{\lambda_B}+\frac{\alpha_B\mathcal{A}\Delta}{\lambda_B}+\frac{\Delta}{\lambda_B}\Big)+o(\mathcal{A}^2)\\=&-p^2\frac{\mathcal{A}^2}{2\lambda_B}\alpha_B^2\Delta+o(\mathcal{A}^2).\end{aligned} \tag{E.76}$$

Combining this with $T_1(\mathcal{A})$, one obtains

$$I(X^b;Y)=\frac{\mathcal{A}^2}{2\lambda_B}\alpha_B^2\,p\,(1-p)\Delta+o(\mathcal{A}^2). \tag{E.77}$$

Hence, in the regime where $\mathcal{A}\to 0$, $C_B\geq\frac{1}{\Delta}I(X^b;Y)\geq\frac{\mathcal{A}^2}{2\lambda_B}\alpha_B^2 p(1-p)$. Note that when only the peak-intensity constraint is active, we choose $p=\frac{1}{2}$. Thus, we have

$$C_B\geq\begin{cases}\frac{\mathcal{A}^2}{8}\frac{\alpha_B^2}{\lambda_B}, & \text{if } \frac{1}{2}\leq p\leq 1,\\ \frac{\mathcal{A}^2}{2}\,p\,(1-p)\frac{\alpha_B^2}{\lambda_B}, & \text{if } 0<p<\frac{1}{2}.\end{cases} \tag{E.78}$$

Next, we observe that C_E can be upper bounded by the capacity of the continuous-time Poisson channel since in this case, the channel input admits infinite bandwidth and is not restricted to be a PAM signal.

Therefore, in the low intensity regime C_E can be upper bounded by [6, Theorem 2]

$$C_E \leq \begin{cases} \frac{\mathcal{A}^2}{8}\frac{\alpha_E^2}{\lambda_E}, & \text{if } \frac{1}{2} \leq p \leq 1, \\ \frac{\mathcal{A}^2}{2}\, p\,(1-p)\frac{\alpha_E^2}{\lambda_E}, & \text{if } 0 < p < \frac{1}{2}. \end{cases} \tag{E.79}$$

Finally, from (E.78) and (E.79), we find that

$$C_S \geq \begin{cases} \frac{\mathcal{A}^2}{8}\left(\frac{\alpha_B^2}{\lambda_B} - \frac{\alpha_E^2}{\lambda_E}\right), & \text{if } \frac{1}{2} \leq p \leq 1, \\ \frac{\mathcal{A}^2}{2}\, p\,(1-p)\left(\frac{\alpha_B^2}{\lambda_B} - \frac{\alpha_E^2}{\lambda_E}\right), & \text{if } 0 < p < \frac{1}{2}. \end{cases} \tag{E.80}$$

This completes the proof of the lemma.

E.8 Upper Bound on the Secrecy Capacity of the DT–PWC in the Low-Intensity Regime

We start the proof by noting that the secrecy capacity of the DT–PWC with peak- and average-intensity constraints is upper bounded by the secrecy capacity of the continuous-time PWC with peak- and average-intensity constraints. This is because in the continuous-time version, the input signals are not restricted to be PAM signals and can admit any waveform with very large transmission bandwidth. Now, we recall the secrecy capacity of the degraded continuous-time PWC with a peak-intensity constraint from [15, Theorem 3]

$$C_S^{CT} = pK(\mathcal{A}) + (1-p)K(0) - K(p\mathcal{A}), \quad 0 \leq p \leq 1, \tag{E.81}$$

where $K(x) = (\alpha_B x + \lambda_B)\log(\alpha_B x + \lambda_B) - (\alpha_E x + \lambda_E)\log(\alpha_E x + \lambda_E)$ and where p is the solution of the equation

$$K(\mathcal{A}) - K(0) = \mathcal{A}K'(p\mathcal{A}). \tag{E.82}$$

The secrecy capacity C_S^{CT} is achieved by a binary input distributions with mass points at $\{0, \mathcal{A}\}$ and respective probabilities $\{(1-p), p\}$. We now find the closed-form expression of C_S^{CT} in the regime where $\mathcal{A} \to 0$. To this end, we expand $K(\mathcal{A})$ around $\mathcal{A} = 0$ and we get

$$K(\mathcal{A}) = \log\frac{\lambda_B^{\lambda_B}}{\lambda_E^{\lambda_E}} + \left((\alpha_B - \alpha_E) + \log\frac{\lambda_B}{\lambda_E}\right)\mathcal{A} + \left(\frac{\alpha_B^2}{\lambda_B} - \frac{\alpha_E^2}{\lambda_E}\right)\frac{\mathcal{A}^2}{2} + o(\mathcal{A}^2). \tag{E.83}$$

Therefore, plugging this expansion into (E.82), the optimal p in the regime where $\mathcal{A} \to 0$ is given by

$$\left(\frac{\alpha_B^2}{\lambda_B} - \frac{\alpha_E^2}{\lambda_E}\right)\frac{\mathcal{A}^2}{2} = \left(\frac{\alpha_B^2}{\lambda_B} - \frac{\alpha_E^2}{\lambda_E}\right)p\mathcal{A}^2 \Rightarrow p = \frac{1}{2}. \tag{E.84}$$

Thus, C_S^{CT} in the regime where $\mathcal{A} \to 0$ is given by

$$C_S^{CT} = \frac{\mathcal{A}^2}{8}\left(\frac{\alpha_B^2}{\lambda_B} - \frac{\alpha_E^2}{\lambda_E}\right). \tag{E.85}$$

Furthermore, we observe that when both peak- and average-intensity constraints are active, the optimal input distribution is also binary with mass points $\{0, \mathcal{A}\}$ and respective probabilities $\{(1-p), p\}$ where $p = \frac{\mathcal{E}}{\mathcal{A}}$. Therefore, in the regime where $\mathcal{A} \to 0$ and $\mathcal{E} \to 0$ with their ratio held fixed at p, C_S^{CT} is

$$C_S^{CT} = \frac{\mathcal{A}^2}{2}p(1-p)\left(\frac{\alpha_B^2}{\lambda_B} - \frac{\alpha_E^2}{\lambda_E}\right). \tag{E.86}$$

Finally, note that since $p = \frac{1}{2}$ is the optimal value of (E.82), we conclude that when $\frac{\mathcal{E}}{\mathcal{A}} \geq \frac{1}{2}$, $p = \frac{1}{2}$ and the average-intensity constraint is not active. This completes the proof of the lemma.

E.9 Upper Bound on the Secrecy Capacity in the High-Intensity Regime For Equal Channel Gains

We start the proof by noting that the output of the eavesdropper's channel Z can be written as $Z = \widetilde{Y} = Y + N_D$, where N_D is defined in the statement of Lemma 3. Therefore, $H_Z(F_X^*) > H_{Z|N_D}(F_X^*) = H_Y(F_X^*)$, and consequently $H_Y(F_X^*) - H_Z(F_X^*) < 0$ for any nontrivial input distribution $F_X^* \in \mathcal{F}^+$. Furthermore, we can expand $H_{Z|X}(F_X^*) - H_{Y|X}(F_X^*)$ as follows

$$\begin{aligned}
H_{Z|X}(F_X^*) - H_{Y|X}(F_X^*) &= \mathbb{E}_{X,Z}\left[-\log p_{Z|X}(z|x)\right] - \mathbb{E}_{X,Y}\left[-\log p_{Y|X}(y|x)\right] \\
&\overset{(a)}{=} \mathbb{E}_{Z|X,Y}\left[\mathbb{E}_{X,Y}\left[\log p_{Y|X}(y|x)\right]\right] - \mathbb{E}_{Y|X,Z}\left[\mathbb{E}_{X,Z}\left[\log p_{Z|X}(z|x)\right]\right] \\
&= \mathbb{E}_{X,Y,Z}\left[\log p_{Y|X}(y|x)\right] - \mathbb{E}_{X,Y,Z}\left[\log p_{Z|X}(z|x)\right] \\
&= \mathbb{E}_{X,Y,Z}\left[\log \frac{p_{Y|X}(y|x)}{p_{Z|X}(z|x)}\right],
\end{aligned} \tag{E.87}$$

where (a) follows as $\log p_{Y|X}(y|x)$ and $\log p_{Z|X}(z|x)$ do no depend on Z and Y, respectively. Plugging (5.1) and (5.2) into (E.87), we get

$$\begin{aligned} H_{Z|X}(F_X^*) - H_{Y|X}(F_X^*) &= \mathbb{E}_{X,Y,Z}\left[\log \frac{e^{-(\alpha_B x+\lambda_B)\Delta}[(\alpha_B x+\lambda_B)\Delta]^{\,y}/y!}{e^{-(\alpha_E x+\lambda_E)\Delta}[(\alpha_E x+\lambda_E)\Delta]^{\,z}/z!}\right] \\ &= \lambda_D\Delta + \mathbb{E}_X\big[(\alpha_B x+\lambda_B)\Delta\log[(\alpha_B x+\lambda_B)\Delta] - (\alpha_E x+\lambda_E)\Delta \\ &\quad \times \log[(\alpha_E x+\lambda_E)\Delta]\big] + \mathbb{E}_{X,Y,Z}\left[\log\frac{Z!}{Y!}\right], \end{aligned} \tag{E.88}$$

Next, we consider the last term in (E.88) and try to find an upper bound on it. To this end, we first note that

$$\mathbb{E}_{X,Y,Z}\left[\log\frac{Z!}{Y!}\right] = \mathbb{E}_X\left[\mathbb{E}_{Y|X}\left[\mathbb{E}_{Z|Y}\left[\log\frac{Z!}{Y!}\right]\right]\right] \tag{E.89}$$

as $X \to Y \to Z$ is a Markov chain. Now, we have to find the conditional PDF of $Z|Y$. We proceed by observing that $Z = Y + N_D$, hence, one can show that

$$p_{Z|Y}(z|y) = \begin{cases} 0, & \text{if } z < y, \\ e^{-\lambda_D\Delta}\frac{(\lambda_D\Delta)^{(z-y)}}{(z-y)!}, & \text{if } z \geq y. \end{cases} \tag{E.90}$$

In what follows, we present chain of inequalities based on (E.90) which leads to the upper bound in (5.20),

$$\begin{aligned}
\mathbb{E}_{X,Y,Z}\left[\log\frac{Z!}{Y!}\right] &= \mathbb{E}_X\left[\mathbb{E}_{Y|X}\left[\sum_{z=0}^{+\infty} p_{Z|Y}(z|y)\log\frac{z!}{y!}\right]\right] \\
&= \mathbb{E}_X\left[\mathbb{E}_{Y|X}\left[\sum_{z=y}^{+\infty} e^{-\lambda_D\Delta}\frac{(\lambda_D\Delta)^{(z-y)}}{(z-y)!}\log\frac{z!}{y!}\right]\right] \\
&= \mathbb{E}_X\left[\mathbb{E}_{Y|X}\left[\sum_{t=0}^{+\infty} e^{-\lambda_D\Delta}\frac{(\lambda_D\Delta)^{t}}{t!}\log\frac{(t+y)!}{y!}\right]\right] \\
&= \mathbb{E}_X\left[\mathbb{E}_{Y|X}\left[\sum_{t=0}^{+\infty} e^{-\lambda_D\Delta}\frac{(\lambda_D\Delta)^{t}}{t!}\sum_{i=1}^{t}\log(y+i)\right]\right] \\
&\overset{(b)}{\leq} \mathbb{E}_X\left[\sum_{t=0}^{+\infty} e^{-\lambda_D\Delta}\frac{(\lambda_D\Delta)^{t}}{t!}\sum_{i=1}^{t}\log[(\alpha_B x+\lambda_B)\Delta+i]\right] \\
&= \mathbb{E}_X\Bigg[\sum_{t=0}^{+\infty} e^{-\lambda_D\Delta}\frac{(\lambda_D\Delta)^{t}}{t!}\Bigg[t\log[(\alpha_B x+\lambda_B)\Delta] \\
&\qquad + \sum_{i=1}^{t}\log\left[1+\frac{i}{(\alpha_B x+\lambda_B)\Delta}\right]\Bigg]\Bigg] \\
&\overset{(c)}{\leq} \mathbb{E}_X\left[\sum_{t=0}^{+\infty} e^{-\lambda_D\Delta}\frac{(\lambda_D\Delta)^{t}}{t!}\left[t\log[(\alpha_B x+\lambda_B)\Delta]+\sum_{i=1}^{t}\frac{i}{(\alpha_B x+\lambda_B)\Delta}\right]\right] \\
&= \mathbb{E}_X\Bigg[\log[(\alpha_B x+\lambda_B)\Delta]\sum_{t=0}^{+\infty} e^{-\lambda_D\Delta}\frac{(\lambda_D\Delta)^{t}}{t!}t+\frac{1}{(\alpha_B x+\lambda_B)\Delta} \\
&\qquad \times\sum_{t=0}^{+\infty} e^{-\lambda_D\Delta}\frac{(\lambda_D\Delta)^{t}}{t!}\frac{t(t+1)}{2}\Bigg] \\
&= \mathbb{E}_X\left[(\lambda_D\Delta)\log[(\alpha_B x+\lambda_B)\Delta]+\frac{1}{(\alpha_B x+\lambda_B)\Delta}\left[\frac{(\lambda_D\Delta)^2}{2}+\lambda_D\Delta\right]\right], \qquad \text{(E.91)}
\end{aligned}$$

where (b) follows from sliding the expectation $\mathbb{E}_{Y|X}$ through the summations and then applying the Jensen's Inequality (as $\log(y+i)$ is a concave function in y), and (c) follows from the fact that $\log(1+x) \leq x,\ \forall x \geq 0$. Now, using the upper bound in (E.91), $H_{Z|X}(F_X^*) - H_{Y|X}(F_X^*)$ can be upper bounded as

$$\begin{aligned}
\frac{1}{\Delta}\left[H_{Z|X}(F_X^*) - H_{Y|X}(F_X^*)\right] &\leq \lambda_D + \mathbb{E}_X\big[(\alpha_B x+\lambda_B)\log[(\alpha_B x+\lambda_B)\Delta] - (\alpha_E x+\lambda_E) \\
&\qquad \times\log[(\alpha_E x+\lambda_E)\Delta]\big] + \mathbb{E}_X\left[\lambda_D\log[(\alpha_B x+\lambda_B)\Delta]\right] \\
&\qquad + \mathbb{E}_X\left[\frac{1}{(\alpha_B x+\lambda_B)\Delta^2}\left[\frac{(\lambda_D\Delta)^2}{2}+\lambda_D\Delta\right]\right] \\
&= \lambda_D + \mathbb{E}_X\left[(\alpha_E x+\lambda_E)\log\frac{\alpha_B x+\lambda_B}{\alpha_E x+\lambda_E}\right] + \left[\frac{\lambda_D^2}{2}+\frac{\lambda_D}{\Delta}\right]\mathbb{E}_X\left[\frac{1}{\alpha_B x+\lambda_B}\right].
\end{aligned} \qquad \text{(E.92)}$$

We note that since $x \geq 0$, $\mathbb{E}_X\left[\frac{1}{\alpha_B x+\lambda_B}\right] \leq \frac{1}{\lambda_B}$. Furthermore, denoting $\psi(x) = (\alpha_E x + \lambda_E)\log\frac{\alpha_B x+\lambda_B}{\alpha_E x+\lambda_E}$, we observe that $\psi(x)$ is strictly negative as $\alpha_B = \alpha_E$ and $\frac{\lambda_E}{\alpha_E} > \frac{\lambda_B}{\alpha_B}$. Furthermore, $\psi(x)$ is a strictly increasing function in x due to the fact that

$$\frac{d\psi(x)}{dx} = \alpha_B\left[-\log\left[1+\frac{\lambda_D/\alpha_B}{x+\lambda_B/\alpha_B}\right]+\frac{\lambda_D/\alpha_B}{x+\lambda_B/\alpha_B}\right] > 0, \quad \forall\ x \geq 0. \tag{E.93}$$

This implies that the maximum value of $\psi(x)$ is located at the end point of the interval $[0, \mathcal{A}]$, if the peak-intensity is active, and is located at $x = +\infty$, if the average-intensity is the only active constraint. In either of these case, we can write

$$\psi(x) \leq \lim_{x\to+\infty}(\alpha_E x+\lambda_E)\log\frac{\alpha_B x+\lambda_B}{\alpha_E x+\lambda_E} = -\lambda_D. \tag{E.94}$$

From the upper bound on $\mathbb{E}_X\left[\frac{1}{\alpha_B x+\lambda_B}\right]$ and (E.94), one can upper bound (E.91) as

$$H_{Z|X}(F_X^*) - H_{Y|X}(F_X^*) \leq \frac{\frac{\lambda_D^2}{2}+\frac{\lambda_D}{\Delta}}{\lambda_B}. \tag{E.95}$$

We note that this upper bound is valid for all values of the peak- and/or average-intensity constraints. This completes the proof of the proposition.

Appendix F: Proof of the Main Results in Chapter 6

We dedicate this section to the proofs of Theorem 9 and Theorem 10. To this end, we first review some preliminaries that are used throughout the proofs. Then, we present a few lemmas that are used to establish the theorems.

F.1 Preliminaries

We start by noting that the channel output distribution $p_Y(y; F_{X_1}, F_{X_2})$ is given by

$$p_Y(y; F_{X_1}, F_{X_2}) = \int_0^{\mathcal{A}_1} \int_0^{\mathcal{A}_2} p(y|x_1, x_2)\, dF_{X_2}\, dF_{X_1},\ y \in \mathbb{R}, \tag{F.1}$$

where $p(y|x_1, x_2)$ is given by (6.3). since the channel input i, $i \in \{1, 2\}$ is constrained by a peak-intensity constraint as (6.1), the following bounds on the conditional channel law can be found [59]

$$k_1 e^{-k_2(y-x_1-x_2)^2} \leq p(y|x_1, x_2) \leq k_3 e^{-k_4(y-x_1-x_2)^2}, \tag{F.2}$$

for some positive finite values k_1, k_2, k_3, k_4, where

$$k_1 \triangleq \frac{1}{\sqrt{2\pi\sigma^2(\mathcal{A}_1, \mathcal{A}_2)}}, \quad k_2 \triangleq \frac{1}{2\sigma_0^2}, \tag{F.3}$$

$$k_3 \triangleq \frac{1}{\sqrt{2\pi\sigma_0^2}}, \qquad k_4 \triangleq \frac{1}{2\sigma^2(\mathcal{A}_1, \mathcal{A}_2)}. \tag{F.4}$$

Next, we define $\gamma(y)$ and $\Gamma(y)$ as follows

$$\gamma(y) \triangleq \min_{\substack{0\leq x_1\leq \mathcal{A}_1 \\ 0\leq x_2\leq \mathcal{A}_2}} k_1 e^{-k_2(y-x_1-x_2)^2}, \tag{F.5}$$

$$\Gamma(y) \triangleq \max_{\substack{0\leq x_1\leq \mathcal{A}_1 \\ 0\leq x_2\leq \mathcal{A}_2}} k_3 e^{-k_4(y-x_1-x_2)^2}. \tag{F.6}$$

It is straightforward to show that

$$\gamma(y) = \begin{cases} k_1 e^{-k_2(y-\mathcal{A}_1-\mathcal{A}_2)^2}, & y \leq \frac{\mathcal{A}_1+\mathcal{A}_2}{2}, \\ k_1 e^{-k_2 y^2}, & \text{otherwise.} \end{cases} \tag{F.7}$$

and

$$\Gamma(y) = \begin{cases} k_3 e^{-k_4 y^2}, & y \leq 0, \\ k_3 e^{-k_4 (y-\mathcal{A}_1-\mathcal{A}_2)^2}, & y > \mathcal{A}_1 + \mathcal{A}_2, \\ k_3, & \text{otherwise.} \end{cases} \tag{F.8}$$

Now, based on (F.1), the bounds in (F.2), and (F.7)–(F.8), one can find that $p_Y(y; F_{X_1}, F_{X_2})$ is bounded as follows

$$\gamma(y) \leq p_Y(y; F_{X_1}, F_{X_2}) \leq \Gamma(y). \tag{F.9}$$

Furthermore, we note that the marginals $p_{Y|X_i}(y|x_i; F_{X_{\bar{i}}})$, $i \in \{1, 2\}$ can be written as

$$p_{Y|X_i}(y|x_i; F_{X_{\bar{i}}}) = \int_0^{\mathcal{A}_{\bar{i}}} p(y|x_1, x_2)\, dF_{X_{\bar{i}}}, \tag{F.10}$$

where $\bar{i} = \{1, 2\} - i$. Similar to the lower and upper bounds given by (F.9) for $p_Y(y; F_{X_1}, F_{X_2})$, one can find the lower bound $\phi(y, x_i)$ and the upper bound $\Phi(y, x_i)$ on $p_{Y|X_i}(y|x_i; F_{X_{\bar{i}}})$ as

$$\phi(y, x_i) \leq p_{Y|X_i}(y|x_i; F_{X_{\bar{i}}}) \leq \Phi(y, x_i),\ i \in \{1, 2\}, \tag{F.11}$$

where

$$\phi(y, x_i) \stackrel{\triangle}{=} \min_{0 \leq x_{\bar{i}} \leq \mathcal{A}_{\bar{i}}} p(y|x_1, x_2),\ i \in \{1, 2\} \tag{F.12}$$

$$\Phi(y, x_i) \stackrel{\triangle}{=} \max_{0 \leq x_{\bar{i}} \leq \mathcal{A}_{\bar{i}}} p(y|x_1, x_2),\ i \in \{1, 2\}. \tag{F.13}$$

It can be easily shown that for $i \in \{1, 2\}$

$$\phi(y, x_i) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2(x_i, \mathcal{A}_{\bar{i}})}} \exp\left(-\frac{(y-x_i-\mathcal{A}_{\bar{i}})^2}{2\sigma^2(x_i,0)}\right), & y - x_i \leq \frac{\mathcal{A}_{\bar{i}}}{2}, \\ \frac{1}{\sqrt{2\pi\sigma^2(x_i, \mathcal{A}_{\bar{i}})}} \exp\left(-\frac{(y-x_i)^2}{2\sigma^2(x_i,0)}\right), & \text{otherwise.} \end{cases} \tag{F.14}$$

and

$$\Phi(y, x_i) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2(x_i, 0)}} \exp\left(-\frac{(y-x_i)^2}{2\sigma^2(x_i, \mathcal{A}_{\bar{i}})}\right), & y - x_i \leq 0, \\ \frac{1}{\sqrt{2\pi\sigma^2(x_i, 0)}} \exp\left(-\frac{(y-x_i-\mathcal{A}_{\bar{i}})^2}{2\sigma^2(x_i, \mathcal{A}_{\bar{i}})}\right), & y - x_i > \mathcal{A}_{\bar{i}}, \\ \frac{1}{\sqrt{2\pi\sigma^2(x_i, 0)}}, & \text{otherwise.} \end{cases} \tag{F.15}$$

Considering the conditional entropy $h(Y;F_{X_1},F_{X_2})$ and $h(Y|X_{\bar{i}};F_{X_i})$, $i\in\{1,2\}$, one can write

$$h(Y;F_{X_1},F_{X_2})=-\int_0^{\mathcal{A}_i}\underbrace{\int_0^{\mathcal{A}_{\bar{i}}}\int_{\mathbb{R}}p(y|x_1,x_2)\log p_Y(y;F_{X_1},F_{X_2})\,dy\,dF_{X_{\bar{i}}}}_{\triangleq\,\Theta(x_i;F_{X_i})}\,dF_{X_i},\ i\in\{1,2\} \tag{F.16}$$

$$h(Y|X_{\bar{i}};F_{X_1},F_{X_2})=-\int_0^{\mathcal{A}_i}\underbrace{\int_0^{\mathcal{A}_{\bar{i}}}\int_{\mathbb{R}}p(y|x_1,x_2)\log p_{Y|X_{\bar{i}}}(y|x_{\bar{i}};F_{X_i})\,dy\,dF_{X_{\bar{i}}}}_{\triangleq\,\Lambda(x_i;F_{X_i})}\,dF_{X_i},\ i\in\{1,2\} \tag{F.17}$$

where $\Theta(x_i;F_{X_i}):[0,\mathcal{A}_i]\times\Omega^+_{\mathcal{A}_i,\mathcal{E}_i}\to\mathbb{R}$, $i\in\{1,2\}$ and $\Lambda(x_i;F_{X_i}):[0,\mathcal{A}_i]\times\Omega^+_{\mathcal{A}_i,\mathcal{E}_i}\to\mathbb{R}$, $i\in\{1,2\}$ and denote the entropy density functionals.

F.2 Proof of Theorem 9

In this section, we present the details for the proof of Theorem 9. First, notice that the optimization problem in (6.8) is symmetric in the input distributions F_{X_1} and F_{X_2} in the sense that one can first consider finding the solution of the optimization problem with respect to one of the distributions, say F_{X_1}, considering the other distribution function is fixed. After finding the optimal solution with respect to F_{X_1}, which we denote by $F^*_{X_1}$, we can choose F_{X_1} to be $F^*_{X_1}$ and solve the optimization problem with respect to F_{X_2}. Therefore, without loss of generality, we will fix F_{X_2} and establish that the answer to the optimization problem in (6.8) with respect to F_{X_1} is discrete and it admits a countably finite number of mass points, i.e., the optimal distribution $F^*_{X_1}$ is discrete with a countably finite support set. We then show that following along similar lines of the proof for the optimality of discrete distributions with a finite support set for $F^*_{X_1}$, the answer to the optimization problem (6.8) with respect to F_{X_2} is also discrete with a finite support set.

We start the proof by defining the mapping $\Xi:(0,\infty)\times\Omega^+_{\mathcal{A}_1,\mathcal{E}_1}\to\mathbb{R}$ such that

$$\Xi(\mu;F_{X_1})=\begin{cases}\mu h(Y;F_{X_1},F_{X_2})+(1-\mu)h(Y|X_2;F_{X_1},F_{X_2})\\ \qquad\qquad -h(Y|X_1,X_2;F_{X_1},F_{X_2}), & 0<\mu<1\\ h(Y;F_{X_1},F_{X_2})-h(Y|X_1,X_2;F_{X_1},F_{X_2}), & \mu=1\\ h(Y;F_{X_1},F_{X_2})+(\mu-1)h(Y|X_1;F_{X_1},F_{X_2})\\ \qquad\qquad -\mu h(Y|X_1,X_2;F_{X_1},F_{X_2}), & \mu>1,\end{cases} \tag{F.18}$$

Next, we extend the approach taken in [36] and show that the optimal input distribution $F^*_{X_1}$ is discrete

with a finite number of mass points. To that end, we first show that the optimization problem in (6.8) is convex and the supremum can be achieved by at least one element $F_{X_1} \in \Omega^+_{\mathcal{A}_1,\mathcal{E}_1}$. To achieve this goal, we will show that: 1) the set $\Omega^+_{\mathcal{A}_1,\mathcal{E}_1}$ is compact and convex; 2) the objective functional $\Xi(\mu; F_{X_1})$ is continuous in F_{X_i} for all $\mu > 0$. Afterwards, we will focus on showing that the objective functional $\Xi(\mu; F_{X_1})$ is weakly differentiable and concave in F_{X_i} for all $\mu > 0$. Taking the weak derivative of $\Xi(\mu; F_{X_1})$ with respect to F_{X_1} and using the concavity, we derive the necessary and sufficient KKT optimality conditions that the optimal distribution $F^*_{X_1}$ must satisfy. We continue the proof by showing that the optimal solution $F_{X_1}*$ must be discrete with a countably finite support set. This is done by proof via a contradiction approach, i.e., we assume to the contrary that the support set of the optimal solution $F^*_{X_1}$ contains an infinite number of elements; then we extend the corresponding rate-region density (defined later in this Section) to the complex plane and observe that it is analytic over some open connected set in the complex plane. Finally, leveraging the Identity Theorem from complex analysis and the Bolzano-Weierstrass Theorem, we will find that a linearly growing function in x would be lower bounded by another function which grows quadratically in x, and thus reaching the desired contradiction. This implies that the support set of $F^*_{X_1}$ cannot have infinitely many elements and therefore, it must be contain a finite number of mass points in the interval $[0, \mathcal{A}_1]$.

For convenience, the proof is streamlined into a few lemmas which we state below.

Lemma 10. *The feasible set $\Omega^+_{\mathcal{A}_1,\mathcal{E}_1}$ is convex and compact in the Levy metric sense.*

Proof. The proof follows along similar lines as [60, Appendix A.1]. ■

Lemma 11. *The functional $\Xi(\mu; F_{X_1})$ is continuous in $F_{X_1} \in \Omega^+_{\mathcal{A}_1,\mathcal{E}_1}$.*

Proof. The proof uses the bounds in (F.2) and (F.9), and follows along similar lines of [25, Section IV]. ■

From Lemma 10 and Lemma 11, $\Xi(\mu; F_{X_1})$ is continuous in F_{X_1} over $\Omega^+_{\mathcal{A}_1,\mathcal{E}_1}$ which itself is a compact and convex set, then by Extreme Value Theorem, $\Xi(\mu; F_{X_1})$ is bounded above and attains its supremum. That is, the supremum in (6.8) is achievable by at least one input distribution $F_{X_1} \in \Omega^+_{\mathcal{A}_1,\mathcal{E}_1}$.

Lemma 12. *The functional $\Xi(F_{X_1})$ is concave in F_{X_1}.*

Proof. It is a well-known fact that $h(Y; F_{X_1}, F_{X_1})$ is concave in F_{X_1} [9]. Furthermore, for a fixed $p(y|x_1, x_2)$, $p_{Y|X_2}(y|x_2; F_{X_1})$ is linear in F_{X_1}. Hence, $h(Y|X_2)$ which is concave in $p_{Y|X_2}(y|x_2; F_{X_1})$, is a concave functional of F_{X_1}. Finally, $h(Y|X_1)$ is a linear functional in F_{X_1}. As a result, $\Xi(\mu; F_{X_1})$ is a concave functional in F_{X_1} for all $\mu > 0$. ■

Lemma 13. *Defining* $F_{X_{1,t}} = (1-t)F^*_{X_1} + tF_{X_1},\ \forall F^*_{X_1}, F_{X_1} \in \Omega^+_{\mathcal{A}_1,\mathcal{E}_1}$, $t \in [0,1]$, *the weak derivative of* $\Xi(F_{X_1})$ *at* $F^*_{X_1}$ *denoted by* $D(\Xi(F^*_{X_1}))$ *exists and is equal to*

$$\begin{aligned} D(\Xi(F^*_{X_1})) &\triangleq \lim_{t\to 0} \frac{\Xi\left((1-t)F^*_{X_1} + tF_{X_1}\right) - \Xi(F^*_{X_1})}{t} \\ &= \int_0^{\mathcal{A}_1} \xi(\mu, x_1; F^*_{X_1})\, dF_{X_1} - \Xi(F^*_{X_1}), \quad \mu > 0 \end{aligned} \tag{F.19}$$

where $\xi(\mu, x_1; F^*_{X_1})$ *is called the rate-region density with respect to* $F^*_{X_1}$ *and is given by*

$$\xi(\mu, x_1; F^*_{X_1}) \triangleq \begin{cases} \mu\,\Theta(x_1; F^*_{X_1}) + (1-\mu)\Lambda(x_1; F^*_{X_1}) - \frac{1}{2}\mathbb{E}_{X_2}[\log 2\pi e\sigma^2(x_1, x_2)], & 0 < \mu < 1 \\ \Theta(x_1; F^*_{X_1}), & \mu = 1 \\ \Theta(x_1; F^*_{X_1}) - (\mu-1)\int_{\mathbb{R}} p_{Y|X_1}(y|x_1)\log p_{Y|X_1}(y|x_1)\, dy & \\ \qquad - \frac{\mu}{2}\mathbb{E}_{X_2}[\log 2\pi e\sigma^2(x_1, x_2)], & \mu > 1 \end{cases} \tag{F.20}$$

Proof. The proof is based on the definition of the weak derivative and follows along similar lines of [36]. ■

Now, from Lemma 10, Lemma 12, and Lemma 13, we have a concave and weakly differentiable functional $\Xi(F_{X_1})$ over $\Omega^+_{\mathcal{A}_1,\mathcal{E}_1}$ which is a convex set, then the necessary and sufficient conditions for an input distribution $F^*_{X_1} \in \Omega^+_{\mathcal{A}_1,\mathcal{E}_1}$ to be optimal is

$$D(\Xi(F^*_{X_1})) \leq 0. \tag{F.21}$$

Furthermore, the mapping defined as $g(F_{X_1}) \triangleq \int_0^{\mathcal{A}_1} x_1\, dF_{X_1}(x_1) - \lambda_1\,\mathbb{E}[X_1]$ from $\Omega^+_{\mathcal{A}_1,\mathcal{E}_1} \to \mathbb{R}$ is continuous in F_{X_1}, concave, and weakly differentiable, where $\lambda_1 > 0$ denotes the Lagrangian multiplier. As a result, following along the similar steps of, e.g., [22],[36, Corollary 1], the necessary and sufficient conditions for the optimality of $F^*_{X_1}$ can be given as

$$\xi(\mu, x_1; F^*_{X_1}) - \lambda_1 x_1 \leq \Xi(F^*_{X_1}) - \lambda_1\,\mathbb{E}[X_1], \quad \forall\, x_1 \in [0, \mathcal{A}_1], \tag{F.22}$$

$$\xi(\mu, x_1; F^*_{X_1}) - \lambda_1 x_1 = \Xi(F^*_{X_1}) - \lambda_1\,\mathbb{E}[X_1], \quad \forall\, x_1 \in \mathcal{S}_{F^*_{X_1}}, \tag{F.23}$$

$$\lambda_1\left(\mathbb{E}[X_1] - \mathcal{E}_1\right) = 0. \tag{F.24}$$

where $\mathcal{S}_{F^*_{X_1}} \subset [0, \mathcal{A}_1]$ is the support set of $F^*_{X_1}$. We are now ready to establish that the optimal input distribution $F^*_{X_1}$ is discrete with a finite number of mass points. To prove the discreteness, we resort to a contradiction argument using the KKT conditions in (F.22)–(F.24). To this end, we first present

a lemma that states that the extension of $\xi(x_1; F^*_{X_1}) - \lambda x_1$ to some open connected set in the complex plane $\mathbb{C}$ is analytic.

Lemma 14. *The extension of $\xi(\mu, x_1; F^*_{X_1}) - \lambda x_1$ to the open connected set $\mathcal{O} \triangleq \{w \in \mathbb{C} : \Re\{w\} \geq 0\}$ is analytic, where $\Re\{w\}$ is the real part of the complex variable w.*

Proof. This can be established by using the bounds on $p_Y(y; F_{X_1}, F_{X_2})$ that are given by (F.9) and following along a similar line of [4, 25]. ■

Next, we assume that $\mathcal{S}_{F^*_{X_1}}$ has an infinite number of elements. In view of the optimality condition (F.23), the analyticity of $\xi(w; F^*_{X_1}) - \lambda_1 w$ over $\mathcal{O}$, and the Identity Theorem from complex analysis along with the Bolzano-Weierstrass Theorem, if $\mathcal{S}_{F^*_{X_1}}$ has an infinite number of elements, we can deduce that $\xi(w; F^*_{X_1}) - \lambda_1 w = \Xi(F^*_{X_1}) - \lambda_1 \mathcal{E}_1, \ \forall w \in \mathcal{O}$. Since $[0, \infty) \subset \mathcal{O}$, we have

$$\xi(x_1; F^*_{X_1}) - \lambda_1 x_1 = \Xi(F^*_{X_1}) - \lambda_1 \mathcal{E}_1, \quad \forall\, x_1 \geq 0. \tag{F.25}$$

We show that the conclusion in (F.25) results in a contradiction. To this end, we fix $\mu \in (0, 1)$ and expand $\xi(\mu, x_1; F^*_{X_1})$ based on (F.20) to get

$$\mu\Theta(x_1; F^*_{X_1}) + (1-\mu)\Lambda(x_1; F^*_{X_1}) = \frac{1}{2}\mathbb{E}_{X_2}\left[\log\left(2\pi e\sigma^2(x_1, x_2)\right)\right] + \Xi(F^*_{X_1}) + \lambda_1 (x_1 - \mathcal{E}_1), \tag{F.26}$$

for all $x_1 \geq 0$. Denoting the left-hand-side of (F.26) by $T(x_1; F^*_{X_1})$ and using the bounds in (F.9) and (F.11), next, we find that for large values of x_1, $T(x_1; F^*_{X_1})$ grows *quadratically* in x_1, i.e., for some positive constants c' and k' we have $T(x_1; F^*_{X_1}) \geq k' x_1^2, \ \forall\, x_1 > c'$, therefore, we have $T(x_1; F^*_{X_1}) = \Omega(x_1^2)$. To this end, we lower bound $\Theta(x_1; F^*_{X_1})$ by upper bounding $p_Y(y; F_{X_1}, F_{X_2})$ based on (F.9). Thus, we

have that

$$\begin{aligned}
\Theta(x_1; F^*_{X_1}) \geq & -\int_0^{\mathcal{A}_2}\int_{\mathbb{R}} p(y|x_1,x_2)\log\Gamma(y)\,dF_{X_2}\,dy \\
&= \int_0^{\mathcal{A}_2}\int_{-\infty}^{0} p(y|x_1,x_2)k_4y^2\,dy\,dF_{X_2} + \int_0^{\mathcal{A}_2}\int_{\mathcal{A}}^{\infty} p(y|x_1,x_2)k_4(y-\mathcal{A})^2\,dy\,dF_{X_2} \\
&\quad - \log k_3 \\
&= \int_0^{\mathcal{A}_2}\int_{-\infty}^{+\infty} k_4\,y^2\,p(y|x_1,x_2)\,dy\,dF_{X_2} - \int_0^{\mathcal{A}_2} \underbrace{\int_0^{\mathcal{A}} k_4\,y^2\,p(y|x_1,x_2)\,dy}_{\leq k_4\,\mathcal{A}^2 \text{ since } \int_0^{\mathcal{A}} p(y|x_1,x_2)\,dy \leq 1} dF_{X_2} \\
&\quad - \int_0^{\mathcal{A}_2}\int_{\mathcal{A}}^{+\infty} 2\mathcal{A}\,k_4y\,p(y|x_1,x_2)\,dy\,dF_{X_2} + \int_0^{\mathcal{A}_2}\int_{\mathcal{A}}^{+\infty} \mathcal{A}^2\,k_4\underbrace{p(y|x_1,x_2)}_{\geq 0}\,dy\,dF_{X_2} \\
&\quad - \log k_3 \\
&= \int_0^{\mathcal{A}_2} k_4\,\mathbb{E}_{Y|X_1,X_2}[y^2]\,dF_{X_2} - \int_0^{\mathcal{A}_2} 2\mathcal{A}k_4\,\mathbb{E}_{Y|X_1,X_2}[y|y\geq\mathcal{A}]\,dF_{X_2} \\
&\quad - \ k_4\,\mathcal{A}^2 - \log k_3 \\
&= k_4\mathbb{E}_{X_2}[\sigma^2(x_1,x_2) + (x_1+x_2)^2] - k_4\,\mathcal{A}^2 - \log k_3 \\
&\quad - \mathbb{E}_{X_2}\left[(x_1+x_2) - \sigma^2(x_1,x_2)\frac{p(+\infty|x_1,x_2) - p(\mathcal{A}|x_1,x_2)}{1-P(\mathcal{A}|x_1,x_2)}\right],
\end{aligned} \tag{F.27}$$

where $\mathcal{A} \triangleq \mathcal{A}_1 + \mathcal{A}_2$ and $P(y|x_1,x_2)$ is the cumulative density function of the Gaussian random variable $Y|X_1,X_2$. Observe that $p(+\infty|x_1,x_2) \geq 0$ and $p(\mathcal{A}|x_1,x_2) \leq k_3$. Furthermore, $P(\mathcal{A}|x_1,x_2) \approx 0$ for large values of x_1. Hence, for large values of x_1, (F.27) can be further lower bounded as

$$\Theta(x_1; F^*_{X_1}) \geq k_4x_1^2 + o(x_1^2), \tag{F.28}$$

where $o(x_1^2)$ is a function which satisfies $\lim_{x_1\to+\infty}\frac{o(x_1^2)}{x_1^2} = 0$. This implies that for some positive constants β_1 and β_2 we have $\Theta(x_1; F^*_{X_1}) \geq \beta_2x_1^2$, $\forall\, x_1 > \beta_1$, therefore, we have $\Theta(x_1; F^*_{X_1}) = \Omega(x_1^2)$.

Next, we show that $\Lambda(x_1; F^*_{X_1})$ also grows quadratically in x_1 for large values of x_1. To this end, we

lower bound $\Lambda(x_1; F^*_{X_1})$ using the upper bound on $p_{Y|X_2}(y|x_2)$ given by (F.15). As such, we can write

$$\begin{aligned}
\Lambda(x_1; F^*_{X_1}) &\geq -\int_0^{\mathcal{A}_2}\int_{\mathbb{R}} p(y|x_1,x_2)\log\Phi(y,x_2)\,dy\,dF_{X_2} \\
&= \int_0^{\mathcal{A}_2}\int_{-\infty}^{x_2} p(y|x_1,x_2)\frac{(y-x_2)^2}{2\sigma^2(\mathcal{A}_1,x_2)}\,dy\,dF_{X_2} + \mathbb{E}_{X_2}\left[\frac{1}{2}\log 2\pi\sigma^2(0,x_2)\right] \\
&\quad + \int_0^{\mathcal{A}_2}\int_{\mathcal{A}_1+x_2}^{\infty} p(y|x_1,x_2)\frac{(y-x_2-\mathcal{A}_1)^2}{2\sigma^2(\mathcal{A}_1,x_2)}\,dy\,dF_{X_2} \\
&\geq \int_0^{\mathcal{A}_2}\int_{-\infty}^{x_2} p(y|x_1,x_2)\frac{(y-x_2)^2}{2\sigma^2(\mathcal{A}_1,\mathcal{A}_2)}\,dy\,dF_{X_2} + \frac{1}{2}\log 2\pi\sigma_0^2 \\
&\quad + \int_0^{\mathcal{A}_2}\int_{\mathcal{A}_1+x_2}^{\infty} p(y|x_1,x_2)\frac{(y-x_2-\mathcal{A}_1)^2}{2\sigma^2(\mathcal{A}_1,\mathcal{A}_2)}\,dy\,dF_{X_2} \\
&= \beta_3\int_0^{\mathcal{A}_2}\int_{\mathbb{R}} p(y|x_1,x_2)y^2\,dy\,dF_{X_2} - \beta_3\int_0^{\mathcal{A}_2}\underbrace{\int_{x_2}^{\mathcal{A}_1+x_2} p(y|x_1,x_2)y^2\,dy}_{\leq(\mathcal{A}_1+x_2)^2}\,dF_{X_2} \\
&\quad - \beta_3\int_0^{\mathcal{A}_2}\underbrace{\int_{-\infty}^{x_2} p(y|x_1,x_2)2x_2y\,dy}_{\leq 2x_2^2}\,dF_{X_2} + \underbrace{\beta_3\int_0^{\mathcal{A}_2}\int_{-\infty}^{x_2} p(y|x_1,x_2)x_2^2\,dy\,dF_{X_2}}_{\geq 0} \\
&\quad - \beta_3\int_0^{\mathcal{A}_2}\int_{\mathcal{A}_1+x_2}^{\infty} p(y|x_1,x_2)(2(\mathcal{A}_1+x_2)y)\,dy\,dF_{X_2} + \frac{1}{2}\log 2\pi\sigma_0^2 \\
&\quad + \underbrace{\beta_3\int_0^{\mathcal{A}_2}\int_{\mathcal{A}_1+x_2}^{\infty} p(y|x_1,x_2)(x_2+\mathcal{A}_1)^2\,dy\,dF_{X_2}}_{\geq 0} \\
&\geq \beta_3\,\mathbb{E}_{X_2}[\sigma^2(x_1,x_2)+(x_1+x_2)^2] - \beta_3\mathcal{A}^2 - 2\beta_3\mathcal{A}_2^2 - 2\beta_3\,\mathbb{E}_{X_2}\Bigg[(\mathcal{A}_1+x_2) \\
&\quad \times\left((x_1+x_2)-\sigma^2(x_1,x_2)\frac{p(\infty|x_1,x_2)-p(\mathcal{A}_1+x_2|x_1,x_2)}{1-P(\mathcal{A}_1+x_2|x_1,x_2)}\right)\Bigg] + \frac{1}{2}\log 2\pi\sigma_0^2, \qquad \text{(F.29)}
\end{aligned}$$

where $\beta_3 \stackrel{\triangle}{=} \frac{1}{2\sigma^2(\mathcal{A}_1,\mathcal{A}_2)}$. Since $p(\infty|x_1,x_2)\geq 0$, $p(\mathcal{A}_1+x_2|x_1,x_2)\leq k_3$ and $P(\mathcal{A}_1+x_2|x_1,x_2)\approx 0$ for large values of x_1, we can write

$$\Lambda(x_1; F^*_{X_1}) \geq \beta_3 x_1^2 + o(x_1^2), \qquad \text{(F.30)}$$

which implies that for some positive constants β_4 and β_5 we have $\Lambda(x_1; F^*_{X_1}) \geq \beta_5 x_1^2$, $\forall\, x_1 > \beta_4$, therefore, we have $\Lambda(x_1; F^*_{X_1}) = \Omega(x_1^2)$. Now, based on (F.28) and (F.30) and the fact that $T(x_1; F^*_{X_1}) = \mu\Theta(x_1; F^*_{X_1}) + (1-\mu)\Lambda(x_1; F^*_{X_1})$, $\mu\in(0,1)$, we conclude that $T(x_1; F^*_{X_1}) = \Omega(x_1^2)$.

However, observe that for large values of x_1, the right-hand-side of (F.26) grows *linearly* in x_1. Hence, the left-hand-side and the right-hand-side of (F.26) do not grow with the same rate and we reach a contradiction. This implies that the support set of the optimal distribution $\mathcal{S}_{F^*_{X_1}} \subset [0, \mathcal{A}_1]$ contains a finite number of element, which in turn, implies that the optimal input distribution $F^*_{X_1}$ is discrete with

a finite number of mass points in the interval $[0, \mathcal{A}_1]$.

We note that the case $m = 1$ (boundary points of the capacity region corresponding to the sum-capacity) can be treated similarly, i.e., following along similar lines of the proof of the discreteness of $F^*_{X_1}$ for $m \in (0, 1)$, one can show that the optimal input distribution $F^*_{X_1}$ which achieves the sum-capacity of the IDGN-OMAC with peak- and average-intensity constraints, is discrete with a finite number of mass points in $[0, \mathcal{A}_1]$.

Finally, for the case $m > 1$, (F.25) is given by

$$\Theta(x_1; F^*_{X_1}) - (\mu - 1) \int_{\mathbb{R}} p_{Y|X_1}(y|x_1) \log p_{Y_1|X_1}(y|x_1)\, dy = \frac{\mu}{2}\, \mathbb{E}_{X_2}\left[\log\left(2\pi e \sigma^2(x_1, x_2)\right)\right] + \Xi(F^*_{X_1}) + \lambda_1 (x_1 - \mathcal{E}_1), \forall\, x_1 \geq 0 \qquad \text{(F.31)}$$

Next, we show that (F.31) results in a contradiction for large values of x_1. To that end, we first observe that since $p_{Y|X_1}(y|x_1) = \int_0^{\mathcal{A}_2} p(y|x_1, x_2)\, dF_{X_2} \leq k_3$, one can find

$$-(\mu - 1) \int_{\mathbb{R}} p_{Y|X_1}(y|x_1) \log p_{Y_1|X_1}(y|x_1)\, dy \geq \frac{(\mu - 1)}{2} \log 2\pi\sigma_0^2, \quad \forall\, x_1 \geq 0. \qquad \text{(F.32)}$$

Since $\Theta(x_1; F^*_{X_1}) = \Omega(x_1^2)$ for large values of x_1, combining this fact along with (F.32), we conclude that for large values of x_1, the left-hand-side of (F.31) grows quadratically in x_1, but the right-hand-side of it grows linearly in x_1. This results in a contradiction, and therefore, the optimal input distribution achieving the boundary of the capacity region for $m > 1$ must be discrete with a finite number of mass points in $[0, \mathcal{A}_1]$. This completes the proof of Theorem 9.

Finally, we note that the growth rate of $\Theta(x_1; F^*_{X_1})$ and $\Lambda(x_1; F^*_{X_1})$ do not change when the variance of the input-dependent noise component σ_1^2 is zero. This implies that invoking the same contradiction argument for IDGN-OMAC will result in establishing that any point on the boundary of the capacity region of the FSO-MAC with nonnegativity, peak- and average-intensity constraints is achieved by discrete distributions with finitely many mass points.

F.3 Proof of Theorem 10

We start the proof of Theorem 10 by noting that the capacity region of the IDGN-OMAC with peak- and average-intensity constraints satisfies $\mathcal{C} \subset [0, C_1] \times [0, C_2]$, where C_i, $i \in \{1, 2\}$ is the single-user capacity of user i. In the regime where $\mathcal{A}_i \to 0$, $i \in \{1, 2\}$ and $\mathcal{E}_i \to 0$, $i \in \{1, 2\}$ while their ratio held fixed at α_i, the single-user capacity C_i is known in a closed-form expression due to [5, Theorem 10] and

is given by

$$C_i = \begin{cases} \dfrac{\mathcal{A}_i^2}{2}\alpha_i(1-\alpha_i)\left(\dfrac{1}{\sigma_0^2}+\dfrac{\sigma_1^4}{2\sigma_0^4}\right), & \alpha_i \in (0,\frac{1}{2}), \\ \dfrac{\mathcal{A}_i^2}{8}\left(\dfrac{1}{\sigma_0^2}+\dfrac{\sigma_1^4}{2\sigma_0^4}\right), & \alpha_i \in [\frac{1}{2},1]. \end{cases} \tag{F.33}$$

where C_i is attained by a binary input distribution with mass points located at $\{0, \mathcal{A}_i\}$, $i \in \{1,2\}$ with corresponding probability masses $\{1-\alpha_i, \alpha_i\}$ when $\alpha_i \in (0, \frac{1}{2})$ and $\{\frac{1}{2}, \frac{1}{2}\}$ when $\alpha_i \in [\frac{1}{2}, 1]$.

Next, we show that $[0, C_1] \times [0, C_2] \subset \mathcal{C}$ in the low-intensity regime. To this end, we need to show for the rate $R_i = C_i$, $i \in \{1,2\}$ to be achievable, it is required that $R_1 + R_2 \leq \sup_{F_{X_i} \in \mathcal{F}_i^+,\, i\in\{1,2\}} I(X_1, X_2; Y)$. We observe that due to (6.4), we can write $I(X_1, X_2; Y) = I(X_1 + X_2; Y)$. Therefore, we have

$$\sup_{F_{X_i} \in \mathcal{F}_i^+,\, i\in\{1,2\}} I(X_1, X_2; Y) = \sup_{F_X \in \mathcal{F}^+} I(X; Y), \tag{F.34}$$

where $X \stackrel{\triangle}{=} X_1 + X_2$ and the feasible set is given by one of the following sets

$$\Omega^+_{\mathcal{A},\,\mathcal{E}} \stackrel{\triangle}{=} \left\{ F_X : \int_0^{\mathcal{A}} dF_X(x) = 1, \int_0^{\mathcal{A}} x\, dF_X(x) \leq \mathcal{E} \right\}, \tag{F.35}$$

$$\Omega^+_{\mathcal{A}} \stackrel{\triangle}{=} \left\{ F_X : \int_0^{\mathcal{A}} dF_X(x) = 1 \right\}, \tag{F.36}$$

where $\mathcal{A} = \mathcal{A}_1 + \mathcal{A}_2$ and $\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2$. Now, we note that in the regime where $\mathcal{A} \to 0$ and $\mathcal{E} \to 0$ with their ratio held fixed at $\alpha = \frac{\mathcal{A}}{\mathcal{E}} = \frac{\mathcal{A}_1+\mathcal{A}_2}{\mathcal{E}_1+\mathcal{E}_2}$, the sum-capacity defined as $C_{\text{Sum}} \stackrel{\triangle}{=} \sup_{F_X \in \mathcal{F}^+} I(X;Y)$ is given by [5, Theorem 10]

$$C_{\text{Sum}} = \begin{cases} \dfrac{\mathcal{A}^2}{2}\alpha(1-\alpha)\left(\dfrac{1}{\sigma_0^2}+\dfrac{\sigma_1^4}{2\sigma_0^4}\right), & \alpha \in (0,\frac{1}{2}) \\ \dfrac{\mathcal{A}^2}{8}\left(\dfrac{1}{\sigma_0^2}+\dfrac{\sigma_1^4}{2\sigma_0^4}\right), & \alpha \in [\frac{1}{2},1] \end{cases} \tag{F.37}$$

and the optimal input distribution F_X^* that attains C_{Sum} is a binary input distributions with mass points at $\{0, \mathcal{A}\}$ with corresponding probability masses $\{1-\alpha, \alpha\}$ when $\alpha \in (0, \frac{1}{2})$ and $\{\frac{1}{2}, \frac{1}{2}\}$ when $\alpha \in [\frac{1}{2}, 1]$. Observe that $\alpha(1-\alpha)\mathcal{A}^2 > \alpha_1(1-\alpha_1)\mathcal{A}_1^2 + \alpha_2(1-\alpha_2)\mathcal{A}_2^2$. This implies that in the low-intensity regime, when $R_i = C_i$, $i \in \{1,2\}$, then $R_1 + R_2 < C_{\text{Sum}}$. Consequently, (C_1, C_2) is achievable and the sum-capacity constraint becomes redundant. That is, in the low-intensity regime $[0, C_1] \times [0, C_2] \subset \mathcal{C}$. This implies that in the low-intensity regime, the capacity region of the IDGN-OMAC with peak- and average-intensity constraints is given by the rate pair (R_1, R_2) satisfying (6.9)–(6.10). This completes the proof of Theorem 10.

Appendix G: Authorization to Reuse IEEE Published Material

Chapter 3 to Chapter 7 are all published by the Institute of Electrical and Electronics Engineering (IEEE) and the following provides the permission to reuse the published material in this dissertation.

G.1 Thesis/Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant.

G.1.1 Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis

- In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line 2011 IEEE.
- In the case of illustrations or tabular material, we require that the copyright line [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

G.1.2 Requirements to be followed when using an entire IEEE copyrighted paper in a thesis

- The following IEEE copyright/ credit notice should be placed prominently in the references: [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
- Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
- In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with

permission in this thesis, the IEEE does not endorse any of the University of Idaho's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to `http://www.ieee.org/publications_standards/publications/rights/rights_link.html` to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.