ABSTRACT

Title of dissertation: RELIABLE COMMUNICATION OVER OPTICAL FADING CHANNELS

Kaushik Chakraborty, Doctor of Philosophy, 2005

Dissertation directed by: Professor Prakash Narayan
Department of Electrical and Computer Engineering
and Institute for Systems Research

In free space optical communication links, atmospheric turbulence causes random fluctuations in the refractive index of air at optical wavelengths, which in turn cause random fluctuations in the intensity and phase of a propagating optical signal. These intensity fluctuations, termed "fading," can lead to an increase in link error probability, thereby degrading communication performance. Two techniques are suggested to combat the detrimental effects of fading, viz., (a) estimation of channel fade and use of these estimates at the transmitter or receiver; and (b) use of multiple transmitter and receiver elements. In this thesis, we consider several key issues concerning reliable transmission over multiple input multiple output (MIMO) optical fading channels. These include the formulation of a block fading channel model that takes into account the slowly varying nature of optical fade; the determination of channel capacity, viz., the maximum achievable rate of reliable

communication, when the receiver has perfect fade information while the transmitter is provided with varying degrees of fade information; characterization of good transmitter power control strategies that achieve capacity; and the capacity in the low and high signal-to-noise ratio (SNR) regimes.

We consider a shot-noise limited, intensity modulated direct detection optical fading channel model in which the transmitted signals are subject to peak and average power constraints. The fading occurs in blocks of duration T_c (seconds) during each of which the channel fade (or channel state) remains constant, and changes across successive such intervals in an independent and identically distributed (i.i.d.) manner. A single-letter characterization of the capacity of this channel is obtained when the receiver is provided with perfect channel state information (CSI) while the transmitter CSI can be imperfect. A two-level signaling scheme ("ON-OFF keying") with arbitrarily fast intertransition times through each of the transmit apertures is shown to achieve channel capacity. Several interesting properties of the optimum transmission strategies for the transmit apertures are discussed. For the special case of a single input single output (SISO) optical fading channel, the behavior of channel capacity in the high and low signal-to-noise ratio (SNR) regimes is explicitly characterized, and the effects of transmitter CSI on capacity are studied.

RELIABLE COMMUNICATION OVER OPTICAL FADING CHANNELS

by

Kaushik Chakraborty

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2005

Advisory Committee:

Professor Prakash Narayan, Chair/Advisor
Professor Alexander Barg
Professor Benjamin Kedem
Professor P. S. Krishnaprasad
Professor Armand Makowski
Professor Adrian Papamarcou
Professor Sennur Ulukus

Dedication

To my son Sattwik.

ACKNOWLEDGMENTS

I express my profound gratitude to my advisor, Professor Prakash Narayan, for his wholehearted support, timely encouragement, and invaluable guidance in every stage of this work. It has been a privilege to be able to work with him; I shall endeavor to follow his exemplary professional and ethical standards.

I am grateful to Professors Alexander Barg, Benjamin Kedem, P. S. Krishnaprasad, Armand Makowski, Adrian Papamarcou and Sennur Ulukus for serving on my dissertation committee. I thank Professor Adrian Papamarcou for his help and guidance from my earliest days at the University of Maryland.

I thank the Institute for Systems Research and the Department of Electrical and Computer Engineering at the University of Maryland for providing the necessary financial support during my graduate studies. I gratefully acknowledge that this research was supported by the Army Research Office under ODDR&E MURI01 Program Grant No. DAAD19-01-1-0465 to the Center for Communicating Networked Control Systems (through Boston University).

I am deeply indebted to my colleagues and friends, Dr. Amit K. Roy Chowdhury, Dr. Damianos Karakos, Dr. Amit Kale, Mr. Chunxuan Ye, and Dr. Onur Kaya for numerous stimulating discussions. I thank Indrajit Bhattacharya, Anubhav Datta, Souvik Mitra, Kaushik Ghosh, Mainak Sen, Ayan Roy Chowdhury and Ayush Gupta for their companionship and support.

I want to take this opportunity to express my gratitude and affection to my parents and brother for their continued encouragement and support. I thank my father-in-law, mother-in-law and sisters-in-law for their affection and faith in me. I thank my dear wife Sharmistha for standing by me through thick and thin; her unconditional love and unwavering faith form the backbone of my life. Finally, I thank my son Sattwik for allowing me to spend much of his share of time towards research. To him, I owe the successful completion of this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

1.1 Motivation

Free space optics (FSO) is emerging as an attractive technology for several applications, e.g., metro network extensions; last mile connectivity; fiber backup; RF-wireless backhaul; and enterprise connectivity [49]. There are many benefits of wireless optical systems, viz., rapid deployment time; high security; inexpensive components; seamless wireless extension of the optical fiber backbone; immunity to RF interference; and lack of licensing regulations, to name a few. Consequently, free space optical communication through the turbulent atmospheric channel has received much attention in recent years [20, 21, 23, 34, 51, 53, 54].

In free space optical communication links, atmospheric turbulence can cause random variations in refractive index of air at optical wavelengths, which, in turn, result in random fluctuations in both the intensity and phase of a propagating optical signal [24, 48]. Such fluctuations can lead to an increase in the link error probability, thereby degrading communication performance [54]. The fluctuations in the intensity of the transmitted optical signal, termed "fading," can be modeled in terms of an ergodic lognormal process with a correlation time of the order of 10^{-3} to 10^{-2} second [44]. In practice, these fades can routinely exceed 10 dB. In the systems under consideration, data rates can typically be of the order of gigabits per second. Therefore, the free space optical channel is a slowly varying fading channel with occasional deep fades that can affect millions of consecutive bits [21].

A technique that is often used to achieve higher rates of reliable communication over fading channels is to use estimates of the channel fade (also referred to as path gain or channel state) at the transmitter and the receiver. See [40] for a comprehensive review of relevant research in radio frequency (RF) communication. In optical fading channels, instantaneous realizations of the channel state can be estimated at the receiver; at typical speeds, nearly 10^6 bits are transmitted during each coherence period, a small fraction of which can be used by the receiver to form good estimates of the channel fade. Depending on the availability of a feedback link and the amount of acceptable delay, the transmitter can be provided with complete or partial knowledge of the channel state, which can be used for adaptive power control, thereby achieving higher throughputs (cf. e.g., [6, 40]). These assumptions lead to a variety of interesting problems which are the subject of current investigation.

Another popular technique to combat the detrimental effects of fading is the use of spatial diversity in the form of multiple transmit and receive elements. In RF communication, the use of multiple transmit and receive antennae has been shown to significantly improve the communication throughput in the presence of channel fade. For a survey of recent results, see [16]. In a recent experimental study [29], multiple laser beams were employed to improve communication performance in a turbulent atmospheric channel. Some attempts have since then been made to characterize analytically the benefits of MIMO communication in optical fading channels [20, 21, 34, 51].

In this dissertation, we consider several important issues concerning reliable communication over MIMO optical fading channels. These include the formulation of a block fading channel model that takes into account the slowly varying nature of optical fade; the determination of channel capacity, viz., the maximum achievable rate of reliable communication, when the receiver has very good estimates of channel state information (CSI) while the transmitter is provided with varying degrees of CSI; characterization of good transmitter power control strategies that achieve capacity; and the limiting behavior of channel capacity in low and high signal-to-noise ratio (SNR) regimes.

1.2 Background

1.2.1 Historical perspective

Historically, conveying information optically through the atmospheric channel is one of the most primitive forms of communication known to mankind. Ever since the discovery of fire, man has used fire beacons and smoke signals for establishing and maintaining contact with other human beings. Classical works, e.g., Euclid's *Optica* and Hero's *Catopricia* discuss how the ancient Greeks used the sun's reflection from metal disks to signal messages over vast distances. Early naval communication relied on signaling flags and shuttered lamps [17]. Claude Chappe invented the "optical telegraph" in the 1790s, a series of semaphores mounted on towers, where human operators relayed messages from one tower to the next [22]. In 1880, Alexander Graham Bell used a "photophone" to transmit telephone signals via an intensity-modulated optical beam 200m through air to a distant receiver [28]. However, Bell's other technology, which evolved to telephony and wireless telegraphy, became the preferred mode of communication.

The resurrection of modern optical communication can be attributed to the invention of lasers in the late 1950s. Laser radiation is monochromatic, coherent, and intense, which makes it an attractive carrier for high data rate communication. Most of the early development of unguided laser communication was directed towards satellite and remote military applications. In the commercial sector, as the popularity of optical fibers soared in the 1970s and 1980s, the interest in wireless optical communication dwindled. However, since the mid-1990s, an increasing demand for broadband communication, fuelled by the explosive growth of the internet, has opened new horizons for optical wireless communication, especially in urban and metropolitan environments. Free space optics was voted as one of the ten "hottest" technologies in 2001 [36].

1.2.2 Atmospheric optical propagation

An optical signal propagating through the free space atmospheric channel undergoes degradations due to many factors. Molecular and aerosol absorbers[1] in the atmosphere, and thermal inhomogeneities in the troposphere[2] cause absorption and

[1] Water vapor, ice, dust, carbon dioxide, nitrogen, oxygen, ozone, organic molecules, etc. contribute to absorption and scattering at optical wavelengths. These effects can be substantially reduced by choosing the operating carrier frequency judiciously. For details, see [27, 50].

[2] The troposphere is the lowest layer of the atmosphere extending from the earth surface up to an elevation of about 15 km (9 miles).

scattering of the propagating optical field. Absorption and scattering can have several detrimental effects on the propagating optical wavefront,viz., attenuation, beam spread, multipath spread, angular spread, Doppler spread and depolarization [45].

We shall focus our attention on atmospheric optical communication under clear weather conditions. The effects of atmospheric absorption are negligible for short to medium range communication links, and will be ignored. We shall assume that typical transmitter beam divergence and receiver field-of-view values are much larger than the beam spread and angular spread induced by turbulence [45]. Under clear weather conditions, multipath spread is of the order of 10^{-12} second for line-of-sight communication [44], and will be ignored henceforth.

Heated air rising from the earth and man-made devices such as heating ducts create thermal inhomogeneities in the troposphere. This leads to random variations in the refractive index of air at optical wavelengths, a phenomenon referred to as "atmospheric turbulence" or "refractive turbulence," which, in turn, result in random fluctuations in both the intensity and phase of a propagating optical signal [24, 48]. Atmospheric turbulence manifests itself in several familiar atmospheric conditions, e.g., the twinkling of stars at night and the shimmering of the horizon on a hot day [28]. The intensity fluctuations of the propagating optical signal, termed "fading," can be modeled in terms of an ergodic lognormal process with a correlation time of the order of 1–10 milliseconds [44]. We shall assume that the fade is frequency nonselective (which is justified by the negligible effect of multipath spread), so that the intensities of the optical signals at the transmit and recieve apertures for each transmit-receive aperture pair are related via a multiplicative fade coefficient [18].

1.2.3 Brief description of the optical communication system

In conventional optical communication systems, the transmitter comprises a photoemitter, e.g., a laser diode (LD) or a light-emitting diode (LED), while the receiver comprises a photodetector, e.g., a semiconductor photodiode, a PIN diode, or an avalanche photodiode (APD). At typical operating frequencies,[3] the simplest and most popular technique for modulating the information onto the emitted light signal is *intensity modulation* [13, 26]. The intensity, and hence the instantaneous power, of the transmitted optical signal is proportional to the modulating current at the photoemitter.[4] At the *direct detection* photodetector, an electrical current is generated by photoabsorption at a rate which is proportional to the instantaneous optical power incident on the active detector surface. The intrinsically discrete nature of the current gives rise to shot noise observed in low light level detection; additional shot noise results from external sources of radiation ("background noise"), as well as from spontaneously generated charge particles ("dark current") [13, 44]. In this dissertation, we restrict our attention to an idealized direct detection optical receiver, in which the photocurrent generated at the detector is modeled as a doubly stochastic Poisson counting process (PCP) whose rate is proportional to the incident optical power at the detector plus a constant (dark current rate) [13, 21]. This simple

[3]Conventional diodes used for wireless optical communication emit light in the infrared spectrum, viz., at wavelengths in the range of 850–950 nanometers [26, 50].

[4]Herein lies an important difference between optical and RF communication; in RF communication, typically the amplitude of the propagating RF signal is proportional to the modulating electrical current, whence the instantaneous power of the transmitted signal is proportional to the *square* of the input current.

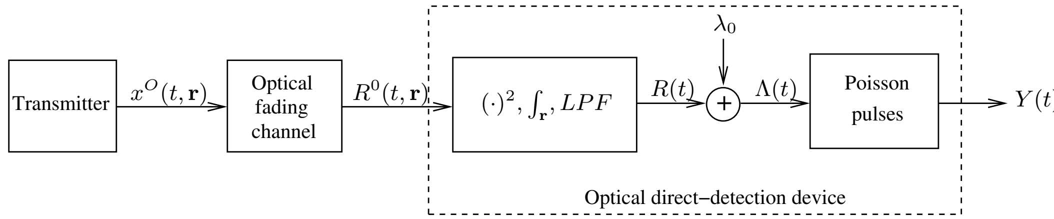

Figure 1.1: An intensity modulation direct detection (IM/DD) optical communication system.

receiver model is valid in the shot noise limited regime, when the effect of background radiation is minimal compared to the incident optical signal at the detector.

A block schematic diagram of a shot noise limited free space optical communication system is given in Figure 1.1. The $\mathbb{R}$-valued amplitude of the transmitted optical waveform is given by

$$x^O(t, \mathbf{r}) \quad = \quad \sqrt{x(t)}\tilde{x}(\mathbf{r}) \cos(2\pi f_0 t), \quad t \geq 0, \quad \mathbf{r} \in \mathbb{R}^3, \tag{1.1}$$

where $\{x(t), t \geq 0\}$ is the $\mathbb{R}_0^+$-valued intensity of the transmitted signal (which is proportional to the modulating electrical current at the photoemitter); $\{\tilde{x}(\mathbf{r}), \mathbf{r} \in \mathbb{R}^3\}$ is the $\mathbb{R}$-valued directional (or spatial) component of the amplitude of the transmitted waveform; and f_0 is the optical carrier frequency. In practice, the light emitting devices are limited in transmission power, which is captured in terms of the following peak and average transmitter power constraints:

$$\begin{aligned} 0 \quad \leq \quad x(t) \quad &\leq \quad A, \quad 0 \leq t \leq T, \\ \tfrac{1}{T} \int_0^T x(\tau) d\tau \quad &\leq \quad \sigma A, \end{aligned} \tag{1.2}$$

where $A > 0$ and $0 \leq \sigma \leq 1$ are fixed.

The optical waveform, propagating through the turbulent atmospheric channel, undergoes frequency nonselective time-varying fading, so that the amplitude of the received optical waveform (in the absence of dark current) is given by

$$R^O(t, \mathbf{r}) \quad = \quad \sqrt{S(t)}x^O(t; \mathbf{r}), \quad t \geq 0, \quad \mathbf{r} \in \mathbb{R}^3, \tag{1.3}$$

where $\{S(t),\, t \geq 0\}$ is the $\mathbb{R}_0^+$-valued multiplicative channel fade (or channel state). The receiver is assumed to possess perfect channel state information (CSI) $\{S(t),\, t \geq 0\}$, while the transmitter CSI is given by $\{U(t) = h(S(t)),\, t \geq 0\}$, with a given mapping $h : \mathbb{R}_0^+ \to \mathcal{U}$, where $\mathcal{U}$ is arbitrary. The electrical current generated at the (idealized) direct detection receiver is a Poisson counting process (PCP) $\{Y(t),\, t \geq 0\}$ with rate (or intensity)

$$\Lambda(t) \quad = \quad R(t) + \lambda_0, \tag{1.4}$$

where $\{R(t) \propto S(t)x(t),\, t \geq 0\}$ is the intensity of the received optical signal, and $\lambda_0 \geq 0$ is the background noise (dark current) rate which is assumed to be constant. Henceforth we refer to this channel model as the *Poisson fading channel*.

It should be noted that our model ignores the bandwidth limitations of the transmitter and receiver devices currently used in practice. We also assume that the effects of infrared and visible background radiation can be effectively captured by a constant rate at the photodetector. These assumptions lead to simple channel models amenable to an exact analysis. Other models have been proposed in the literature that describe the background noise as additive white Gaussian (cf. e.g., [20, 23, 26, 51, 53, 54] and the references therein).

In this dissertation, we provide a systematic treatment of the problem of the determination of the capacity of the multiple input multiple output (MIMO) Poisson fading channel, and analyze how the knowledge of varying levels of CSI at the transmitter with perfect CSI at the receiver can be favorably used to enhance capacity.

1.2.4 Previous work

The direct detection photon channel *without any fading* has been studied extensively in the literature. Most of these studies focus on the Poisson channel with transmitter power constraints. Kabanov [25] used martingale techniques[5] to determine the capacity of the Poisson channel subject to a peak transmitter signal power constraint. Davis [11] considered the same channel subject to both peak and average transmitter power constraints. In [52], Wyner employed more direct methods to derive not only the capacity, but also an exact expression for the error exponent of the Poisson channel subject to peak and average power constraints. A strong converse for the coding theorem using a sphere packing bound was derived in [4]. The capacity of the Poisson channel with random noise intensity subject to time-varying peak and average power constraints was derived in [12]. The capacity region of the Poisson multiple-access channel was obtained in [32], and the error exponents for this channel were determined in [2]. The Poisson broadcast channel has been addressed in [33]. Capacity and decoding rules for the Poisson arbitrarily varying channel were discussed in [3].

[5] See [1] for a comprehensive treatment of the application of martingale techniques to the theory of point processes. In this dissertation, we do not employ this powerful analytical tool to prove our results, but rely instead on more elementary methods.

In practice, the transmitting devices are limited in bandwidth. This important issue was considered in [41, 42], where the author discussed the Poisson channel with constraints on the intertransition times of the transmitted signal. More general spectral constraints on the transmitted signals were considered in [43]. In [31], bounds on channel capacity were obtained for a discrete-time Poisson channel.

In the last decade, several researchers have investigated the use of multiple transmit and receive apertures in optical channels. In [34], the authors have demonstrated that use of multiple transmit and receive apertures can lead to reduction in bit error rate. Of direct relevance to our work are the recent results of Haas and Shapiro [19, 21]. Upper and lower bounds on the capacity of the MIMO Poisson channel with deterministic channel fade are derived in [19], whereas ergodic and outage capacity issues for the MIMO Poisson channel with random channel fade are addressed in [21].

1.3 Overview of the dissertation

This dissertation is organized as follows. We begin with the single input single output (SISO) Poisson fading channel in Chapter 2. We propose a block fading channel model that takes into account the slowly varying nature of optical fade. The channel coherence time T_c is used as a measure of the intermittent coherence of the time-varying fade. The channel fade is assumed to remain fixed over time intervals of width T_c, and change in an independent and identically distributed (i.i.d.) manner across successive such intervals. The receiver is assumed to possess perfect CSI, while the transmitter CSI may be imperfect. Under these assumptions, we characterize the capacity of the Poisson fading channel when the transmitted signals are subject to peak and average power constraints, and analyze properties of optimal transmission strategies that achieve channel capacity. Two extreme cases of this general formulation are of special interest, viz., (a) perfect CSI at transmitter, and (b) no CSI at transmitter. We also study the behavior of channel capacity in the high and low signal-to-noise regimes, and identify situations when the availability of good estimates of CSI at the transmitter can lead to increased throughput.

In Chapter 3, we consider a MIMO Poisson channel with constant channel fade, i.e., the channel fade is assumed to remain constant throughout the duration of transmission. Though this assumption of deterministic channel fade does not capture the practical channel conditions effectively, this model is rich enough to reveal some of the complexities of communicating over MIMO Poisson channel with random fade (which is discussed in Chapter 4), and to provide useful insights into the structure of efficient transmission strategies. In this setting, we allow for two kinds of transmitter power constraints: (a) peak and average transmitter power constraints on individual transmit apertures; and (b) individual peak power constraints and an average power constraint on the sum of the transmitted signals from all transmit apertures. We characterize the properties of optimal transmission strategies that achieve capacity, and discuss their implications on communication system design. The average sum power constraint, although apparently artificially introduced, finds an application in Chapter 4, where a simplified expression for channel capacity is derived for the symmetric MIMO Poisson channel with isotropic channel fade.

The MIMO Poisson channel with random channel fade is addressed in Chapter 4. A block fading channel model is considered; the channel fade matrix is assumed to remain unvarying on intervals of duration T_c, and vary in an i.i.d. manner across successive intervals. The receiver is assumed to possess perfect CSI, while the transmitter CSI can be imperfect. The transmit apertures are subject to individual peak and average power constraints. In this setting, we investigate the general capacity problem for the MIMO Poisson fading channel. Several important properties of optimal transmission strategies are identified for the symmteric MIMO channel when the channel fade is isotropically distributed and the transmitter has perfect CSI.

Finally, in the concluding section in Chapter 5, a new RF/optical sum channel model is introduced, in which information is transmitted using one of two available wireless channels, viz., an RF fading channel and an optical channel (without fading), depending on the instantaneous RF channel conditions. Furthermore, by randomly switching between the two channels, additional information can be conveyed. We outline some of the information theoretic issues which might arise in communication over such a hybrid RF/optical wireless channel.

1.4 Contributions of this dissertation

We conclude this chapter by outlining the main contributions of this dissertation.

(i) A block fading channel model has been introduced for the slowly varying free space optical fading channel. The channel fade is assumed to remain unvarying on intervals of duration T_c, where T_c is the channel coherence time, and is assumed to vary in an i.i.d. manner across successive such intervals. This simplistic abstraction of a complicated physical phenomenon allows us to derive exact expressions for channel capacity and to identify properties of optimal transmission strategies for reliable communication over a shot-noise limited Poisson fading channel.

(ii) We have analyzed two techniques to combat fading in optical channels, viz., (a) estimation of channel state, and use of channel state information (CSI) at the transmitter and the receiver; and (b) use of multiple apertures at the transmitter and receiver. The receiver is assumed to possess perfect CSI, while the transmitter CSI can be imperfect. In this setting, we have obtained a single-letter characterization of the capacity of a block fading MIMO Poisson channel subject to peak and average transmitter power constraints, and discussed several properties of optimal transmission and reception strategies.

(iii) We have demonstrated that a two-level signaling scheme ("ON-OFF keying"), in which each transmit aperture either remains silent ("OFF") or transmits at its peak power level ("ON"), with arbitrarily fast intertransition times can achieve channel capacity. Furthermore, the capacity of the block fading Poisson channel *with perfect receiver CSI* does not depend on the coherence time T_c.

(iv) For the SISO Poisson fading channel, it has been established that a knowledge of CSI at the transmitter leads to higher channel capacity in the high signal-to-noise (SNR) regime when the average power constraint is stringent, while in the low SNR regime, the transmitter CSI does not increase capacity. This is in contrast to the results cited in [21], where the authors have claimed that at the high SNR regime, a knowledge of CSI at transmitter does not improve capacity.

(v) For the MIMO Poisson channel with constant channel fade, a key property of the correlation structure of transmitted signals from all the transmit apertures has been identified: it been established that whenever a transmit aperture remains ON, all the transmit apertures with higher allowable average power levels must also remain ON. Under a symmetric average power constraint, it follows that the "simultaneous ON-OFF keying" strategy, which dictates all the transmit apertures to simultaneously remain either ON or OFF, can achieve channel capacity, thereby establishing the tightness of the lower bound proposed in [19].

(vi) For the symmetric MIMO Poisson channel with isotropically distributed random fade, the notion of "mirror states" has been introduced to obtain a simplified expression for channel capacity. We have identified a partitioning of the channel state set, which separates the channel states according to the relative ordering of the transmit apertures' optimal average conditional duty cycles. The simultaneous ON-OFF keying lower bound is shown to be strictly suboptimal for the MIMO Poisson channel with isotropic fade, even when the power constraints are symmetric. An improved lower bound is proposed based on our results for the MIMO Poisson channel with deterministic fade and a sum average power constraint.

(vii) We have studied a combined RF/optical sum channel with perfect RF fade information at the receiver and partial RF fade information at the transmitter. We have determined an expression for capacity of this channel in terms of the capacities of the individual RF and optical channels, and a gain in the code rate due to random switching. A complete characterization of the optimal power control and switching strategies, which are functions of transmitter RF fade information, remains unresolved, primarily due to the lack of an exact expression for capacity of the discrete-time Poisson channel.

Chapter 2

SISO Poisson fading channel

2.1 Introduction

In this chapter, we consider a single-user single input single output (SISO) Poisson fading channel limited by shot noise. Information is transmitted over this channel by modulating the intensity of an optical signal, and the receiver performs direct detection, which in effect, counts individual photon arrivals at the photodetector. The nonnegative transmitted signal is constrained in its peak and average power. A block fading channel model is introduced that accounts for the slowly varying nature of optical fade; the channel fade remains constant for a coherence interval of a fixed duration T_c (seconds), and changes across successive such intervals in an independent and identically distributed (i.i.d.) fashion. We then consider situations in which varying levels of information regarding the channel fade, i.e., channel state information (CSI) can be provided to the transmitter, while the receiver has perfect CSI.

We provide a systematic treatment of the problem of the determination of the capacity of a Poisson fading channel, and analyze how the knowledge of varying levels of CSI at the transmitter with perfect CSI at the receiver can favorably be used to enhance capacity. Our model is similar to the one studied in [21], but in this chapter is limited to the case of SISO single-user channel. In this setting,

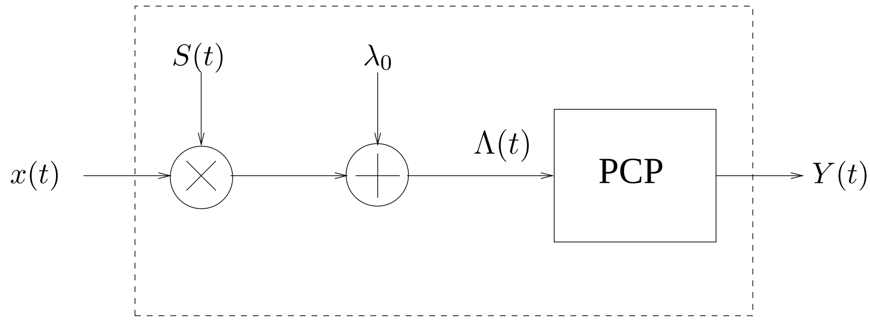

Figure 2.1: Poisson fading channel.

we consider a more general class of problems than those considered in [21]. From our new capacity results, conclusions are drawn that differ considerably from those reported in [21].

The remainder of this chapter is organized as follows. Section 2.2 deals with the problem formulation. Our results are stated in Section 2.3 and proved in Section 2.4. An illustrative example of a channel with a lognormal fade is considered in Section 2.5. Finally, some concluding remarks are provided in Section 2.6.

2.2 Problem formulation

The following notation will be used throughout this dissertation. Random variables (rvs) are denoted by upper-case letters and random vectors by bold upper-case letters. We use the notation $\mathbf{X}_i^j$ to denote a sequence of rvs $\{X_i, X_{i+1}, \cdots, X_j\}$; when $i = 1$, we use $\mathbf{X}^j = \{X_1, \cdots, X_j\}$. A continuous time random process $\{X(t), a \leq t \leq b\}$ is denoted in shorthand notation by X_a^b; when $a = 0$, we use $X^b = \{X(t), 0 \leq t \leq b\}$. Realizations of rvs and random processes, which are denoted in lower-case letters, follow the same convention.

A block schematic diagram of the channel model is given in Figure 2.1. For a given $\mathbb{R}_0^+$-valued[1] transmitted signal $\{x(t),\, t \geq 0\}$, the received signal $Y^\infty = \{Y(t),\, t \geq 0\}$ is a $\mathbb{Z}_0^+$-valued nondecreasing (left-continuous) Poisson counting process (PCP) with rate (or intensity) equal to

$$\Lambda(t) = S(t)x(t) + \lambda_0, \quad t \geq 0,$$

where $\{S(t),\, t \geq 0\}$ is the $\mathbb{R}_0^+$-valued random fade, and $\lambda_0 \geq 0$ is the background noise (dark current) rate which is assumed to be constant. Note that Y^∞ is an independent increments process with $Y(0) = 0$, such that for $0 \leq \tau,\, t < \infty$,

$$\Pr\left\{Y(t+\tau) - Y(t) = j | \Lambda_t^{t+\tau} = \lambda_t^{t+\tau}\right\} = \frac{1}{j!} e^{-\Gamma(\lambda_t^{t+\tau})} \Gamma^j(\lambda_t^{t+\tau}), \quad j = 0, 1, \cdots,$$

where $\Gamma(\lambda_t^{t+\tau}) = \int_t^{t+\tau} \lambda(u) du$. Physically, the jumps in Y^∞ correspond to the arrival of photons in the receiver. Let $\Sigma(T)$ denote the space of nondecreasing, left-continuous, piecewise-constant, $\mathbb{Z}_0^+$-valued functions $\{g(t),\, 0 \leq t \leq T\}$ with $g(0) = 0$. Then the output process $Y^T = \{Y(t),\, 0 \leq t \leq T\}$ takes values in $\Sigma(T)$.

The input to the channel is a $\mathbb{R}_0^+$-valued signal $x^T = \{x(t),\, 0 \leq t \leq T\}$ which is proportional to the transmitted optical power, and which satisfies peak and average power constraints of the form:

$$\begin{aligned} 0 \;\leq\; x(t) \;&\leq\; A, \qquad 0 \leq t \leq T, \\ \tfrac{1}{T}\int_0^T x(t)dt \;&\leq\; \sigma A, \end{aligned} \tag{2.1}$$

where the peak power $A > 0$ and the ratio of average-to-peak power σ, $0 \leq \sigma \leq 1$, are fixed.

[1]The set of nonnegative real numbers is denoted by $\mathbb{R}_0^+$, the set of positive integers by $\mathbb{Z}^+$, and the set of nonnegative integers by $\mathbb{Z}_0^+$.

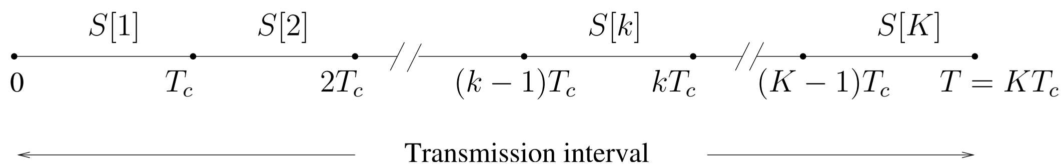

Figure 2.2: Block fading channel.

The channel fade, i.e., path gain, is modeled by a $\mathbb{R}_0^+$-valued random process $\{S(t),\, t \geq 0\}$. The channel coherence time T_c is a measure of the intermittent coherence of the time-varying channel fade. We assume that the channel fade remains fixed over time intervals of width T_c, and changes in an i.i.d. manner across successive such intervals. For $k = 1, 2, \cdots$, let the channel fade on $[(k-1)T_c, kT_c)$ be denoted by the rv $S[k]$ (see Figure 2.2); in other words,

$$S(t) = S[k], \quad t \in [(k-1)T_c, kT_c), \quad k = 1, 2, \cdots .$$

The channel fade is then described by the random sequence $\mathbf{S}^\infty = \{S[1], S[2], \cdots\}$ of i.i.d. repetitions of a rv S with known distribution. Our general results hold for a broad class of distributions for S which satisfy the following technical conditions: $\Pr\{S > 0\} = 1$, $\mathbb{E}[S] < \infty$ and[2] $\mathbb{E}[S \log S] < \infty$. Note that the rate of the received signal Y^∞ is now given by[3]

$$\Lambda(t) = S[\lceil t/T_c \rceil]x(t) + \lambda_0, \quad t \geq 0.$$

In an illustrative example discussed in Section 2.5 below, we shall use the lognormal distribution with normalized channel fade, as advocated in [21].

[2]Unless mentioned otherwise, all logarithms are natural logarithms.

[3]The expression $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.

Various degrees of CSI can be made available to the transmitter and the receiver. *We shall assume throughout that the receiver has perfect CSI*[4]. In general, we can model the CSI available at the transmitter in terms of a given mapping $h : \mathbb{R}_0^+ \to \mathcal{U}$, where $\mathcal{U}$ is an arbitrary subset of $\mathbb{R}_0^+$, not necessarily finite. For $\mathbf{S}^\infty = \mathbf{s}^\infty$, the transmitter (resp. receiver) is provided with CSI $u[k] = h(s[k])$ (resp. $s[k]$) on $[(k-1)T_c, kT_c)$, $k = 1, 2, \cdots$. Let $\{U[k] = h(S[k]), k = 1, 2, \cdots\}$ denote the CSI at the transmitter, hereafter referred to as the *transmitter CSI* h. We shall be particularly interested in two special cases of this general framework. In the first case, the transmitter has perfect CSI, i.e., h is the identity mapping. In the second case, the transmitter is provided no CSI, i.e., h is the trivial (constant) mapping.

We assume, without loss of generality, that the message transmission duration T is an integer multiple of the channel coherence time T_c, i.e., $T = KT_c$, $K \in \mathbb{Z}^+$. For the channel under consideration, a (W, T)–code (f, ϕ) is defined as follows.

1. For each $\mathbf{u}^K \in \mathcal{U}^K$, the codebook comprises a set of W waveforms $f(w, \mathbf{u}^K) = \{x_w(t, \mathbf{u}^{\lceil t/T_c \rceil}), 0 \le t \le T\}$, $w \in \mathcal{W} = \{1, \cdots, W\}$, satisfying peak and average power constraints which follow from (2.1):

$$\begin{aligned} 0 \ \le \ x_w(t, \mathbf{u}^{\lceil t/T_c \rceil}) \ &\le \ A, \quad 0 \le t \le T, \quad w \in \mathcal{W}; \\ \frac{1}{T}\int_0^T x_w\left(t, \mathbf{u}^{\lceil t/T_c \rceil}\right) dt \ &\le \ \sigma A, \quad w \in \mathcal{W}. \end{aligned} \tag{2.2}$$

   Note that the transmitter output, i.e., the signal $x_w(t)$ at time t is allowed to be a function of $\mathbf{u}^{\lceil t/T_c \rceil}$.

[4]Typically T_c is of the order of $1-10$ ms, during which $1-10$ Mb can be transmitted at a rate of 1 Gbps. A fraction of the bits transmitted and received during a coherence interval can be used by the receiver to estimate the prevailing fade value.

2. The decoder is a mapping $\phi : \Sigma(T) \times (\mathbb{R}_0^+)^K \to \mathcal{W}$.

For each message $w \in \mathcal{W}$ and transmitter CSI $\mathbf{u}^K \in \mathcal{U}^K$ corresponding to the fade vector $\mathbf{s}^K \in (\mathbb{R}_0^+)^K$, the transmitter sends a waveform $x_w(t,\, \mathbf{u}^{\lceil t/T_c \rceil})$, $0 \le t \le T$, over the channel. The receiver, upon observing y^T and being provided with $\mathbf{s}^K$, produces an output $\hat{w} = \phi(y^T,\, \mathbf{s}^K)$. The rate of this $(W,\, T)$–code is given by $R = \frac{1}{T} \log W$ nats/sec., and its average probability of decoding error is given by

$$P_e(f,\, \phi) = \frac{1}{W} \sum_{w=1}^{W} \mathbb{E}\left[\Pr\left\{ \phi(Y^T, \mathbf{S}^K) \neq w \left| x_w^T(\mathbf{U}^K), \mathbf{S}^K \right. \right\} \right],$$

where we have used the shorthand notation

$$x_w^T\left(\mathbf{U}^K\right) = \left\{ x_w\left(t, \mathbf{U}^{\lceil t/T_c \rceil}\right),\ \ 0 \le t \le T \right\}, \quad w \in \mathcal{W}.$$

Definition 1 *Let A, λ_0, σ, T_c be fixed. Given $0 < \epsilon < 1$, a number $R > 0$ is an ϵ-achievable rate if for every $\delta > 0$ and for all T sufficiently large, there exist $(W,\, T)$–codes $(f,\, \phi)$ such that $\frac{1}{T} \log W > R - \delta$ and $P_e(f,\, \phi) < \epsilon$; R is an achievable rate if it is ϵ-achievable for all $0 < \epsilon < 1$. The supremum of all achievable rates is the capacity of the channel, and will be denoted by C.*

2.3 Statement of results

Our first and main result provides a single–letter characterization of the capacity of the Poisson fading channel model described above. Recall our persistent assumption that the receiver has perfect CSI. The two special cases of perfect and no CSI at the transmitter are considered next. Finally, we analyze the limiting behavior of channel capacity in the high and low signal-to-shot-noise ratio (SNR) regimes, viz. in the limits as $\lambda_0 \to 0$ and $\lambda_0 \to \infty$, respectively.

For stating our results, it is convenient to set

$$\zeta(x, y) \triangleq (x+y)\log(x+y) - y\log y, \quad x \geq 0, \quad y \geq 0 \tag{2.3}$$

(with $0 \log 0 \triangleq 0$), and

$$\alpha(x) \triangleq \frac{1}{x}\left(e^{-1}(1+x)^{(1+1/x)} - 1\right), \quad x \geq 0, \tag{2.4}$$

whence it can be verified that

$$\zeta(x, y) - x(1+\log y) = x\log(1+\alpha(x/y)x/y), \quad x \geq 0, \quad y \geq 0. \tag{2.5}$$

Note that $\zeta(\cdot, y)$ is strictly convex on $[0, \infty)$ for every $y \geq 0$.

Theorem 1 *Let A, λ_0, σ, T_c be fixed. The capacity for transmitter CSI h is given by*

$$C = \max_{\substack{\mu:\mathcal{U}\to[0,1] \\ \mathbf{E}[\mu(U)]\leq\sigma}} \mathbb{E}\left[\mu(U)\zeta(SA, \lambda_0) - \zeta(\mu(U)SA, \lambda_0)\right], \tag{2.6}$$

where $U = h(S)$.

The capacities for the special cases of perfect and no CSI at the transmitter follow directly from Theorem 1.

Corollary 1 *The capacity for perfect CSI at the transmitter and the receiver is given by*

$$C_P = \max_{\substack{\mu:\mathbb{R}_0^+\to[0,1] \\ \mathbf{E}[\mu(S)]\leq\sigma}} \mathbb{E}\left[\mu(S)\zeta(SA, \lambda_0) - \zeta(\mu(S)SA, \lambda_0)\right]. \tag{2.7}$$

The capacity for no CSI at the transmitter and perfect CSI at the receiver is given by

$$C_N = \max_{0 \le \mu \le \sigma} \mathbb{E}\left[\mu\zeta(SA,\, \lambda_0) - \zeta(\mu SA,\, \lambda_0)\right]. \tag{2.8}$$

Remarks: (i) The optimization in (2.6) (as also in (2.7), (2.8)) is that of a concave functional over a convex compact set, so that the maximum clearly exists.

(ii) From Theorem 1, we see that the channel capacity does not depend on the coherence time T_c. Conditioned on the transmitted signal $\{x_t,\, 0 \le t \le T\}$, and perfect receiver CSI $\mathbf{s}^K$, the received signal $\{Y_t,\, 0 \le t \le T\}$ is independent across coherence intervals; hence it suffices to look at a single coherence interval in the mutual information computations. Furthermore, in a single coherence interval, conditioned on perfect receiver CSI, the optimality of i.i.d. transmitted signals leads to a lack of dependence of capacity on T_c. The fact that the channel capacity of a block fading channel with perfect receiver CSI does not depend on the block size has been reported in the literature in various settings in other contexts (cf., e.g., [37] for such a result on block interference channels).

(iii) Our proof of the achievability part of Theorem 1 shows that $\{0,\, A\}$-valued transmitted signals, which are i.i.d. (conditioned on the current transmitter CSI) with arbitrarily fast intertransition times, can achieve channel capacity. This is in concordance with previous results (cf., e.g., [52]), where the optimality of binary signaling for Poisson channels has been established.

(iv) The optimizing "power control law" μ in (2.6), (2.7) and (2.8) depicts the probability with which the transmitter picks the signal level A depending on the available transmitter CSI. Thus, it can be interpreted as the optimal average "conditional duty cycle" of the transmitted signal as a function of transmitter CSI.

Theorem 2 *The optimal power control law $\mu^* : \mathcal{U} \to [0,\, 1]$ that achieves the maximum in (2.6) is given as follows. For $\rho \geq 0$, $u \in \mathcal{U}$, let $\mu = \mu_\rho(u)$ be the solution of the equation*

$$\mathbb{E}\left[SA \log \frac{(1 + \alpha(SA/\lambda_0) SA/\lambda_0)}{(1 + \mu SA/\lambda_0)} \middle| U = u \right] \;=\; \rho. \tag{2.9}$$

If $\sigma_0 \triangleq \mathbb{E}[\mu_0(U)] > \sigma$, let $\rho = \rho^* > 0$ be the solution of the equation[5]

$$\mathbb{E}\left[[\mu_\rho(U)]^+ \right] \;=\; \sigma. \tag{2.10}$$

Then the optimal power control law μ^* in (2.6) is given by

$$\mu^*(u) \;=\; \begin{cases} \mu_0(u), & \sigma_0 \leq \sigma, \\ [\mu_{\rho^*}(u)]^+, & \sigma_0 > \sigma, \quad u \in \mathcal{U}. \end{cases} \tag{2.11}$$

The following corollary particularizes the previous optimal power control law in the special cases of perfect and no CSI at the transmitter.

Corollary 2 *For perfect CSI at the transmitter, the optimal power control law μ^* : $\mathbb{R}_0^+ \to [0,\, 1]$ that achieves the maximum in (2.7) is given as follows. For $\rho \geq 0$, let*

$$\mu_\rho(s) \;\triangleq\; \frac{\lambda_0}{sA}\left(e^{-\left(1 + \frac{\rho}{sA}\right)} \left(1 + \frac{sA}{\lambda_0}\right)^{\left(1 + \frac{\lambda_0}{sA}\right)} - 1 \right), \quad s \in \mathbb{R}_0^+. \tag{2.12}$$

Let $\sigma_0 \triangleq \mathbb{E}[\mu_0(S)]$, and if $\sigma_0 > \sigma$, let $\rho = \rho^* > 0$ be the solution of the equation $\mathbb{E}\left[[\mu_\rho(S)]^+\right] = \sigma$. Then the optimal power control law μ^* in (2.7) is given by

$$\mu^*(s) \;=\; \begin{cases} \mu_0(s), & \sigma_0 \leq \sigma, \\ [\mu_{\rho^*}(s)]^+, & \sigma_0 > \sigma, \quad s \in \mathbb{R}_0^+. \end{cases} \tag{2.13}$$

[5] We denote $\max\{x, 0\}$ by $[x]^+$.

For no CSI at the transmitter, the maximum in (2.8) is achieved by $\mu^* = \min\{\sigma,\, \mu_0\}$*, with* μ_0 *being the solution of the equation*

$$\mathbb{E}\left[SA \log \frac{(1+\alpha(SA/\lambda_0)SA/\lambda_0)}{(1+\mu_0 SA/\lambda_0)}\right] \;=\; 0. \tag{2.14}$$

Remarks: (i) It can be verified that the left-side of (2.10) decreases monotonically from σ_0 to 0 as ρ increases from 0 to ∞, so that for each $0 \le \sigma < \sigma_0$, there exists a unique $\rho = \rho^* > 0$ that solves (2.10). Furthermore, it can be shown that for each $u \in \mathcal{U}$, $0 \le \mu^*(u) \le 1/2$, so that the power control law given by (2.11) is well-defined.

(ii) For the case of perfect transmitter CSI, the optimal power control law in (2.13) differs from, and yields a higher capacity value than, the claimed optimal power control law in ([21], eq. 4). An example in Section 2.5 shows the difference in the values of channel capacity when computed using the two power control laws for a range of values of A/λ_0 and σ.

The peak signal-to-noise ratio, ,denoted SNR, is defined as $\mathrm{SNR} = A/\lambda_0$. We characterize next the capacity in the low and high shot-noise regimes (equivalently the high and low SNR regimes) when the peak signal power A is fixed.

Theorem 3 *In the high SNR regime in the limit as* $\lambda_0 \to 0$*, the capacity for transmitter CSI* h *is*

$$C^H \;=\; \mathbb{E}\left[\mu^H(U)\zeta(SA,\,0) - \zeta(\mu^H(U)SA,\,0)\right], \tag{2.15}$$

with $U = h(S)$*, where the optimal power control law* $\mu^H : \mathcal{U} \to [0,\,1]$ *is given as follows: if* $\sigma < 1/e$*, let* $\rho = \rho^* > 0$ *be the solution of the equation* $\mathbb{E}\left[e^{-\rho/A\,\mathbf{E}[S|U]}\right] =$

$e\sigma$; then μ^H is given by

$$\mu^H(u) = \begin{cases} e^{-1-\rho^*/A\,\mathbf{E}[S|U=u]}, & \sigma < 1/e, \\ e^{-1}, & \sigma \geq 1/e, \quad u \in \mathcal{U}. \end{cases} \tag{2.16}$$

In the low SNR regime for $\lambda_0 \gg 1$, the capacity is[6]

$$C^L = \mu^L(1-\mu^L)\,\mathbb{E}[S^2]A^2/2\lambda_0 + O\left(\lambda_0^{-2}\right), \tag{2.17}$$

where $\mu^L = \min\{\sigma,\, 1/2\}$.

Remark: In the low SNR regime for $\lambda_0 \gg 1$, we see from (2.17) that transmitter CSI does not improve capacity. The capacity increases with SNR (approximately) linearly with a slope proportional to $\mu^L(1-\mu^L)$, where the optimal conditional duty cycle $\mu^L = \min\{1/2,\, \sigma\}$ does not depend on the transmitter CSI. However, in the high SNR regime as $\lambda_0 \to 0$, if $\sigma < 1/e$, the optimal conditional duty cycle μ^H (and hence the capacity in (2.15)) depends on transmitter CSI. This is in contrast with the results in [21], where the authors have argued that at high SNR, a knowledge of the fade at the transmitter does not improve capacity.

In the next two corollaries of Theorem 3, we characterize the capacity in the high and low SNR regimes, in the special cases when the transmitter is provided with perfect and no CSI, respectively.

Corollary 3 *For perfect CSI at transmitter, in the high SNR regime in the limit as $\lambda_0 \to 0$, the capacity is*

$$C_P^H = \mathbb{E}\left[\mu^H(S)\zeta(SA,\, 0) - \zeta(\mu^H(S)SA,\, 0)\right], \tag{2.18}$$

[6]By the standard notation $f(x) = O(g(x))$, we mean that there exists a number $0 \leq A < \infty$, not depending on x, such that $f(x) \leq Ag(x)\ \forall\ x \geq x_0$, where $x_0 = x_0(A) < \infty$.

where the optimal power control law $\mu^H : \mathbb{R}_0^+ \to [0,\, 1]$ is given as follows: if $\sigma < 1/e$, let $\rho = \rho^* > 0$ be the solution of $\mathbb{E}\left[e^{-\rho/SA}\right] = e\sigma$; then μ^H is given by

$$\mu^H(s) \;=\; \begin{cases} e^{-1-\rho^*/sA}, & \sigma < 1/e, \\ e^{-1}, & \sigma \geq 1/e, \quad s \in \mathbb{R}_0^+. \end{cases} \tag{2.19}$$

In the low SNR regime for $\lambda_0 \gg 1$, the capacity is the same as in (2.17), i.e.,

$$C_P^L \;=\; \mu^L(1-\mu^L)\,\mathbb{E}[S^2]A^2/2\lambda_0 + O\left(\lambda_0^{-2}\right), \tag{2.20}$$

where $\mu^L = \min\{\sigma,\, 1/2\}$.

Corollary 4 *For no CSI at transmitter, in the high SNR regime in the limit as $\lambda_0 \to 0$, the capacity is*

$$C_N^H \;=\; \mathbb{E}\left[\mu^H \zeta(SA,\, 0) - \zeta(\mu^H SA,\, 0)\right], \tag{2.21}$$

where $\mu^H = \min\{\sigma,\, 1/e\}$.

In the low SNR regime for $\lambda_0 \gg 1$, the capacity is the same as in (2.17), i.e.,

$$C_N^L \;=\; \mu^L(1-\mu^L)\,\mathbb{E}[S^2]A^2/2\lambda_0 + O\left(\lambda_0^{-2}\right), \tag{2.22}$$

where $\mu^L = \min\{\sigma,\, 1/2\}$.

2.4 Proofs

We begin this section with some additional definitions that will be needed in our proofs. First, observe that for any $\tau > 0$, the number of photon arrivals N_τ on $[0,\, \tau]$ together with the corresponding (ordered) arrival times $\mathbf{T}^{N_\tau} = (T_1,\, \cdots,\, T_{N_\tau})$

are sufficient statistics for Y^τ, so that the random vector $(N_T,\, \mathbf{T}^{N_T})$ is a complete description of the random process Y^T.

The channel is characterized as follows. For an input signal x^T satisfying (2.1) and a fade vector $\mathbf{S}^K = \mathbf{s}^K$, the channel output $(N_T,\, \mathbf{T}^{N_T})$ has the "conditional sample function density" (cf. e.g., [47])

$$f_{N_T, \mathbf{T}^{N_T} | X^T, \mathbf{S}^K} \left(n_T,\, \mathbf{t}^{n_T} \,\middle|\, x^T,\, \mathbf{s}^K \right) \;=\; \exp\left(-\int_0^T \lambda(\tau) d\tau \right) \cdot \prod_{i=1}^{n_T} \lambda(t_i), \quad (2.23)$$

where

$$\lambda(\tau) = s[\lceil \tau / T_c \rceil] x(\tau) + \lambda_0, \quad 0 \leq \tau \leq T.$$

In order to write the channel output sample function density conditioned only on the fade for a given joint distribution of $(X^T,\, \mathbf{S}^K)$, consider the conditional mean of X^T (conditioned causally on the channel output and the fade) by

$$\hat{X}(\tau) \;\triangleq\; \mathbb{E}\left[X(\tau) \,\middle|\, N_\tau,\, \mathbf{T}^{N_\tau},\, \mathbf{S}^{\lceil \tau / T_c \rceil} \right], \quad 0 \leq \tau \leq T, \quad (2.24)$$

where we have suppressed the dependence of $\hat{X}(\tau)$ on $(N_\tau,\, \mathbf{T}^{N_\tau},\, \mathbf{S}^{\lceil \tau / T_c \rceil})$ for notational convenience; and define

$$\Lambda(\tau) \;\triangleq\; S[\lceil \tau / T_c \rceil] X(\tau) + \lambda_0, \quad 0 \leq \tau \leq T, \quad (2.25)$$

and

$$\begin{aligned} \hat{\Lambda}(\tau) \;&\triangleq\; \mathbb{E}\left[\Lambda(\tau) \,\middle|\, N_\tau,\, \mathbf{T}^{N_\tau},\, \mathbf{S}^{\lceil \tau / T_c \rceil} \right] \\ &=\; S_{\lceil \tau / T_c \rceil} \hat{X}(\tau) + \lambda_0, \quad 0 \leq \tau \leq T. \quad (2.26) \end{aligned}$$

From ([47], Theorem 7.2.1), it follows that conditioned on $\mathbf{S}^K$, the process $(N_T,\, \mathbf{T}^{N_T})$ is a self-exciting PCP with rate process $\hat{\Lambda}^T$, and the (conditional) sample function density is given by

$$f_{N_T,\, \mathbf{T}^{N_T}|\mathbf{S}^K}\left(n_T,\, \mathbf{t}^{n_T}\,\middle|\mathbf{s}^K\right) \;=\; \exp\left(-\int_0^T \hat{\lambda}(\tau)d\tau\right)\cdot\prod_{i=1}^{n_T}\hat{\lambda}(t_i), \qquad (2.27)$$

where

$$\hat{\lambda}(\tau) = s[\lceil \tau/T_c\rceil]\hat{x}(\tau) + \lambda_0,$$

with

$$\hat{x}(\tau) \;=\; \mathbb{E}\left[X(\tau)\,\middle|N_\tau = n_\tau,\, \mathbf{T}^{N_\tau} = \mathbf{t}^{n_\tau},\, \mathbf{S}^{\lceil \tau/T_c\rceil} = \mathbf{s}^{\lceil \tau/T_c\rceil}\right], \quad 0 \le \tau \le T.$$

Proof of Theorem 1:

Converse part: Let the rv W be uniformly distributed on (the message set) $\mathcal{W} = \{1, \cdots, W\}$, and independent of $\mathbf{S}^K$. With $T = KT_c$, $K \in \mathbb{Z}^+$, consider a (W, T)–code (f, ϕ) of rate $R = \frac{1}{T}\log W$, and with $P_e(f, \phi) \le \epsilon$, where $0 \le \epsilon \le 1$ is given. Denote $X(t) \triangleq x_W(t, \mathbf{U}^{\lceil t/T_c\rceil})$, $0 \le t \le T$. Note that (2.2) then implies that

$$\begin{aligned} 0 \;\le\; X(t) \;&\le\; A, \qquad 0 \le t \le T, \\ \tfrac{1}{T}\textstyle\int_0^T \mathbb{E}[X(\tau)]d\tau \;&\le\; \sigma A. \end{aligned} \qquad (2.28)$$

Let $(N_T,\, \mathbf{T}^{N_T})$ be the channel output when X^T is transmitted and the channel fade is $\mathbf{S}^K$. Clearly, the following Markov condition holds:

$$W -\!\circ\!- X^t\,\mathbf{S}^{\lceil t/T_c\rceil} -\!\circ\!- N_t\mathbf{T}^{N_t}, \quad 0 \le t \le T. \qquad (2.29)$$

By a standard argument,

$$\begin{aligned} R \;&=\; \frac{1}{T}H(W) \\ &=\; \frac{1}{T}\left[I(W \wedge \phi(N_T,\, \mathbf{T}^{N_T},\, \mathbf{S}^K)) + H(W|\phi(N_T,\, \mathbf{T}^{N_T},\, \mathbf{S}^K))\right], \end{aligned}$$

which, upon using Fano's inequality

$$\begin{aligned} H(W|\phi(N_T,\, \mathbf{T}^{N_T},\, \mathbf{S}^K)) &\leq \epsilon \log W + h_b(\epsilon) \\ &= \epsilon T R + h_b(\epsilon), \end{aligned}$$

leads to

$$R \;\leq\; \frac{1}{(1-\epsilon)}\left[\frac{1}{T} I(W \wedge \phi(N_T,\, \mathbf{T}^{N_T},\, \mathbf{S}^K)) + \frac{1}{T} h_b(\epsilon)\right], \tag{2.30}$$

where h_b denotes binary entropy. Since $0 < \epsilon < 1$ was arbitrary, we get the standard converse result that the rate R of the $(W,\, T)$–code $(f,\, \phi)$ with $P_e(f,\, \phi) \leq \epsilon$ must satisfy

$$R \;\lessapprox\; \frac{1}{T} I\left(W \wedge \phi(N_T,\, \mathbf{T}^{N_T},\, \mathbf{S}^K)\right). \tag{2.31}$$

Proceeding further with the right side of (2.31),

$$\begin{aligned} I\left(W \wedge \phi(N_T,\, \mathbf{T}^{N_T},\, \mathbf{S}^K)\right) &\leq I\left(W \wedge N_T,\, \mathbf{T}^{N_T},\, \mathbf{S}^K\right) && (2.32) \\ &= I\left(W \wedge N_T,\, \mathbf{T}^{N_T} | \mathbf{S}^K\right) && (2.33) \\ &= h(N_T,\, \mathbf{T}^{N_T}|\mathbf{S}^K) - h(N_T,\, \mathbf{T}^{N_T}|W,\, \mathbf{S}^K) \\ &\leq h(N_T,\, \mathbf{T}^{N_T}|\mathbf{S}^K) - h(N_T,\, \mathbf{T}^{N_T}|X^T, \mathbf{S}^K) && (2.34) \end{aligned}$$

where (2.32), (2.33) and (2.34) follow, respectively, by the data processing result for mixed rvs[7], the independence of W from $\mathbf{S}^K$, and (2.29). The difference between the conditional entropies of the mixed rvs[8] on the right side of (2.34) is

$$h(N_T,\, \mathbf{T}^{N_T}|\mathbf{S}^K) - h(N_T,\, \mathbf{T}^{N_T}|X^T,\, \mathbf{S}^K)$$

[7] This result can be deduced, for instance, from [39], the last paragraph of Section 3.4, p. 36, and Kolmogorov's formula (3.6.3) on p. 37.

[8] Our definition of the conditional entropy of mixed rvs is consistent with the general formulation developed, for instance, in [39], Chapter 3.

$$
\begin{aligned}
&= \mathbb{E}\left[-\log f_{N_T,\,\mathbf{T}^{N_T}|\mathbf{S}^K}(N_T,\ \mathbf{T}^{N_T}|\mathbf{S}^K)\right] \\
&\quad -\mathbb{E}\left[-\log f_{N_T,\,\mathbf{T}^{N_T}|X^T,\,\mathbf{S}^K}(N_T,\ \mathbf{T}^{N_T}|X^T,\ \mathbf{S}^K)\right] \\
&= \mathbb{E}\left[\log \frac{\exp\left(-\int_0^T \Lambda(\tau)d\tau\right)\prod_{i=1}^{N_T}\Lambda(T_i)}{\exp\left(-\int_0^T \hat{\Lambda}(\tau)d\tau\right)\prod_{i=1}^{N_T}\hat{\Lambda}(T_i)}\right] && (2.35) \\
&= \mathbb{E}\left[\int_0^T \left(\hat{\Lambda}(\tau)-\Lambda(\tau)\right)d\tau\right] + \mathbb{E}\left[\sum_{i=1}^{N_T}\left(\log\Lambda(T_i)-\log\hat{\Lambda}(T_i)\right)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{N_T}\left(\log\Lambda(T_i)-\log\hat{\Lambda}(T_i)\right)\right] && (2.36) \\
&= \int_0^T \left\{\mathbb{E}\left[\zeta\left(S[\lceil\tau/T_c\rceil]X(\tau),\ \lambda_0\right)\right] - \mathbb{E}\left[\zeta\left(S[\lceil\tau/T_c\rceil]\hat{X}(\tau),\ \lambda_0\right)\right]\right\}d\tau, && (2.37)
\end{aligned}
$$

where (2.35) is by (2.23), (2.27); (2.36) holds by an interchange of operations[9] to get

$$
\mathbb{E}\left[\int_0^T (\hat{\Lambda}(\tau)-\Lambda(\tau))d\tau\right] = \int_0^T \mathbb{E}\left[\hat{\Lambda}(\tau)-\Lambda(\tau)\right]d\tau,
$$

followed by noting that $\mathbb{E}[\hat{\Lambda}(\tau)] = \mathbb{E}[\Lambda(\tau)]$, $0 \le \tau \le T$, by (2.26); and (2.37) is proved in Appendix A.1.

Next, in the right side of (2.37),

$$
\begin{aligned}
&\mathbb{E}\left[\zeta\left(S[\lceil\tau/T_c\rceil]\hat{X}(\tau),\ \lambda_0\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\zeta\left(S[\lceil\tau/T_c\rceil]\hat{X}(\tau),\ \lambda_0\right)\middle|\, S[\lceil\tau/T_c\rceil]\right]\right] \\
&\ge \mathbb{E}\left[\zeta\left(\mathbb{E}\left[S[\lceil\tau/T_c\rceil]\hat{X}(\tau)\middle|\, S[\lceil\tau/T_c\rceil]\right],\ \lambda_0\right)\right] && (2.38) \\
&= \mathbb{E}\left[\zeta\left(S[\lceil\tau/T_c\rceil]\,\mathbb{E}\left[\hat{X}(\tau)\middle|\, S[\lceil\tau/T_c\rceil]\right],\ \lambda_0\right)\right] \\
&= \mathbb{E}\left[\zeta\left(S[\lceil\tau/T_c\rceil]\,\mathbb{E}\left[X(\tau)|\, S[\lceil\tau/T_c\rceil]\right],\ \lambda_0\right)\right] && (2.39) \\
&= \mathbb{E}\left[\zeta\left(S[\lceil\tau/T_c\rceil]\,\mathbb{E}\left[X(\tau)|\, U[\lceil\tau/T_c\rceil]\right],\ \lambda_0\right)\right], && (2.40)
\end{aligned}
$$

[9]The interchange is permissible as the assumed condition $\mathbb{E}[S] < \infty$ implies the integrability of $\{\Lambda(\tau), 0 \le \tau \le T\}$ and $\{\hat{\Lambda}(\tau), 0 \le \tau \le T\}$.

where (2.38) is by Jensen's inequality applied to the convex function $\zeta(\cdot,\,\lambda_0)$; (2.39) is from

$$\begin{aligned}\mathbb{E}\left[\hat{X}(\tau)\,|S[\lceil\tau/T_c\rceil]\right] &= \mathbb{E}\left[\mathbb{E}\left[X(\tau)\,\middle|N_\tau,\,\mathbf{T}^{N_\tau},\,\mathbf{S}^{\lceil\tau/T_c\rceil}\right]|S[\lceil\tau/T_c\rceil]\right] \\ &= \mathbb{E}\left[X(\tau)\,|S[\lceil\tau/T_c\rceil]\right];\end{aligned}$$

and (2.40) holds as

$$\begin{aligned}\mathbb{E}\left[X(\tau)\,|S[\lceil\tau/T_c\rceil]\right] &= \mathbb{E}\left[X(\tau)\,|S[\lceil\tau/T_c\rceil],\,U[\lceil\tau/T_c\rceil]\right] \\ &= \mathbb{E}\left[X(\tau)\,|U[\lceil\tau/T_c\rceil]\right] \qquad (2.41)\end{aligned}$$

by virtue of the Markov condition

$$X(\tau) -\!\!\circ\!\!- U[\lceil\tau/T_c\rceil] -\!\!\circ\!\!- S[\lceil\tau/T_c\rceil], \quad 0\le t\le T, \qquad (2.42)$$

which is established in Appendix A.2. Summarizing collectively (2.34), (2.37), (2.40), we get that

$$\begin{aligned}&I\left(W\wedge\phi(N_T,\,\mathbf{T}^{N_T},\,\mathbf{S}^K)\right) \\ &\quad\le \int_0^T\{\mathbb{E}\left[\zeta(S[\lceil\tau/T_c\rceil]X(\tau),\,\lambda_0)\right. \\ &\qquad\left.-\zeta(S[\lceil\tau/T_c\rceil]\,\mathbb{E}[X(\tau)|U[\lceil\tau/T_c\rceil]],\,\lambda_0)\right]\}\,d\tau. \qquad (2.43)\end{aligned}$$

The right side of (2.43) is further bounded above by a suitable modification of the argument of Davis [11]. Considering the integrand in (2.43), fix $0\le\tau\le T$ and condition on $S[\lceil\tau/T_c\rceil]=s$, $s\in\mathbb{R}_0^+$. Then

$$\begin{aligned}&\mathbb{E}\left[\zeta\left(S[\lceil\tau/T_c\rceil]X(\tau),\,\lambda_0\right)-\zeta\left(S[\lceil\tau/T_c\rceil]\,\mathbb{E}[X(\tau)|U[\lceil\tau/T_c\rceil]],\,\lambda_0\right)|S[\lceil\tau/T_c\rceil]=s\right] \\ &= \mathbb{E}\left[\zeta(sX(\tau),\,\lambda_0)|S[\lceil\tau/T_c\rceil]=s\right]-\zeta(s\,\mathbb{E}[X(\tau)|U[\lceil\tau/T_c\rceil]=h(s)],\,\lambda_0)\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\left[\zeta(sX(\tau),\, \lambda_0)|U[\lceil\tau/T_c\rceil] = h(s)\right] \\
&\quad -\zeta(s\,\mathbb{E}[X(\tau)|U[\lceil\tau/T_c\rceil] = h(s)],\, \lambda_0),
\end{aligned} \tag{2.44}
$$

by (2.42). Consider maximizing the right side of (2.44) over all conditional distributions of $X(\tau)$ conditioned on $U[\lceil\tau/T_c\rceil] = h(s)$ with a fixed conditional mean $\mathbb{E}[X(\tau)|U[\lceil\tau/T_c\rceil] = h(s)] = \pi_\tau(h(s))$, say, and subject to the first constraint (alone) in (2.28). Then, the right side of (2.44) equals

$$
\mathbb{E}[\zeta(sX(\tau),\, \lambda_0)|U[\lceil\tau/T_c\rceil] = h(s)] - \zeta(s\pi_\tau(h(s)),\, \lambda_0), \tag{2.45}
$$

and is maximized by considering the first term above. Using the strict convexity of $\zeta(\cdot,\, \lambda_0)$, this term is largest (see [11], proof of Theorem 1, or [43], Lemma 1)[10] iff $X(\tau)$ is a $\{0,\, A\}$-valued rv with

$$
\begin{aligned}
\Pr(X(\tau) = A|U[\lceil\tau/T_c\rceil] = h(s)) &= 1 - \Pr(X(\tau) = 0|U[\lceil\tau/T_c\rceil] = h(s)) \\
&= \pi_\tau(h(s))/A,
\end{aligned} \tag{2.46}
$$

and the corresponding largest value of (2.45) is

$$
\pi_\tau(h(s))\zeta(sA,\, \lambda_0)/A - \zeta(s\pi_\tau(h(s)),\, \lambda_0). \tag{2.47}
$$

[10]An alternative proof can be gleaned from the following simple observation. Let X be a $[0,\, A]$-valued rv, $A > 0$, of arbitrary distribution but with fixed mean $\mu = \mathbb{E}[X]$. Let $g : [0,\, A] \to \mathbb{R}_0^+$ be a strictly convex mapping, with $g(0) = 0$. Then, $g(X) \leq \frac{g(A)}{A}X$, whence $\mathbb{E}[g(X)] \leq \frac{g(A)\mu}{A}$. It is readily seen that this upper bound on $\mathbb{E}[g(X)]$ is achieved if $X \in \{0,\, A\}$ with $\Pr\{X = 0\} = 1 - \Pr\{X = A\} = 1 - \frac{\mu}{A}$. The necessity of this choice of (optimal) X follows from the strict convexity of $g(\cdot)$ on $[0,\, A]$.

Noting that (2.28) implies that

$$\begin{aligned} 0 &\le \pi_\tau(h(s)) &\le A, \quad 0 \le \tau \le T, \quad s \in \mathbb{R}_0^+, \\ \tfrac{1}{T}\int_0^T \mathbb{E}[\pi_\tau(h(S[\lceil \tau/T_c \rceil]))]d\tau &\le \sigma A, \end{aligned} \tag{2.48}$$

we thus see from (2.43), (2.44), (2.45), (2.47), (2.48) that

$$\begin{aligned} &\frac{1}{T} I\left(W \wedge \phi(N_T, \mathbf{T}^{N_T}, \mathbf{S}^K)\right) \\ &\le \max_{\substack{\pi_\tau : \mathcal{U} \to [0, A], \ 0 \le \tau \le T \\ \frac{1}{T}\int_0^T \mathbf{E}[\pi_\tau(h(S[\lceil \tau/T_c \rceil]))]d\tau \le \sigma A}} \frac{1}{T}\int_0^T \{\mathbb{E}[\pi_\tau(h(S[\lceil \tau/T_c \rceil]))\zeta(S[\lceil \tau/T_c \rceil]A, \lambda_0)/A] \\ &\qquad - \mathbb{E}[\zeta(S[\lceil \tau/T_c \rceil]\pi_\tau(h(S[\lceil \tau/T_c \rceil])), \lambda_0)]\} \, d\tau \\ &= \max_{\substack{\pi_\tau : \mathcal{U} \to [0, A], \ 0 \le \tau \le T \\ \frac{1}{T}\int_0^T \mathbf{E}[\pi_\tau(U[\lceil \tau/T_c \rceil])]d\tau \le \sigma A}} \frac{1}{T}\int_0^T \{\mathbb{E}[\pi_\tau(U[\lceil \tau/T_c \rceil])\zeta(S[\lceil \tau/T_c \rceil]A, \lambda_0)/A] \\ &\qquad - \mathbb{E}[\zeta(S[\lceil \tau/T_c \rceil]\pi_\tau(U[\lceil \tau/T_c \rceil]), \lambda_0)]\} \, d\tau. \end{aligned} \tag{2.49}$$

In order to simplify the right side of (2.49), for every $u \in \mathcal{U}$, define

$$\nu_k(u) = \frac{1}{T_c}\int_{(k-1)T_c}^{kT_c} \mathbb{E}[\pi_\tau(U[k])|U[k] = u]d\tau, \quad k = 1, \cdots, K, \tag{2.50}$$

$$\mu(u) = \frac{1}{K}\sum_{k=1}^{K} \frac{\nu_k(u)}{A}. \tag{2.51}$$

From (2.48), we get

$$0 \le \mu(u) \le 1, \quad u \in \mathcal{U}, \tag{2.52}$$

and

$$\begin{aligned} \sigma &\ge \frac{1}{A}\frac{1}{T}\int_0^T \mathbb{E}[\pi_\tau(h(S[\lceil \tau/T_c \rceil]))]d\tau \\ &= \frac{1}{A}\frac{1}{K}\sum_{k=1}^{K}\frac{1}{T_c}\int_{(k-1)T_c}^{kT_c} \mathbb{E}[\pi_\tau(U[k])]d\tau \\ &= \frac{1}{A}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[\nu_k(U[k])] \end{aligned} \tag{2.53}$$

$$= \frac{1}{A}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[\nu_k(U)] \quad (2.54)$$

$$= \mathbb{E}[\mu(U)], \quad (2.55)$$

where (2.53) is by (2.50); (2.54) holds by the i.i.d. nature of the channel fade sequence $\mathbf{S}^\infty$; and (2.55) is by (2.51). The time-averaged integral on the right side of (2.49) can be written as

$$\frac{1}{T}\int_0^T \{\mathbb{E}[\pi_\tau(U[\lceil\tau/T_c\rceil])\zeta(S[\lceil\tau/T_c\rceil]A,\, \lambda_0)/A] - \mathbb{E}[\zeta(S[\lceil\tau/T_c\rceil]\pi_\tau(U[\lceil\tau/T_c\rceil]),\, \lambda_0)]\}\, d\tau$$

$$= \frac{1}{K}\sum_{k=1}^{K}\frac{1}{T_c}\int_{(k-1)T_c}^{kT_c}\{\mathbb{E}[\pi_\tau(U[k])\zeta(S[k]A,\, \lambda_0)/A - \zeta(S[k]\pi_\tau(U[k]),\, \lambda_0)]\}\, d\tau$$

$$\leq \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[\nu_k(U[k])\zeta(S[k]A,\, \lambda_0)/A - \zeta(S[k]\nu_k(U[k]),\, \lambda_0)] \quad (2.56)$$

$$= \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[\nu_k(U)\zeta(SA,\, \lambda_0)/A - \zeta(S\nu_k(U),\, \lambda_0)] \quad (2.57)$$

$$\leq \mathbb{E}[\mu(U)\zeta(SA,\, \lambda_0) - \zeta(S\mu(U)A,\, \lambda_0)], \quad (2.58)$$

where (2.56) is by Jensen's inequality applied to the convex function $\zeta(\cdot,\, \lambda_0)$ and (2.50); (2.57) holds by the i.i.d. nature of $\mathbf{S}^\infty$; and (2.58) is by Jensen's inequality and (2.51). Summarizing collectively (2.31), (2.49), (2.52), (2.55), (2.58), we get that

$$R \lessapprox \max_{\substack{\mu:\mathcal{U}\to[0,\,1] \\ \mathbf{E}[\mu(U)]\leq\sigma}} \mathbb{E}\left[\mu(U)\zeta(SA,\, \lambda_0) - \zeta(\mu(U)SA,\, \lambda_0)\right]. \quad (2.59)$$

This concludes the proof of the converse part of Theorem 1.

Achievability part: We closely follow Wyner's approach [52].

Fix $L \in \mathbb{Z}^+$ and set $\Delta \triangleq T_c/L$. Divide the time interval $[0,\, T]$, where $T = KT_c$ with $K \in \mathbb{Z}^+$, into KL equal subintervals, each of duration Δ. See Figure 2.3. Then, in the channel fade sequence $\mathbf{S}^\infty = \{S[k]\}_{k=1}^\infty$, each $S[k]$ remains unvarying for a

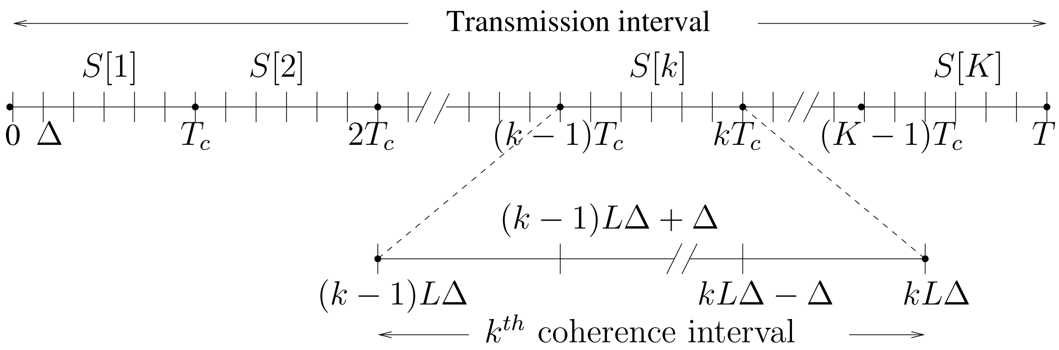

Figure 2.3: Subintervals of $[0, T]$.

block of L consecutive Δ-duration subintervals within $[(k-1)T_c, kT_c]$, and $\{S[k]\}_{k=1}^{\infty}$ varies across such blocks in an i.i.d. manner.

Now, consider the situation in which for each message $w \in \mathcal{W}$, the channel input waveform $f(w, \mathbf{u}^K) = \{x_w(t, \mathbf{u}^{\lceil t/T_c \rceil}), 0 \le t \le T\}$ is restricted to be $\{0, A\}$-valued and piecewise constant in each of the KL time slots of duration Δ. Define

$$\tilde{x}_n(w, \mathbf{u}^{\lceil n/L \rceil}) \triangleq \begin{cases} 0, & \text{if } x_w(t, \mathbf{u}^{\lceil t/T_c \rceil}) = 0, \\ 1, & \text{if } x_w(t, \mathbf{u}^{\lceil t/T_c \rceil}) = A, \\ & t \in [(n-1)\Delta, n\Delta), \quad n = 1, \cdots, KL. \end{cases} \tag{2.60}$$

Note that the condition (2.2) requires that

$$\frac{1}{KL} \sum_{n=1}^{KL} \tilde{x}_n(w, \mathbf{u}^{\lceil n/L \rceil}) \le \sigma, \quad w \in \mathcal{W}, \quad \mathbf{u}^K \in \mathcal{U}^K. \tag{2.61}$$

Next, consider a decoder $\phi : \{0, 1\}^{KL} \times (\mathbb{R}_0^+)^K \to \mathcal{W}$ based on restricted observations over the KL time slots, comprising

$$\tilde{Y}_n = 1(Y(n\Delta) - Y((n-1)\Delta) = 1), \quad n = 1, \cdots, KL, \tag{2.62}$$

(with $Y(0) = 0$), and $\mathbf{S}^K$. The largest achievable rate of restricted (W, T)-codes as above – and, hence, the capacity C for transmitter CSI h – is clearly no smaller

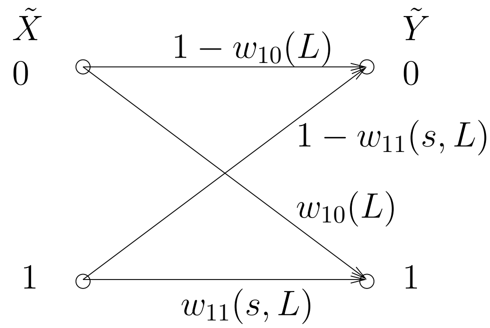

Figure 2.4: Discrete channel approximation.

than $\frac{C(L)}{T_c}$, where $C(L)$ is the capacity of a (L-) block discrete memoryless channel (in nats per block channel use) with input alphabet $\tilde{\mathcal{X}}^L = \{0,\, 1\}^L$; output alphabet $\tilde{\mathcal{Y}}^L = \{0,\, 1\}^L$; state alphabet $\mathbb{R}_0^+$; transition probability mass function (pmf)

$$\begin{aligned} W^{(L)}(\tilde{\mathbf{y}}^L|\tilde{\mathbf{x}}^L,\, s) &= \prod_{l=1}^{L} W_{\tilde{Y}|\tilde{X},S}(\tilde{y}_l|\tilde{x}_l,\, s), \\ &\qquad \tilde{\mathbf{x}}^L \in \{0,\, 1\}^L, \quad \tilde{\mathbf{y}}^L \in \{0,\, 1\}^L, \quad s \in \mathbb{R}_0^+, \end{aligned} \tag{2.63}$$

where $\tilde{X}$, S, $\tilde{Y}$, respectively, are $\tilde{\mathcal{X}}$-, $\mathbb{R}_0^+$- and $\tilde{\mathcal{Y}}$-valued rvs and $W_{\tilde{Y}|\tilde{X},S}(\cdot|\cdot,\, s)$, $s \in \mathbb{R}_0^+$, is given by (see Figure 2.4)

$$\begin{aligned} W_{\tilde{Y}|\tilde{X},S}(1|0,\, s) &= 1 - W_{\tilde{Y}|\tilde{X},S}(0|0,\, s) \\ &= \lambda_0 \tfrac{T_c}{L} \exp\left(-\lambda_0 \tfrac{T_c}{L}\right) \;\triangleq\; w_{10}(L), \\ W_{\tilde{Y}|\tilde{X},S}(1|1,\, s) &= 1 - W_{\tilde{Y}|\tilde{X},S}(0|1,\, s) \\ &= (\lambda_0 + sA)\tfrac{T_c}{L} \exp\left(-(\lambda_0 + sA)\tfrac{T_c}{L}\right) \;\triangleq\; w_{11}(s,\, L), \end{aligned} \tag{2.64}$$

with transmitter CSI h (and perfect receiver CSI); and under constraint (2.61). From ([9], Remark A2 following Proposition 1), it is readily obtained[11] that

$$C(L) = \max_{P_{\tilde{\mathbf{X}}^L|U}: \sum_{l=1}^{L} \mathbf{E}[\tilde{X}_l] \le L\sigma} I(\tilde{\mathbf{X}}^L \wedge \tilde{\mathbf{Y}}^L|S), \tag{2.65}$$

[11] In [9], S is taken to be finite-valued; however the result (2.65) can be seen to hold also when S is $\mathbb{R}_0^+$-valued.

with $U = h(S)$, where the joint conditional pmf $P_{\tilde{\mathbf{X}}^L,\,\tilde{\mathbf{Y}}^L|S}$ is given by

$$
\begin{aligned}
P_{\tilde{\mathbf{X}}^L,\,\tilde{\mathbf{Y}}^L|S}(\tilde{\mathbf{x}}^L,\,\tilde{\mathbf{y}}^L|s) &= P_{\tilde{\mathbf{X}}^L|U}(\tilde{\mathbf{x}}^L|h(s))W^{(L)}(\tilde{\mathbf{y}}^L|\tilde{\mathbf{x}}^L,\,s), \\
&\qquad \tilde{\mathbf{x}}^L \in \{0,\,1\}^L, \quad \tilde{\mathbf{y}}^L \in \{0,\,1\}^L, \quad s \in \mathbb{R}_0^+. \qquad (2.66)
\end{aligned}
$$

By a standard argument which uses (2.63), the maximum in (2.65) is achieved by

$$
P_{\tilde{\mathbf{X}}^L|U}(\tilde{\mathbf{x}}^L|h(s)) = \prod_{l=1}^{L} P_{\tilde{X}_l|U}(\tilde{x}_l|h(s)), \quad \tilde{\mathbf{x}}^L \in \{0,\,1\}^L, \quad s \in \mathbb{R}_0^+, \qquad (2.67)
$$

so that from (2.65),

$$
C(L) = L \max_{P_{\tilde{X}|U}:\, \mathbf{E}[\tilde{X}] \le \sigma} I(\tilde{X} \wedge \tilde{Y}|S), \qquad (2.68)
$$

where $P_{\tilde{X},\,\tilde{Y}|S}$ is given by

$$
\begin{aligned}
P_{\tilde{X},\,\tilde{Y}|S}(\tilde{x},\,\tilde{y}|s) &= P_{\tilde{X}|U}(\tilde{x}|h(s))W_{\tilde{Y}|\tilde{X},\,S}(\tilde{y}|\tilde{x},\,s), \\
&\qquad \tilde{x} \in \{0,\,1\}, \quad \tilde{y} \in \{0,\,1\}, \quad s \in \mathbb{R}_0^+. \qquad (2.69)
\end{aligned}
$$

Setting

$$
\mu(u) \triangleq \Pr\{\tilde{X} = 1|U = u\} = \mathbb{E}[\tilde{X}|U = u], \quad u \in \mathcal{U}, \qquad (2.70)
$$

and

$$
\begin{aligned}
\beta_L(s) &\triangleq I(\tilde{X} \wedge \tilde{Y}|S = s) \\
&= H(\tilde{Y}|S = s) - H(\tilde{Y}|\tilde{X},\,S = s) \\
&= h_b\left(\mu(h(s))w_{11}(s,\,L) + (1 - \mu(h(s)))w_{10}(L)\right) \\
&\quad - \left(\mu(h(s))h_b(w_{11}(s,\,L)) + (1 - \mu(h(s)))h_b(w_{10}(L))\right), \quad s \in \mathbb{R}_0^+, \qquad (2.71)
\end{aligned}
$$

where $h_b(\cdot)$ is the binary entropy function, we can express (2.68) as

$$
C(L) = \max_{\substack{\mu:\mathcal{U}\to[0,\,1] \\ \mathbf{E}[\mu(U)] \le \sigma}} L\, \mathbb{E}[\beta_L(S)]. \qquad (2.72)
$$

Since $L \in \mathbb{Z}^+$ was arbitrary, we have

$$\begin{aligned} C &\geq \lim_{L\to\infty} \frac{C(L)}{T_c} \\ &\geq \max_{\substack{\mu:\mathcal{U}\to[0,\,1]\\ \mathbf{E}[\mu(U)]\leq\sigma}} \lim_{L\to\infty} \frac{\mathbb{E}[\beta_L(S)]}{T_c/L} \end{aligned} \tag{2.73}$$

by (2.72). Finally, it is shown in Appendix A.3 that

$$\lim_{L\to\infty} \frac{\mathbb{E}[\beta_L(S)]}{T_c/L} = \mathbb{E}[\mu(U)\zeta(SA,\,\lambda_0) - \zeta(S\mu(U)A,\,\lambda_0)], \tag{2.74}$$

whence

$$C \geq \max_{\substack{\mu:\mathcal{U}\to[0,\,1]\\ \mathbf{E}[\mu(U)]\leq\sigma}} \mathbb{E}[\mu(U)\zeta(SA,\,\lambda_0) - \zeta(S\mu(U)A,\,\lambda_0)], \tag{2.75}$$

thereby completing the proof of the achievability part. □

Remarks: (i) In Section 2.3, remark (iii) following Theorem 1 constitutes an interpretation of (2.67) when $L \to \infty$.

(ii) In the proof above of the achievability part of Theorem 1, we could also have considered a restricted decoder ϕ with $\tilde{Y}_n$ in (2.62) replaced by

$$\tilde{Y}_n = 1 - 1(Y(n\Delta) - Y((n-1)\Delta) = 0), \quad n = 1, \cdots, KL.$$

Proof of Theorem 2:

From 2.6,

$$C = \max_{\substack{\mu:\mathcal{U}\to[0,\,1]\\ \mathbf{E}[\mu(U)]\leq\sigma}} \mathbb{E}[\psi(U,\,\mu(U))], \tag{2.76}$$

where

$$\psi(u,\,\mu(u)) \triangleq \mathbb{E}[\mu(u)\zeta(SA,\,\lambda_0) - \zeta(S\mu(u)A,\,\lambda_0)|U = u], \quad u \in \mathcal{U}. \tag{2.77}$$

Note that

$$
\begin{aligned}
\frac{\partial\psi}{\partial\mu} &= \mathbb{E}\left[\zeta(SA,\, \lambda_0) - SA(1+\log(S\mu(u)A+\lambda_0)) \middle| U=u\right] \\
&= \mathbb{E}\left[SA\log\frac{(1+\alpha(SA/\lambda_0)SA/\lambda_0)}{(1+\mu(u)SA/\lambda_0)}\middle| U=u\right], \qquad (2.78) \\
\frac{\partial^2\psi}{\partial\mu^2} &= -\,\mathbb{E}\left[\frac{S^2A^2}{(S\mu(u)A+\lambda_0)}\middle| U=u\right], \quad u\in\mathcal{U}.
\end{aligned}
$$

In order to determine the optimal power control law $\mu^* : \mathcal{U} \to [0,\, 1]$ we use variational calculus.

First consider the "unconstrained optimization" problem, i.e., without the constraints $\mu : \mathcal{U} \to [0,\, 1]$ and $\mathbb{E}[\mu(U)] \le \sigma$:

$$
\max_{\mu:\mathcal{U}\to\mathbb{R}} \mathbb{E}[\psi(U,\, \mu(U))], \qquad (2.79)
$$

and let $\mu_0 : \mathcal{U} \to \mathbb{R}$ denote the maximizer. Then μ_0 must satisfy the (necessary) Euler-Lagrange condition (cf. e.g., [15])

$$
\left.\frac{\partial\psi}{\partial\mu}\right|_{\mu_0} = 0, \qquad (2.80)
$$

which, by (2.78) is

$$
\mathbb{E}\left[SA\log\frac{(1+\alpha(SA/\lambda_0)SA/\lambda_0)}{(1+\mu_0(u)SA/\lambda_0)}\middle| U=u\right] = 0, \quad u\in\mathcal{U}. \qquad (2.81)
$$

Furthermore, by (2.78), $\frac{\partial^2\psi}{\partial\mu^2} < 0$, $u \in \mathcal{U}$, i.e., $\psi(\cdot)$ is a strictly concave function of μ, so that (2.80) also constitutes a sufficient condition. It can be verified (cf. e.g., [11], Figure 2) that $\alpha(\cdot)$ is monotone decreasing on $[0,\, \infty)$ with $\alpha(0) = \frac{1}{2}$ and $\alpha(\infty) = \frac{1}{e}$. Therefore, from (2.81), it follows that $\frac{1}{e} \le \mu_0(u) \le \frac{1}{2}$, $u \in \mathcal{U}$. With $\sigma_0 \triangleq \mathbb{E}[\mu_0(U)]$, we see that $\frac{1}{e} \le \sigma_0 \le \frac{1}{2}$.

Consider next the constrained optimization problem on the right side of (2.76), i.e., now with the inclusion of the constraints $\mu : \mathcal{U} \to [0,\, 1]$ and $\mathbb{E}[\mu(U)] \le \sigma$. If $\sigma \ge$

σ_0, then the "unconstrained" maximizer μ_0 satisfies the previous two constraints, and hence is the solution of the constrained problem in (2.76) as well. Suppose next that $\sigma < \sigma_0$. First ignore the (local) constraint $0 \le \mu(u) \le 1$, $u \in \mathcal{U}$, and define the Lagrangian functional

$$L(\mu) \quad = \quad \mathbb{E}[\xi(U,\, \mu(U))], \tag{2.82}$$

where

$$\xi(u,\, \mu(u)) \quad \triangleq \quad \psi(u,\, \mu(u)) - \rho\mu(u), \quad u \in \mathcal{U}, \tag{2.83}$$

and $\rho \ge 0$ is a Lagrange multiplier. Using the strict concavity of $L(\cdot)$, we conclude that a necessary and sufficient condition for $\mu_\rho : \mathcal{U} \to \mathbb{R}$ to be the maximizer in (2.76) is given by the Euler-Lagrange equation

$$\left.\frac{\partial \xi}{\partial \mu}\right|_{\mu_\rho} \quad = \quad 0. \tag{2.84}$$

By (2.78), (2.83) and (2.84), we then see that μ_ρ satisfies

$$\mathbb{E}\left[SA \log \frac{(1 + \alpha(SA/\lambda_0)SA/\lambda_0)}{(1 + \mu_\rho(u)SA/\lambda_0)} \right| U = u \bigg] \quad = \quad \rho, \quad u \in \mathcal{U}. \tag{2.85}$$

We now impose the constraint $0 \le \mu(u) \le 1$, $u \in \mathcal{U}$. Note that $\mu_\rho(u) \le \mu_0(u) \le \frac{1}{2}$, $u \in \mathcal{U}$ for all $\rho \ge 0$. However, given $\rho > 0$, $\mu_\rho(u)$ can be < 0 for some $u \in \mathcal{U}$. By the strict concavity of $\xi(u,\, \cdot)$, it follows that if for some $u \in \mathcal{U}$, $\mu_\rho(u) < 0$, then for all $\omega \ge 0$, $\xi(u,\, \cdot)$ is a strictly decreasing function of ω. Therefore, if $\mu_\rho(u) < 0$ for some $u \in \mathcal{U}$, the constraint $\mu(u) \ge 0$ dictates the maximizing solution to be $\mu^*(u) = 0$; otherwise, the maximizing solution is given by $\mu^*(u) = \mu_\rho(u)$. Summarizing the previous observations, we get that $\mu^*(u) = [\mu_\rho(u)]^+$, $u \in \mathcal{U}$. Finally, the optimal

Lagrange multiplier ρ^* is chosen to satisfy the power constraint $\mathbb{E}[[\mu_{\rho^*}(U)]^+] = \sigma$. This concludes the proof of Theorem 2. $\square$

Proof of Theorem 3:

First consider the limit as $\lambda_0 \to 0$. Let $C^H = \lim_{\lambda_0 \to 0} C$, $\mu_\rho^H(u) = \lim_{\lambda_0 \to 0} \mu_\rho(u)$ and $\mu^H(u) = \lim_{\lambda_0 \to 0} \mu^*(u)$, $u \in \mathcal{U}$, $\rho \geq 0$, where C, $\mu_\rho(\cdot)$ and $\mu^*(\cdot)$ are as defined in (2.6), (2.9) and (2.11) respectively. Since

$$\begin{aligned} &\lim_{\lambda_0 \to 0} \mathbb{E}\left[SA \log \frac{(1 + \alpha(SA/\lambda_0)SA/\lambda_0)}{(1 + \mu SA/\lambda_0)} \middle| U = u \right] \\ &= \mathbb{E}[SA \log(\alpha(0)/\mu) | U = u], \quad 0 \leq \mu \leq 1, \end{aligned}$$

by (2.9), it follows that $\mu_\rho^H(u) = e^{-1-\rho/A\,\mathbf{E}[S|U=u]}$. Clearly, $0 \leq \mu_\rho^H(u) \leq e^{-1}$, $\rho \geq 0$, $u \in \mathcal{U}$. Furthermore, $\mu_0^H(u) = e^{-1}$, $u \in \mathcal{U}$, so that $\sigma_0 = e^{-1}$. By (2.11), it follows that

$$\mu^H(u) = \begin{cases} e^{-1}, & \sigma \geq e^{-1}, \\ e^{-1-\rho^*/A\,\mathbf{E}[S|U=u]}, & \sigma < e^{-1}, \end{cases}$$

where for $\sigma < e^{-1}$, $\rho^* \geq 0$ satisfies $\mathbb{E}[e^{-1-\rho^*/A\,\mathbf{E}[S|U]}] = \sigma$. Finally, by (2.6), in the limit as $\lambda_0 \to 0$, we get

$$C^H = \mathbb{E}\left[\mu^H(U)\zeta(SA, 0) - \zeta(\mu^H(U)SA, 0)\right].$$

This concludes the proof of the first part of Theorem 3.

Next consider the case $\lambda_0 \gg 1$. Let $\mu_\rho^L(u) = \lim_{\lambda_0 \to \infty} \mu_\rho(u)$ and $\mu^L(u) = \lim_{\lambda_0 \to \infty} \mu^*(u)$, $u \in \mathcal{U}$, $\rho \geq 0$. Given $s \geq 0$, $a \geq 0$,

$$\begin{aligned} &\zeta(sA, \lambda_0) - sA(1 + \log(\lambda_0 + asA)) \\ &= sA \log \lambda_0 + (sA + \lambda_0) \log(1 + sA/\lambda_0) - sA\left(1 + \log \lambda_0 + \log(1 + asA/\lambda_0)\right), \end{aligned}$$

which, for $\lambda_0 \gg 1$, with the approximation $\log x = x - x^2/2 + O(x^3)$ for $x \ll 1$, equals

$$\begin{aligned}
&(sA+\lambda_0)\left(sA/\lambda_0 - (sA/\lambda_0)^2/2 + O(\lambda_0^{-3})\right) \\
&\qquad -sA\left(1 + asA/\lambda_0 - (asA/\lambda_0)^2/2 + O(\lambda_0^{-3})\right) \\
&= (1-2a)s^2A^2/2\lambda_0 + O(\lambda_0^{-2}) \quad \to \quad 0 \quad \text{as } \lambda_0 \to \infty, \qquad (2.86)
\end{aligned}$$

Hence, for $\lambda_0 \gg 1$, by (2.9), (2.78) and (2.86), it follows that

$$\begin{aligned}
\rho &= \mathbb{E}[\zeta(SA,\, \lambda_0) - SA(1 + \log(\lambda_0 + \mu^L_\rho(u)SA))|U=u] \\
&= (1 - 2\mu^L_\rho(u))\, \mathbb{E}[S^2|U=u]A^2/2\lambda_0 + O(\lambda_0^{-2}),
\end{aligned}$$

so that as $\lambda_0 \to \infty$, $\rho \to 0$ and $\mu^L_\rho(u) = 1/2 + O(\lambda_0^{-1}) \to 1/2$, $u \in \mathcal{U}$. In this case $\sigma_0 = 1/2$, and therefore $\mu^L(u) = \min\{\sigma, 1/2\}$, $u \in \mathcal{U}$, a constant which we denote by μ^L.

Finally, we compute C^L. Given $s \geq 0$, for $\lambda_0 \gg 1$, we have

$$\begin{aligned}
&\mu^L\zeta(sA,\, \lambda_0) - \zeta(\mu^L sA,\, \lambda_0) \\
&\quad = \mu^L(sA+\lambda_0)\log(1 + sA/\lambda_0) - (\mu^L sA + \lambda_0)\log(1 + \mu^L sA/\lambda_0) \\
&\quad = \mu^L(sA+\lambda_0)\left(sA/\lambda_0 - (sA/\lambda_0)^2/2 + O(\lambda_0^{-3})\right) \\
&\qquad -(\mu^L sA + \lambda_0)\left(\mu^L sA/\lambda_0 - (\mu^L sA/\lambda_0)^2/2 + O(\lambda_0^{-3})\right) \qquad (2.87) \\
&\quad = \mu^L(1-\mu^L)s^2A^2/2\lambda_0 + O\left(\lambda_0^{-2}\right),
\end{aligned}$$

where we have used the approximation $\log x = x - \frac{x^2}{2} + O(x^3)$ for $x \ll 1$ in (2.87). By (2.6), we then obtain

$$C^L = \mu^L(1-\mu^L)\, \mathbb{E}[S^2]A^2/2\lambda_0 + O(\lambda_0^{-2}).$$

This completes the proof of Theorem 3. $\square$

2.5 Numerical example

In this example, we consider the fade to have a lognormal distribution. We determine the capacity for the two special cases discussed in Section 2.3, and compare the values with the capacity as determined by the results of [21] as well as for the channel without fading.

The channel fade is an i.i.d. lognormal process, i.e., $S_k \sim S = \exp(2G)$, $k = 1, 2, \cdots$, where G is Gaussian with mean μ_G and variance σ_G^2. We choose $\mu_G = -\sigma_G^2$ so that the fade is normalized, i.e., $\mathbb{E}[S] = 1$. This implies that the atmosphere, on an average, does not attenuate or amplify the transmitted signal. The log-amplitude variance σ_G^2 can vary from 0 (negligible fading) to 0.5 (severe turbulence) [21]. We pick $\sigma_G^2 = 0.1$, which corresponds to a moderately turbulent Poisson channel. Denote the (peak) signal-to-noise ratio by $\text{SNR} = 10\log_{10}(A/\lambda_0)$ (in dB). We fix $A = 1$ and vary the parameters σ and λ_0 to study the effect of the average power constraint and SNR on the channel capacity.

Figure 2.5 shows the behavior of the optimal power control law $\mu_P^*(\cdot)$ (see (2.13)) with the channel fade for different values of σ for SNR = 0 dB. The power law in ([21], eq. 4) is also plotted for comparison. Note that the average power constraint becomes ineffective if $\sigma > \sigma_0$. If $\sigma < \sigma_0$, the optimal power control law dictates that the transmitter should not transmit when the channel fade is very bad (i.e., for small values of s). This behavior is similar to the *waterfilling* power law in

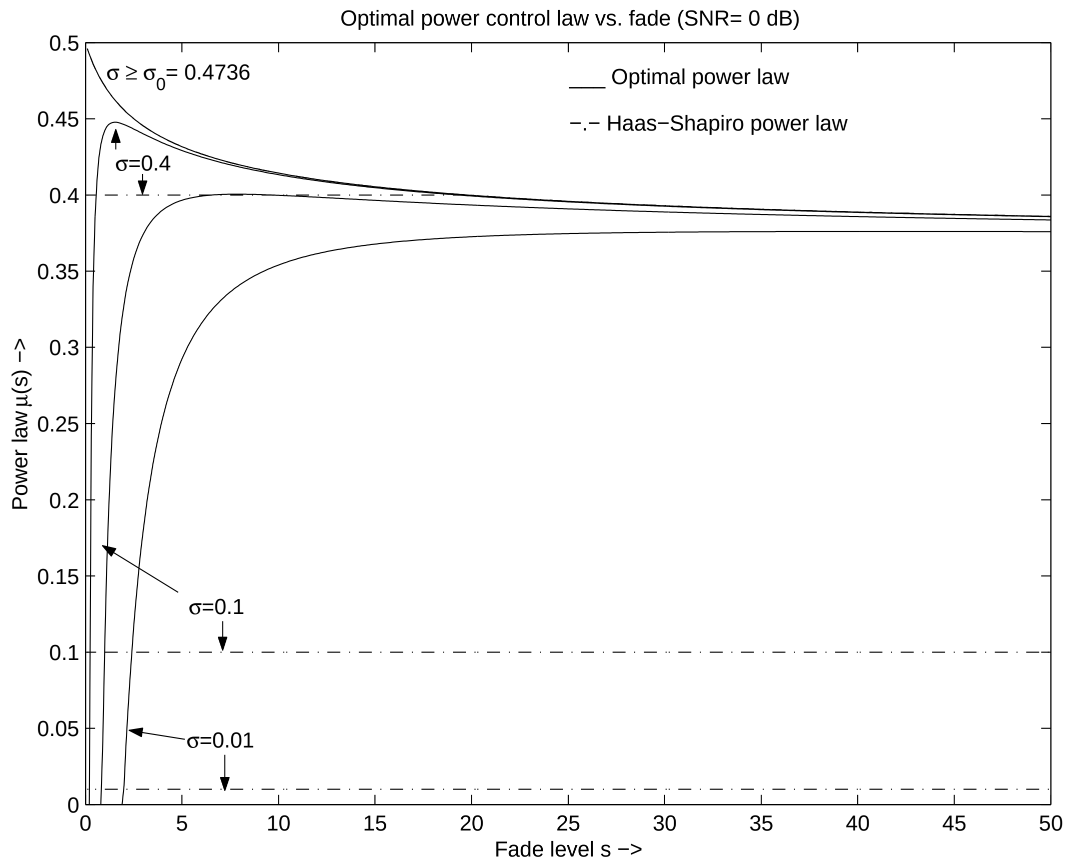

Figure 2.5: Optimal power control law with perfect CSI at transmitter and receiver.

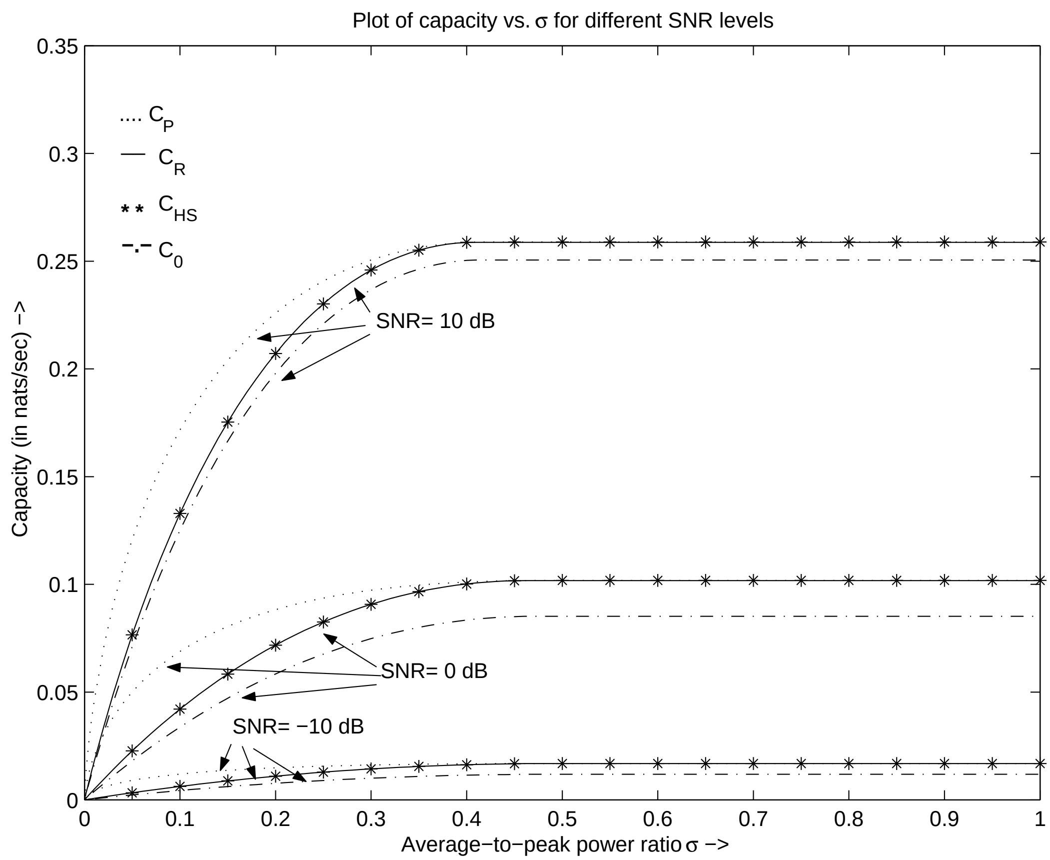

Figure 2.6: Comparison of capacity versus σ for various assumptions on transmitter CSI.

RF Gaussian fading channels (cf. e.g., [40]). On the other hand, the power control law due to ([21], eq. 4) is constant over a wide range of values of σ, and does not properly exploit the channel fade.

In Figures 2.6 and 2.7, we compare the capacity values obtained under different assumptions. In particular, we plot the capacity for perfect transmitter and receiver CSI (C_P), the capacity for no transmitter CSI (C_R), the capacity obtained from [21] (C_{HS}) and the capacity of the Poisson channel without fading (C_0) (cf. e.g. [52], eq. 1.5) for various values of σ and SNR. From the figures, we conclude

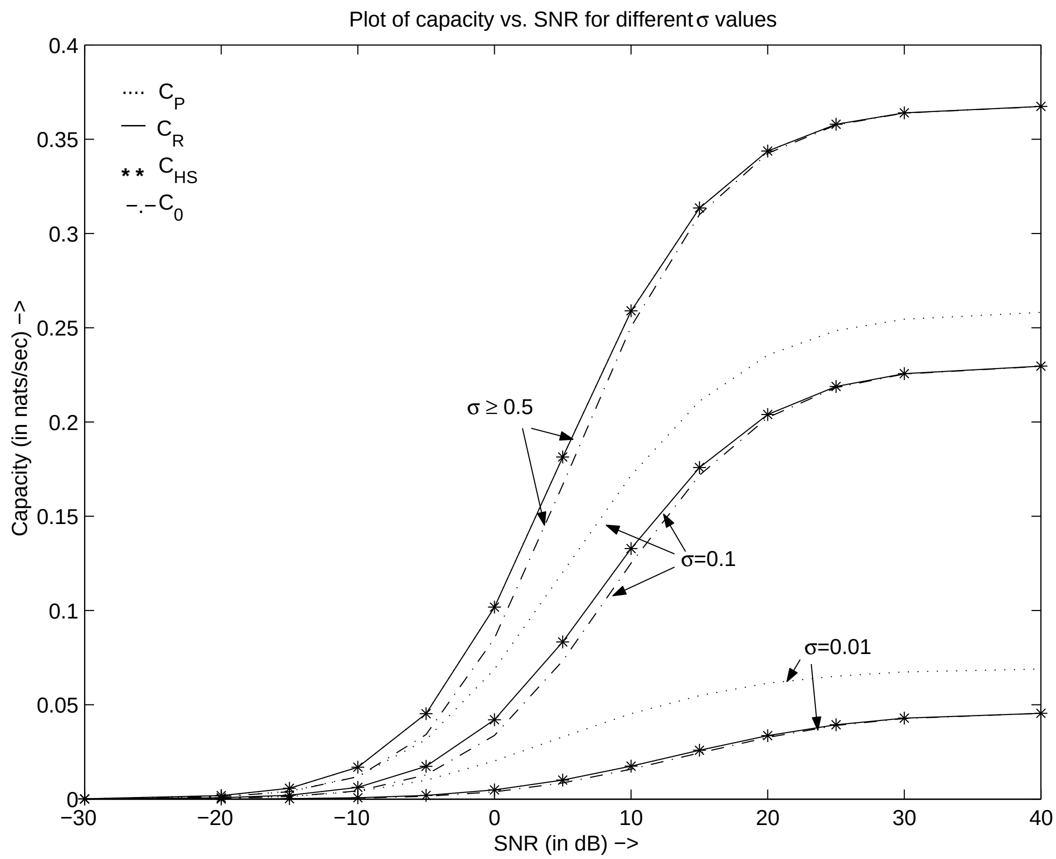

Figure 2.7: Comparison of capacity versus SNR for various assumptions on transmitter CSI.

that the knowledge of CSI at the transmitter can increase the channel capacity. The improvement is greater for small σ values, i.e., when the average power constraint is severe. Furthermore, knowledge of CSI at the transmitter can greatly improve capacity at high SNR when σ is small. It is clear that C_{HS}, the capacity computed from [21], is very close to C_R, the capacity when no CSI is available at the transmitter. This can be understood from the power law in Figure 2.5, which does not depend on the fade for a large range of values of σ.

2.6 Discussion

We have studied the capacity problem for a shot-noise limited direct detection block-fading Poisson channel. A novel channel model for the free-space Poisson fading channel has been proposed in which the channel fade remains unvarying in intervals of duration T_c, and vary across successive such intervals in an i.i.d. fashion. Under the assumptions of perfect CSI at the receiver while the transmitter CSI can be imperfect, a single-letter characterization of the capacity has been obtained when the transmitted signal is constrained in its peak and average power levels.

Binary signaling with arbitrarily fast intertransition times is shown to be optimal for this channel. The two signaling levels correspond to no transmission ("OFF" state) and transmission at the peak power level ("ON" state). Furthermore, with perfect CSI at the receiver, the channel capacity does not depend on the channel coherence time T_c. The i.i.d. nature of channel fade allows us to focus on a single coherence interval, while inside a single interval, the optimality of i.i.d. input signaling (as a function of current transmitter CSI) leads to a lack of dependence of the capacity formula on T_c. An exact characterization of the "optimal power control law," which represents the conditional probability of the transmit aperture remaining in the ON state as a function of current transmitter CSI, and can be viewed as the optimal average conditional duty cycle of the transmitted signal, has been obtained.

We have analyzed the effects of varying degrees of CSI at the transmitter on channel capacity as a function of the peak signal-to-noise power ratio (SNR). In the high SNR regime, a knowledge of varying degrees of CSI at the transmitter can lead to a significant gain in capacity when the average power constraint is stringent. On the other hand, in the low SNR regime, a knowledge of CSI at the transmitter does not provide any additional advantage.

Chapter 3

MIMO Poisson channel with constant channel fade

3.1 Introduction

In this chapter, we focus on the shot-noise limited, single-user MIMO Poisson channel with N transmit and M receive apertures. We refer to this channel as the $N \times M$ MIMO Poisson channel. We assume, for the time being, that the channel fade remains unvarying for the duration of transmission and reception, with the fade coefficients being known to the transmitter and the receiver. We consider two variants of transmitter power constraints: (a) peak and average power constraints on individual transmit apertures; and (b) peak power constraints on individual transmit apertures and a constraint on the sum of the average powers from all transmit apertures. In this setting, we provide a "single-letter characterization" of the channel capacity and outline key properties of optimal transmission strategies. The results of this chapter provide useful insights into the capacity problem of the MIMO Poisson fading channel with a random fade, which is addressed in the next chapter. Several properties of optimal transmission strategies are derived for the MIMO Poisson channel with constant fade, many of which are relevant even when the channel fade is random. For the MIMO channel with constant fade and a sum average power constraint, a key property of optimum transmission strategy is identified, which helps us to derive a simplified expression for channel capacity of a

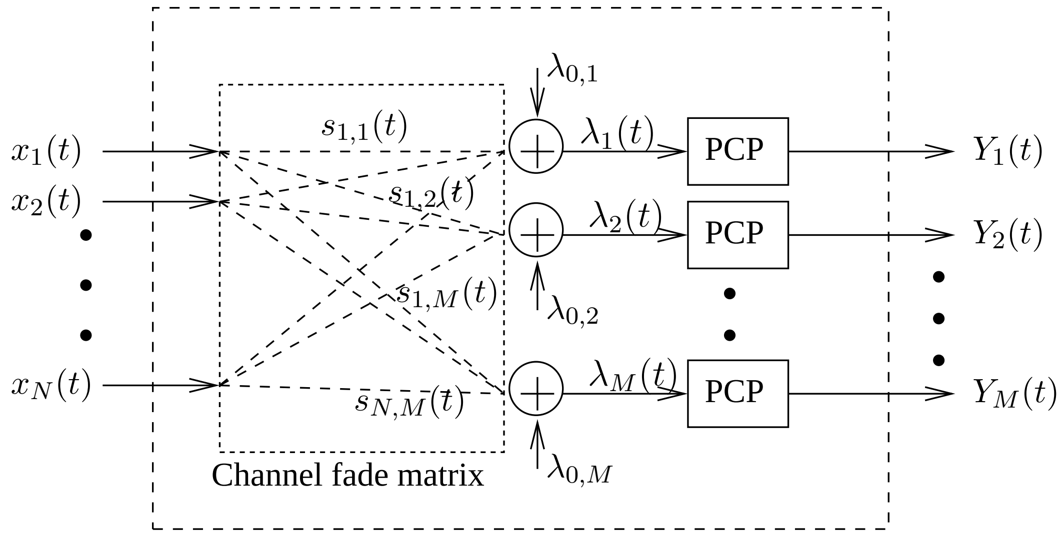

Figure 3.1: $N \times M$ MIMO Poisson channel with deterministic channel fade.

symmetric MIMO Poisson channel with isotropically distributed channel fade.

The remainder of the chapter is organized as follows. In Section 3.2, we introduce the channel model and state the problem formulation. In Section 3.3, we state our results, which are proved in Section 3.4. A few illustrative examples are discussed in Section 3.5. We close the chapter with some concluding remarks in Section 3.6.

3.2 Problem formulation

A block schematic diagram of the channel model is given in Figure 3.1. For a given set of N $\mathbb{R}_0^+$–valued transmitted signals $\{x_n(t), t \geq 0\}_{n=1}^N$, the received signal $\{Y_m(t), t \geq 0\}$ at the m^{th} receive aperture is a $\mathbb{Z}_0^+$–valued nondecreasing, (left-continuous) Poisson counting process (PCP) with rate (or intensity) equal to

$$\lambda_m(t) \;\; = \;\; \sum_{n=1}^{N} s_{n,m} x_n(t) + \lambda_{0,m}, \quad t \geq 0, \tag{3.1}$$

where $s_{n,m} \geq 0$ is the (deterministic) fade (or path gain) from the n^{th} transmit aperture to the m^{th} receive aperture, $n = 1, \cdots, N$, $m = 1, \cdots, M$, and $\lambda_{0,m} \geq 0$ is the background noise (dark current) rate at the m^{th} receive aperture, $m = 1, \cdots, M$. We assume that the channel fade coefficients $\{s_{n,m}\}_{n=1, m=1}^{N, M}$ and the dark current rates $\{\lambda_{0,m}\}_{m=1}^{M}$ remain unvarying throughout the entire duration of transmission and reception, and are known to both the transmitter and the receiver. We also assume that the receive apertures are sufficiently separated in space, so that given the knowledge of $\mathbf{x}^{\infty} = \{x_n(t),\, t \geq 0\}_{n=1}^{N}$, the processes $Y_1^{\infty}, \cdots, Y_M^{\infty}$ are conditionally mutually independent [47, 19].

The input to the channel is a set of N $\mathbb{R}_0^+$–valued transmitted signals, one corresponding to each transmit aperture, collectively denoted by $\mathbf{x}^T = \{x_n(t),\, 0 \leq t \leq T\}_{n=1}^{N}$, each of which is proportional to the transmitted optical power from the respective transmit aperture, and which satisfy peak and average power constraints of the form:

$$\begin{aligned} 0 \quad \leq \quad x_n(t) \quad &\leq \quad A_n, \quad 0 \leq t \leq T, \\ \tfrac{1}{T} \int_0^T x_n(t) dt \quad &\leq \quad \sigma_n A_n, \end{aligned} \tag{3.2}$$

where the peak powers $A_n > 0$ and the average-to-peak power ratios σ_n, $0 \leq \sigma_n \leq 1$, $n = 1, \cdots, N$ are fixed.

Definition 2 *We say that the transmitted signals are subject to a* **symmetric** *average (resp. peak) power constraint σ (resp. A) if $\sigma_n = \sigma$ (resp. $A_n = A$) for all $n = 1, \cdots, N$; else the average (resp. peak) power constraints are* **asymmetric**.

In practice, the transmit apertures may belong to a laser array with similar specifications, or they may be a collection of a diverse group of devices (lasers or LEDs) with different operating characteristics. A symmetric power constraint is a reasonable assumption for the former case, while asymmetric constraints are more appropriate for the latter. In [19], the authors addressed the special case of MIMO Poisson channel with a symmetric average power constraint. Here we also allow for asymmetric average power constraints, with significantly different consequences.

We consider another variant of the average transmit power constraint, viz., a constraint on the *sum* of the average powers of the transmitted signals from all the transmit apertures. We say that the transmitted signals are subject to peak power constraints A_n, $n = 1, \cdots, N$, and an average sum power constraint σ if

$$\begin{aligned} 0 \quad \leq \quad x_n(t) \quad &\leq \quad A_n, \quad n = 1, \cdots, N, \\ \tfrac{1}{T}\int_0^T \sum_{n=1}^N x_n(t)dt \quad &\leq \quad \sigma \sum_{n=1}^N A_n, \end{aligned} \tag{3.3}$$

where $0 \leq \sigma \leq 1$ is fixed. We shall see later in Chapter 4 that the average sum power constraint plays an important role in the characterization of channel capacity of the MIMO Poisson fading channel with random channel fade.

For the channel under consideration, a (W, T)–code (f, ϕ) is defined as follows.

1. The codebook comprises a set of W waveform vectors $f(w) = \{x_n(w, t), 0 \leq t \leq T\}_{n=1}^N$, $w \in \mathcal{W} = \{1, \cdots, W\}$, satisfying the following peak and average power constraints which follow from (3.2):

$$\begin{aligned} 0 \quad \leq \quad x_n(w, t) \quad &\leq \quad A_n, \quad 0 \leq t \leq T, \quad w \in \mathcal{W}; \\ \tfrac{1}{T}\int_0^T x_n(w, t)dt \quad &\leq \quad \sigma_n A_n, \quad w \in \mathcal{W}. \end{aligned} \tag{3.4}$$

   If the transmitted signals are subject to individual peak power constraints A_n, $n = 1, \cdots, N$, and an average sum power constraint σ, the waveform vectors $f(w)$, $w \in \mathcal{W}$ satisfy the following peak and average power constraints which follow from (3.3):

$$\begin{array}{rcll} 0 & \leq & x_n(w,\, t) \quad \leq \quad A_n, & 0 \leq t \leq T, \quad w \in \mathcal{W}; \\ \frac{1}{T}\int_0^T \sum_{n=1}^N x_n(w,\, t)dt & \leq & \sigma \sum_{n=1}^N A_n, & w \in \mathcal{W}. \end{array} \tag{3.5}$$

2. The decoder is a mapping $\phi : (\Sigma(T))^M \rightarrow \mathcal{W}$, where $\Sigma(T)$ has been defined earlier in Chapter 2, Section 2.

For each message $w \in \mathcal{W}$, the transmitter sends N waveforms, one from each transmit aperture; the waveform $x_n^T(w) = \{x_n(w,\, t),\, 0 \leq t \leq T\}$ is sent from the n^{th} transmit aperture, $n = 1, \cdots, N$. The receiver, upon observing y_m^T from the m^{th} receive aperture, $m = 1, \cdots, M$, produces an output $\hat{w} = \phi(y_1^T, \cdots, y_M^T)$. The rate of this (W, T)–code is given by $R = \frac{1}{T}\log W$ nats/sec., and its average probability of decoding error is given by

$$P_e(f,\, \phi) = \frac{1}{W}\sum_{w=1}^{W} \Pr\left\{\phi(Y_1^T, \cdots, Y_M^T) \neq w \,\middle|\, x_1^T(w), \cdots, x_N^T(w)\right\}.$$

In this chapter, we provide a "single-letter characterization" of the channel capacity C (see Definition 1) in terms of signal and channel parameters, and examine some properties of the optimum transmission strategies. *We use the notation C_{ind} and C_{sum} to denote the capacity with individual and sum average power constraints respectively.*

3.3 Statement of results

The following inequality, which can be verified by simple algebra, will be used in our results:

$$\begin{aligned}\zeta\left(\sum_{n=1}^{N} x_n,\, y\right) &= \sum_{n=1}^{N} \zeta\left(x_n,\, y+\sum_{k=1}^{n-1} x_k\right) \\ &\geq \sum_{n=1}^{N} \zeta(x_n,\, y), \quad x_n \geq 0, \quad n=1,\,\cdots,\,N, \quad y \geq 0, \end{aligned} \tag{3.6}$$

with $x_0 = 0$.

3.3.1 Channel capacity

Our main result characterizes the capacity of the $N \times M$ MIMO Poisson channel.

Theorem 4 *The capacity of the $N \times M$ MIMO Poisson channel with peak and average power constraints A_n, $n = 1, \cdots, N$ and σ_n, $n = 1, \cdots, N$ respectively, is given by*

$$C_{ind} = \max_{\substack{0 \leq \mu_n \leq \sigma_n, \\ n=1,\,\cdots,\,N}} I(\boldsymbol{\mu}^N,\, \mathbf{s}), \tag{3.7}$$

where

$$\begin{aligned} I(\boldsymbol{\mu}^N,\, \mathbf{s}) &\triangleq \sum_{m=1}^{M}\left[\sum_{n=1}^{N} \nu_n \zeta\left(\sum_{k=\Pi(1)}^{\Pi(n)} s_{k,\,m} A_k,\, \lambda_{0,\,m}\right)\right. \\ &\qquad \left. -\zeta\left(\sum_{n=1}^{N} \nu_n \sum_{k=\Pi(1)}^{\Pi(n)} s_{k,\,m} A_k,\, \lambda_{0,\,m}\right)\right], \end{aligned} \tag{3.8}$$

with $\Pi : \{1,\,\cdots,\,N\} \to \{1,\,\cdots,\,N\}$ being a permutation of $\{1,\,\cdots,\,N\}$ such that

$$\mu_{\Pi(n)} \geq \mu_{\Pi(n+1)}, \quad n = 1,\,\cdots,\,N-1, \tag{3.9}$$

and

$$\nu_n \triangleq \begin{cases} \mu_{\Pi(n)} - \mu_{\Pi(n+1)}, & n = 1, \cdots, N-1, \\ \mu_{\Pi(N)}, & n = N. \end{cases} \tag{3.10}$$

Corollary 5 *The capacity with peak power constraints A_n, $n = 1, \cdots, N$ and an average sum power constraint σ, is given by*

$$C_{sum} = \max_{\substack{0 \le \mu_n \le 1,\, n=1,\, \cdots,\, N \\ \sum_{n=1}^{N} \mu_n A_n \le \sigma \sum_{n=1}^{N} A_n}} I(\boldsymbol{\mu}^N, \mathbf{s}), \tag{3.11}$$

where $I(\cdot, \cdot)$ is as defined in (3.8).

Remarks: (i) The optimization in (3.7) (as well as in (3.11)) is of a concave functional over a convex compact set, so that the maximum clearly exists.

(ii) Our proof of the achievability part of Theorem 4 (as well as Corollary 5) shows that binary signaling from each transmit aperture, with arbitrarily fast intertransition times, can achieve channel capacity. The two signaling levels at each transmit aperture correspond to no transmission ("OFF" state), and transmission at the peak power level ("ON" state). The parameter μ_n in (3.7) (as well as in (3.11)) is the probability of the transmitted signal through transmit aperture n remaining in the level A_n (ON state), and can be interpreted as the *average duty cycle* of the n^{th} aperture's transmitted signal, $n = 1, \cdots, N$.

(iii) In general, the capacity-achieving transmitted signals through the N transmit apertures are correlated across apertures but i.i.d. in time. A *transmission event* is an assignment of ON and OFF states to the N transmit apertures at each time instant. We show that the optimum set of transmission events can take at most $N + 1$ values (out of a total of 2^N possible values), which correspond to k transmit apertures being ON, $k = 0, 1, \cdots, N$. Whenever a transmit aperture is ON, all the transmit apertures with higher average duty cycles (in the sense of the permutation Π) must also remain ON.

3.3.2 Optimum transmission strategy

The optimum transmission strategy can be examined upon solving the optimization problems in (3.7) and (3.11), which involve the computation of the maximum of a concave function of N variables, viz., $\{\mu_n, n = 1, \cdots, N\}$, over linear constraint sets. The optimization problem in (3.11) differs from the problem in (3.7) *only* in the constraint set, with the individual constraints on the average duty cycles, viz., $\mu_n \leq \sigma_n$, $n = 1, \cdots, N$ being replaced by a constraint on their sum, viz., $\sum_{n=1}^{N} \mu_n A_n \leq \sigma \sum_{n=1}^{N} A_n$. Optimal solutions of both the problems can be computed using Kuhn-Tucker conditions (cf. e.g., [35], p. 233).

We present below the structure of the optimal solutions of (3.7) and (3.11) for the special case of $N = 2$ transmit apertures and M receive apertures, where $M \in \mathbb{Z}^+$ is arbitrary. The optimal solutions for general $N \in \mathbb{Z}^+$ can be obtained along similar lines, but are omitted due to notational complications.

3.3.2.1 Individual average power constraints

First consider the MIMO Poisson channel with peak and average power constraints on individual transmit apertures. We begin with some notation. Let

$$\begin{aligned} b_{n,\,m} &\triangleq \tfrac{s_{n,\,m}A_n}{\lambda_{0,\,m}}, \qquad n = 1,\,2, \\ B_m &= \textstyle\sum_{n=1}^{2} b_{n,m}, \quad m = 1,\,\cdots,\,M. \end{aligned} \tag{3.12}$$

Let $\rho = \rho_n$ be the solution of

$$\sum_{m=1}^{M} \lambda_{0,\,m} b_{n,\,m} \log\left(\frac{1+\alpha(b_{n,\,m})b_{n,\,m}}{1+\rho B_m}\right) \;=\; 0, \quad n = 1,\,2, \tag{3.13}$$

where $\alpha(\cdot)$ is as defined in (2.4), and let $\rho = \bar{\rho}$ be the solution of

$$\sum_{m=1}^{M} \lambda_{0,\,m} B_m \log\left(\frac{1+\alpha\left(B_m\right)B_m}{1+\rho B_m}\right) \;=\; 0. \tag{3.14}$$

It can be verified that[1] $\rho_1 \geq 0$, $\rho_2 \geq 0$, $\frac{1}{e} \leq \bar{\rho} \leq \frac{1}{2}$ and $\bar{\rho} \geq \max\{\rho_1,\,\rho_2\}$. For $0 \leq x \leq \rho_1$, let $\beta_1 = \beta_1(x)$ solve

$$\sum_{m=1}^{M} \lambda_{0,\,m} b_{1,\,m} \log\left(\frac{1+\alpha(b_{1,\,m})b_{1,\,m}}{1+b_{1,\,m}\beta_1 + b_{2,\,m}x}\right) \;=\; 0, \tag{3.15}$$

and for $0 \leq x \leq \rho_2$, let $\beta_2 = \beta_2(x)$ solve

$$\sum_{m=1}^{M} \lambda_{0,\,m} b_{2,\,m} \log\left(\frac{1+\alpha(b_{2,\,m})b_{2,\,m}}{1+b_{1,\,m}x + b_{2,\,m}\beta_2}\right) \;=\; 0. \tag{3.16}$$

It can be verified that $\beta_1(\cdot)$ and $\beta_2(\cdot)$ are monotone decreasing, $\frac{1}{e} \leq \beta_n(0) \leq \frac{1}{2}$, $\beta_n(\rho_n) = \rho_n$, $n = 1,\,2$. The parameters ρ_1, ρ_2, $\bar{\rho}$ and the functions $\beta_1(\cdot)$ and $\beta_2(\cdot)$ are obtained in course of evaluation of the Kuhn-Tucker conditions.

We are now ready to characterize the optimum transmission strategy for the $2 \times M$ MIMO Poisson channel with peak and average power constraints.

[1]The following fact is useful for the purpose of the said verification: $\alpha(\cdot)$ is monotone decreasing on $[0,\,\infty)$ with $\alpha(0) = \frac{1}{2}$ and $\alpha(\infty) = \frac{1}{e}$.

Theorem 5 *The pair* (μ_1^*, μ_2^*) *that maximizes the optimization problem in (3.7) for the special case of* $N = 2$ *is given as follows.*

1. *If* $\max\{\sigma_1, \sigma_2\} \geq \bar{\rho}$*, then* $\mu_1^* = \mu_2^* = \bar{\rho}$.

2. *If* $\sigma_1 \geq \sigma_2$ *and* $\rho_1 \leq \sigma_2 \leq \bar{\rho}$*, then* $\mu_1^* = \mu_2^* = \sigma_2$.

3. *If* $\sigma_1 \leq \sigma_2$ *and* $\rho_2 \leq \sigma_1 \leq \bar{\rho}$*, then* $\mu_1^* = \mu_2^* = \sigma_1$.

4. *If* $\sigma_2 \leq \rho_1$ *and* $\sigma_1 \geq \beta_1(\sigma_2)$*, then* $\mu_1^* = \beta_1(\sigma_2)$*,* $\mu_2^* = \sigma_2$.

5. *If* $\sigma_1 \leq \rho_2$ *and* $\sigma_2 \geq \beta_2(\sigma_1)$*, then* $\mu_1^* = \sigma_1$*,* $\mu_2^* = \beta_2(\sigma_1)$.

6. *For all other* (σ_1, σ_2) *pairs,* $\mu_1^* = \sigma_1$*,* $\mu_2^* = \sigma_2$.

Remarks: (i) In Figure 3.2, we partition the unit square into regions corresponding to the conditions (1) – (6) of Theorem 5. Specifically, condition (k) of Theorem 5 corresponds to (σ_1, σ_2) pairs in region R_k in Figure 3.2, $k = 1, \cdots, 6$. The optimum pair (μ_1^*, μ_2^*) lies either inside or on the boundary of R_6, which is characterized by one of the lines specified by the equations (i) $\mu_1^* = \mu_2^* = \rho$, $\min\{\rho_1, \rho_2\} \leq \rho \leq \bar{\rho}$, (ii) $\mu_1^* = \beta_1(\mu_2^*)$, $0 \leq \mu_2^* \leq \rho_1$, (iii) $\mu_2^* = \beta_2(\mu_1^*)$, $0 \leq \mu_1^* \leq \rho_2$. The average power constraints involving both σ_1 and σ_2 are active at optimality[2] for points in R_6, while neither constraint is active at optimality for points in R_1. For points in R_3 and R_5, only the constraint involving σ_1 is active at optimality, while for points in R_2 and R_4, only the constraint involving σ_2 is active at optimality. This is a generalization of the SISO Poisson channel capacity result [11, 52], in which the average power

[2]We say that the power constraint $\mu_i \leq \sigma_i$ is active at optimality if $\mu_i^* = \sigma_i$, $i = 1, 2$. See [35], p. 220 for a formal definition of active constraints.

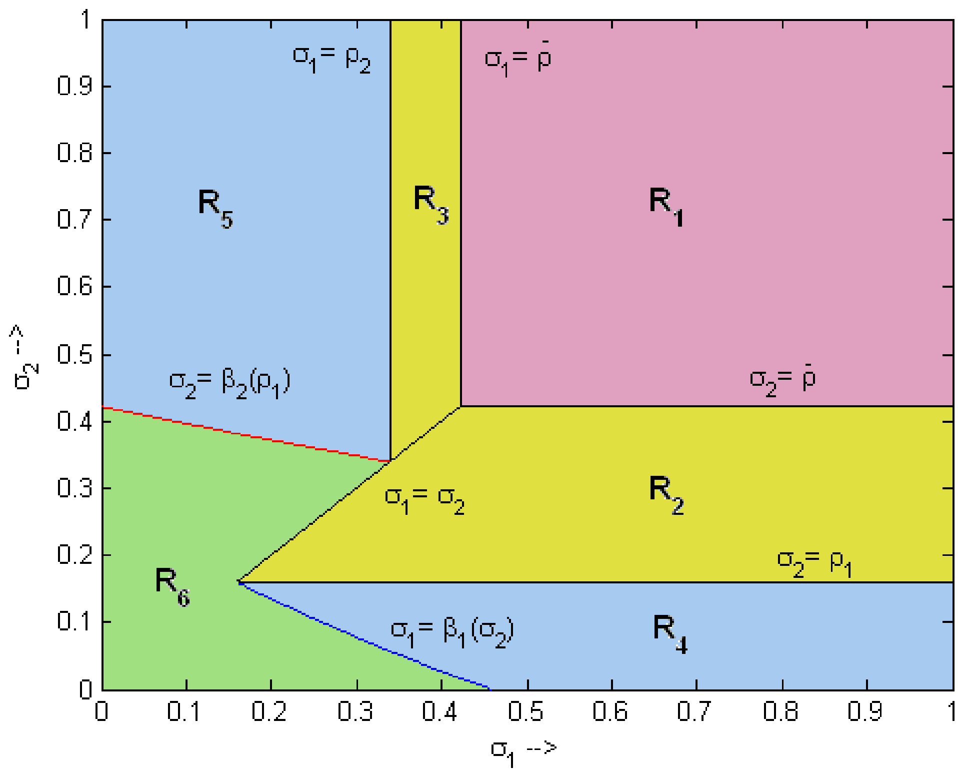

Figure 3.2: The structure of the optimal solution for $N = 2$. Region R_k corresponds to condition (k) in Theorem 5, $k = 1, \cdots, 6$.

constraint is active at optimality only below a threshold (k_m in [11], Theorem 1) which is determined from the system parameters.

(ii) It follows from Theorem 5 that if $\sigma_1 = \sigma_2 = \sigma$, then $\mu_1^* = \mu_2^* = \min\{\sigma, \bar{\rho}\}$. This implies that for the $2 \times M$ MIMO Poisson channel with a *symmetric* average power constraint, the optimum transmission strategy assigns nonzero probability to only 2 transmission events, viz., (a) both transmit apertures in the ON state (with probability $\min\{\sigma, \bar{\rho}\}$), and (b) both apertures in the OFF state (with probability $1 - \min\{\sigma, \bar{\rho}\}$). This *simultaneous ON-OFF keying* strategy can achieve capacity for an arbitrary number of transmit apertures, under a symmetric average power constraint. This special case corresponds to the problem addressed in [19], wherein the lower bound derived is *always* tight.

(iii) For the $2 \times M$ MIMO Poisson channel with *asymmetric* average power constraints, the optimum set of transmission events has at most 3 values. In particular, if $\sigma_1 > \sigma_2$, the following events can have nonzero probability: (a) both transmit apertures in the ON state (with probability μ_2^*), (b) aperture 1 in the ON state and aperture 2 in the OFF state (with probability $\mu_1^* - \mu_2^*$), and (c) both apertures in the OFF state (with probability $1 - \mu_1^*$). If $\sigma_1 < \sigma_2$, the corresponding events are: (a') both transmit apertures in the ON state (with probability μ_1^*), (b') aperture 1 in the OFF state and aperture 2 in the ON state (with probability $\mu_2^* - \mu_1^*$), and (c') both apertures in the OFF state (with probability $1 - \mu_2^*$). However, for a range of values of σ_1 and σ_2 (corresponding to regions R_1, R_2, R_3 in Figure 3.2), the optimum number of transmission events is still 2, and the simultaneous ON-OFF keying strategy is optimal.

3.3.2.2 Average sum power constraint

Consider next the MIMO Poisson channel with peak average power constraints on individual transmit apertures and a constraint on the sum of the average powers from all the transmit apertures. We begin with some notation. Define

$$\begin{aligned} K_1 &\stackrel{\Delta}{=} \textstyle\sum_{m=1}^{M} \lambda_{0,m} \left(B_m \log\left(1+\bar{\rho}B_m\right) - (1+a)b_{2,m}\log(1+\rho_2 B_m) \right), \\ K_2 &\stackrel{\Delta}{=} \textstyle\sum_{m=1}^{M} \lambda_{0,m} \left((1+a)b_{1,m}\log(1+\rho_1 B_m) - aB_m \log\left(1+\bar{\rho}B_m\right) \right), \end{aligned} \tag{3.17}$$

where $a = \frac{A_1}{A_2}$. Using (3.6), it is not difficult to verify that $K_1 \geq K_2$. For $x, y \in \mathbb{R}$, define

$$d(x, y) \stackrel{\Delta}{=} \sum_{m=1}^{M} \lambda_{0,m}(b_{1,m} - ab_{2,m}) \log(1 + (b_{1,m} - ab_{2,m})x + (1+a)b_{2,m}y). \tag{3.18}$$

Note that $d(\cdot, y)$ is nondecreasing on $\mathbb{R}$ for every $y \in \mathbb{R}$. Let $x = \gamma_n(y)$ be the solution of the equation

$$d(x, y) = K_n, \quad n = 1, 2, \tag{3.19}$$

whence it can be verified that $\gamma_1(y) \geq \gamma_2(y)$, $y \in \mathbb{R}$.

We are now ready to characterize the optimum transmission strategy for the $2 \times M$ MIMO Poisson channel with peak power constraints and an average sum power constraint.

Theorem 6 *The pair (μ_1^*, μ_2^*) that maximizes the optimization problem in (3.11) for the special case of $N = 2$ is given as follows.*

1. *If $\sigma \geq \bar{\rho}$, then $\mu_1^* = \mu_2^* = \bar{\rho}$.*

2. If $\sigma < \bar{\rho}$, then $\mu_2^* = (1+a)\sigma - a\mu_1^*$, where

$$\mu_1^* \;=\; \begin{cases} \max\{0,\, \gamma_1(\sigma)\}, & \text{if} \quad K_1 \le d(\sigma,\, \sigma), \\ \min\{\gamma_2(\sigma),\, (1+1/a)\sigma\}, & \text{if} \quad K_2 \ge d(\sigma,\, \sigma), \\ \sigma, & \text{otherwise.} \end{cases} \tag{3.20}$$

A key property of the optimum pair $(\mu_1^*,\, \mu_2^*)$ is discussed in the following corollary. We show that the relative ordering of μ_1^* and μ_2^*, i.e., whether $\mu_1^* \ge \mu_2^*$ or $\mu_1^* \le \mu_2^*$, does not depend on σ. An application of this property is shown in Chapter 4 where we consider the MIMO Poisson channel with isotropically distributed random channel fade.

Corollary 6 *For all $\sigma \in [0,\, 1]$,*

1. *if $K_1 \ge 0$ and $K_2 \le 0$, then $\mu_1^* = \mu_2^*$;*
2. *if $K_2 > 0$, then $\mu_1^* \ge \mu_2^*$;*
3. *if $K_1 < 0$, then $\mu_1^* \le \mu_2^*$.*

Remarks: (i) From Corollary 6, it follows that regardless of the value of $0 \le \sigma \le 1$, $\mu_1^* \ge \mu_2^*$ iff $K_1 + K_2 \ge 0$. Thus, the relative ordering of μ_1^* and μ_2^* does not depend on σ. This implies that given a set of fade coefficients $\{s_{n,\,m}\}_{n=1,\,m=1}^{N,\,M}$, peak power constraints $\{A_n\}_{n=1}^{N}$, and background noise rates $\{\lambda_{0,\,m}\}_{m=1}^{M}$, one transmit aperture experiences better channel conditions (which is reflected by the condition $1(K_1 + K_2 \ge 0)$), and hence is assigned a higher optimal average duty cycle, than the other transmit aperture, for all $0 \le \sigma \le 1$.

(ii) The remark above can be generalized to the case of an arbitrary number of transmit apertures. The relative ordering of $\{\mu_n^*\}_{n=1}^{N}$ does not depend on σ.

3.4 Proofs

We begin this section with some additional definitions that will be needed in our proofs. For any $\tau > 0$, the number of photon arrivals $N_m(\tau)$ on $[0, \tau]$ together with the corresponding (ordered) arrival times $\mathbf{T}_m^{N_m(\tau)} = \{T_{m,1}, \cdots, T_{m,N_m(\tau)}\}$ are sufficient statistics for Y_m^τ, $m = 1, \cdots, M$. Therefore, the random vector $\mathbf{Y}_{ss}(\tau) = \{(N_m(\tau), \mathbf{T}_m^{N_m(\tau)})\}_{m=1}^M$ is a complete description of the random processes $\mathbf{Y}^\tau = \{Y_1^\tau, \cdots, Y_M^\tau\}$, $\tau \geq 0$.

The channel is characterized as follows. For an input signal $\mathbf{x}^T = \{x_n(t), 0 \leq t \leq T\}_{n=1}^N$ satisfying (3.2), the channel output $(N_m(T), \mathbf{T}_m^{N_m(T)})$ at the m^{th} receive aperture has the "conditional sample function density" (cf. e.g., [47])

$$\begin{aligned} & f_{N_m(T),\, \mathbf{T}_m^{N_m(T)}|\mathbf{X}^T}\left(n_m,\, \mathbf{t}_m^{n_m} \left|\mathbf{x}^T\right.\right) \\ &= \exp\left(-\int_0^T \lambda_m(\tau)d\tau\right) \cdot \prod_{i=1}^{n_m} \lambda_m(t_{m,i}), \end{aligned} \tag{3.21}$$

where $\lambda_m(\cdot)$ is as defined in (3.1). Recall that the processes $(N_m(T), \mathbf{T}_m^{N_m(T)})$, $m = 1, \cdots, M$ are conditionally mutually independent given the knowledge of $\mathbf{x}^T$. Therefore, the conditional sample function density of $\mathbf{Y}_{ss}(T)$ given $\mathbf{x}^T$ is given by

$$\begin{aligned} f_{\mathbf{Y}_{ss}(T)|\mathbf{X}^T}(\mathbf{y_{ss}}|\mathbf{x}^T) &= \prod_{m=1}^M f_{N_m(T),\, \mathbf{T}_m^{N_m(T)}|\mathbf{X}^T}(n_m,\, \mathbf{t}_m^{n_m}|\mathbf{x}^T) \\ &= \prod_{m=1}^M \exp\left(-\int_0^T \lambda_m(\tau)d\tau\right) \cdot \prod_{i=1}^{n_m} \lambda_m(t_{m,i}), \end{aligned} \tag{3.22}$$

where $\mathbf{y}_{ss} = \{(n_m, \mathbf{t}_m^{n_m})\}_{m=1}^M$. In order to write the channel output sample function density for a given distribution of $\mathbf{X}^T$, consider the conditional mean of X_n^T (conditioned causally on the channel output) by

$$\hat{X}_n(\tau) \triangleq \mathbb{E}[X_n(\tau)|\mathbf{Y}_{ss}(\tau)], \quad 0 \leq \tau \leq T, \quad n = 1, \cdots, N, \tag{3.23}$$

where we have suppressed the dependence of $\hat{X}_n(\tau)$, $n = 1, \cdots, N$ on $\mathbf{Y}_{ss}(\tau)$ for notational convenience; and define

$$\Lambda_m(\tau) \triangleq \sum_{n=1}^{N} s_{n,m} X_n(\tau) + \lambda_{0,m}, \quad 0 \le \tau \le T, \quad m = 1, \cdots, M, \tag{3.24}$$

and

$$\begin{aligned}\hat{\Lambda}_m(\tau) &\triangleq \mathbb{E}[\Lambda_m(\tau)|\mathbf{Y}_{ss}(\tau)] \\ &= \sum_{n=1}^{N} s_{n,m}\hat{X}_n(\tau) + \lambda_{0,m}, \quad 0 \le \tau \le T, \quad m = 1, \cdots, M. \end{aligned}\tag{3.25}$$

From ([47], pp. 425-427), it follows that the process $(N_m(T), \mathbf{T}_m^{N_m(T)})$ is a self-exciting PCP with rate process $\hat{\Lambda}_m^T$, $m = 1, \cdots, M$, and the output sample function density is given by

$$f_{\mathbf{Y}_{ss}(T)}(\mathbf{y}_{ss}) = \prod_{m=1}^{M} \exp\left(-\int_0^T \hat{\lambda}_m(\tau)d\tau\right) \cdot \prod_{i=1}^{n_m} \hat{\lambda}_m(t_{m,i}), \tag{3.26}$$

where $\mathbf{y}_{ss} = \{(n_m, \mathbf{t}_m^{n_m})\}_{m=1}^{M}$, and

$$\hat{\lambda}_m(\tau) = \sum_{n=1}^{N} s_{n,m}\hat{x}_n(\tau) + \lambda_{0,m}, \quad m = 1, \cdots, M,$$

with

$$\hat{x}_n(\tau) = \mathbb{E}[X_n(\tau)|\mathbf{Y}_{ss}(\tau) = \mathbf{y}_{ss}], \quad 0 \le \tau \le T, \quad n = 1, \cdots, N.$$

Proof of Theorem 4:

Converse part: Let the rv W be uniformly distributed on (the message set) $\mathcal{W} = \{1, \cdots, W\}$. Consider a (W, T)–code (f, ϕ) of rate $R = \frac{1}{T}\log W$, and with $P_e(f, \phi) \le \epsilon$, where $0 \le \epsilon \le 1$ is given. Denote $X_n(t) \triangleq x_n(W, t)$, $0 \le t \le T$, $n = 1, \cdots, N$. Note that (3.4) then implies that

$$\begin{aligned} 0 &\le X_n(t) \le A_n, \quad 0 \le t \le T, \quad n = 1, \cdots, N, \\ \tfrac{1}{T}\int_0^T \mathbb{E}[X_n(\tau)]d\tau &\le \sigma_n A_n, \quad n = 1, \cdots, N. \end{aligned}\tag{3.27}$$

Let $(N_m(T),\, \mathbf{T}_m^{N_m(T)})$ be the channel output at the m^{th} receive aperture when X_n^T is transmitted from the n^{th} transmit aperture, $n = 1, \cdots, N$, $m = 1, \cdots, M$. Clearly, the following Markov condition holds:

$$W \multimap \mathbf{X}^\tau \multimap \mathbf{Y}_{ss}(\tau), \quad 0 \le \tau \le T. \tag{3.28}$$

By a standard argument, the rate R of the $(W,\, T)$–code $(f,\, \phi)$ with $P_e(f,\, \phi) \le \epsilon$ must satisfy

$$R \ \lessapprox \ \frac{1}{T} I\left(W \wedge \phi(\mathbf{Y}_{ss}(T))\right). \tag{3.29}$$

Proceeding further with the right side of (3.29),

$$\begin{aligned} I\left(W \wedge \phi(\mathbf{Y}_{ss}(T))\right) \ &\le \ I\left(W \wedge \mathbf{Y}_{ss}(T)\right) && (3.30) \\ &= \ h(\mathbf{Y}_{ss}(T)) - h(\mathbf{Y}_{ss}(T)|W) \\ &\le \ h(\mathbf{Y}_{ss}(T)) - h(\mathbf{Y}_{ss}(T)|X^T) && (3.31) \end{aligned}$$

where (3.30) is by the data processing result for mixed rvs[3]; and (3.31) is by (3.28). The difference between the entropies of mixed rvs[4] on the right side of (3.31) is

$$\begin{aligned} & h(\mathbf{Y}_{ss}(T)) - h(\mathbf{Y}_{ss}(T)|X^T) \\ = \ & \mathbb{E}\left[-\log f_{\mathbf{Y}_{ss}(T)}(\mathbf{Y}_{ss}(T))\right] - \mathbb{E}\left[-\log f_{\mathbf{Y}_{ss}(T)|\mathbf{X}^T}(\mathbf{Y}_{ss}(T)|\mathbf{X}^T)\right] \\ = \ & \mathbb{E}\left[\log \frac{\prod_{m=1}^M \exp\left(-\int_0^T \Lambda_m(\tau)d\tau\right) \prod_{i=1}^{N_m(T)} \Lambda_m(T_{m,i})}{\prod_{m=1}^M \exp\left(-\int_0^T \hat{\Lambda}_m(\tau)d\tau\right) \prod_{i=1}^{N_m(T)} \hat{\Lambda}_m(T_{m,i})}\right] \end{aligned} \tag{3.32}$$

[3]This result can be deduced, for instance, from [39], the last paragraph of Section 3.4, p. 36, and Kolmogorov's formula (3.6.3) on p. 37.

[4]Our definition of the conditional entropy of mixed rvs is consistent with the general formulation developed, for instance, in [39], Chapter 3.

$$
\begin{aligned}
&= \sum_{m=1}^{M} \Bigg\{ \mathbb{E}\left[\int_0^T \left(\hat{\Lambda}_m(\tau) - \Lambda_m(\tau)\right) d\tau\right] \\
&\qquad + \mathbb{E}\left[\sum_{i=1}^{N_m(T)} \left(\log \Lambda_m(T_{m,i}) - \log \hat{\Lambda}_m(T_{m,i})\right)\right]\Bigg\} \\
&= \sum_{m=1}^{M} \mathbb{E}\left[\sum_{i=1}^{N_m(T)} \left(\log \Lambda_m(T_{m,i}) - \log \hat{\Lambda}_m(T_{m,i})\right)\right] \qquad (3.33) \\
&= \sum_{m=1}^{M} \int_0^T \left\{ \mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} s_{n,m} X_n(\tau), \lambda_{0,m}\right) - \zeta\left(\sum_{n=1}^{N} s_{n,m} \hat{X}_n(\tau), \lambda_{0,m}\right)\right]\right\} d\tau, \qquad (3.34)
\end{aligned}
$$

where (3.32) is by (3.22), (3.26); (3.33) holds by an interchange of operations[5] to get

$$
\mathbb{E}\left[\int_0^T (\hat{\Lambda}_m(\tau) - \Lambda_m(\tau)) d\tau\right] = \int_0^T \mathbb{E}\left[\hat{\Lambda}_m(\tau) - \Lambda_m(\tau)\right] d\tau,
$$

followed by noting that $\mathbb{E}[\hat{\Lambda}_m(\tau)] = \mathbb{E}[\Lambda_m(\tau)]$, $0 \le \tau \le T$, $m = 1, \cdots, M$, by (3.25); and (3.34) is proved in Appendix B.1.

Next, in the right side of (3.34),

$$
\begin{aligned}
\mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} s_{n,m} \hat{X}_n(\tau),\, \lambda_{0,m}\right)\right] &\geq \zeta\left(\sum_{n=1}^{N} s_{n,m}\, \mathbb{E}[\hat{X}_n(\tau)],\, \lambda_{0,m}\right) \qquad (3.35) \\
&= \zeta\left(\sum_{n=1}^{N} s_{n,m}\, \mathbb{E}[X_n(\tau)],\, \lambda_{0,m}\right), \qquad (3.36)
\end{aligned}
$$

where (3.35) is by Jensen's inequality applied to the convex function $\zeta(\cdot,\, \lambda_0)$; and (3.36) holds as

$$
\mathbb{E}[\hat{X}_n(\tau)] = \mathbb{E}[\mathbb{E}[X_n(\tau) | \mathbf{Y}_{ss}(\tau)]] = \mathbb{E}[X_n(\tau)].
$$

Summarizing collectively (3.31), (3.34), (3.36), we get that

$$
I\left(W \wedge \phi(\mathbf{Y}_{ss}(T))\right) \leq \int_0^T \sum_{m=1}^{M} \left\{ \mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} s_{n,m} X_n(\tau),\, \lambda_{0,m}\right)\right.\right.
$$

[5] The interchange is permissible by the integrability of $\{\Lambda_m(\tau), 0 \le \tau \le T\}$ and $\{\hat{\Lambda}_m(\tau), 0 \le \tau \le T\}$, $m = 1, \cdots, M$.

$$-\zeta\left(\sum_{n=1}^{N} s_{n,m}\,\mathbb{E}[X_n(\tau)],\ \lambda_{0,m}\right)\Bigg]\Bigg\}\,d\tau. \quad (3.37)$$

The right side of (3.37) is further bounded above by a suitable modification of Footnote 10, Chapter 2. Fixing $0 \leq \tau \leq T$, consider maximizing the integrand on the right side of (3.37) over all joint distributions of $\{X_n(\tau),\ n = 1, \cdots, N\}$ with fixed means $\mathbb{E}[X_n(\tau)] = \chi_n(\tau)A_n$, $n = 1, \cdots, N$, say, and subject to the first constraint (alone) in (3.27). Then, the integrand on the right side of (3.37) at time instant τ equals

$$\sum_{m=1}^{M}\left\{\mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} s_{n,m}X_n(\tau), \lambda_{0,m}\right)\right] - \zeta\left(\sum_{n=1}^{N} s_{n,m}\chi_n(\tau)A_n, \lambda_{0,m}\right)\right\},$$

and is maximized by considering the first term above. Consider the optimization problem

$$h(\tau, \boldsymbol{\chi}^N(\tau)) \;=\; \max_{\substack{0\leq X_n(\tau)\leq A_n\\ \mathbb{E}[X_n(\tau)]=\chi_n(\tau)A_n\\ n=1,\cdots,N}} \mathbb{E}\left[\sum_{m=1}^{M}\zeta\left(\sum_{n=1}^{N} s_{n,m}X_n(\tau), \lambda_{0,m}\right)\right]. \quad (3.38)$$

We now establish the fact that the optimal marginal distributions of $\{X_n(\tau)\}_{n=1}^N$ that maximize the right-side of (3.38) are binary $\{0,\ A_n\}$-valued. To this end, consider the following alternate formulation of (3.38):

$$h(\tau,\ \boldsymbol{\chi}^N(\tau))$$
$$= \max_{\substack{0\leq X_n(\tau)\leq A_n\\ \mathbb{E}[X_n(\tau)]=\chi_n(\tau)A_n\\ n=2,\,\cdots,\,N\\ \mathbf{E}[\eta_1(\mathbf{X}_2^N(\tau))]=\chi_1(\tau)A_n}} \ \max_{\substack{0\leq X_1(\tau)\leq A_1\\ \mathbf{E}[X_1(\tau)|\mathbf{X}_2^N(\tau)]=\eta_1(\mathbf{X}_2^N(\tau))}} \mathbb{E}\left[\sum_{m=1}^{M}\zeta\left(\sum_{n=1}^{N} s_{n,m}X_n(\tau), \lambda_{0,m}\right)\right],$$

where the inner optimization is performed over the class of conditional distributions of $X_1(\tau)$ (conditioned on $\mathbf{X}_2^N(\tau)$) with a finite support $[0,\ A_1]$ and fixed conditional mean $\mathbb{E}[X_1(\tau)|\mathbf{X}_2^N(\tau) = \mathbf{x}_2^N] = \eta_1(\mathbf{x}_2^N)$, $\mathbf{x}_2^N \in \prod_{n=2}^N[0,\ A_n]$. Using the strict convexity of $\zeta(\cdot,\ y)$, the inner maximum is achieved (see footnote 10, Chapter 2) iff $X_1(\tau)$

is a $\{0,\ A_1\}$–valued rv with

$$\begin{aligned}
\Pr\{X_1(\tau) = A_1 | \mathbf{X}_2^N(\tau) = \mathbf{x}_2^N\} &= 1 - \Pr\{X_1(\tau) = 0 | \mathbf{X}_2^N(\tau) = \mathbf{x}_2^N\} \\
&= \frac{\eta_1(\mathbf{x}_2^N)}{A_1}, \quad \mathbf{x}_2^N \in \prod_{n=2}^{N} [0,\ A_n].
\end{aligned}$$

Returning to (3.38), we see that the optimal marginal pmf of $X_1(\tau)$ that maximizes the right side of (3.38) is $\{0,\ A_1\}$–valued with

$$\Pr\{X_1(\tau) = A_1\} = \frac{\mathbb{E}[\eta_1(\mathbf{X}_2^N(\tau))]}{A_1} = \chi_1(\tau).$$

By symmetry, it follows that the optimal marginal pmf of $X_n(\tau)$ that maximizes the right side of (3.38) is $\{0,\ A_n\}$–valued with

$$\begin{aligned}
\Pr\{X_n(\tau) = A_n\} &= 1 - \Pr\{X_n(\tau) = 0\} \\
&= \chi_n(\tau), \quad n = 1,\ \cdots,\ N. \qquad (3.39)
\end{aligned}$$

It suffices, therefore, to focus on the joint pmf of binary $\{0,\ A_n\}$–valued rvs $X_n(\tau)$, $n = 1,\ \cdots,\ N$, in order to compute the right side of (3.38). To this end, we introduce the following notation. Let $b_n(i) \in \{0,\ 1\}$ denote the n^{th} bit in the binary representation of i, $i = 0,\ 1,\ \cdots,\ 2^N - 1$, $n = 1,\ \cdots,\ N$; in other words, $i = \sum_{n=1}^{N} b_n(i) 2^{n-1}$. Let $I_i = \{n \in \{1,\ \cdots,\ N\} : b_n(i) = 1\}$ denote the set of nonzero bit positions in the binary representation of i, $i = 0,\ \cdots,\ 2^N - 1$. Then

$$q_i(\tau) \triangleq \Pr\{X_n(\tau) = b_n(i) A_n, \quad n = 1,\ \cdots,\ N\}, \quad i = 0,\ \cdots,\ 2^N - 1, \quad (3.40)$$

denotes the (joint) pmf of n^{th} transmit aperture remaining in state $b_n(i)$, $n = 1,\ \cdots,\ N$, at time instant τ, $0 \le \tau \le T$. By (3.39) and (3.40), the joint pmf vector $\mathbf{q}(\tau) = \{q_i(\tau)\}_{i=0}^{2^N-1}$ satisfies

$$\begin{aligned} q_i(\tau) &\geq 0, \quad i = 0, \cdots, 2^N - 1, \\ \textstyle\sum_{i=0}^{2^N-1} q_i(\tau) &= 1, \\ \textstyle\sum_{i:b_n(i)=1} q_i(\tau) &= \chi_n(\tau), \quad n = 1, \cdots, N. \end{aligned} \tag{3.41}$$

Summarizing collectively (3.38), (3.39), (3.40), we get that

$$h(\tau, \boldsymbol{\chi}^N(\tau)) = \max_{\mathbf{q}(\tau)} \sum_{i=0}^{2^N-1} q_i(\tau) \sum_{m=1}^{M} \zeta\left(\sum_{n=1}^{N} s_{n,m} b_n(i) A_n, \lambda_{0,m}\right), \tag{3.42}$$

where $\mathbf{q}(\tau) = \{q_i(\tau)\}_{i=0}^{2^N-1}$ satisfies (3.41). In Appendix B.2, we show that the maximum in the right side of (3.42) is achieved by

$$q_i^*(\tau) = \begin{cases} \chi_{P_\tau(k)}(\tau) - \chi_{P_\tau(k+1)}(\tau), & \text{if } i = \sum_{n=1}^{k} 2^{P_\tau(n)-1}, \quad k = 1, \cdots, N-1, \\ \chi_{P_\tau(N)}(\tau), & \text{if } i = 2^N - 1, \\ 1 - \chi_{P_\tau(1)}(\tau), & \text{if } i = 0, \\ 0, & \text{otherwise,} \end{cases} \tag{3.43}$$

where $P_\tau : \{1, \cdots, N\} \to \{1, \cdots, N\}$ is a permutation of $\{1, \cdots, N\}$ such that

$$\chi_{P_\tau(n)}(\tau) \geq \chi_{P_\tau(n+1)}(\tau), \quad n = 1, \cdots, N-1. \tag{3.44}$$

By (3.42) and (3.43), we then get

$$\begin{aligned} h(\tau, \boldsymbol{\chi}^N(\tau)) &= \sum_{n=1}^{N-1} \left(\chi_{P_\tau(n)}(\tau) - \chi_{P_\tau(n+1)}(\tau)\right) \sum_{m=1}^{M} \zeta\left(\sum_{k=1}^{n} s_{P_\tau(k),m} A_{P_\tau(k)}, \lambda_{0,m}\right) \\ &\quad + \chi_{P_\tau(N)}(\tau) \sum_{m=1}^{M} \zeta\left(\sum_{k=1}^{N} s_{P_\tau(k),m} A_{P_\tau(k)}, \lambda_{0,m}\right) \\ &= \sum_{n=1}^{N} \psi_{n,P_\tau} \chi_{P_\tau(n)}(\tau), \end{aligned} \tag{3.45}$$

where we have defined[6]

$$\psi_{n,P} \triangleq \sum_{m=1}^{M} \left\{ \zeta\left(\sum_{k=1}^{n} s_{P(k),m} A_{P(k)}, \lambda_{0,m}\right) \right.$$

[6] We take $\sum_{k=1}^{0} x_k = 0$.

$$-\zeta\left(\sum_{k=1}^{n-1} s_{P(k),\,m} A_{P(k)},\ \lambda_{0,\,m}\right)\Bigg\} \tag{3.46}$$

for $n = 1, \cdots, N$ and any permutation P of $\{1, \cdots, N\}$. Noting that (3.27) implies that

$$\begin{aligned} 0 \ \leq \ \chi_n(\tau) \ &\leq \ 1, \quad 0 \leq \tau \leq T, \\ \tfrac{1}{T}\textstyle\int_0^T \chi_n(\tau) d\tau \ &\leq \ \sigma_n, \quad n = 1, \cdots, N, \end{aligned} \tag{3.47}$$

we thus see from (3.37), (3.38), (3.45), (3.47) that

$$\frac{1}{T} I(W \wedge \phi(\mathbf{Y}_{ss}(T))) \ \leq \ \max_{\substack{0 \leq \chi_n(\tau) \leq 1 \\ \frac{1}{T}\int_0^T \chi_n(\tau)d\tau \leq \sigma_n \\ n=1,\cdots,N}} \frac{1}{T}\int_0^T g(\tau,\, \boldsymbol{\chi}^N(\tau)) d\tau, \tag{3.48}$$

where we have defined

$$g(\tau,\, \boldsymbol{\chi}^N(\tau)) \ \triangleq \ \sum_{n=1}^{N} \psi_{n,\, P_\tau} \chi_{P_\tau(n)}(\tau) - \sum_{m=1}^{M} \zeta\left(\sum_{n=1}^{N} s_{n,\,m} \chi_n(\tau) A_n,\ \lambda_{0,\,m}\right). \tag{3.49}$$

Setting

$$\mu_n \ \triangleq \ \frac{1}{T}\int_0^T \chi_n(\tau) d\tau, \quad n = 1, \cdots, N, \tag{3.50}$$

from (3.49), we then get

$$\begin{aligned} &\frac{1}{T}\int_0^T g(\tau,\, \boldsymbol{\chi}^N(\tau)) d\tau \\ &\quad = \ \frac{1}{T}\int_0^T \left\{\sum_{n=1}^{N} \psi_{n,\, P_\tau} \chi_{P_\tau(n)}(\tau) - \sum_{m=1}^{M} \zeta\left(\sum_{n=1}^{N} s_{n,\,m} \chi_n(\tau) A_n,\ \lambda_{0,\,m}\right)\right\} d\tau \\ &\quad \leq \ \sum_{n=1}^{N} \frac{1}{T}\int_0^T \psi_{n,P_\tau} \chi_{P_\tau(n)}(\tau) d\tau - \sum_{m=1}^{M} \zeta\left(\sum_{n=1}^{N} s_{n,m} \mu_n A_n, \lambda_{0,m}\right) && (3.51) \\ &\quad \leq \ \sum_{n=1}^{N} \psi_{n,\Pi} \mu_{\Pi(n)} - \sum_{m=1}^{M} \zeta\left(\sum_{n=1}^{N} s_{n,m} \mu_n A_n, \lambda_{0,m}\right) && (3.52) \\ &\quad = \ \sum_{n=1}^{N} \mu_{\Pi(n)} \psi_{n,\,\Pi} - \sum_{m=1}^{M} \zeta\left(\sum_{n=1}^{N} \mu_{\Pi(n)} s_{\Pi(n),\,m} A_{\Pi(n)},\ \lambda_{0,\,m}\right), && (3.53) \end{aligned}$$

where (3.51) is by Jensen's inequality applied to the convex function $\zeta(\cdot,\, y)$ for all $y \geq 0$ and (3.50); (3.52) (with Π as defined in (3.9)) is proved in Appendix B.3; and (3.53) is obtained by a rearrangement of terms. Next, we use the following identity to simplify (3.53). For sequences of real numbers $\{x_n\}_{n=1}^N$ and $\{y_n\}_{n=1}^N$,

$$\sum_{n=1}^{N} x_n y_n \;=\; \sum_{n=1}^{N-1}(x_n - x_{n+1})\sum_{k=1}^{n} y_k + x_N \sum_{k=1}^{N} y_k. \tag{3.54}$$

Setting $x_n = \mu_{\Pi(n)}$ and $y_n = \psi_{n,\,\Pi}$, $n = 1,\, \cdots,\, N$, by (3.54), we get

$$\begin{aligned}\sum_{n=1}^{N} \mu_{\Pi(n)}\psi_{n,\,\Pi} \;&=\; \sum_{n=1}^{N-1}(\mu_{\Pi(n)} - \mu_{\Pi(n+1)})\sum_{k=1}^{n} \psi_{k,\,\Pi} + \mu_{\Pi(N)}\sum_{k=1}^{n} \psi_{k,\,\Pi} \\ &=\; \sum_{n=1}^{N} \nu_n \sum_{m=1}^{M} \zeta\left(\sum_{k=1}^{n} s_{\Pi(k),m} A_{\Pi(k)},\, \lambda_{0,m}\right),\end{aligned} \tag{3.55}$$

by (3.10), (3.46). Similarly, by setting $x_n = \mu_{\Pi(n)}$ and $y_n = s_{\Pi(n),\,m}A_{\Pi(n)}$, $n = 1,\, \cdots,\, N$, by (3.10), (3.54), we get

$$\sum_{n=1}^{N} \mu_{\Pi(n)} s_{\Pi(n),\,m} A_{\Pi(n)} \;=\; \sum_{n=1}^{N} \nu_n \sum_{k=1}^{N} s_{\Pi(k),\,m} A_{\Pi(k)}. \tag{3.56}$$

Summarizing collectively (3.29), (3.48), (3.53), (3.55), (3.56),

$$\begin{aligned} R \;\lessapprox\; \max_{\substack{0 \leq \mu_n \leq \sigma_n \\ n=1,\,\cdots,\,N}} \sum_{m=1}^{M} \Bigg[& \sum_{n=1}^{N} \nu_n \zeta\left(\sum_{k=1}^{n} s_{\Pi(k),\,m} A_{\Pi(k)},\; \lambda_{0,\,m}\right) \\ & -\zeta\left(\sum_{n=1}^{N} \nu_n \sum_{k=1}^{n} s_{\Pi(k),\,m} A_{\Pi(k)},\; \lambda_{0,\,m}\right)\Bigg]. \end{aligned} \tag{3.57}$$

This concludes the proof of the converse part of Theorem 4.

Achievability part: Fix $L \in \mathbb{Z}^+$ and set $\Delta = T/L$. Divide the time interval $[0,\, T]$ into L equal subintervals, each of duration Δ. Now, consider the situation in which for each message $w \in \mathcal{W}$, the channel input waveform at the n^{th} transmit aperture, $\{x_n(w,\, t),\, 0 \leq t \leq T\}$, is restricted to be $\{0,\, A_n\}$–valued and piecewise constant in each of the L time slots of duration Δ. Define

$$\tilde{x}_n(w,\, l) \;\triangleq\; \begin{cases} 0, & \text{if } x_n(w,\, t) = 0, \\ 1, & \text{if } x_n(w,\, t) = A_n, \\ & t \in [(l-1)\Delta,\, l\Delta), \quad l = 1, \cdots, L. \end{cases} \tag{3.58}$$

Note that the condition (3.2) requires that

$$\frac{1}{L}\sum_{l=1}^{L} \tilde{x}_n(w,\, l) \;\le\; \sigma_n, \quad w \in \mathcal{W}, \quad n = 1, \cdots, N. \tag{3.59}$$

Next, consider a decoder $\phi : \{0,\, 1\}^{ML} \to \mathcal{W}$ based on restricted observations over the L time slots, comprising

$$\tilde{Y}_m(l) \;=\; 1(Y_m(l\Delta) - Y_m((l-1)\Delta) = 1), \quad l = 1, \cdots, L, \tag{3.60}$$

(with $Y_m(0) = 0$) at the m^{th} receive aperture, $m = 1, \cdots, M$. The largest achievable rate of restricted $(W,\, T)$–codes as above – and, hence, the capacity C_{ind} – is clearly no smaller than $\frac{C(L)}{T}$, where $C(L)$ is the capacity of a $(L-)$ block discrete memoryless channel (in nats per block channel use) with input alphabet $\tilde{\mathcal{X}}^{NL} = \{0,\, 1\}^{NL}$; output alphabet $\tilde{\mathcal{Y}}^{ML} = \{0,\, 1\}^{ML}$; transition probability mass function (pmf)

$$W^{(L)}(\tilde{\mathbf{y}}^{ML} | \tilde{\mathbf{x}}^{NL}) \;=\; \prod_{l=1}^{L}\prod_{m=1}^{M} W^{(m)}_{\tilde{Y}_m | \tilde{\mathbf{X}}^N}(\tilde{y}_m(l) | \tilde{\mathbf{x}}^N(l)), \qquad \tilde{\mathbf{x}}^{NL} \in \{0,\, 1\}^{NL}, \quad \tilde{\mathbf{y}}^{ML} \in \{0,\, 1\}^{ML}, \tag{3.61}$$

where $\tilde{X}_n$, $n = 1, \cdots, N$ and $\tilde{Y}_m$, $m = 1, \cdots, M$ are $\tilde{\mathcal{X}}$– and $\tilde{\mathcal{Y}}$–valued rvs respectively, and the channel transition pmf $W^{(m)}_{\tilde{Y}_m | \tilde{\mathbf{X}}^N}(\cdot|\cdot)$ (corresponding to the m^{th} receive aperture) is given by

$$\begin{aligned} W^{(m)}_{\tilde{Y}_m | \tilde{\mathbf{X}}^N}(1 | \tilde{\mathbf{x}}^N) &= w_m(\tilde{\mathbf{x}}^N)\tfrac{T}{L}\exp\left(-w_m(\tilde{\mathbf{x}}^N)\tfrac{T}{L}\right) \;\triangleq\; \omega_m(\tilde{\mathbf{x}}^N,\, L), \\ W^{(m)}_{\tilde{Y}_m | \tilde{\mathbf{X}}^N}(0 | \tilde{\mathbf{x}}^N) &= 1 - \omega_m(\tilde{\mathbf{x}}^N,\, L), \end{aligned} \tag{3.62}$$

with

$$w_m(\tilde{\mathbf{x}}^N) \quad \triangleq \quad \sum_{n=1}^{N} s_{n,m}\tilde{x}_n A_n + \lambda_{0,m}, \quad \tilde{\mathbf{x}}^N \in \{0, 1\}^N, \quad m = 1, \cdots, M, \tag{3.63}$$

and under constraint (3.59). From [14], it is readily obtained that

$$C(L) \quad = \quad \max_{\substack{P_{\tilde{\mathbf{X}}^{NL}}: \sum_{l=1}^{L} \mathbf{E}[\tilde{X}_n(l)] \leq L\sigma_n \\ n=1, \cdots, N}} I(\tilde{\mathbf{X}}^{NL} \wedge \tilde{\mathbf{Y}}^{ML}), \tag{3.64}$$

where the joint pmf $P_{\tilde{\mathbf{X}}^{NL}, \tilde{\mathbf{Y}}^{ML}}$ is given by

$$\begin{aligned} P_{\tilde{\mathbf{X}}^{NL}, \tilde{\mathbf{Y}}^{ML}}(\tilde{\mathbf{x}}^{NL}, \tilde{\mathbf{y}}^{ML}) \quad &= \quad P_{\tilde{\mathbf{X}}^{NL}}(\tilde{\mathbf{x}}^{NL}) W^{(L)}(\tilde{\mathbf{y}}^{ML}|\tilde{\mathbf{x}}^{NL}), \\ &\qquad \tilde{\mathbf{x}}^{NL} \in \{0, 1\}^{NL}, \quad \tilde{\mathbf{y}}^{ML} \in \{0, 1\}^{ML}. \end{aligned} \tag{3.65}$$

By a standard argument which uses (3.61), the maximum in (3.64) is achieved by

$$P_{\tilde{\mathbf{X}}^{NL}}(\tilde{\mathbf{x}}^{NL}) \quad = \quad \prod_{l=1}^{L} P_{\tilde{\mathbf{X}}^{N}(l)}(\tilde{\mathbf{x}}^N(l)), \quad \tilde{\mathbf{x}}^{NL} \in \{0, 1\}^{NL}, \tag{3.66}$$

so that from (3.64),

$$C(L) \quad = \quad L \max_{\substack{P_{\tilde{\mathbf{X}}^{N}}: \mathbf{E}[\tilde{X}_n] \leq \sigma_n \\ n=1, \cdots, N}} \sum_{m=1}^{M} I(\tilde{\mathbf{X}}^N, \tilde{Y}_m), \tag{3.67}$$

where $P_{\tilde{\mathbf{X}}^N, \tilde{Y}_m}$ is given by

$$\begin{aligned} P_{\tilde{\mathbf{X}}^N, \tilde{Y}_m}(\tilde{\mathbf{x}}^N, \tilde{y}) \quad &= \quad P_{\tilde{\mathbf{X}}^N}(\tilde{\mathbf{x}}^N) W^{(m)}_{\tilde{Y}_m|\tilde{\mathbf{X}}^N}(\tilde{y}|\tilde{\mathbf{x}}^N), \\ &\qquad \tilde{\mathbf{x}}^N \in \{0, 1\}^N, \quad \tilde{y} \in \{0, 1\}. \end{aligned} \tag{3.68}$$

Let the joint pmf of $\tilde{\mathbf{X}}^N$ be represented by the vector $\mathbf{p} = \{p_i\}_{i=0}^{2^N-1}$, where

$$p_i \quad = \quad \Pr\{\tilde{X}_n = b_n(i), \, n = 1, \cdots, N\}, \quad i = 0, \cdots, 2^N - 1, \tag{3.69}$$

where $b_n(i)$ is the n^{th} bit in the binary representation of i. Clearly, $\mathbf{p} = \{p_i\}_{i=0}^{2^N-1}$ satisfies the constraints

$$\begin{aligned} p_i &\geq 0, \quad i = 0, \cdots, 2^N - 1, \\ \textstyle\sum_{i=0}^{2^N-1} p_i &= 1, \\ \textstyle\sum_{i:b_n(i)=1} p_i &\leq \sigma_n, \quad n = 1, \cdots, N, \end{aligned} \tag{3.70}$$

where the last inequality follows from the constraints $\mathbb{E}[\tilde{X}_n] \leq \sigma_n$, $n = 1, \cdots, N$. From (3.62) and (3.69), it follows that for $m = 1, \cdots, M$,

$$\begin{aligned} \Pr\{\tilde{Y}_m = 1\} &= \textstyle\sum_{i=0}^{2^N-1} p_i \omega_m \left(\mathbf{b}^N(i), L\right), \\ H\left(\tilde{Y}_m | \tilde{\mathbf{X}}^N\right) &= \textstyle\sum_{i=0}^{2^N-1} p_i h_b \left(\omega_m \left(\mathbf{b}^N(i), L\right)\right), \\ H(\tilde{Y}_m) &= h_b \left(\textstyle\sum_{i=0}^{2^N-1} p_i \omega_m \left(\mathbf{b}^N(i), L\right)\right). \end{aligned} \tag{3.71}$$

Therefore, by (3.67), (3.71), it follows that

$$C(L) = \max_{\mathbf{p}} L\beta_L(\mathbf{p}), \tag{3.72}$$

with

$$\beta_L(\mathbf{p}) \triangleq \sum_{m=1}^{M} \left[h_b \left(\sum_{i=0}^{2^N-1} p_i \omega_m \left(\mathbf{b}^N(i)\right) \right) - \sum_{i=0}^{2^N-1} p_i h_b \left(\omega_m \left(\mathbf{b}^N(i)\right)\right) \right], \tag{3.73}$$

where $\mathbf{p} = \{p_i\}_{i=0}^{2^N-1}$ satisfies (3.70). Since $L \in \mathbb{Z}^+$ was arbitrary, we have

$$\begin{aligned} C_{ind} &\geq \lim_{L\to\infty} \frac{C(L)}{T} \\ &\geq \max_{\mathbf{p}} \lim_{L\to\infty} \frac{\beta_L(\mathbf{p})}{T/L} \qquad (3.74) \\ &= \max_{\mathbf{p}} \sum_{m=1}^{M} \left[\sum_{i=0}^{2^N-1} p_i \zeta_m \left(\sum_{n=1}^{N} s_{n,m} b_n(i) A_n, \lambda_{0,m} \right) \right. \\ &\qquad \left. - \zeta_m \left(\sum_{n=1}^{N} s_{n,m} A_n \sum_{i=0}^{2^N-1} p_i b_n(i), \lambda_{0,m} \right) \right], \qquad (3.75) \end{aligned}$$

with $\mathbf{p} = \{p_i\}_{i=0}^{2^N-1}$ satisfying (3.70), where (3.74) is by (3.72); and (3.75) is established in Appendix B.4. Setting

$$\begin{aligned}\mu_n &= \sum_{i=0}^{2^N-1} p_i b_n(i) \\ &= \sum_{i:b_n(i)=1} p_i, \quad n = 1, \cdots, N,\end{aligned} \tag{3.76}$$

by (3.70) and (3.75), it follows that

$$C_{ind} \geq \max_{\substack{0 \leq \mu_n \leq \sigma_n \\ n=1,\cdots,N}} \left[h(\boldsymbol{\mu}^N) - \sum_{m=1}^{M} \zeta_m \left(\sum_{n=1}^{N} s_{n,m} \mu_n A_n, \, \lambda_{0,m} \right) \right], \tag{3.77}$$

where

$$h(\boldsymbol{\mu}^N) = \max_{\mathbf{p}} \sum_{i=0}^{2^N-1} p_i \sum_{m=1}^{M} \zeta_m \left(\sum_{n=1}^{N} s_{n,m} b_n(i) A_n, \, \lambda_{0,m} \right), \tag{3.78}$$

with $\mathbf{p} = \{p_i\}_{i=0}^{2^N-1}$ satisfying the constraints

$$\begin{aligned} p_i &\geq 0, \quad i = 0, \cdots, 2^N - 1, \\ \textstyle\sum_{i=0}^{2^N-1} p_i &= 1, \\ \textstyle\sum_{i:b_n(i)=1} p_i &= \mu_n, \quad n = 1, \cdots, N. \end{aligned} \tag{3.79}$$

Comparing (3.42) and (3.78), we immediately see that the optimum $\mathbf{p}^* = \{p_i^*\}_{i=0}^{2^N-1}$ that obtains the maximum in (3.78) is given by

$$p_i^* = \begin{cases} \mu_{\Pi(k)} - \mu_{\Pi(k+1)}, & \text{if } i = \sum_{n=1}^{k} 2^{\Pi(n)-1}, \quad k = 1, \cdots, N-1, \\ \mu_{\Pi(N)}, & \text{if } i = 2^N - 1, \\ 1 - \mu_{\Pi(1)}, & \text{if } i = 0, \\ 0, & \text{otherwise}, \end{cases} \tag{3.80}$$

where $\Pi(\cdot)$ is as defined in (3.9), and the corresponding largest value is

$$h(\boldsymbol{\mu}^N) = \sum_{n=1}^{N} \mu_{\Pi(n)} \psi_{n,\Pi}, \tag{3.81}$$

with $\psi_{n,\Pi}$ as defined in (3.46). Summarizing collectively (3.74), (3.77), (3.81), we get

$$\begin{aligned} C_{ind} &\geq \max_{\substack{0\leq\mu_n\leq\sigma_n \\ n=1,\cdots,N}} \left[\sum_{n=1}^{N} \mu_{\Pi(n)}\psi_{n,\Pi} - \sum_{m=1}^{M} \zeta\left(\sum_{n=1}^{N} s_{n,m}\mu_n A_n,\ \lambda_{0,m} \right) \right] \\ &= \max_{\substack{0\leq\mu_n\leq\sigma_n \\ n=1,\cdots,N}} \sum_{m=1}^{M} \left[\sum_{n=1}^{N} \nu_n \zeta\left(\sum_{k=1}^{n} s_{\Pi(k),m} A_{\Pi(k)},\ \lambda_{0,m} \right) \right. \\ &\qquad \left. -\zeta\left(\sum_{n=1}^{N} \nu_n \sum_{k=1}^{n} s_{\Pi(k),m} A_{\Pi(k)},\ \lambda_{0,m} \right) \right], \end{aligned} \tag{3.82}$$

by (3.55), (3.56). This concludes the proof of the achievability part. □

Proof of Corollary 5:

The proof follows the same arguments as the proof of Theorem 4, taking into account the difference between individual average power constraints at each transmit aperture (see (3.4)) and an average sum power constraint (see (3.5)). In the following, we identify only the differences.

Consider first the proof of the converse part. Note that instead of (3.27), from (3.5), it follows that

$$\begin{aligned} 0 \ \leq\ X_n(t) &\leq A_n, \quad 0\leq t\leq T, \quad n=1,\cdots,N, \\ \textstyle\sum_{n=1}^{N} \frac{1}{T}\int_0^T \mathbb{E}[X_n(\tau)]d\tau &\leq \sigma \textstyle\sum_{n=1}^{N} A_n, \end{aligned} \tag{3.83}$$

whence instead of (3.47), we now get

$$\begin{aligned} 0 \ \leq\ \chi_n(\tau) &\leq 1, \quad 0\leq \tau\leq T, \quad n=1,\cdots,N, \\ \textstyle\sum_{n=1}^{N} \frac{1}{T}\int_0^T \chi_n(\tau)A_n d\tau &\leq \sigma \textstyle\sum_{n=1}^{N} A_n, \end{aligned} \tag{3.84}$$

whence by (3.50), it follows that

$$\begin{aligned} 0 \ \leq\ \mu_n &\leq 1, \quad n=1,\cdots,N, \\ \textstyle\sum_{n=1}^{N} \mu_n A_n &\leq \sigma \textstyle\sum_{n=1}^{N} A_n. \end{aligned} \tag{3.85}$$

By (3.85), instead of (3.57), we now get

$$R \lessapprox \max_{\substack{0 \leq \mu_n \leq 1,\, n=1,\, \cdots,\, N \\ \sum_{n=1}^{N} \mu_n A_n \leq \sigma \sum_{n=1}^{N} A_n}} \sum_{m=1}^{M} \left[\sum_{n=1}^{N} \nu_n \zeta \left(\sum_{k=1}^{n} s_{\Pi(k),\, m} A_{\Pi(k)},\ \lambda_{0,\, m} \right) \right.$$
$$\left. - \zeta \left(\sum_{n=1}^{N} \nu_n \sum_{k=1}^{n} s_{\Pi(k),\, m} A_{\Pi(k)},\ \lambda_{0,\, m} \right) \right], \quad (3.86)$$

which concludes the proof of the converse part.

Consider next the achievability part. Instead of (3.59), by (3.5), we now have

$$\frac{1}{L} \sum_{l=1}^{L} \sum_{n=1}^{N} \tilde{x}_n(w,\, l) A_n \ \leq\ \sigma \sum_{n=1}^{N} A_n, \qquad w \in \mathcal{W}. \quad (3.87)$$

Instead of (3.64), we now have

$$C(L) \ =\ \max_{P_{\tilde{\mathbf{X}}^{NL}} : \sum_{n=1}^{N} \mathbf{E}[\tilde{\mathbf{X}}_n^L] A_n \leq L\sigma \sum_{n=1}^{N} A_n} I(\tilde{\mathbf{X}}^{NL} \wedge \tilde{\mathbf{Y}}^{ML}), \quad (3.88)$$

which leads to

$$C(L) \ =\ L \max_{P_{\tilde{\mathbf{X}}^N} : \sum_{n=1}^{N} \mathbf{E}[\tilde{X}_n] A_n \leq \sigma \sum_{n=1}^{N} A_n} \sum_{m=1}^{M} I(\tilde{\mathbf{X}}^N,\, \tilde{Y}_m), \quad (3.89)$$

instead of (3.67). With $\mathbf{p} = \{p_i\}_{i=0}^{2^N-1}$ as defined in (3.69), the last inequality in (3.70) is now replaced by

$$\sum_{n=1}^{N} A_n \sum_{i : b_n(i)=1} p_i \ \leq\ \sigma \sum_{n=1}^{N} A_n, \quad (3.90)$$

which implies the following constraints on $\{\mu_n\}_{n=1}^N$:

$$\begin{aligned} 0 \ \leq\ \mu_n \ &\leq\ 1, \quad n = 1,\, \cdots,\, N, \\ \textstyle\sum_{n=1}^{N} \mu_n A_n \ &\leq\ \sigma \textstyle\sum_{n=1}^{N} A_n. \end{aligned} \quad (3.91)$$

The proof of the achievability part is completed by replacing the constraints on $\{\mu_n\}_{n=1}^N$ in (3.82) by (3.91). □

Proof of Theorem 5:

The proof is an exercise in convex optimization. Without loss of generality, assume that $\sigma_1 \geq \sigma_2$. The case of $\sigma_1 \leq \sigma_2$ follows by symmetry.

For the special case of $N = 2$ transmit apertures, from (3.7), we have

$$C_{ind} = \max_{\substack{0 \leq \mu_1 \leq \sigma_1 \\ 0 \leq \mu_2 \leq \sigma_2}} I(\mu_1,\, \mu_2,\, \mathbf{s}), \tag{3.92}$$

with

$$I(\mu_1,\, \mu_2,\, \mathbf{s}) = \begin{cases} I_{1>2}(\mu_1 - \mu_2,\, \mu_2,\, \mathbf{s}), & \text{if } \mu_1 \geq \mu_2, \\ I_{1<2}(\mu_2 - \mu_1,\, \mu_1,\, \mathbf{s}), & \text{otherwise}, \end{cases} \tag{3.93}$$

where for $x_1 \geq 0$, $x_2 \geq 0$, $\mathbf{s} \in (\mathbb{R}_0^+)^{2\times M}$, we have defined

$$\begin{aligned} I_{1>2}(x_1,\, x_2,\, \mathbf{s}) &= \textstyle\sum_{m=1}^{M} \Big[x_1 \zeta(s_{1,\,m} A_1,\, \lambda_{0,\,m}) + x_2 \zeta\left(\sum_{n=1}^{2} s_{n,\,m} A_n,\, \lambda_{0,\,m}\right) \\ &\quad -\zeta\left(x_1 s_{1,\,m} A_1 + x_2 \sum_{n=1}^{2} s_{n,\,m} A_n,\, \lambda_{0,\,m}\right)\Big], \\ I_{1<2}(x_1,\, x_2, \mathbf{s}) &= \textstyle\sum_{m=1}^{M} \Big[x_1 \zeta(s_{2,\,m} A_2,\, \lambda_{0,\,m}) + x_2 \zeta\left(\sum_{n=1}^{2} s_{n,\,m} A_n,\, \lambda_{0,\,m}\right) \\ &\quad -\zeta\left(x_1 s_{2,\,m} A_2 + x_2 \sum_{n=1}^{2} s_{n,\,m} A_n,\, \lambda_{0,\,m}\right)\Big]. \end{aligned} \tag{3.94}$$

By (2.5), (3.12), it follows that

$$\begin{aligned} \tfrac{\partial I_{1>2}}{\partial x_1} &= \textstyle\sum_{m=1}^{M} \lambda_{0,m} b_{1,m} \log\left(\frac{1+\alpha(b_{1,m}) b_{1,m}}{1+x_1 b_{1,m} + x_2 B_m}\right), \\ \tfrac{\partial I_{1<2}}{\partial x_1} &= \textstyle\sum_{m=1}^{M} \lambda_{0,m} b_{2,m} \log\left(\frac{1+\alpha(b_{2,m}) b_{2,m}}{1+x_1 b_{2,m} + x_2 B_m}\right), \\ \tfrac{\partial I_{1>2}}{\partial x_2} = \tfrac{\partial I_{1<2}}{\partial x_2} &= \textstyle\sum_{m=1}^{M} \lambda_{0,m} B_m \log\left(\frac{1+\alpha(B_m) B_m}{1+x_1 b_{1,m} + x_2 B_m}\right). \end{aligned} \tag{3.95}$$

Observe that $I(x_1,\, x_2,\, \mathbf{s})$ is nondifferentiable for any $\mathbf{s} \in (\mathbb{R}_0^+)^{2\times M}$ if $x_1 = x_2$. In order to determine the optimal pair (μ_1^*, μ_2^*) that achieves the maximum in (3.92), we divide the constraint set $\{0 \leq \mu_1 \leq \sigma_1,\ 0 \leq \mu_2 \leq \sigma_2\}$ into two half sets as follows. Setting

$$\begin{aligned} \mathcal{F}_{1>2} &= \{(x_1,\, x_2) : 0 \leq x_1 \leq \sigma_1,\, 0 \leq x_2 \leq \sigma_2,\, x_1 \geq x_2\}, \\ \mathcal{F}_{1<2} &= \{(x_1,\, x_2) : 0 \leq x_1 \leq \sigma_1,\, 0 \leq x_2 \leq \sigma_2,\, x_1 \leq x_2\}, \end{aligned} \tag{3.96}$$

by (3.92), we get

$$C_{ind} = \max\{C_{1>2},\, C_{1<2}\}, \tag{3.97}$$

where

$$C_{1>2} = \max_{(\mu_1,\,\mu_2)\in\mathcal{F}_{1>2}} I_{1>2}(\mu_1-\mu_2,\,\mu_2,\,\mathbf{s}), \tag{3.98}$$

$$C_{1<2} = \max_{(\mu_1,\,\mu_2)\in\mathcal{F}_{1<2}} I_{1<2}(\mu_2-\mu_1,\,\mu_1,\,\mathbf{s}). \tag{3.99}$$

We apply the transformation $\nu_1 = \mu_1 - \mu_2$ (resp. $\nu_1 = \mu_2 - \mu_1$) and $\nu_2 = \mu_2$ (resp. $\nu_2 = \mu_1$) in order to compute $C_{1>2}$ (resp. $C_{1<2}$). By (3.96), (3.98), (3.99),

$$C_{1>2} = \max_{(\nu_1,\,\nu_2)\in\mathcal{G}_{1>2}} I_{1>2}(\nu_1,\,\nu_2,\,\mathbf{s}), \tag{3.100}$$

$$C_{1<2} = \max_{(\nu_1,\,\nu_2)\in\mathcal{G}_{1<2}} I_{1<2}(\nu_1,\,\nu_2,\,\mathbf{s}), \tag{3.101}$$

where

$$\begin{aligned} \mathcal{G}_{1>2} &= \{(x_1,\,x_2) \,:\, x_1 \geq 0,\, 0 \leq x_2 \leq \sigma_2,\, x_1 + x_2 \leq \sigma_1\}, \\ \mathcal{G}_{1<2} &= \{(x_1,\,x_2) \,:\, x_1 \geq 0,\, x_2 \geq 0,\, x_1 + x_2 \leq \sigma_2\}. \end{aligned} \tag{3.102}$$

We first compute $C_{1<2}$ using the Lagrangian

$$J_{1<2}(\nu_1,\,\nu_2) = I_{1<2}(\nu_1,\,\nu_2,\,\mathbf{s}) + \eta_1\nu_1 + \eta_2\nu_2 + \eta_3(\sigma_2 - \nu_1 - \nu_2), \tag{3.103}$$

where $\eta_k \geq 0$, $k = 1,\,2,\,3$ are Lagrange multipliers to be determined. The optimal pair $(\nu_1^*,\,\nu_2^*)$ that achieves the maximum in (3.101) satisfies the following Kuhn-

Tucker conditions (cf. e.g., [35], p. 233):

$$\begin{aligned}
\left.\frac{\partial I_{1<2}}{\partial \nu_1}\right|_{(\nu_1^*,\nu_2^*)} + \eta_1 - \eta_3 &= 0, \\
\left.\frac{\partial I_{1<2}}{\partial \nu_2}\right|_{(\nu_1^*,\nu_2^*)} + \eta_2 - \eta_3 &= 0, \\
\eta_1 \nu_1^* &= 0, \qquad (3.104) \\
\eta_2 \nu_2^* &= 0, \\
\eta_3(\sigma_2 - \nu_1^* - \nu_2^*) &= 0,
\end{aligned}$$

whose solution yields

$$\begin{aligned}
\nu_1^* &= 0, \\
\nu_2^* &= \min\{\sigma_2, \bar{\rho}\},
\end{aligned} \qquad (3.105)$$

where $\bar{\rho}$ is as defined in (3.14), so that by (3.101),

$$C_{1<2} = I_{1<2}(0, \min\{\sigma_2, \bar{\rho}\}, \mathbf{s}). \qquad (3.106)$$

We now compute $C_{1>2}$ using the Lagrangian

$$\begin{aligned}
J_{1>2}(\kappa_1, \kappa_2) &= I_{1>2}(\kappa_1, \kappa_2, \mathbf{s}) + \eta_1\kappa_1 + \eta_2\kappa_2 \\
&\quad + \eta_3(\sigma_2 - \kappa_2) + \eta_4(\sigma_1 - \kappa_1 - \kappa_2),
\end{aligned} \qquad (3.107)$$

where $\eta_k \geq 0$, $k = 1, \cdots, 4$ are Lagrange multipliers to be determined. The optimal pair (κ_1^*, κ_2^*) that achieves the maximum in (3.100) satisfies the following Kuhn-

Tucker conditions:

$$\begin{aligned}
\left.\frac{\partial I_{1<2}}{\partial \kappa_1}\right|_{(\kappa_1^*,\kappa_2^*)} + \eta_1 - \eta_4 &= 0, \\
\left.\frac{\partial I_{1<2}}{\partial \kappa_2}\right|_{(\kappa_1^*,\kappa_2^*)} + \eta_2 - \eta_3 - \eta_4 &= 0, \\
\eta_1 \kappa_1^* &= 0, \\
\eta_2 \kappa_2^* &= 0, \\
\eta_3(\sigma_2 - \kappa_2^*) &= 0, \\
\eta_4(\sigma_1 - \kappa_1^* - \kappa_2^*) &= 0,
\end{aligned} \tag{3.108}$$

whose solution yields

$$\begin{aligned}
\kappa_1^* &= \begin{cases} 0, & \text{if } \sigma_2 \geq \rho_1, \\ \min\{\sigma_1,\, \beta_1(\sigma_2)\} - \sigma_2, & \text{otherwise,} \end{cases} \\
\kappa_2^* &= \min\{\sigma_2,\, \bar{\rho}\},
\end{aligned} \tag{3.109}$$

where ρ_1, $\bar{\rho}$ and $\beta_1(\cdot)$ are as defined in (3.13), (3.14) and (3.15) respectively, so that by (3.100), we get

$$C_{1>2} = \begin{cases} I_{1>2}(0,\, \min\{\sigma_2,\, \bar{\rho}\},\, \mathbf{s}), & \text{if } \sigma_2 \geq \rho_1, \\ I_{1>2}(\min\{\sigma_1,\, \beta_1(\sigma_2)\} - \sigma_2,\, \min\{\sigma_2,\, \bar{\rho}\}, \mathbf{s}), & \text{otherwise.} \end{cases} \tag{3.110}$$

Comparing (3.106) and (3.110), we see that if $\sigma_1 \geq \sigma_2$, then $C_{1>2} \geq C_{1<2}$. Therefore, $C_{ind} = C_{1>2}$, and by the inverse transformation, the optimal pair $\mu_1^* = \kappa_1^* + \kappa_2^*$ and $\mu_2^* = \kappa_2^*$ satisfies one of the conditions (1), (2), (4), (6) of Theorem 5. Conditions (1), (3), (5), (6) corresponding to the case $\sigma_1 \leq \sigma_2$ can be verified using similar arguments. This concludes the proof of Theorem 5. □

Proof of Theorem 6:

First, note that standard Lagrangian techniques (as used in the proof of Theorem 5), can be used to show that at optimality,

$$\mu_1^* A_1 + \mu_2^* A_2 \quad = \quad \sigma(A_1 + A_2) \tag{3.111}$$

holds iff $\sigma \leq \bar{\rho}$, where $\bar{\rho}$ is as defined in (3.14). If $\sigma > \bar{\rho}$, then the optimal pair $(\mu_1^*,\, \mu_2^*)$ that achieves the maximum in (3.11) is given by $\mu_1^* = \mu_2^* = \bar{\rho}$, the "unconstrained" optimum pair.

Next, we consider the case $\sigma < \bar{\rho}$, when (3.111) holds at optimality. Substituting $\mu_1 = x$ and $\mu_2 = (1+a)\sigma - ax$, with $a = \frac{A_1}{A_2}$, from (3.11), we get

$$C_{sum} \quad = \quad \max_{0 \leq x \leq (1+1/a)\sigma} I_0(x), \tag{3.112}$$

where

$$I_0(x) \quad = \quad \begin{cases} \sum_{m=1}^{M} [(1+a)(\sigma - x)\zeta(s_{2,\,m}A_2,\, \lambda_{0,\,m}) + x\zeta(\sum_{n=1}^{2} s_{n,\,m}A_n,\, \lambda_{0,\,m}) \\ -\zeta(xs_{1,\,m}A_1 + ((1+a)\sigma - ax)s_{2,\,m}A_2,\, \lambda_{0,\,m})], \quad \text{if} \quad x \leq \sigma, \\ \sum_{m=1}^{M} [(1+a)(x - \sigma)\zeta(s_{1,\,m}A_1,\, \lambda_{0,\,m}) \\ +((1+a)\sigma - ax)\zeta(\sum_{n=1}^{2} s_{n,\,m}A_n,\, \lambda_{0,\,m}) \\ -\zeta(xs_{1,\,m}A_1 + ((1+a)\sigma - ax)s_{2,\,m}A_2,\, \lambda_{0,\,m})], \quad \text{otherwise}, \end{cases} \tag{3.113}$$

and $\mu_1^* = \arg\max_{0 \leq x \leq (1+1/a)\sigma} I_0(x)$.

Observe that $I_0(\cdot)$ is a continuous, concave, piecewise differentiable function on $[0,\, (1+1/a)\sigma]$, with a nondifferentiable point at σ. From (3.113), using (2.5) and (3.12), it can be verified that

$$I_0^{'}(x) \quad = \quad \frac{dI_0(x)}{dx} \quad = \quad \begin{cases} K_1 - d(x,\, \sigma), & \text{if} \quad x < \sigma, \\ K_2 - d(x,\, \sigma), & \text{if} \quad x > \sigma, \end{cases} \tag{3.114}$$

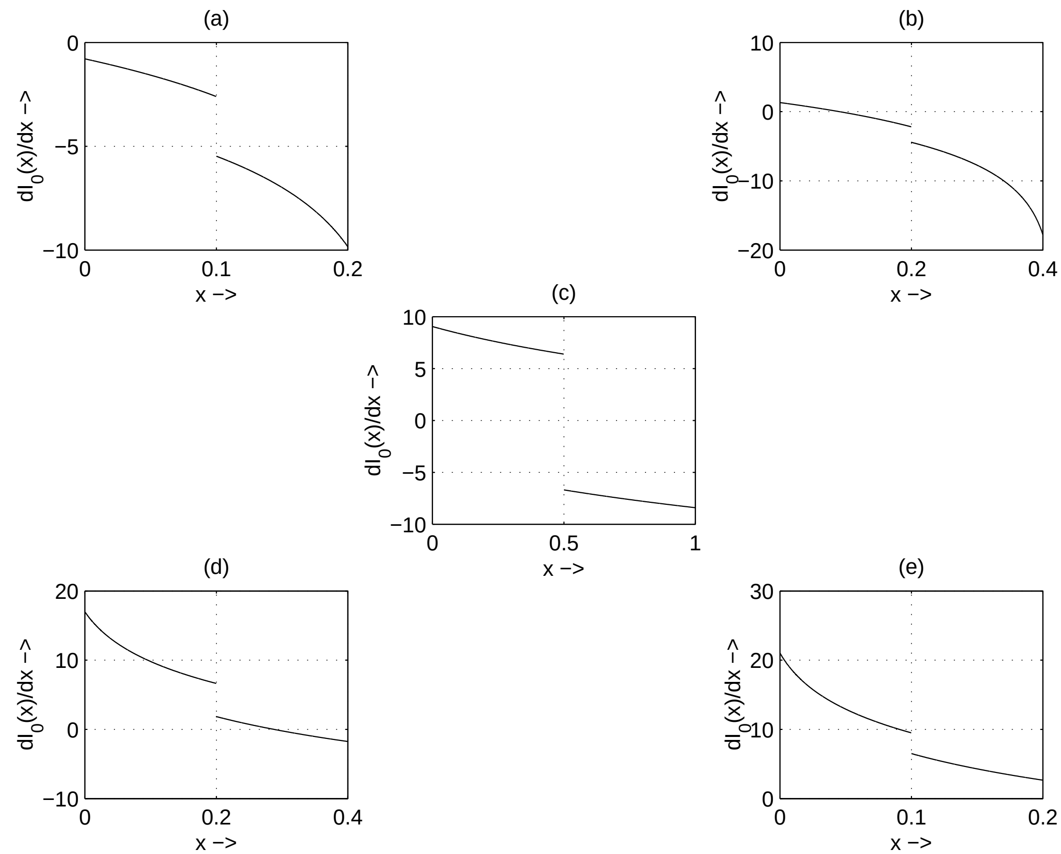

Figure 3.3: The possible variations of $I_0'(x)$ versus x.

(a) $I_0'(0) < 0$;

(b) $I_0'(x_0) = 0$ for some $0 \leq x_0 < \sigma$;

(c) $I_0'(\sigma^-) > 0$, $I_0'(\sigma^+) < 0$;

(d) $I_0'(x_0) = 0$ for some $\sigma < x_0 \leq (1 + 1/a)\sigma$;

(e) $I_0'((1 + 1/a)\sigma) > 0$.

where K_1, K_2 and $d(\cdot,\cdot)$ are as defined in (3.17) and (3.18) respectively. Keeping in mind that $K_1 \geq K_2$ and $d(\cdot, y)$ is non-decreasing for any $y \in \mathbb{R}$, we conclude that (see Figure 3.3)

1. if $I_0'(0) < 0$, i.e., if $\gamma_1(\sigma) < 0$, then $\mu_1^* = 0$;

2. if $I_0'(x_0) = 0$ for some $0 \leq x_0 < \sigma$, i.e., if $0 \leq \gamma_1(\sigma) < \sigma$, then $\mu_1^* = x_0 = \gamma_1(\sigma)$;

3. if $\lim_{x\to\sigma^-} I_0'(x) > 0$ and $\lim_{x\to\sigma^+} I_0'(x) < 0$, then $\mu_1^* = \sigma$;

4. if $I_0'(x_0) = 0$ for some $\sigma < x_0 \leq (1+1/a)\sigma$, i.e., if $\sigma < \gamma_2(\sigma) \leq (1+1/a)\sigma$, then $\mu_1^* = x_0 = \gamma_2(\sigma)$;

5. if $I_0'((1+1/a)\sigma) > 0$, i.e., if $\gamma_2(\sigma) > (1+1/a)\sigma$, then $\mu_1^* = (1+1/a)\sigma$.

The statement of Theorem 6 is a rearrangement of these observations. □

Proof of Corollary 6:

For $0 \leq \sigma \leq \bar{\rho}$, define $D_1(\sigma) \triangleq \lim_{x\to\sigma^-} I_0'(x)$, and $D_2(\sigma) \triangleq \lim_{x\to\sigma^+} I_0'(x)$. By (3.114),

$$\begin{aligned} D_1(\sigma) &= K_1 - d(\sigma, \sigma) \\ &= \sum_{m=1}^{M} \lambda_{0,m} \left[B_m \log\left(\frac{1+\bar{\rho}B_m}{1+\sigma B_m}\right) + (1+a)b_{2,m} \log\left(\frac{1+\sigma B_m}{1+\rho_2 B_m}\right)\right], \end{aligned} \tag{3.115}$$

and

$$\begin{aligned} D_2(\sigma) &= K_2 - d(\sigma,\sigma) \\ &= -\sum_{m=1}^{M} \lambda_{0,m} \left[aB_m \log\left(\frac{1+\bar{\rho}B_m}{1+\sigma B_m}\right) + (1+a)b_{1,m} \log\left(\frac{1+\sigma B_m}{1+\rho_1 B_m}\right)\right], \end{aligned} \tag{3.116}$$

where (3.115), (3.116) are by (3.17), (3.18). Note that $D_1(0) = K_1$, $D_2(0) = K_2$, and $D_1(\sigma) - D_2(\sigma) = K_1 - K_2$ for all $0 \leq \sigma \leq 1$. By (3.115), (3.116), and the fact that $\bar{\rho} \geq \max\{\rho_1, \rho_2\}$, it follows that $D_1(\bar{\rho}) \geq 0$ and $D_2(\bar{\rho}) \leq 0$.

Case 1: $K_1 < 0$: By the continuity of $D_1(\cdot)$ on $[0, \bar{\rho}]$, and the facts that $D_1(0) = K_1 < 0$, $D_1(\bar{\rho}) \geq 0$, we conclude that there exists $0 \leq \sigma_c \leq \bar{\rho}$ such that $D_1(\sigma_c) = 0$. If $\sigma \leq \sigma_c$, then $I_0'(x_0) = 0$ for some $0 \leq x_0 \leq \sigma$. By Theorem 6, it follows that $\mu_1^* = x_0 \leq \sigma$, and $\mu_2^* = (1+a)\sigma - a\mu_1^* \geq \sigma$. On the other hand, if $\sigma > \sigma_c$, then $D_1(\sigma) > 0$ and $D_2(\sigma) < 0$, so that by Theorem 6, it follows that $\mu_1^* = \mu_2^* = \sigma$. In summary, $\mu_1^* \leq \sigma \leq \mu_2^*$ for all $0 \leq \sigma \leq 1$.

Case 2: $K_2 > 0$: By the continuity of $D_2(\cdot)$ on $[0, \bar{\rho}]$, and the facts that $D_2(0) = K_2 > 0$, $D_2(\bar{\rho}) \leq 0$, we conclude that there exists $0 \leq \sigma_c \leq \bar{\rho}$ such that $D_2(\sigma_c) = 0$. If $\sigma \geq \sigma_c$, then $I_0'(x_0) = 0$ for some $\sigma \leq x_0 \leq (1 + 1/a)\sigma$. By Theorem 6, it follows that $\mu_1^* = x_0 \geq \sigma$, and $\mu_2^* = (1+a)\sigma - a\mu_1^* \leq \sigma$. On the other hand, if $\sigma < \sigma_c$, then $D_1(\sigma) > 0$ and $D_2(\sigma) < 0$, so that by Theorem 6, it follows that $\mu_1^* = \mu_2^* = \sigma$. In summary, $\mu_1^* \geq \sigma \geq \mu_2^*$ for all $0 \leq \sigma \leq 1$.

Case 3: $K_1 \geq 0 \geq K_2$: For all $0 \leq \sigma \leq \bar{\rho}$, $D_1(\sigma) \geq 0$ and $D_2(\sigma) \leq 0$, so that $\mu_1^* = \mu_2^* = \sigma$. □

3.5 Numerical examples

In this section, we provide some illustrative examples to supplement the results described in this chapter. We determine the capacity of a 2×2 MIMO Poisson channel with constant channel fade. We consider symmetric and asymmetric peak

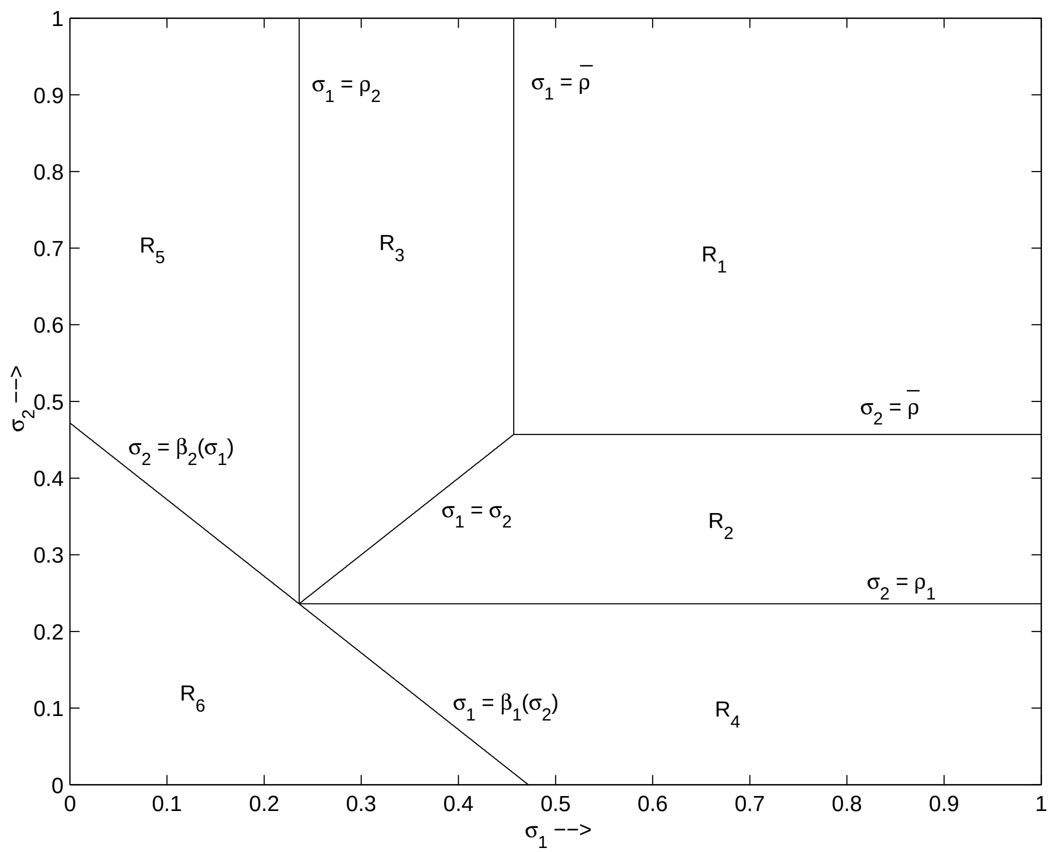

Figure 3.4: Decision region of optimal duty cycles for individual average power constraints σ_1, σ_2 for Example 3.1.

and average power constraints, as well as average sum power constraint on the transmitted signals.

Example 3.1: Consider a 2×2 MIMO Poisson channel with $A_1 = A_2 = 1.0$, $\lambda_{0,1} = \lambda_{0,2} = 1.0$, $\mathbf{s} = [1.0\ 0.1;\ 1.0\ 0.1]$. It can be verified that $\rho_1 = \rho_2 = 0.236$, $\bar{\rho} = 0.457$, $K_1 = 0.5324$, $K_2 = -0.5324$. The channel conditions experienced by the two transmit apertures are identical; therefore, it is not surprising that the decision region of the optimal duty cycles is symmetric when the transmit apertures are subject to individual average power constraints (see Figure 3.4). Furthermore,

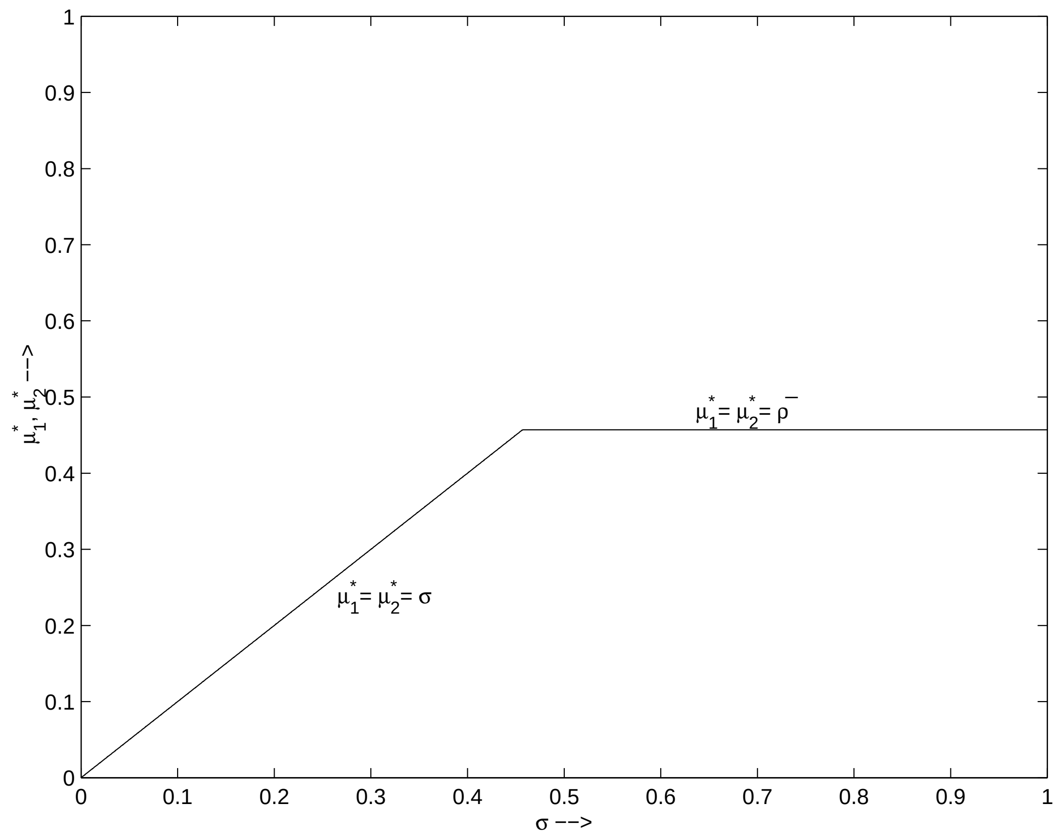

Figure 3.5: Optimal duty cycles for an average sum power constraint σ for Example 3.1.

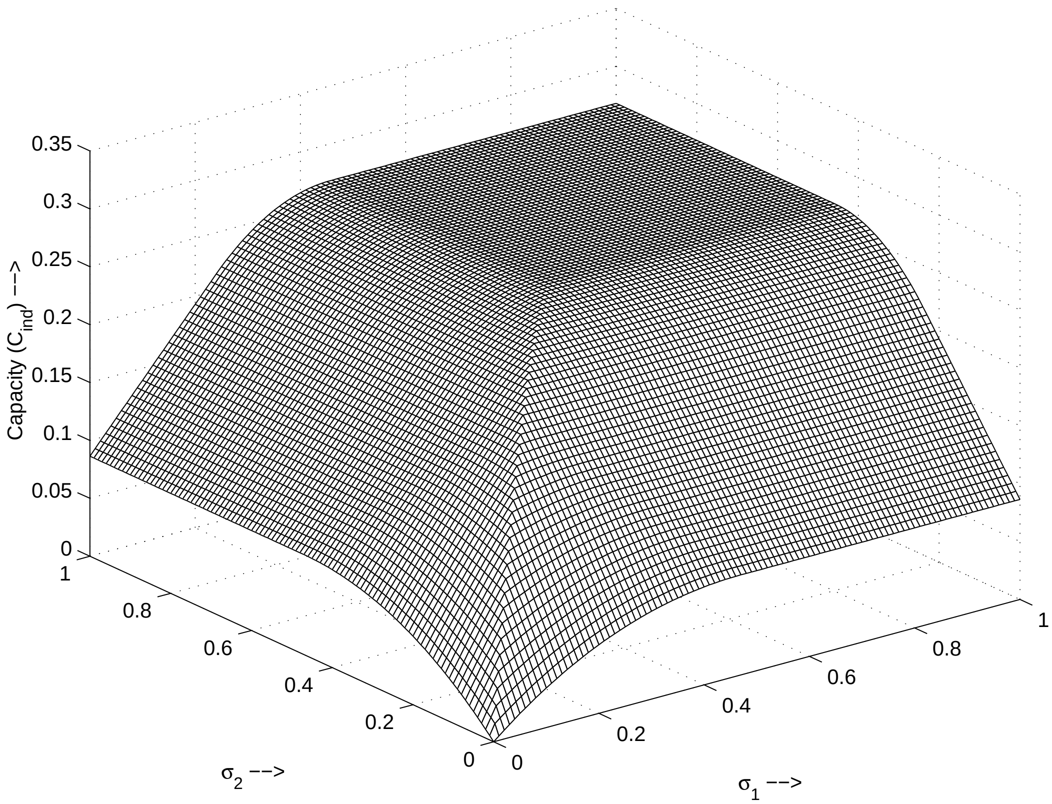

Figure 3.6: Plot of capacity (C_{ind}) versus σ_1, σ_2 for Example 3.1.

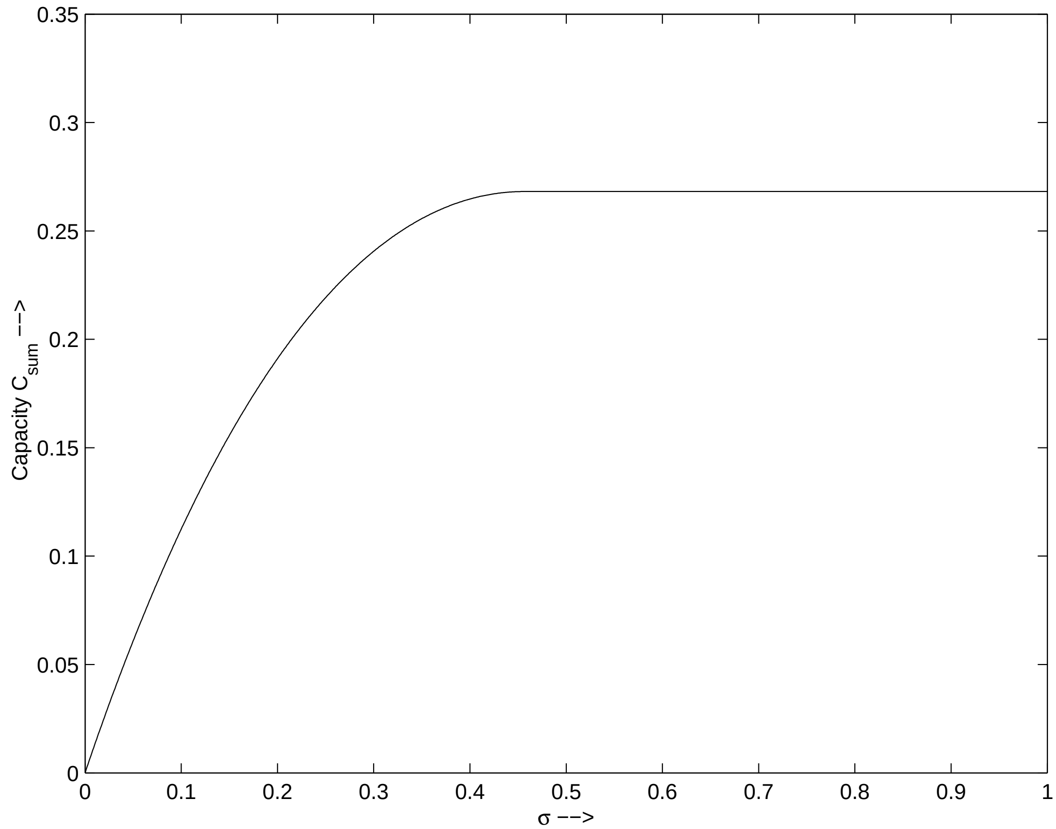

Figure 3.7: Plot of capacity (C_{sum}) versus σ for Example 3.1.

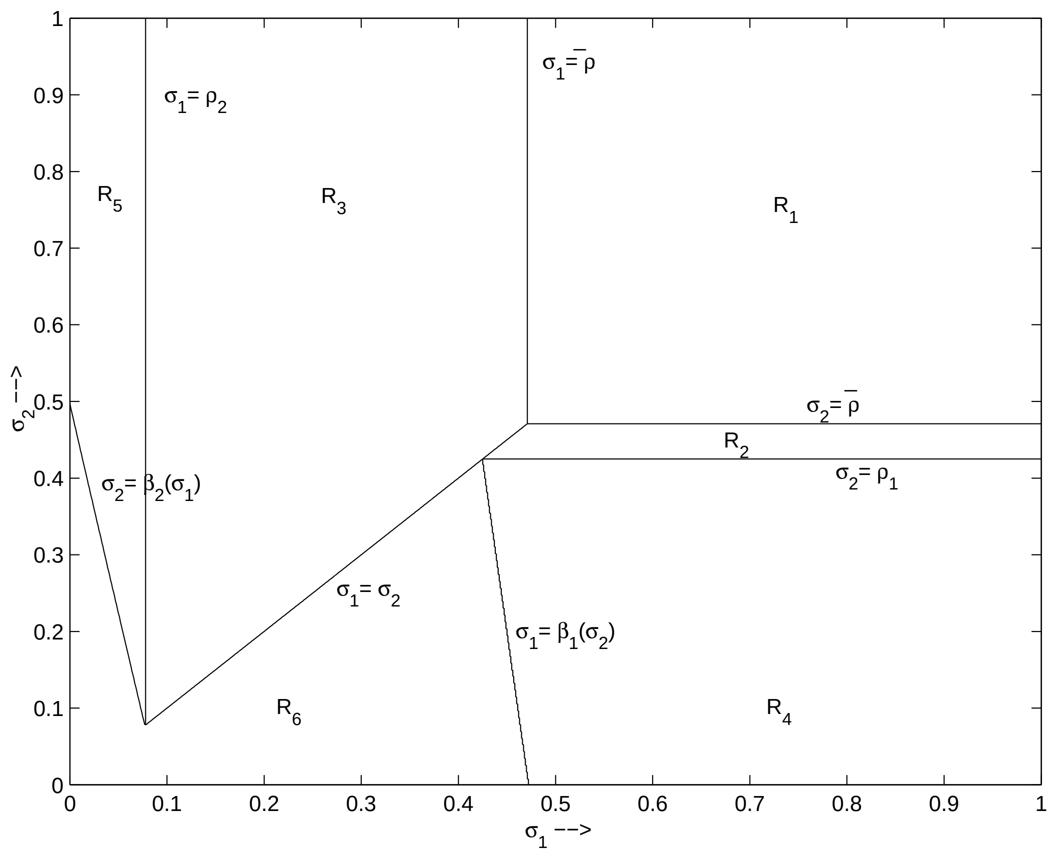

Figure 3.8: Decision region of optimal duty cycles for individual average power constraints σ_1, σ_2 for Example 3.2.

when the transmit apertures are subject to an average sum power constraint σ, the optimal average duty cycles of both transmit apertures are equal for all $0 \leq \sigma \leq 1$ (see Figure 3.5). In Figures 3.6 and 3.7, we plot respectively the capacities for individual and sum average power constraints versus the average power constraint parameters.

Example 3.2: Consider a 2×2 MIMO Poisson channel with $A_1 = A_2 = 1.0$, $\lambda_{0,1} = \lambda_{0,2} = 1.0$, $\mathbf{s} = [1.0\ 0.1;\ 0.1\ 0.1]$. It can be verified that $\rho_1 = 0.425$, $\rho_2 = 0.078$, $\bar{\rho} = 0.471$, $K_1 = 0.4575$, $K_2 = 0.3054$. In this example, it can be seen

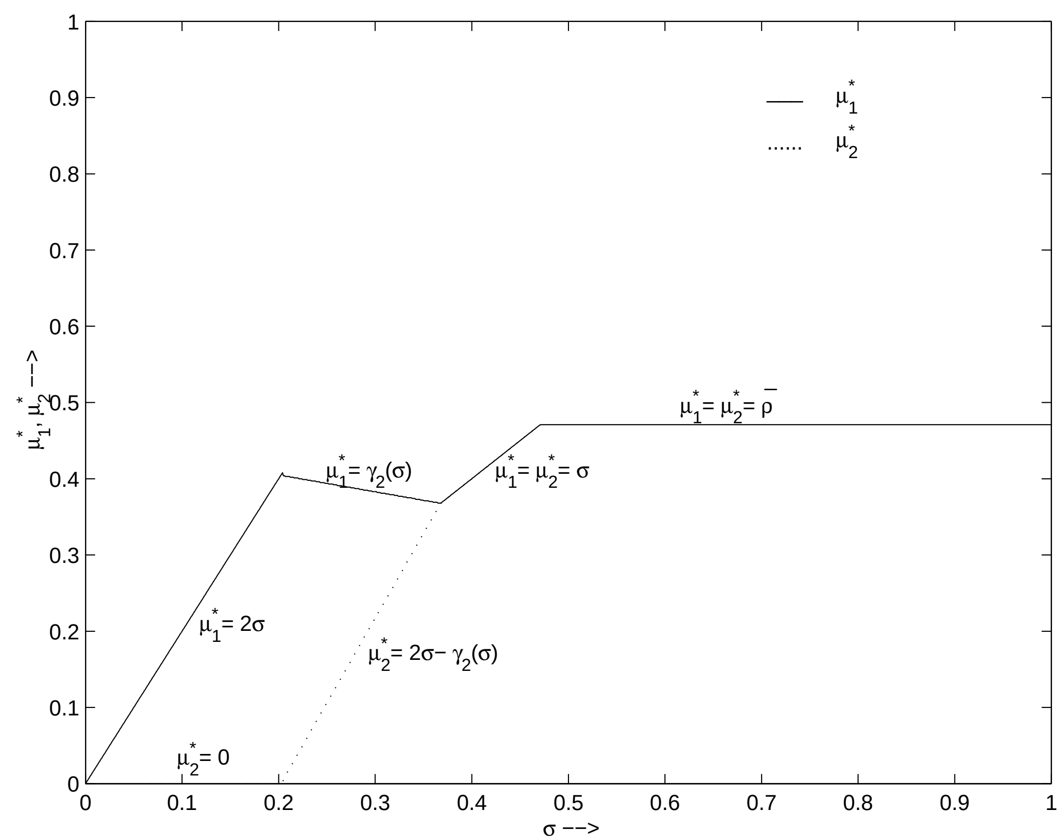

Figure 3.9: Optimal duty cycles for an average sum power constraint σ for Example 3.2.

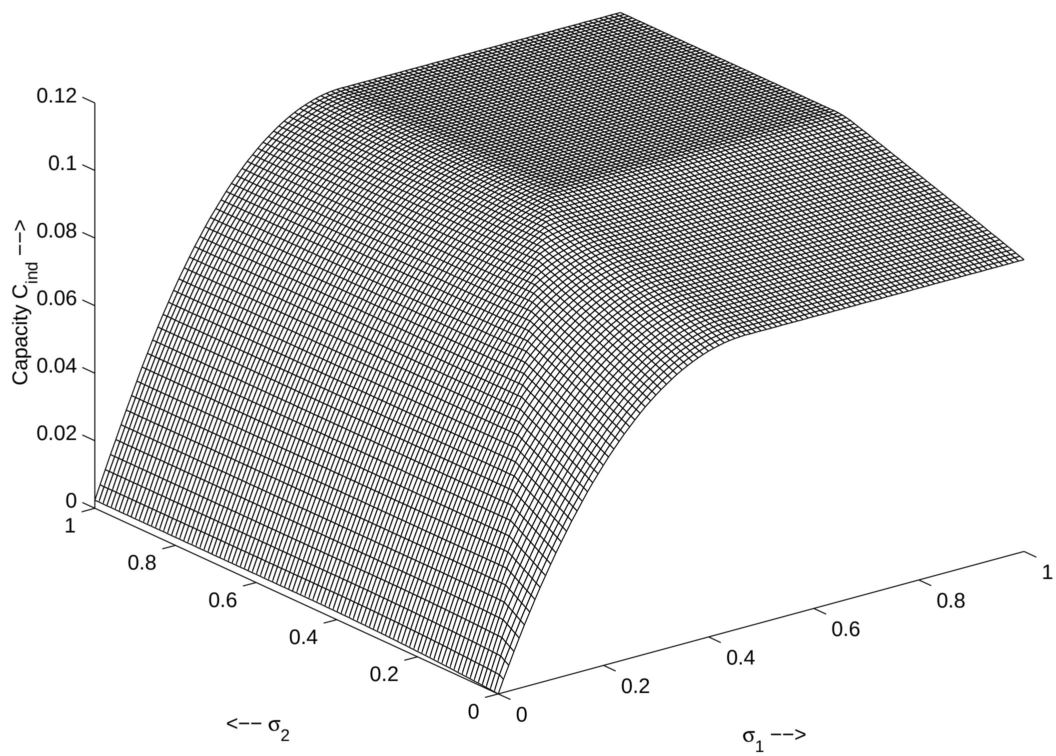

Figure 3.10: Plot of capacity (C_{ind}) versus σ_1, σ_2 for Example 3.2.

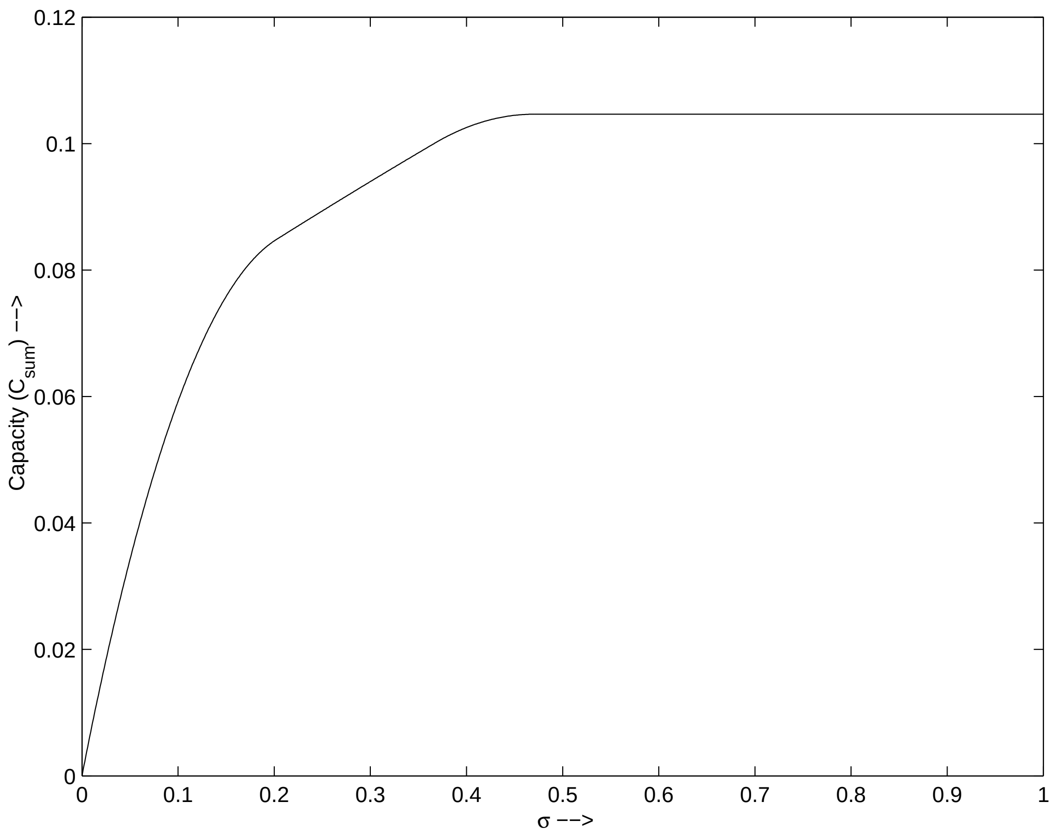

Figure 3.11: Plot of capacity (C_{sum}) versus σ for Example 3.2.

that transmit aperture 1 experiences significantly better channel conditions than transmit aperture 2, which is captured in the condition $K_1 > K_2 > 0$. The decision region for optimal duty cycles when the transmit apertures are subject to individual average power constraints (see Figure 3.8) is clearly asymmetric, with a bias towards transmit aperture 1. When the transmit apertures are subject to a sum average power constraint σ, we see that (see Figure 3.9) for a range of values of σ, the optimal average duty cycle of transmit aperture 1 is strictly larger than that of transmit aperture 2. In Figure 3.10 (resp. Figure 3.11), the variation of capacity with average power constraint parameters σ_1 and σ_2 (resp. σ) is plotted; the asymmetric nature of channel conditions with respect to the two transmit apertures is reflected in the asymmetry of these plots.

3.6 Discussion

We have studied the shot-noise limited $N \times M$ MIMO Poisson channel with constant channel fade and background noise rates. At each receive aperture, the optical fields received from different transmit apertures are assumed to be sufficiently separated in frequency or angle of arrival such that the received total power is the sum of powers from individual transmit apertures, scaled by the respective channel fade. The transmit apertures are subject to individual peak power constraints, and one of two types of average power constraints: (a) individual average power constraints, or (b) a constraint on the sum of the average powers from all transmit apertures. The capacity of this channel is explicitly characterized and optimum strategies for transmission are determined.

It has been established that a two-level signaling strategy through each transmit aperture with arbitrarily fast intertransition times is capacity-achieving. The two levels correspond to no transmission (OFF state) and transmission at the peak power level (ON state). The transmitted signals through the N apertures are in general correlated across apertures and i.i.d. in time. Furthermore, the optimum set of transmission events has at most $N + 1$ values (out of 2^N possible values). Each of these values correspond to events in which exactly k transmit apertures are ON and the remaining $N - k$ are OFF, where $k = 0, 1, \cdots, N$. There is a specific ordering of the transmit apertures (denoted by Π) which dictates that whenever a transmit aperture is ON, all the apertures with stronger average power constraints (in the sense of Π) must also remain ON.

For the special case of a symmetric average power constraint, the optimum transmission strategy corresponds to the simultaneous ON-OFF keying events, in which all the transmit apertures are simultaneously ON or OFF. This implies that the lower bound derived in [19] for the MIMO Poisson channel with a symmetric average power constraint is always tight, a fact not mentioned therein.

On the other hand, in the case of asymmetric average power constraints, the optimum set of transmission events may have more than 2 values, as was shown in Theorem 5 for $N = 2$ transmit apertures. Specifically, the transmit apertures with higher average power levels may remain ON even when the ones with lower average power levels are OFF. However, for a range of values of average power levels, the optimum strategy assigns nonzero mass to only the simultaneous ON-OFF keying transmission events.

We have identified a key property of the optimal transmission strategy for communication over a MIMO Poisson channel subject to individual peak power constraints and a constraint on the sum of the average power constraints. It has been demonstrated that the relative ordering of the optimal average duty cycles across the transmit apertures does not depend on the average sum power constraint σ.

In the next chapter, these results will be used to characterize the capacity of a MIMO Poisson channel with random channel fade. Several properties of optimal transmission strategies, e.g., optimality of i.i.d. binary signaling from each transmit aperture with arbitrarily fast intertransition times, and a relative ordering of transmit apertures according to optimal average duty cycles will be shown to extend to the random fading channel as well. Finally, the result of Corollary 6 will be used to obtain a simplified expression for channel capacity of the symmetric MIMO channel with isotropically distributed channel fade.

Chapter 4

MIMO Poisson channel with random channel fade

4.1 Introduction

In this chapter, we consider the $N \times M$ MIMO Poisson channel with random channel fade. We assume that the channel fade matrix remains unvarying over intervals of duration T_c, and changes across successive intervals in an i.i.d. manner. The receiver is assumed to possess perfect CSI, while the transmitter CSI can be imperfect. The transmitted signals from each transmit aperture are subject to peak and average power constraints. We provide a single-letter characterization of the channel capacity in terms of the channel and signal parameters. While an explicit characterization of the capacity-achieving optimal power control law in its full generality is yet to be determined, we are able to identify interesting properties of the optimal transmission strategy for the special case of a symmetric MIMO Poisson channel with isotropically distributed channel fade when the transmitter has perfect CSI.

The remainder of this chapter is organized as follows. In Section 4.2, we describe the problem formulation. The results are discussed in Section 4.3, and the proofs are outlined in Section 4.4. An illustrative example is discussed in Section 4.5. Finally, the main contributions are summarized in Section 4.6.

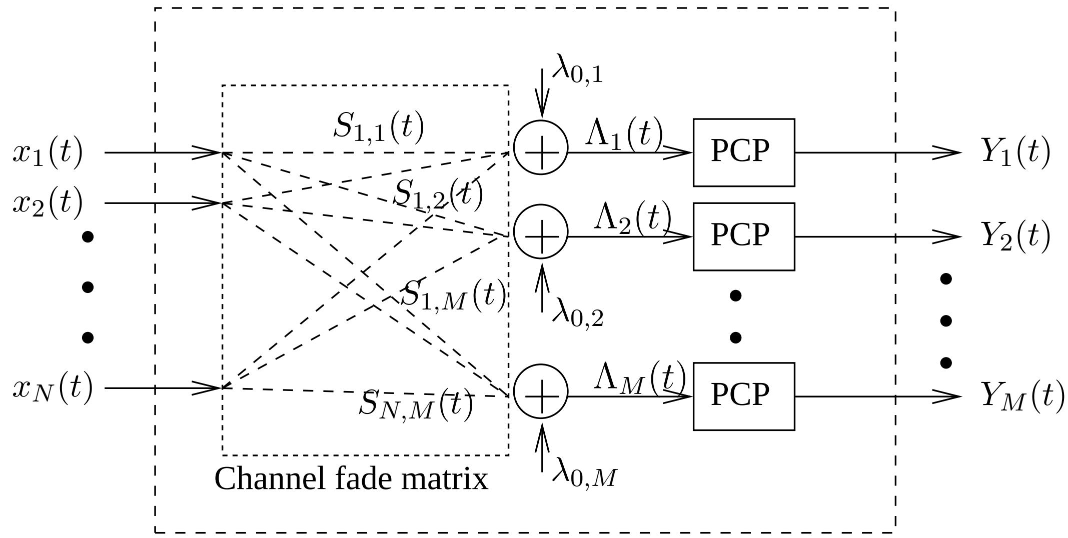

Figure 4.1: $N \times M$ MIMO Poisson channel with random channel fade.

4.2 Problem formulation

A block schematic diagram of the channel model is given in Figure 4.1. For a given set of N $\mathbb{R}_0^+$–valued transmitted signals $\{x_n(t),\, t \geq 0\}_{n=1}^N$, the received signal $\{Y_m(t),\, t \geq 0\}$ at the m^{th} receive aperture is a $\mathbb{Z}_0^+$-valued nondecreasing (left-continuous) PCP with rate equal to

$$\Lambda_m(t) \;=\; \sum_{n=1}^{N} S_{n,m}(t) x_n(t) + \lambda_{0,m}, \quad t \geq 0, \tag{4.1}$$

where $\{S_{n,m}(t),\, 0 \leq t \leq T\}$ is the $\mathbb{R}_0^+$-valued random fade from the n^{th} transmit aperture to the m^{th} receive aperture, $n = 1, \cdots, N$, $m = 1, \cdots, M$, and $\lambda_{0,m} \geq 0$ is the (constant) background noise rate at the m^{th} receive aperture, $m = 1, \cdots, M$. We also assume that the receive apertures are sufficiently separated in space, so that conditioned on the knowledge of transmitted signals and the instantaneous realization of channel fade at the receiver, the processes $Y_m^\infty = \{Y_m(t),\, t \geq 0\}$, $m = 1, \cdots, M$, are taken to be mutually independent [47, 19].

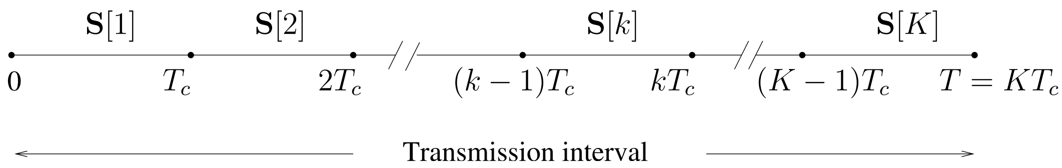

Figure 4.2: Block fading channel.

The input to the channel is a set of N $\mathbb{R}_0^+$–valued transmitted signals, one corresponding to each transmit aperture, collectively denoted by $\mathbf{x}^T = \{x_n(t),\, 0 \leq t \leq T\}_{n=1}^N$, each of which is proportional to the transmitted optical power from the respective transmit aperture, and which satisfy peak and average power constraints of the form

$$\begin{aligned} 0 \;\leq\; x_n(t) \;&\leq\; A_n, \quad 0 \leq t \leq T, \\ \frac{1}{T}\int_0^T x_n(t)dt \;&\leq\; \sigma_n A_n, \end{aligned} \tag{4.2}$$

where the peak powers $A_n > 0$ and the ratios of average-to-peak power σ_n, $0 \leq \sigma_n \leq 1$, $n = 1, \cdots, N$, are fixed. Also, symmetric and asymmetric peak and average power constraints can be defined in a manner similar to Definition 2.

The channel fade, i.e., path gain, from the n^{th} transmit aperture to the m^{th} receive aperture is a $\mathbb{R}_0^+$-valued random process $\{S_{n,m}(t),\, t \geq 0\}$. The channel coherence time T_c is a measure of the intermittent coherence of the time-varying channel fade. We assume that the channel fade remain fixed over time intervals of duration T_c, and change in an i.i.d. manner across successive such intervals. For $k = 1, 2, \cdots$, let the channel fade from the n^{th} transmit aperture to the m^{th} receive aperture on $[(k-1)T_c,\, kT_c)$ be denoted by the rv $S_{n,m}[k]$; in other words, for $n = 1, \cdots, N$, $m = 1, \cdots, M$,

$$S_{n,m}[k] = S_{n,m}(t), \quad t \in [(k-1)T_c,\, kT_c), \quad k = 1, 2, \cdots.$$

The channel fade matrix on $[(k-1)T_c,\, kT_c)$ is denoted by $\mathbf{S}[k] = \{S_{n,m}[k],\ n = 1, \cdots, N,\ m = 1, \cdots, M\}$, $k = 1, 2, \cdots$ (see Figure 4.2). The channel fade is then described by the random sequence $\{\mathbf{S}[k], k \in \mathbb{Z}^+\}$ of i.i.d. repetitions of a random matrix $\mathbf{S}$ with known distribution $F_{\mathbf{S}}$. Note that the rate of $\{Y_m(t), t \geq 0\}$ is thus given by

$$\Lambda_m(t) \;=\; \sum_{n=1}^{N} S_{n,m}[\lceil t/T_c \rceil] x_n(t) + \lambda_{0,m}, \quad t \geq 0.$$

Our general results below hold for a broad class of distributions for $\mathbf{S}$ which satisfy the following technical conditions: (a) $\Pr\{S_{n,m} > 0\} = 1$, (b) $\mathbb{E}[S_{n,m}] < \infty$ and (c) $\mathbb{E}\left[S_{n,m} \log S_{n,m}\right] < \infty$, $n = 1, \cdots, N$, $m = 1, \cdots, M$.

Various degrees of channel state information (CSI) can be made available to the transmitter and the receiver. As before, we shall assume that the receiver has perfect CSI. In general, we can model the CSI available at the transmitter in terms of a given mapping $h : (\mathbb{R}_0^+)^{N\times M} \to \mathcal{U}$, where $\mathcal{U}$ is an arbitrary subset of $(\mathbb{R}_0^+)^{N\times M}$, not necessarily finite. For $\mathbf{S}[k] = \mathbf{s}[k]$, $\mathbf{s}[k] \in (\mathbb{R}_0^+)^{N\times M}$, the transmitter (resp. receiver) is provided with CSI $\mathbf{u}[k] = h(\mathbf{s}[k])$ (resp. $\mathbf{s}[k]$) on $[(k-1)T_c,\, kT_c)$, $k \in \mathbb{Z}^+$. Let $\{\mathbf{U}[k] = h(\mathbf{S}[k]),\ k \in \mathbb{Z}^+\}$ denote the CSI at the transmitter, hereafter referred to as the *transmitter CSI* h. We shall be particularly interested in two special cases of this general framework. In the first case, the transmitter has perfect CSI, i.e., h is the identity mapping. In the second case, the transmitter has no CSI, i.e., h is the trivial (constant) mapping.

We assume, without loss of generality, that the message transmission duration T is an integer multiple of the channel coherence time, i.e., $T = KT_c$, $K \in \mathbb{Z}^+$.

Then the receiver CSI is given by the collection of matrices $\mathbf{S}^K = \{\mathbf{S}[1], \cdots, \mathbf{S}[K]\}$ and the transmitter CSI is given by $\mathbf{U}^K$. For the channel under consideration, a (W, T)–code (f, ϕ) is defined as follows.

1. For each $\mathbf{u}^K \in \mathcal{U}^K$, the codebook comprises a set of W waveform vectors $f(w, \mathbf{u}^K) = \{x_n(w, t, \mathbf{u}^{\lceil t/T_c \rceil}), 0 \le t \le T\}_{n=1}^N$, $w \in \mathcal{W} = \{1, \cdots, W\}$, satisfying peak and average power constraints which follow from (4.2):

$$\begin{aligned} 0 \quad \le \quad x_n(w, t, \mathbf{u}^{\lceil t/T_c \rceil}) \quad &\le \quad A_n, \quad 0 \le t \le T, \quad w \in \mathcal{W}, \\ \tfrac{1}{T} \int_0^T x_n(w, t, \mathbf{u}^{\lceil t/T_c \rceil}) dt \quad &\le \quad \sigma_n A_n, \quad w \in \mathcal{W}. \end{aligned} \tag{4.3}$$

   Note that the transmitted signals $\{x_n(w, t)\}_{n=1}^N$ at time t are allowed to depend on $\mathbf{u}^{\lceil t/T_c \rceil}$, the causal transmitter CSI.

2. The decoder is a mapping

$$\phi : (\Sigma(T))^M \times ((\mathbb{R}_0^+)^{N \times M})^K \quad \to \quad \mathcal{W}.$$

For each message $w \in \mathcal{W}$ and transmitter CSI $\mathbf{u}^K \in \mathcal{U}^K$ corresponding to the fade vector $\mathbf{s}^K \in ((\mathbb{R}_0^+)^{N \times M})^K$, the transmitter sends a waveform $\{x_n(w, t, \mathbf{u}^{\lceil t/T_c \rceil}), 0 \le t \le T\}$ from the n^{th} transmit aperture, $n = 1, \cdots, N$. The receiver, upon observing y_m^T from the m^{th} receive aperture, $m = 1, \cdots, M$, and being provided with $\mathbf{s}^K$, produces an output $\hat{w} = \phi(y_1^T, \cdots, y_M^T, \mathbf{s}^K)$. The rate of this (W, T)–code is given by $R = \frac{1}{T} \log W$ nats/sec., and the average probability of decoding error is given by

$$\begin{aligned} P_e(f, \phi) \quad = \quad \frac{1}{W} \sum_{w=1}^{W} \mathbb{E} \big[\Pr \big\{ \phi(Y_1^T, \cdots, Y_M^T, \mathbf{S}^K) \neq w \big| \\ x_1^T(w, \mathbf{U}^K), \cdots, x_N^T(w, \mathbf{U}^K), \mathbf{S}^K \big\} \big], \end{aligned}$$

where we have used the shorthand notation

$$x_n^T\left(w,\, \mathbf{U}^K\right) = \left\{x_n\left(w,\, t,\, \mathbf{U}^{\lceil t/T_c \rceil}\right),\, 0 \le t \le T\right\},\ \ w \in \mathcal{W},\ \ n = 1,\, \cdots,\, N.$$

In the subsequent sections, we provide a "single-letter characterization" of the channel capacity C (see Definition 1) in terms of signal and channel parameters, and examine some properties of the associated optimal transmission strategies.

4.3 Statement of results

4.3.1 Channel capacity

We begin with a single-letter characterization of the capacity of the MIMO Poisson channel with block fading.

Theorem 7 *The capacity of the MIMO Poisson fading channel for transmitter CSI h is given by*

$$C = \max_{\substack{\mu_n:\mathcal{U}\to[0,\,1] \\ \mathbf{E}[\mu_n(\mathbf{U})]\le\sigma_n \\ n=1,\cdots,N}} \mathbb{E}[I(\boldsymbol{\mu}^N(\mathbf{U}),\, \mathbf{S})], \tag{4.4}$$

where $\mathbf{U} = h(\mathbf{S})$, and $I(\cdot,\,\cdot)$ is as defined in (3.8).

The capacities for the special cases of perfect and no CSI at the transmitter follow directly from Theorem 7.

Corollary 7 *The capacity for perfect CSI at the transmitter is given by*

$$C_P = \max_{\substack{\mu_n:(\mathbb{R}_0^+)^{N\times M}\to[0,\,1] \\ \mathbf{E}[\mu_n(\mathbf{S})]\le\sigma_n \\ n=1,\cdots,N}} \mathbb{E}[I(\boldsymbol{\mu}^N(\mathbf{S}),\, \mathbf{S})]. \tag{4.5}$$

The capacity for no CSI at the transmitter is given by

$$C_N = \max_{\substack{0 \le \mu_n \le \sigma_n \\ n=1,\cdots,N}} \mathbb{E}[I(\boldsymbol{\mu}^N, \mathbf{S})]. \tag{4.6}$$

Remarks: (i) The optimization in (4.4) (as well as in (4.5) and (4.6)) is that of a concave functional over a convex compact set, so that the maximum clearly exists.

(ii) From Theorem 7, we see that the capacity of the MIMO Poisson fading channel with perfect receiver CSI *does not depend on the coherence time* T_c. Conditioned on the transmitted signals $\mathbf{x}^T$, and perfect receiver CSI $\mathbf{s}^K$, the received signals $\{Y_m^T\}_{m=1}^M$ are independent across coherence intervals; hence it suffices to look at a single coherence interval in the mutual information computations. Furthermore, within a coherence interval, the optimality of i.i.d. (in time) transmitted signals leads to a lack of dependence of capacity on T_c.

(iii) Our proof of the achievability part shows that $\{0, A_n\}$–valued transmitted signals (corresponding to ON and OFF signal levels) from the n^{th} transmit aperture, $n = 1, \cdots, N$, with arbitrarily fast intertransition times, can achieve channel capacity. The signal characteristics during each coherence interval depend only on the current transmitter CSI and not on the past transmitter CSI.

(iv) The optimizing "power control law" μ_n in (4.4), (4.5), (4.6) specifies the conditional probability of the n^{th} transmit aperture assuming the level A_n (ON state) depending on the available transmitter CSI. Thus, it can be interpreted as the average "conditional duty cycle" of the n^{th} aperture's transmitted signal as a function of transmitter CSI.

(v) In general, the optimal transmission strategy will comprise at most $N+1$ transmission events, where each event corresponds to a set of k transmit apertures remaining in the ON state and the remaining $N-k$ transmit apertures simultaneously remaining in the OFF state. The active set of transmit apertures at any given time instant will depend on the current transmitter CSI.

(vi) Theorem 1 (resp. Corollary 1) is obviously a special case of Theorem 7 (resp. Corollary 7), with $N = M = 1$.

4.3.2 Optimum power control strategy

It remains to determine the optimal power control law parameters $\{\mu_n^*(\cdot)\}_{n=1}^N$ that achieve the maximum in (4.4). One difficulty in obtaining a closed form analytical solution for $\{\mu_n^*(\cdot)\}_{n=1}^N$ is the nondifferentiability of $I(\cdot, \mathbf{s})$; standard variational techniques for differentiable functions (cf. e.g., [15]) cannot be applied here. Nonsmooth optimization techniques (cf. e.g., [7]) can be used to determine the solution computationally. While the analytical structure of the optimal solution remains to be characterized in full generality, we provide solutions in some special cases below. For the sake of notational simplicity, we focus on $N = 2$ transmit apertures.

We begin with some notation. For $\mathbf{s} \in (\mathbb{R}_0^+)^{2\times M}$, let $\rho = \rho_n(\mathbf{s})$ be the solution of

$$\sum_{m=1}^{M} \lambda_{0,\,m} b_{n,\,m} \log\left(\frac{1+\alpha(b_{n,\,m})b_{n,\,m}}{1+\rho B_m}\right) \;=\; 0, \quad n = 1, 2, \tag{4.7}$$

where $\{b_{n,\,m}\}_{n=1,\,m=1}^{2,\,M}$ and $\{B_m\}_{m=1}^{M}$ are as defined in (3.12); and let $\rho = \bar{\rho}(\mathbf{s})$ be the solution of

$$\sum_{m=1}^{M} \lambda_{0,\,m} B_m \log\left(\frac{1+\alpha\,(B_m)\,B_m}{1+\rho B_m}\right) \;=\; 0. \tag{4.8}$$

From Section 3.3, recall that $\rho_1(\mathbf{s}) \geq 0$, $\rho_2(\mathbf{s}) \geq 0$, $\frac{1}{e} \leq \bar{\rho}(\mathbf{s}) \leq \frac{1}{2}$ and $\bar{\rho}(\mathbf{s}) \geq \max\{\rho_1(\mathbf{s}), \rho_2(\mathbf{s})\}$ for all $\mathbf{s} \in (\mathbb{R}_0^+)^{2\times M}$. Let

$$\begin{aligned} K_1(\mathbf{s}) &= \textstyle\sum_{m=1}^{M} \lambda_{0,\,m}\left[B_m \log\left(1+\bar{\rho}(\mathbf{s})B_m\right) - (1+a)b_{2,\,m}\log\left(1+\rho_2(\mathbf{s})b_{2,\,m}\right)\right], \\ K_2(\mathbf{s}) &= \textstyle\sum_{m=1}^{M} \lambda_{0,\,m}\left[(1+a)b_{1,\,m}\log\left(1+\rho_1(\mathbf{s})b_{1,\,m}\right) - aB_m\log\left(1+\bar{\rho}(\mathbf{s})B_m\right)\right], \end{aligned} \tag{4.9}$$

with $a = \frac{A_1}{A_2}$. From Section 3.3, recall that $K_1(\mathbf{s}) \geq K_2(\mathbf{s})$, $\mathbf{s} \in (\mathbb{R}_0^+)^{2\times M}$. For $x,\, y \in \mathbb{R}$ and $\mathbf{s} \in (\mathbb{R}_0^+)^{2\times M}$, define

$$\begin{aligned} d(x,\, y,\, \mathbf{s}) &= \sum_{m=1}^{M} \lambda_{0,\,m}(b_{1,\,m} - ab_{2,\,m}) \times \\ &\quad \log(1+(b_{1,\,m}-ab_{2,\,m})x + (1+a)b_{2,\,m}y), \end{aligned} \tag{4.10}$$

and let $x = \gamma_j(y,\, \mathbf{s})$ be the solution of

$$d(x,\, y,\, \mathbf{s}) \;=\; K_j(\mathbf{s}), \quad j = 1,\, 2, \tag{4.11}$$

whence it can be verified that $\gamma_1(y,\, \mathbf{s}) \geq \gamma_2(y,\, \mathbf{s})$ for all $y \in \mathbb{R}$ and $\mathbf{s} \in (\mathbb{R}_0^+)^{2\times M}$.

4.3.3 Symmetric MIMO channel with isotropic fade

We now consider a *symmetric* channel model with random *isotropic* channel fade. We assume that the transmit apertures are subject to symmetric peak and average power constraints, i.e., $A_n = A$, and $\sigma_n = \sigma$, $n = 1, 2$. Furthermore, we assume that the receive apertures experience similar background radiation, so that $\lambda_{0,\,m} = \lambda_0$, $m = 1, \cdots, M$. The channel fade coefficients are assumed to be

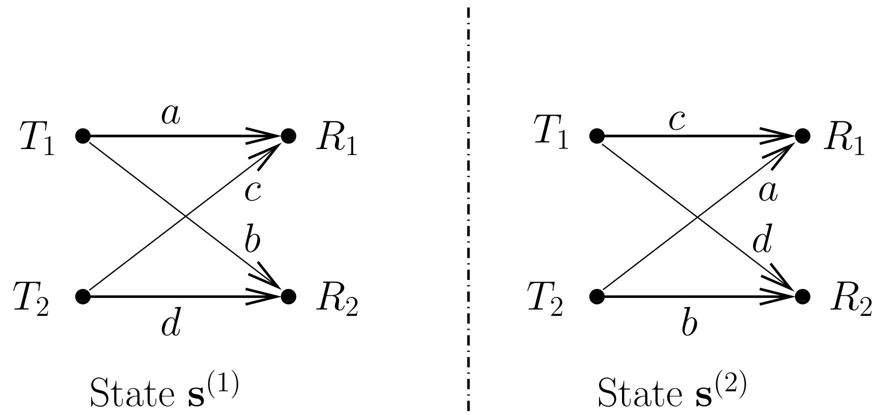

Figure 4.3: Mirror states of a 2×2 MIMO Poisson channel.

i.i.d., across both time and apertures, i.e., $S_{n,m}[k] \sim S$, $k = 1, \cdots, K$, $n = 1, 2$, $m = 1, \cdots, M$, with common cdf[1] F_S. For this special channel model, we provide a simplified characterization of channel capacity. We begin with the following definition.

Definition 3 *The states* $\mathbf{s}^{(1)} \in (\mathbb{R}_0^+)^{2\times M}$ *and* $\mathbf{s}^{(2)} \in (\mathbb{R}_0^+)^{2\times M}$ *are called mirror states if the rows of the respective channel fade matrices are permutations of each other, i.e.,* $s_{1,m}^{(1)} = s_{2,m}^{(2)}$ *and* $s_{2,m}^{(1)} = s_{1,m}^{(2)}$*,* $m = 1, \cdots, M$*.*

An example of mirror states for a 2×2 MIMO Poisson channel is given in Figure 4.3. The channel characteristics experienced by the transmit apertures are interchanged in the mirror states. We now summarize some interesting properties of mirror states.

Lemma 1 *If* $\mathbf{s}^{(1)}$ *and* $\mathbf{s}^{(2)}$ *are mirror states, then*

1. $F_{\mathbf{S}}(\mathbf{s}^{(1)}) = F_{\mathbf{S}}(\mathbf{s}^{(2)})$*;*

[1] Recall that the distribution function of $\mathbf{S}[k]$ is denoted by $F_{\mathbf{S}}$. The distinction between F_S and $F_{\mathbf{S}}$ should be clear from context.

2. $K_1(\mathbf{s}^{(1)}) + K_2(\mathbf{s}^{(2)}) = 0 = K_1(\mathbf{s}^{(2)}) + K_2(\mathbf{s}^{(1)})$*;*

3. $I(a,\, b,\, \mathbf{s}^{(1)}) = I(b,\, a,\, \mathbf{s}^{(2)}),\ a,\, b \geq 0$*;*

4. $I(a_1,\, b_1,\, \mathbf{s}^{(1)}) + I(a_2,\, b_2,\, \mathbf{s}^{(2)}) \leq 2I(\frac{a_1+b_2}{2},\, \frac{a_2+b_1}{2},\, \mathbf{s}^{(1)}),\ a_1,\, a_2,\, b_1,\, b_2 \geq 0.$

From the symmetry of the channel model, we conclude that

Lemma 2 *If* $\mathbf{s}^{(1)}$ *and* $\mathbf{s}^{(2)}$ *are mirror states, then at optimality,*

$$\begin{aligned} \mu_1^*(\mathbf{s}^{(1)}) &= \mu_2^*(\mathbf{s}^{(2)}), \\ \mu_1^*(\mathbf{s}^{(2)}) &= \mu_2^*(\mathbf{s}^{(1)}). \end{aligned} \tag{4.12}$$

Remarks: (i) The channel fades experienced by the two transmit apertures are interchanged in the mirror states. Given the symmetry of the channel model, it is, therefore, not surprising that the role reversal of transmit apertures in the mirror states is reflected in the optimum power control law.

(ii) We can generalize the notion of mirror states and interchange of roles of transmit apertures for the case of an arbitrary number N of transmit apertures. Any permutation of the rows of the channel fade matrix leads to a mirror state. We state without proof the following generalization of Lemma 2: at optimality, $\mu_n^*(\mathbf{s}) = \mu_{P(n)}^*(\mathbf{s}^{(P)}),\ n = 1,\, \cdots,\, N$, where P is any permutation of $\{1,\, \cdots,\, N\}$ and $\mathbf{s}^{(P)}$ is obtained by applying P on the rows of $\mathbf{s}$.

Consider next the following partitioning of the state set that separates the mirror states. Let

$$\begin{aligned} \mathcal{S}_0 &= \{\mathbf{s} \in (\mathbb{R}_0^+)^{2\times M} : \mathbf{s} = \mathbf{s}^{mir}\}, \\ \mathcal{S}_1 &= \{\mathbf{s} \in (\mathbb{R}_0^+)^{2\times M} : K_1(\mathbf{s}) + K_2(\mathbf{s}) \geq 0,\ \text{and } \mathbf{s}^{mir} \in \mathcal{S}_2\}, \\ \mathcal{S}_2 &= \{\mathbf{s} \in (\mathbb{R}_0^+)^{2\times M} : K_1(\mathbf{s}) + K_2(\mathbf{s}) \leq 0,\ \text{and } \mathbf{s}^{mir} \in \mathcal{S}_1\}, \end{aligned} \tag{4.13}$$

where $\mathbf{s}^{mir}$ is the mirror state of $\mathbf{s}$. By Lemma 1, property 2, we see that if $\mathbf{s} \in \mathcal{S}_1$, then $\mathbf{s}^{mir} \in \mathcal{S}_2$, and vice versa. Let

$$\hat{F}_{\mathbf{S}}(\mathbf{s}) = \begin{cases} F_{\mathbf{S}}(\mathbf{s}), & \text{if } \mathbf{s} \in \mathcal{S}_0, \\ 2F_{\mathbf{S}}(\mathbf{s}), & \text{if } \mathbf{s} \in \mathcal{S}_1, \\ 0, & \text{if } \mathbf{s} \in \mathcal{S}_2. \end{cases} \tag{4.14}$$

By Lemma 1, property 1, and (4.13), it follows that $\hat{F}_{\mathbf{S}}$ is a valid pdf.

Theorem 8 *The capacity of the $2 \times M$ MIMO isotropic block fading channel for perfect CSI at the transmitter is given by*

$$C = \max_{\substack{\mu_1 : (\mathbb{R}_0^+)^{2\times M} \to [0,\,1] \\ \mu_2 : (\mathbb{R}_0^+)^{2\times M} \to [0,\,1] \\ \mu_1(\mathbf{S}) \geq \mu_2(\mathbf{S}) \\ \mathbf{E}_{\hat{F}_{\mathbf{S}}}[\mu_1(\mathbf{S}) + \mu_2(\mathbf{S})] \leq 2\sigma}} \mathbb{E}_{\hat{F}_{\mathbf{S}}}[I_{1>2}(\mu_1(\mathbf{S}) - \mu_2(\mathbf{S}),\, \mu_2(\mathbf{S}),\, \mathbf{S})], \tag{4.15}$$

where $I_{1>2}(\cdot,\,\cdot,\,\cdot)$ is as defined in (3.94).

Remark: For the symmetric MIMO channel with isotropic channel fade, we show that a necessary and sufficient condition for $\mu_1^*(\mathbf{s}) \geq \mu_2^*(\mathbf{s})$ is $K_1(\mathbf{s}) + K_2(\mathbf{s}) \geq 0$, $\mathbf{s} \in (\mathbb{R}_0^+)^{2\times M}$. Heuristically, the condition $1(K_1(\mathbf{s}) + K_2(\mathbf{s}) \geq 0)$ allows us to partition the state set into channel states for which the channel conditions are "more favorable" for one transmit aperture than the other. Clearly, the optimal power control law assigns higher average power to the "stronger" transmit aperture, i.e., the one which experiences better channel conditions. Since the channel fade is isotropically distributed, it follows that for each fade realization, there exists another fade realization with the same probability of occurrence (which corresponds exactly to the mirror state), for which the channel conditions experienced by the transmit apertures are reversed. By defining $\hat{F}_{\mathbf{S}}$ as in (4.14), we are able to reduce the state set by picking the states for which the channel conditions are more favorable for transmit aperture 1, and hence $\mu_1^*(\cdot) \geq \mu_2^*(\cdot)$.

Note that the individual average power constraints on $\mu_1(\cdot)$ and $\mu_2(\cdot)$ in the right side of (4.5) (with $N = 2$) are replaced by a constraint on their sum in (4.15). We now invoke the result of Theorem 6 to further simplify (4.15).

Theorem 9 *The capacity of the $2 \times M$ MIMO isotropic block fading channel for perfect CSI at the transmitter is given by*

$$C = \max_{\substack{\mu:(\mathbb{R}_0^+)^{2\times M}\to[0,\,1] \\ \mathbf{E}_{\hat{F}_{\mathbf{S}}}[\mu(\mathbf{S})]\leq\sigma}} \mathbb{E}_{\hat{F}_{\mathbf{S}}}[I_{1>2}(\kappa_1(\mathbf{S},\mu(\mathbf{S})) - \kappa_2(\mathbf{S},\mu(\mathbf{S})),\, \kappa_2(\mathbf{S},\mu(\mathbf{S})),\, \mathbf{S})], \quad (4.16)$$

where $\kappa_1(\cdot,\cdot)$ and $\kappa_2(\cdot,\cdot)$ are given as follows:

1. *if $\mu(\mathbf{s}) \geq \bar{\rho}(\mathbf{s})$, then*

$$\kappa_1(\mathbf{s},\mu(\mathbf{s})) = \kappa_2(\mathbf{s},\mu(\mathbf{s})) = \bar{\rho}(\mathbf{s}); \quad (4.17)$$

2. *if $\mu(\mathbf{s}) < \bar{\rho}(\mathbf{s})$, then*

$$\kappa_1(\mathbf{s},\mu(\mathbf{s})) = \begin{cases} \min\{\gamma_2(\mu(\mathbf{s}),\,\mathbf{s}),\, 2\mu(\mathbf{s})\}, & \text{if } K_2(\mathbf{s}) \geq d(\mu(\mathbf{s}),\,\mu(\mathbf{s}),\,\mathbf{s}), \\ \mu(\mathbf{s}), & \text{otherwise}, \end{cases} \quad (4.18)$$

$$\kappa_2(\mathbf{s},\mu(\mathbf{s})) = 2\mu(\mathbf{s}) - \kappa_1(\mathbf{s},\mu(\mathbf{s})). \quad (4.19)$$

Remark: The optimization problem in (4.16) involves a single mapping $\mu : \mathcal{S}_0 \cup \mathcal{S}_1 \to [0,\,1]$, as opposed to the bivariate optimization problem in (4.15). It is difficult to obtain a closed form expression for the optimal $\mu^*(\cdot)$ that maximizes the right side of (4.16), primarily due to the following reason. For any $\mathbf{s} \in \mathcal{S}_0 \cup \mathcal{S}_1$, even though

$I_{1>2}(\cdot,\,\cdot,\,\mathbf{s})$ is differentiable everywhere with respect to its arguments (see (3.95)), $\kappa_1(\mathbf{s}, y)$ and $\kappa_2(\mathbf{s}, y)$ are nondifferentiable functions of y, $0 \leq y \leq 1$.

We now determine a lower bound on channel capacity, based on a suboptimal power control law, and demonstrate that this new lower bound improves on the "simultaneous ON-OFF keying" lower bound proposed in [21].

Corollary 8 *Suppose* $\mu_L : (\mathbb{R}_0^+)^{2\times M} \to [0,\,1]$ *be the power control law that achieves the maximum in*

$$C_{SOOK} \;=\; \max_{\substack{\mu:(\mathbb{R}_0^+)^{2\times M}\to[0,\,1]\\ \mathbf{E}[\mu(\mathbf{S})]\leq\sigma}} \mathbb{E}[I(\mu(\mathbf{S}),\,\mu(\mathbf{S}),\,\mathbf{S})], \tag{4.20}$$

with $I(\cdot,\,\cdot,\,\cdot)$ *as defined in (3.8) with* $N = 2$*. Let*

$$C_{LB} \;\stackrel{\Delta}{=}\; \mathbb{E}_{\hat{F}_{\mathbf{S}}}[I_{1>2}(\kappa_1(\mathbf{S},\,\mu_L(\mathbf{S})) - \kappa_2(\mathbf{S},\,\mu_L(\mathbf{S})),\,\kappa_2(\mathbf{S},\,\mu_L(\mathbf{S})),\,\mathbf{S})], \tag{4.21}$$

where $\kappa_1(\cdot,\,\cdot)$*,* $\kappa_2(\cdot,\,\cdot)$ *are as defined in (4.17)–(4.19). Then,*

$$C \;\geq\; C_{LB} \;\geq\; C_{SOOK}. \tag{4.22}$$

Remark: In an illustrative example below, we show that for some values of σ, the simultaneous ON-OFF keying transmission strategy is strictly suboptimal for the symmetric MIMO channel with isotropic fade. This is in contrast to the optimum power control strategy for the symmetric MIMO Poisson channel with constant fades (see Remark (ii) following Theorem 5).

4.4 Proofs

We begin this section with some additional definitions that will be needed in our proofs. First, observe that for any $\tau > 0$, the number of photon arrivals $N_m(\tau)$

during $[0, \tau]$ together with the corresponding (ordered) arrival times $\mathbf{T}_m^{N_m(\tau)} = \{T_{m,1}, \cdots, T_{m, N_m(\tau)}\}$ are sufficient statistics for Y_m^τ, $m = 1, \cdots, M$. Therefore, the random vector $\mathbf{Y}_{ss}(\tau) = \{(N_m(\tau), \mathbf{T}_m^{N_m(\tau)})\}_{m=1}^M$ is a complete description of the random processes $\mathbf{Y}^\tau = \{Y_1^\tau, \cdots, Y_M^\tau\}$, $\tau \geq 0$.

The channel of interest is characterized as follows. For an input signal $\mathbf{x}^T = \{x_n(t), 0 \leq t \leq T\}_{n=1}^N$ satisfying (4.2) and a fade vector $\mathbf{S}^K = \mathbf{s}^K$, the channel output $(N_m(T), \mathbf{T}_m^{N_m(T)})$ at the m^{th} receive aperture has the "conditional sample function density" (cf. e.g., [47])

$$\begin{aligned} & f_{N_m(T), \mathbf{T}_m^{N_m(T)} | \mathbf{X}^T, \mathbf{S}^K} \left(n_m, \mathbf{t}_m^{n_m} \left| \mathbf{x}^T, \mathbf{s}^K \right.\right) \\ &= \exp\left(-\int_0^T \lambda_m(\tau) d\tau\right) \prod_{i=1}^{n_m} \lambda_m(t_{m,i}), \end{aligned} \tag{4.23}$$

where

$$\lambda_m(t) = \sum_{n=1}^N s_{n,m}[\lceil t/T_c \rceil] x_n(t) + \lambda_{0,m}, \quad 0 \leq t \leq T.$$

Recall our persistent assumption that the receive apertures are sufficiently separated in space, so that conditioned on $\mathbf{x}^T$ and $\mathbf{s}^K$, the processes $Y_1^T, \cdots, Y_M^T$ are conditionally mutually independent [47, 19]. Therefore, the conditional sample function density of $\mathbf{Y}_{ss}(T)$ given $\mathbf{x}^T$ and $\mathbf{s}^K$ is given by

$$\begin{aligned} & f_{\mathbf{Y}_{ss}(T) | \mathbf{X}^T, \mathbf{S}^K} (\mathbf{y_{ss}} | \mathbf{x}^T, \mathbf{s}^K) \\ &= \prod_{m=1}^M f_{N_m(T), \mathbf{T}_m^{N_m(T)} | \mathbf{X}^T, \mathbf{S}^K} (n_m, \mathbf{t}_m^{n_m} | \mathbf{x}^T, \mathbf{s}^K) \\ &= \prod_{m=1}^M \exp\left(-\int_0^T \lambda_m(\tau) d\tau\right) \prod_{i=1}^{n_m} \lambda_m(t_{m,i}), \end{aligned} \tag{4.24}$$

where $\mathbf{y}_{ss} = \{(n_m, \mathbf{t}_m^{n_m})\}_{m=1}^M$. In order to write the channel output sample function density conditioned only on the fade for a given joint distribution of $(\mathbf{X}^T, \mathbf{S}^K)$, consider the conditional mean of X_n^T (conditioned causally on the channel output and the fade) by

$$\hat{X}_n(\tau) \triangleq \mathbb{E}\left[X_n(\tau)|\mathbf{Y}_{ss}(\tau),\, \mathbf{S}^{\lceil \tau/T_c \rceil}\right], \quad 0 \le \tau \le T, \quad n = 1, \cdots, N, \tag{4.25}$$

where we have suppressed the dependence of $\{\hat{X}_n(\tau)\}_{n=1}^N$ on $(\mathbf{Y}_{ss}(\tau),\, \mathbf{S}^{\lceil \tau/T_c \rceil})$ for notational convenience; and define

$$\Lambda_m(\tau) \triangleq \sum_{n=1}^N S_{n,\,m}[\lceil \tau/T_c \rceil] X_n(\tau) + \lambda_{0,\,m}, \quad 0 \le \tau \le T, \quad m = 1, \cdots, M, \tag{4.26}$$

and

$$\begin{aligned} \hat{\Lambda}_m(\tau) &\triangleq \mathbb{E}\left[\Lambda_m(\tau)|\mathbf{Y}_{ss}(\tau), \mathbf{S}^{\lceil \tau/T_c \rceil}\right] \\ &= \sum_{n=1}^N S_{n,\,m}[\lceil \tau/T_c \rceil] \hat{X}_n(\tau) + \lambda_{0,\,m}, \quad 0 \le \tau \le T, \quad m = 1, \cdots, M. \end{aligned} \tag{4.27}$$

By ([47], pp. 425-427), it follows that conditioned on $\mathbf{S}^K$, the process $(N_m(T),$ $\mathbf{T}_m^{N_m(T)})$ is a self-exciting PCP with rate process $\hat{\Lambda}_m^T$, $m = 1, \cdots, M$, so that the output (conditional) sample function density is given by

$$\begin{aligned} f_{\mathbf{Y}_{ss}(T)|\mathbf{S}^K}(\mathbf{y}_{ss}|\mathbf{s}^K) &= \prod_{m=1}^M f_{N_m(T),\, \mathbf{T}_m^{N_m(T)}|\mathbf{S}^K}(n_m,\, \mathbf{t}_m^{n_m}|\mathbf{s}^K) \\ &= \prod_{m=1}^M \exp\left(-\int_0^T \hat{\lambda}_m(\tau) d\tau\right) \cdot \prod_{i=1}^{n_m} \hat{\lambda}_m(t_{m,\,i}), \end{aligned} \tag{4.28}$$

where $\mathbf{y}_{ss} = \{(n_m,\ \mathbf{t}_m^{n_m})\}_{m=1}^M$, and

$$\hat{\lambda}_m(\tau) = \sum_{n=1}^N s_{n,\,m}[\lceil \tau/T_c \rceil] \hat{x}_n(\tau) + \lambda_{0,\,m}, \quad 0 \le \tau \le T, \quad m = 1, \cdots, M,$$

with

$$\hat{x}_n(\tau) = \mathbb{E}\left[X_n(\tau)|\mathbf{Y}_{ss}(\tau) = \mathbf{y}_{ss}(\tau),\, \mathbf{S}^{\lceil \tau/T_c \rceil} = \mathbf{s}^{\lceil \tau/T_c \rceil}\right],$$

$$0 \le \tau \le T, \quad n = 1, \cdots, N.$$

Proof of Theorem 7:

Converse part: Let the rv W be uniformly distributed on the message set $\mathcal{W} = \{1, \cdots, W\}$, and independent of $\mathbf{S}^K$. With $T = KT_c$, $K \in \mathbb{Z}^+$, consider a (W, T)–code (f, ϕ) of rate $R = \frac{1}{T}\log W$, and with $P_e(f, \phi) \leq \epsilon$, where $0 \leq \epsilon \leq 1$ is given. Denote $X_n(t) \stackrel{\Delta}{=} x_n(W, t, \mathbf{U}^{\lceil t/T_c \rceil}), \quad 0 \leq t \leq T, \quad n = 1, \cdots, N$. Note that (4.3) then implies that

$$\begin{aligned} 0 \;\leq\; X_n(t) \;&\leq\; A_n, \quad n = 1, \cdots, N, \quad 0 \leq t \leq T, \\ \tfrac{1}{T}\int_0^T \mathbb{E}[X_n(\tau)]d\tau \;&\leq\; \sigma_n A_n, \quad n = 1, \cdots, N. \end{aligned} \tag{4.29}$$

Let $(N_m(T), \mathbf{T}_m^{N_m(T)})$ be the channel output at the m^{th} receive aperture when X_n^T is transmitted from the n^{th} transmit aperture, $n = 1, \cdots, N$, $m = 1, \cdots, M$. Clearly, the following Markov condition holds:

$$W -\!\circ\!- \mathbf{X}^t\,\mathbf{S}^{\lceil t/T_c \rceil} -\!\circ\!- \mathbf{Y}_{ss}(t), \quad 0 \leq t \leq T, \tag{4.30}$$

where $\mathbf{Y}_{ss}(t) = \{(N_m(t), \mathbf{T}_m^{N_m(t)})\}_{m=1}^M$, $0 \leq t \leq T$. By a standard argument, the rate R of the (W, T)–code (f, ϕ) with $P_e(f, \phi) \leq \epsilon$ must satisfy

$$R \;\lessapprox\; \frac{1}{T} I\left(W \wedge \phi(\mathbf{Y}_{ss}(T), \mathbf{S}^K)\right). \tag{4.31}$$

Proceeding further with the right side of (4.31),

$$\begin{aligned} I\left(W \wedge \phi(\mathbf{Y}_{ss}(T), \mathbf{S}^K)\right) \;&\leq\; I\left(W \wedge \mathbf{Y}_{ss}(T), \mathbf{S}^K\right) && (4.32) \\ &=\; I\left(W \wedge \mathbf{Y}_{ss}(T) | \mathbf{S}^K\right) && (4.33) \\ &=\; h(\mathbf{Y}_{ss}(T)|\mathbf{S}^K) - h(\mathbf{Y}_{ss}(T)|W, \mathbf{S}^K) \\ &\leq\; h(\mathbf{Y}_{ss}(T)|\mathbf{S}^K) - h(\mathbf{Y}_{ss}(T)|X^T, \mathbf{S}^K), && (4.34) \end{aligned}$$

where (4.32) is by the data processing result for mixed rvs (cf. (2.32)); (4.33) is by the independence of W from $\mathbf{S}^K$; and (4.34) is by (4.30). The difference between the conditional entropies of mixed rvs on the right side of (4.34) is

$$\begin{aligned}
&h(\mathbf{Y}_{ss}(T)|\mathbf{S}^K) - h(\mathbf{Y}_{ss}(T)|X^T,\, \mathbf{S}^K) \\
&\quad = \mathbb{E}\left[-\log f_{\mathbf{Y}_{ss}(T)|\mathbf{S}^K}(\mathbf{Y}_{ss}(T)|\mathbf{S}^K)\right] \\
&\qquad - \mathbb{E}\left[-\log f_{\mathbf{Y}_{ss}(T)|\mathbf{X}^T,\,\mathbf{S}^K}(\mathbf{Y}_{ss}(T)|\mathbf{X}^T,\, \mathbf{S}^K)\right] \\
&\quad = \mathbb{E}\left[\log \frac{\prod_{m=1}^{M} \exp\left(-\int_0^T \Lambda_m(\tau)d\tau\right) \prod_{i=1}^{N_m(T)} \Lambda_m(T_{m,\,i})}{\prod_{m=1}^{M} \exp\left(-\int_0^T \hat{\Lambda}_m(\tau)d\tau\right) \prod_{i=1}^{N_m(T)} \hat{\Lambda}_m(T_{m,\,i})}\right] \qquad (4.35) \\
&\quad = \sum_{m=1}^{M} \mathbb{E}\left[\int_0^T \left(\hat{\Lambda}_m(\tau) - \Lambda_m(\tau)\right) d\tau\right] \\
&\qquad + \sum_{m=1}^{M} \mathbb{E}\left[\sum_{i=1}^{N_m(T)} \left(\log \Lambda_m(T_{m,\,i}) - \log \hat{\Lambda}_m(T_{m,\,i})\right)\right] \\
&\quad = \sum_{m=1}^{M} \mathbb{E}\left[\sum_{i=1}^{N_m(T)} \left(\log \Lambda_m(T_{m,\,i}) - \log \hat{\Lambda}_m(T_{m,\,i})\right)\right] \qquad (4.36) \\
&\quad = \sum_{m=1}^{M} \int_0^T \left\{\mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} S_{n,\,m}[\lceil \tau/T_c \rceil] X_n(\tau),\, \lambda_{0,\,m}\right)\right]\right. \\
&\qquad \left. - \mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} S_{n,\,m}[\lceil \tau/T_c \rceil] \hat{X}_n(\tau),\, \lambda_{0,\,m}\right)\right]\right\} d\tau, \qquad (4.37)
\end{aligned}$$

where (4.35) is by (4.24), (4.28); (4.36) holds by an interchange of operations[2] to get

$$\mathbb{E}\left[\int_0^T (\hat{\Lambda}_m(\tau) - \Lambda_m(\tau))d\tau\right] = \int_0^T \mathbb{E}\left[\hat{\Lambda}_m(\tau) - \Lambda_m(\tau)\right] d\tau,$$

followed by noting that $\mathbb{E}[\hat{\Lambda}_m(\tau)] = \mathbb{E}[\Lambda_m(\tau)]$, $0 \le \tau \le T$, $m = 1, \cdots, M$, by (4.27); and (4.37) is proved in Appendix C.1.

[2]The interchange is permissible as the assumed condition $\mathbb{E}[S_{n,\,m}] < \infty$, $n = 1, \cdots, N$, $m = 1, \cdots, M$ implies the integrability of $\{\Lambda_m(\tau),\, 0 \le \tau \le T\}$ and $\{\hat{\Lambda}_m(\tau),\, 0 \le \tau \le T\}$, $m = 1, \cdots, M$.

Next, in the right side of (4.37),

$$
\begin{aligned}
&\mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} S_{n,m}[\lceil\tau/T_c\rceil]\hat{X}_n(\tau),\, \lambda_{0,m}\right)\right] \\
&\quad= \mathbb{E}\left[\mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} S_{n,m}[\lceil\tau/T_c\rceil]\hat{X}_n(\tau),\, \lambda_{0,m}\right)\middle| \mathbf{S}[\lceil\tau/T_c\rceil]\right]\right] \\
&\quad\geq \mathbb{E}\left[\zeta\left(\mathbb{E}\left[\sum_{n=1}^{N} S_{n,m}[\lceil\tau/T_c\rceil]\hat{X}_n(\tau)\middle| \mathbf{S}[\lceil\tau/T_c\rceil]\right],\, \lambda_{0,m}\right)\right] && (4.38)\\
&\quad= \mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} S_{n,m}[\lceil\tau/T_c\rceil]\,\mathbb{E}\left[\hat{X}_n(\tau)\middle| \mathbf{S}[\lceil\tau/T_c\rceil]\right],\, \lambda_{0,m}\right)\right] \\
&\quad= \mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} S_{n,m}[\lceil\tau/T_c\rceil]\,\mathbb{E}\left[X_n(\tau)\middle| \mathbf{S}[\lceil\tau/T_c\rceil]\right],\, \lambda_{0,m}\right)\right] && (4.39)\\
&\quad= \mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} S_{n,m}[\lceil\tau/T_c\rceil]\,\mathbb{E}\left[X_n(\tau)\middle| \mathbf{U}[\lceil\tau/T_c\rceil]\right],\, \lambda_{0,m}\right)\right] && (4.40)
\end{aligned}
$$

where (4.38) is by Jensen's inequality applied to the convex function $\zeta(\cdot,\, y)$, $y \geq 0$; (4.39) is from

$$
\begin{aligned}
\mathbb{E}\left[\hat{X}_n(\tau)\left|\mathbf{S}[\lceil\tau/T_c\rceil]\right.\right] &= \mathbb{E}\left[\mathbb{E}\left[X_n(\tau)\left|\mathbf{Y}_{ss}(\tau),\, \mathbf{S}^{\lceil\tau/T_c\rceil}\right.\right]\middle| \mathbf{S}[\lceil\tau/T_c\rceil]\right] \\
&= \mathbb{E}\left[X_n(\tau)\left|\mathbf{S}[\lceil\tau/T_c\rceil]\right.\right];
\end{aligned}
$$

and (4.40) holds as

$$
\begin{aligned}
\mathbb{E}\left[X_n(\tau)\left|\mathbf{S}[\lceil\tau/T_c\rceil]\right.\right] &= \mathbb{E}\left[X_n(\tau)\left|\mathbf{S}[\lceil\tau/T_c\rceil],\, \mathbf{U}[\lceil\tau/T_c\rceil]\right.\right] \\
&= \mathbb{E}\left[X_n(\tau)\left|\mathbf{U}[\lceil\tau/T_c\rceil]\right.\right] && (4.41)
\end{aligned}
$$

by virtue of the Markov condition

$$
X_n(\tau) -\!\circ\!- \mathbf{U}[\lceil\tau/T_c\rceil] -\!\circ\!- \mathbf{S}[\lceil\tau/T_c\rceil], \quad 0 \leq t \leq T, \quad n = 1,\, \cdots,\, N, \tag{4.42}
$$

which is obtained from Appendix A.2 by setting $A = (W,\, \mathbf{U}^{\lceil\tau/T_c\rceil-1})$, $B = \mathbf{S}[\lceil\tau/T_c\rceil]$, $C = \mathbf{U}[\lceil\tau/T_c\rceil]$ and $X = X_n(\tau)$. Summarizing collectively (4.34), (4.37), (4.40), we get that

$$
\begin{aligned}
& I\left(W \wedge \phi(\mathbf{Y}_{ss}(T),\, \mathbf{S}^K)\right) \\
& \quad \le \int_0^T \sum_{m=1}^{M} \left\{ \mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} S_{n,m}[\lceil \tau/T_c \rceil] X_n(\tau),\, \lambda_{0,m} \right) \right] \right. \\
& \qquad \left. - \mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} S_{n,m}[\lceil \tau/T_c \rceil]\, \mathbb{E}\left[X_n(\tau) | \mathbf{U}[\lceil \tau/T_c \rceil] \right],\, \lambda_{0,m} \right) \right] \right\} d\tau. \qquad (4.43)
\end{aligned}
$$

The right side of (4.43) is further bounded above by a suitable modification of Footnote 10, Chapter 2. Considering the integrand in (4.43), fix $0 \le \tau \le T$ and condition on $\mathbf{S}[\lceil \tau/T_c \rceil] = \mathbf{s}$, $\mathbf{s} \in (\mathbb{R}_0^+)^{N \times M}$. Then

$$
\begin{aligned}
& \sum_{m=1}^{M} \left\{ \mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} S_{n,m}[\lceil \tau/T_c \rceil] X_n(\tau),\, \lambda_{0,m} \right) \right. \right. \\
& \left. \left. - \zeta\left(\sum_{n=1}^{N} S_{n,m}[\lceil \tau/T_c \rceil]\, \mathbb{E}[X_n(\tau) | \mathbf{U}[\lceil \tau/T_c \rceil]],\, \lambda_{0,m} \right) \right| \mathbf{S}[\lceil \tau/T_c \rceil] = \mathbf{s} \right] \right\} \\
& \quad = \sum_{m=1}^{M} \left\{ \mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} s_{n,m} X_n(\tau),\, \lambda_{0,m} \right) \right| \mathbf{S}[\lceil \tau/T_c \rceil] = \mathbf{s} \right] \\
& \qquad \left. - \zeta\left(\sum_{n=1}^{N} s_{n,m}\, \mathbb{E}[X_n(\tau) | \mathbf{U}[\lceil \tau/T_c \rceil] = h(\mathbf{s})],\, \lambda_{0,m} \right) \right\} \\
& \quad = \sum_{m=1}^{M} \left\{ \mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} s_{n,m} X_n(\tau),\, \lambda_{0,m} \right) \right| \mathbf{U}[\lceil \tau/T_c \rceil] = h(\mathbf{s}) \right] \\
& \qquad \left. - \zeta\left(\sum_{n=1}^{N} s_{n,m}\, \mathbb{E}[X_n(\tau) | \mathbf{U}[\lceil \tau/T_c \rceil] = h(\mathbf{s})],\, \lambda_{0,m} \right) \right\} \qquad (4.44)
\end{aligned}
$$

by (4.42). Consider maximizing the right side of (4.44) over all (conditional) joint distributions of $\{X_n(\tau),\, n = 1, \cdots, N\}$ conditioned on $\mathbf{U}[\lceil \tau/T_c \rceil] = h(\mathbf{s})$ with fixed conditional means $\mathbb{E}[X_n(\tau) | \mathbf{U}[\lceil \tau/T_c \rceil] = h(\mathbf{s})] = \chi_n(\tau,\, h(\mathbf{s})) A_n$, $n = 1, \cdots, N$, say, and subject to the first constraint (alone) in (4.29). The right side of (4.44) then equals

$$
\sum_{m=1}^{M} \left\{ \mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} s_{n,m} X_n(\tau), \lambda_{0,m} \right) \right| \mathbf{U}[\lceil \tau/T_c \rceil] = h(\mathbf{s}) \right]
$$

$$- \zeta \left(\sum_{n=1}^{N} s_{n,\,m} \chi_n(\tau,\, h(\mathbf{s})) A_n,\, \lambda_{0,\,m} \right) \Bigg\}, \tag{4.45}$$

and is maximized by considering the first term above. Let

$$\begin{aligned} & h(\tau,\, \mathbf{s},\, \boldsymbol{\chi}^N(\tau,\, h(\mathbf{s}))) \\ & = \max_{\substack{0 \le X_n(\tau) \le A_n \\ \mathbb{E}[X_n(\tau) | \mathbf{U}[\lceil \tau/T_c \rceil] = h(\mathbf{s})] = \chi_n(\tau,\, h(\mathbf{s})) A_n \\ n=1,\, \cdots,\, N}} \\ & \qquad \mathbb{E}\left[\sum_{m=1}^{M} \zeta \left(\sum_{n=1}^{N} s_{n,\,m} X_n(\tau),\, \lambda_{0,\,m} \right) \middle| \mathbf{U}[\lceil \tau/T_c \rceil] = h(\mathbf{s}) \right]. \end{aligned} \tag{4.46}$$

Noting the structural similarity of the optimization problems in (4.46) and (3.38), by (3.45), we conclude that

$$h(\tau,\, \mathbf{s},\, \boldsymbol{\chi}^N(\tau,\, h(\mathbf{s}))) \;\; = \;\; \sum_{n=1}^{N} \psi_{n,\, P_\tau}(\mathbf{s}) \chi_{P_\tau(n)}(\tau,\, h(\mathbf{s})), \tag{4.47}$$

where P_τ is a permutation of $\{1,\, \cdots,\, N\}$ such that

$$\chi_{P_\tau(n)}(\tau,\, h(\mathbf{s})) \;\; \ge \;\; \chi_{P_\tau(n+1)}(\tau,\, h(\mathbf{s})), \quad n = 1,\, \cdots,\, N-1, \tag{4.48}$$

and

$$\begin{aligned} \psi_{n,\, P}(\mathbf{s}) \;\; \triangleq \;\; \sum_{m=1}^{M} \Bigg\{ & \zeta \left(\sum_{k=1}^{n} s_{P(k),\, m} A_{P(k)},\, \lambda_{0,\, m} \right) \\ & - \zeta \left(\sum_{k=1}^{n-1} s_{P(k),\, m} A_{P(k)},\, \lambda_{0,\, m} \right) \Bigg\} \end{aligned} \tag{4.49}$$

for $n = 1,\, \cdots,\, N$ and any permutation P of $\{1,\, \cdots,\, N\}$. For notational brevity, we have suppressed the dependence of P_τ on $h(\mathbf{s})$. Since (4.29) implies that

$$\begin{aligned} 0 \;\; \le \;\; \chi_n(\tau,\, h(\mathbf{s})) \;\; & \le \;\; 1, \quad 0 \le \tau \le T, \quad \mathbf{s} \in (\mathbb{R}_0^+)^{N \times M}, \\ \tfrac{1}{T} \int_0^T \mathbb{E}[\chi_n(\tau,\, h(\mathbf{S}[\lceil \tau/T_c \rceil]))] d\tau \;\; & \le \;\; \sigma_n, \quad n = 1,\, \cdots,\, N, \end{aligned} \tag{4.50}$$

we thus see from (4.43)–(4.47), (4.50) that

$$
\begin{aligned}
&\frac{1}{T} I(W \wedge \phi(\mathbf{Y}_{ss}(T), \mathbf{S}^K)) \\
&\quad \leq \max_{\substack{\chi_n:[0,\,T]\times\mathcal{U}\to[0,\,1] \\ \frac{1}{T}\int_0^T \mathbb{E}[\chi_n(\tau, h(\mathbf{S}[\lceil\tau/T_c\rceil]))]d\tau \leq \sigma_n \\ n=1,\cdots,N}} \frac{1}{T}\int_0^T \mathbb{E}[g(\tau,\, \mathbf{S}[\lceil\tau/T_c\rceil],\, \boldsymbol{\chi}^N(\tau,\, h(\mathbf{S}[\lceil\tau/T_c\rceil])))]d\tau \\
&\quad = \max_{\substack{\chi_n:[0,\,T]\times\mathcal{U}\to[0,\,1] \\ \frac{1}{T}\int_0^T \mathbb{E}[\chi_n(\tau, \mathbf{U}[\lceil\tau/T_c\rceil])]d\tau \leq \sigma_n \\ n=1,\cdots,N}} \frac{1}{T}\int_0^T \mathbb{E}[g(\tau,\, \mathbf{S}[\lceil\tau/T_c\rceil],\, \boldsymbol{\chi}^N(\tau,\, \mathbf{U}[\lceil\tau/T_c\rceil]))]d\tau, \quad (4.51)
\end{aligned}
$$

where for $0 \leq \tau \leq T$, $\mathbf{s} \in (\mathbb{R}_0^+)^{N\times M}$, we have defined

$$
\begin{aligned}
g(\tau,\, \mathbf{s},\, \boldsymbol{\chi}^N(\tau,\, h(\mathbf{s}))) \;\triangleq\; & \sum_{n=1}^{N} \psi_{n,\,P_\tau}(\mathbf{s})\chi_{P_\tau(n)}(\tau,\, h(\mathbf{s})) \\
& - \sum_{m=1}^{M} \zeta\left(\sum_{n=1}^{N} s_{n,\,m}\chi_n(\tau,\, h(\mathbf{s}))A_n,\, \lambda_{0,\,m}\right). \quad (4.52)
\end{aligned}
$$

In order to simplify the right side of (4.51), for $\mathbf{u} \in \mathcal{U}$, define

$$
\eta_n(k, \mathbf{u}) \;=\; \frac{1}{T_c}\int_{(k-1)T_c}^{kT_c} \mathbb{E}[\chi_n(\tau,\, \mathbf{U}[k])|\mathbf{U}[k] = \mathbf{u}]d\tau, \quad k = 1,\, \cdots,\, K, \quad (4.53)
$$

$$
\mu_n(\mathbf{u}) \;=\; \frac{1}{K}\sum_{k=1}^{K} \eta_n(k, \mathbf{u}), \quad n = 1,\, \cdots,\, N, \quad (4.54)
$$

whence by (4.50), we get

$$
0 \;\leq\; \mu_n(\mathbf{u}) \;\leq\; 1, \quad \mathbf{u} \in \mathcal{U}, \quad n = 1,\, \cdots,\, N, \quad (4.55)
$$

and

$$
\begin{aligned}
\sigma_n \;\geq\; & \frac{1}{T}\int_0^T \mathbb{E}[\chi_n(\tau,\, \mathbf{U}[\lceil\tau/T_c\rceil])]d\tau \\
=\; & \frac{1}{K}\sum_{k=1}^{K}\frac{1}{T_c}\int_{(k-1)T_c}^{kT_c} \mathbb{E}[\chi_n(\tau,\, \mathbf{U}[k])]d\tau \\
=\; & \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}[\eta_n(k, \mathbf{U}[k])] \quad (4.56) \\
=\; & \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}[\eta_n(k, \mathbf{U})] \quad (4.57) \\
=\; & \mathbb{E}[\mu_n(\mathbf{U})], \quad (4.58)
\end{aligned}
$$

where (4.56) is by (4.53); (4.57) holds by the i.i.d. nature of channel fade $\mathbf{S}^{\infty}$; and (4.58) is by (4.54). The time-averaged integral on the right side of (4.51) can be written as

$$\begin{aligned}
&\frac{1}{T}\int_0^T \mathbb{E}[g(\tau,\, \mathbf{S}[\lceil \tau/T_c\rceil],\, \boldsymbol{\chi}^N(\tau,\, \mathbf{U}[\lceil \tau/T_c\rceil]))]d\tau \\
&\quad = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{T_c}\int_{(k-1)T_c}^{kT_c} \mathbb{E}[g(\tau,\, \mathbf{S}[k],\, \boldsymbol{\chi}^N(\tau,\, \mathbf{U}[k]))]d\tau \\
&\quad = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{T_c}\int_{(k-1)T_c}^{kT_c} \left\{\mathbb{E}\left[\sum_{n=1}^{N}\psi_{n,\,P_\tau}(\mathbf{S}[k])\chi_{P_\tau(n)}(\tau,\, \mathbf{U}[k])\right.\right. \\
&\qquad\qquad \left.\left. -\sum_{m=1}^{M}\zeta\left(\sum_{n=1}^{N} S_{n,\,m}[k]\chi_n(\tau,\, \mathbf{U}[k])A_n,\, \lambda_{0,\,m}\right)\right]\right\}d\tau, \qquad (4.59)
\end{aligned}$$

by (4.52). In Appendix C.2, we show that

$$\begin{aligned}
&\frac{1}{K}\sum_{k=1}^{K}\frac{1}{T_c}\int_{(k-1)T_c}^{kT_c} \mathbb{E}\left[\sum_{n=1}^{N}\psi_{n,\,P_\tau}(\mathbf{S}[k])\chi_{P_\tau(n)}(\tau,\, \mathbf{U}[k])\right]d\tau \\
&\quad \leq \mathbb{E}\left[\sum_{n=1}^{N}\nu_n(\mathbf{U})\sum_{m=1}^{M}\zeta\left(\sum_{j=1}^{n} S_{\Pi(j),\,m}A_{\Pi(j)},\, \lambda_{0,\,m}\right)\right], \qquad (4.60)
\end{aligned}$$

where Π is a permutation of $\{1,\, \cdots,\, N\}$ so that for $\mathbf{u}\in\mathcal{U}$,

$$\mu_{\Pi(n)}(\mathbf{u}) \;\geq\; \mu_{\Pi(n+1)}(\mathbf{u}), \quad n = 1,\, \cdots,\, N-1, \qquad (4.61)$$

(where for notational brevity, we have suppressed the dependence of Π on $\mathbf{u}$), and

$$\nu_n(\mathbf{u}) \;=\; \begin{cases} \mu_{\Pi(n)}(\mathbf{u}) - \mu_{\Pi(n+1)}(\mathbf{u}), & \text{if } n = 1,\, \cdots,\, N-1, \\ \mu_{\Pi(N)}(\mathbf{u}), & \text{if } n = N. \end{cases} \qquad (4.62)$$

Furthermore, note that

$$\begin{aligned}
&\frac{1}{K}\sum_{k=1}^{K}\frac{1}{T_c}\int_{(k-1)T_c}^{kT_c} \mathbb{E}\left[\sum_{m=1}^{M}\zeta\left(\sum_{n=1}^{N} S_{n,\,m}[k]\chi_n(\tau,\, \mathbf{U}[k])A_n,\, \lambda_{0,\,m}\right)\right]d\tau \\
&\quad \geq \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}\left[\sum_{m=1}^{M}\zeta\left(\sum_{n=1}^{N} S_{n,\,m}[k]\eta_n(k,\, \mathbf{U}[k])A_n,\, \lambda_{0,\,m}\right)\right] \qquad (4.63)
\end{aligned}$$

$$= \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}\left[\sum_{m=1}^{M}\zeta\left(\sum_{n=1}^{N}S_{n,\,m}\eta_n(k,\,\mathbf{U})A_n,\,\lambda_{0,\,m}\right)\right] \tag{4.64}$$

$$\geq \mathbb{E}\left[\sum_{m=1}^{M}\zeta\left(\sum_{n=1}^{N}S_{n,\,m}\mu_n(\mathbf{U})A_n,\,\lambda_{0,\,m}\right)\right] \tag{4.65}$$

$$= \mathbb{E}\left[\sum_{m=1}^{M}\zeta\left(\sum_{n=1}^{N}S_{\Pi(n),\,m}\mu_{\Pi(n)}(\mathbf{U})A_{\Pi(n)},\,\lambda_{0,\,m}\right)\right]$$

$$= \mathbb{E}\left[\sum_{m=1}^{M}\zeta\left(\sum_{n=1}^{N}\nu_n(\mathbf{U})\sum_{j=1}^{n}S_{\Pi(j),\,m}A_{\Pi(j)},\,\lambda_{0,\,m}\right)\right], \tag{4.66}$$

where (4.63) is by Jensen's inequality applied to the convex function $\zeta(\cdot,\,y)$ and (4.53); (4.64) holds by i.i.d. nature of $\mathbf{S}^\infty$; (4.65) is by Jensen's inequality and (4.54); and (4.66) is by application of (3.54) with the substitution $x_n = \mu_{\Pi(n)}(\mathbf{U})$, $y_n = S_{\Pi(n),\,m}A_{\Pi(n)}$. Summarizing collectively (4.31), (4.51), (4.55), (4.58), (4.59), (4.60), (4.66), we get

$$R \lessapprox \max_{\substack{\mu_n:\mathcal{U}\to[0,\,1]\\ \mathbb{E}[\mu_n(\mathbf{U})]\leq\sigma_n\\ n=1,\,\cdots,\,N}} \mathbb{E}\left[\sum_{n=1}^{N}\nu_n(\mathbf{U})\sum_{m=1}^{M}\zeta\left(\sum_{j=1}^{n}S_{\Pi(j),\,m}A_{\Pi(j)},\,\lambda_{0,\,m}\right)\right.$$
$$\left. -\sum_{m=1}^{M}\zeta\left(\sum_{n=1}^{N}\nu_n(\mathbf{U})\sum_{j=1}^{n}S_{\Pi(j),\,m}A_{\Pi(j)},\,\lambda_{0,\,m}\right)\right]. \tag{4.67}$$

This concludes the proof of the converse part of Theorem 7.

Achievability part: We closely follow Wyner's approach [52].

Fix $L \in \mathbb{Z}^+$ and set $\Delta \triangleq T_c/L$. Divide the time interval $[0,\,T]$, where $T = KT_c$ with $K \in \mathbb{Z}^+$, into KL equal subintervals, each of duration Δ. Then, in the channel fade sequence $\mathbf{S}^\infty = \{\mathbf{S}[k]\}_{k=1}^\infty$, each $\mathbf{S}[k]$ remains unvarying for a block of L consecutive Δ-duration subintervals within $[(k-1)T_c,\,kT_c]$, and $\{\mathbf{S}[k]\}_{k=1}^\infty$ varies across such blocks in an i.i.d. manner.

Now, consider the situation in which for each message $w \in \mathcal{W}$ and transmitter CSI $\mathbf{u}^K \in \mathcal{U}^K$, the channel input waveform at the n^{th} transmit aperture,

$\{x_n(w, t, \mathbf{u}^{\lceil t/T_c \rceil}), 0 \leq t \leq T\}$, is restricted to be $\{0, A_n\}$–valued and piecewise constant in each of the KL time slots of duration Δ. Define

$$\tilde{x}_n(w, l, \mathbf{u}^{\lceil l/L \rceil}) \triangleq \begin{cases} 0, & \text{if } x_n(w, t, \mathbf{u}^{\lceil t/T_c \rceil}) = 0, \\ 1, & \text{if } x_n(w, t, \mathbf{u}^{\lceil t/T_c \rceil}) = A_n, \\ & t \in [(l-1)\Delta, l\Delta), \quad l = 1, \cdots, KL. \end{cases} \tag{4.68}$$

Note that the condition (4.3) requires that

$$\frac{1}{KL} \sum_{l=1}^{KL} \tilde{x}_n(w, l, \mathbf{u}^{\lceil l/L \rceil}) \leq \sigma_n, \quad w \in \mathcal{W}, \quad \mathbf{u}^K \in \mathcal{U}^K, \quad n = 1, \cdots, N. \tag{4.69}$$

Next, consider a decoder $\phi : (\{0, 1\}^M)^{KL} \times ((\mathbb{R}_0^+)^{N \times M})^K \rightarrow \mathcal{W}$ based on restricted observations over the KL time slots, comprising

$$\tilde{Y}_m(l) = 1(Y_m(l\Delta) - Y_m((l-1)\Delta) = 1), \quad l = 1, \cdots, KL, \tag{4.70}$$

(with $Y_m(0) = 0$) and $\mathbf{S}^K$, at the m^{th} receive aperture, $m = 1, \cdots, M$. The largest achievable rate of restricted (W, T)–codes as above – and, hence, the capacity C – is clearly no smaller than $\frac{C(L)}{T_c}$, where $C(L)$ is the capacity of a $(L$ –) block discrete memoryless channel (in nats per block channel use) with input alphabet $\tilde{\mathcal{X}}^{NL} = \{0, 1\}^{NL}$; output alphabet $\tilde{\mathcal{Y}}^{ML} = \{0, 1\}^{ML}$; state alphabet $(\mathbb{R}_0^+)^{N \times M}$; transition pmf

$$\begin{aligned} W^{(L)}(\tilde{\mathbf{y}}^{ML} | \tilde{\mathbf{x}}^{NL}, \mathbf{s}) &= \prod_{l=1}^{L} \prod_{m=1}^{M} W^{(m)}_{\tilde{Y}_m | \tilde{\mathbf{X}}^N, \mathbf{S}}(\tilde{y}_m(l) | \tilde{\mathbf{x}}^N(l), \mathbf{s}), \\ & \tilde{\mathbf{x}}^{NL} \in \{0, 1\}^{NL}, \quad \tilde{\mathbf{y}}^{ML} \in \{0, 1\}^{ML}, \quad \mathbf{s} \in (\mathbb{R}_0^+)^{N \times M}, \end{aligned} \tag{4.71}$$

where $\tilde{X}_n$, $n = 1, \cdots, N$, $\tilde{Y}_m$, $m = 1, \cdots, M$, and $\mathbf{S}$ are $\tilde{\mathcal{X}}$ –, $\tilde{\mathcal{Y}}$ –, and $(\mathbb{R}_0^+)^{N \times M}$ –valued rvs respectively, and the channel transition pmf (corresponding to the m^{th}

receive aperture) $W^{(m)}_{\tilde{Y}_m|\tilde{\mathbf{X}}^N,\mathbf{S}}(\cdot|\cdot,\,\mathbf{s})$, $\mathbf{s} \in (\mathbb{R}_0^+)^{N\times M}$, is given by

$$\begin{aligned} W^{(m)}_{\tilde{Y}_m|\tilde{\mathbf{X}}^N,\mathbf{S}}(1|\tilde{\mathbf{x}}^N,\,\mathbf{s}) &= w_m(\tilde{\mathbf{x}}^N,\,\mathbf{s})\tfrac{T_c}{L}\exp\left(-w_m(\tilde{\mathbf{x}}^N,\,\mathbf{s})\tfrac{T_c}{L}\right) \triangleq \omega_m(\tilde{\mathbf{x}}^N,\,\mathbf{s},\,L), \\ W^{(m)}_{\tilde{Y}_m|\tilde{\mathbf{X}}^N,\mathbf{S}}(0|\tilde{\mathbf{x}}^N,\,\mathbf{s}) &= 1-\omega_m(\tilde{\mathbf{x}}^N,\,\mathbf{s},\,L), \end{aligned} \tag{4.72}$$

where

$$w_m(\tilde{\mathbf{x}}^N,\,\mathbf{s}) \triangleq \sum_{n=1}^{N} s_{n,m}\tilde{x}_n A_n + \lambda_{0,m}, \quad \tilde{\mathbf{x}}^N \in \{0,\,1\}^N, \quad \mathbf{s} \in (\mathbb{R}_0^+)^{N\times M}, \tag{4.73}$$

with transmitter CSI h (and perfect receiver CSI); and under constraint (4.69). From ([9], Remark A2 following Proposition 1), it is readily obtained[3] that

$$C(L) = \max_{\substack{P_{\tilde{\mathbf{X}}^{NL}|\mathbf{U}}:\sum_{l=1}^{L}\mathbf{E}[\tilde{X}_n(l)]\le L\sigma_n \\ n=1,\cdots,N}} I(\tilde{\mathbf{X}}^{NL}\wedge\tilde{\mathbf{Y}}^{ML}|\mathbf{S}), \tag{4.74}$$

with $\mathbf{U} = h(\mathbf{S})$, where the joint conditional pmf $P_{\tilde{\mathbf{X}}^{NL},\tilde{\mathbf{Y}}^{ML}|\mathbf{S}}$ is given by

$$\begin{aligned} & P_{\tilde{\mathbf{X}}^{NL},\tilde{\mathbf{Y}}^{ML}|\mathbf{S}}(\tilde{\mathbf{x}}^{NL},\,\tilde{\mathbf{y}}^{ML}|\mathbf{s}) \\ &= P_{\tilde{\mathbf{X}}^{NL}|\mathbf{U}}(\tilde{\mathbf{x}}^{NL}|h(\mathbf{s}))W^{(L)}(\tilde{\mathbf{y}}^{ML}|\tilde{\mathbf{x}}^{NL},\,\mathbf{s}), \\ &\quad \tilde{\mathbf{x}}^{NL} \in \{0,\,1\}^{NL}, \quad \tilde{\mathbf{y}}^{ML} \in \{0,\,1\}^{ML}, \quad \mathbf{s} \in (\mathbb{R}_0^+)^{N\times M}. \end{aligned} \tag{4.75}$$

By a standard argument which uses (4.71), the maximum in (4.74) is achieved by

$$\begin{aligned} P_{\tilde{\mathbf{X}}^{NL}|\mathbf{U}}(\tilde{\mathbf{x}}^{NL}|h(\mathbf{s})) &= \prod_{l=1}^{L} P_{\tilde{\mathbf{X}}^N(l)|\mathbf{U}}(\tilde{\mathbf{x}}^N(l)|h(\mathbf{s})), \\ &\qquad \tilde{\mathbf{x}}^{NL} \in \{0,\,1\}^{NL}, \quad \mathbf{s} \in (\mathbb{R}_0^+)^{N\times M}, \end{aligned} \tag{4.76}$$

so that from (4.74),

$$C(L) = L \max_{\substack{P_{\tilde{\mathbf{X}}^N|\mathbf{U}}:\mathbf{E}[\tilde{X}_n]\le\sigma_n \\ n=1,\cdots,N}} \sum_{m=1}^{M} I(\tilde{\mathbf{X}}^N\wedge\tilde{Y}_m|\mathbf{S}), \tag{4.77}$$

[3] In [9], S is taken to be finite-valued; however the result (4.74) can be seen to hold also when S is $\mathbb{R}_0^+$-valued.

where $P_{\tilde{\mathbf{X}}^N,\,\tilde{Y}_m|\mathbf{S}}$ is given by

$$P_{\tilde{\mathbf{X}}^N,\,\tilde{Y}_m|\mathbf{S}}(\tilde{\mathbf{x}}^N,\,\tilde{y}|\mathbf{s}) = P_{\tilde{\mathbf{X}}^N|\mathbf{U}}(\tilde{\mathbf{x}}^N|h(\mathbf{s}))W_{\tilde{Y}_m|\tilde{\mathbf{X}}^N,\mathbf{S}}(\tilde{y}|\tilde{\mathbf{x}}^N,\,\mathbf{s}),$$
$$\tilde{\mathbf{x}}^N \in \{0,\,1\}^N, \quad \tilde{y} \in \{0,\,1\}, \quad \mathbf{s} \in (\mathbb{R}_0^+)^{N\times M}. \tag{4.78}$$

For $\mathbf{u} \in \mathcal{U}$, $i = 0, \cdots, 2^N - 1$, define

$$p_i(\mathbf{u}) \triangleq \Pr\{\tilde{X}_n = b_n(i),\, n = 1, \cdots, N|\mathbf{U} = \mathbf{u}\}, \tag{4.79}$$

where $b_n(i)$ is the n^{th} bit in the binary representation of i. Clearly, $\mathbf{p}(\cdot) = \{p_i(\cdot)\}_{i=0}^{2^N-1}$ satisfies the constraints

$$\begin{aligned} p_i(\mathbf{u}) &\geq 0, \quad i = 0, \cdots, 2^N - 1, \\ \textstyle\sum_{i=0}^{2^N-1} p_i(\mathbf{u}) &= 1, \\ \textstyle\sum_{i:b_n(i)=1} \mathbb{E}[p_i(\mathbf{U})] &\leq \sigma_n, \quad n = 1, \cdots, N, \end{aligned} \tag{4.80}$$

where the last inequality follows from the constraints $\mathbb{E}[\tilde{X}_n] \leq \sigma_n$, $n = 1, \cdots, N$. By (4.72) and (4.79), it then follows that for $m = 1, \cdots, M$,

$$\begin{aligned} \Pr\{\tilde{Y}_m = 1|\mathbf{S} = \mathbf{s}\} &= \textstyle\sum_{i=0}^{2^N-1} p_i(h(\mathbf{s}))\omega_m\left(\mathbf{b}^N(i),\,\mathbf{s}\right), \\ H\left(\tilde{Y}_m|\tilde{\mathbf{X}}^N,\,\mathbf{S}\right) &= \mathbb{E}\left[\textstyle\sum_{i=0}^{2^N-1} p_i(h(\mathbf{S}))h_b\left(\omega_m\left(\mathbf{b}^N(i),\,\mathbf{S}\right)\right)\right], \\ H(\tilde{Y}_m|\mathbf{S}) &= \mathbb{E}\left[h_b\left(\textstyle\sum_{i=0}^{2^N-1} p_i(h(\mathbf{S}))\omega_m\left(\mathbf{b}^N(i),\,\mathbf{S}\right)\right)\right]. \end{aligned} \tag{4.81}$$

By (4.77) and (4.81), it follows that

$$C(L) = \max_{\substack{p_i:\mathcal{U}\to[0,\,1] \\ i=0,\,\cdots,\,2^N-1}} L\,\mathbb{E}[\beta_L(\mathbf{p}(\mathbf{U}),\,\mathbf{S})], \tag{4.82}$$

where $\mathbf{p}(\cdot) = \{p_i(\cdot)\}_{i=0}^{2^N-1}$ satisfies (4.80), and

$$\begin{aligned} \beta_L(\mathbf{p},\,\mathbf{s}) \triangleq \sum_{m=1}^{M}\Bigg[& h_b\left(\sum_{i=0}^{2^N-1} p_i\omega_m\left(\mathbf{b}^N(i),\,\mathbf{s},\,L\right)\right) \\ & - \sum_{i=0}^{2^N-1} p_i h_b\left(\omega_m\left(\mathbf{b}^N(i),\,\mathbf{s},\,L\right)\right)\Bigg], \\ & \mathbf{p} \in [0,\,1]^{2^N-1}, \ \ \mathbf{s} \in (\mathbb{R}_0^+)^{N\times M}. \end{aligned} \tag{4.83}$$

Since $L \in \mathbb{Z}^+$ was arbitrary, we have

$$\begin{aligned} C &\geq \lim_{L\to\infty} \frac{C(L)}{T_c} \\ &\geq \max_{\substack{p_i:\mathcal{U}\to[0,\,1] \\ i=0,\,\cdots,\,2^N-1}} \lim_{L\to\infty} \frac{\mathbb{E}[\beta_L(\mathbf{p}(\mathbf{U}),\,\mathbf{S})]}{T_c/L} \qquad (4.84) \\ &= \max_{\substack{p_i:\mathcal{U}\to[0,\,1] \\ i=0,\,\cdots,\,2^N-1}} \sum_{m=1}^{M} \mathbb{E}\left[\sum_{i=0}^{2^N-1} p_i(\mathbf{U})\zeta_m\left(\sum_{n=1}^{N} S_{n,m}b_n(i)A_n, \lambda_{0,m}\right)\right. \\ &\qquad \left. -\zeta_m\left(\sum_{n=1}^{N} S_{n,m}A_n \sum_{i=0}^{2^N-1} p_i(\mathbf{U})b_n(i), \lambda_{0,m}\right)\right], \qquad (4.85) \end{aligned}$$

with $\mathbf{p}(\cdot) = \{p_i(\cdot)\}_{i=0}^{2^N-1}$ satisfying (4.80), where (4.84) is by (4.82); and (4.85) is established in Appendix C.3. Setting

$$\begin{aligned} \mu_n(\mathbf{u}) &= \sum_{i=0}^{2^N-1} p_i(\mathbf{u})b_n(i) \\ &= \sum_{i:b_n(i)=1} p_i(\mathbf{u}), \quad n = 1,\,\cdots,\,N, \quad \mathbf{u} \in \mathcal{U}, \qquad (4.86) \end{aligned}$$

by (4.80) and (4.85), it follows that

$$C \geq \max_{\substack{\mu_n:\mathcal{U}\to[0,\,1] \\ \mathbf{E}[\mu_n(\mathbf{U})]\leq\sigma_n \\ n=1,\,\cdots,\,N}} \mathbb{E}\left[h(\boldsymbol{\mu}^N(\mathbf{U}),\,\mathbf{S}) - \sum_{m=1}^{M}\zeta_m\left(\sum_{n=1}^{N} S_{n,\,m}\mu_n(\mathbf{U})A_n,\,\lambda_{0,\,m}\right)\right], \qquad (4.87)$$

where

$$\begin{aligned} h(\boldsymbol{\mu}^N,\,\mathbf{s}) &= \max_{\mathbf{p}} \sum_{i=0}^{2^N-1} p_i \sum_{m=1}^{M} \zeta_m\left(\sum_{n=1}^{N} s_{n,\,m}b_n(i)A_n,\,\lambda_{0,\,m}\right), \\ &\qquad \boldsymbol{\mu}^N \in [0,\,1]^N, \;\; \mathbf{s} \in (\mathbb{R}_0^+)^{N\times M}, \qquad (4.88) \end{aligned}$$

with $\mathbf{p} = \{p_i\}_{i=0}^{2^N-1}$ satisfying the constraints

$$\begin{aligned} p_i &\geq 0, \quad i = 0,\,\cdots,\,2^N-1, \\ \textstyle\sum_{i=0}^{2^N-1} p_i &= 1, \qquad (4.89) \\ \textstyle\sum_{i:b_n(i)=1} p_i &= \mu_n, \quad n = 1,\,\cdots,\,N. \end{aligned}$$

Comparing (3.78) and (4.88), it follows that the optimum joint pmf $\mathbf{p}^*$ that obtains the maximum in (4.88) is given by

$$p_i^* = \begin{cases} \mu_{Q(k)} - \mu_{Q(k+1)}, & \text{if } i = \sum_{n=1}^{k} 2^{Q(n)-1}, \quad k = 1, \cdots, N-1, \\ \mu_{Q(N)}, & \text{if } i = 2^N - 1, \\ 1 - \mu_{Q(1)}, & \text{if } i = 0, \\ 0, & \text{otherwise,} \end{cases} \tag{4.90}$$

where Q is a permutation of $\{1, \cdots, N\}$ such that

$$\mu_{Q(n)} \geq \mu_{Q(n+1)}, \quad n = 1, \cdots, N-1,$$

and the corresponding largest value is

$$h(\boldsymbol{\mu}^N, \mathbf{s}) = \sum_{n=1}^{N} \mu_{Q(n)} \psi_{n,Q}(\mathbf{s}), \tag{4.91}$$

where $\psi_{n,P}(\cdot)$ is as defined in (4.49). Summarizing collectively (4.87), (4.88), (4.91), we get

$$\begin{aligned} C \geq & \max_{\substack{\mu_n: \mathcal{U} \to [0,1] \\ \mathbf{E}[\mu_n(\mathbf{U})] \leq \sigma_n \\ n=1, \cdots, N}} \mathbb{E}\left[\sum_{n=1}^{N} \mu_{\Pi(n)}(\mathbf{U}) \psi_{n,\Pi}(\mathbf{S}) \right. \\ & \left. - \sum_{m=1}^{M} \zeta\left(\sum_{n=1}^{N} S_{n,m} \mu_n(\mathbf{U}) A_n, \lambda_{0,m} \right) \right] \qquad (4.92) \\ = & \max_{\substack{\mu_n: \mathcal{U} \to [0,1] \\ \mathbf{E}[\mu_n(\mathbf{U})] \leq \sigma_n \\ n=1, \cdots, N}} \sum_{m=1}^{M} \mathbb{E}\left[\sum_{n=1}^{N} \nu_n(\mathbf{U}) \zeta\left(\sum_{k=1}^{n} S_{\Pi(k),m} A_{\Pi(k)}, \lambda_{0,m} \right) \right. \\ & \left. - \zeta\left(\sum_{n=1}^{N} \nu_n(\mathbf{U}) \sum_{k=1}^{n} S_{\Pi(k),m} A_{\Pi(k)}, \lambda_{0,m} \right) \right], \qquad (4.93) \end{aligned}$$

where $\Pi(\cdot)$, $\{\nu_n(\cdot)\}_{n=1}^N$ are as defined in (4.61), (4.62) respectively, and (4.93) is by (3.54) (see (3.55), (3.56) for similar applications of (3.54)). This concludes the proof of the achievability part of Theorem 7. $\square$

Proof of Lemma 1:

Property 1: The first property follows directly from the definition of mirror states and the isotropic nature of channel fade matrix $\mathbf{S}$.

Property 2: We prove $K_1(\mathbf{s}^{(1)}) = -K_2(\mathbf{s}^{(2)})$. The proof of $K_1(\mathbf{s}^{(2)}) = -K_2(\mathbf{s}^{(1)})$ follows in a similar manner. Noting that $A_1 = A_2 = A$, $\lambda_{0,m} = \lambda_0$, $m = 1, \cdots, M$, by the definition of mirror states,

$$
\begin{aligned}
b_{1,m}^{(1)} &= s_{1,m}^{(1)}\tfrac{A}{\lambda_0} = s_{2,m}^{(2)}\tfrac{A}{\lambda_0} = b_{2,m}^{(2)}, \\
b_{2,m}^{(1)} &= s_{2,m}^{(1)}\tfrac{A}{\lambda_0} = s_{1,m}^{(2)}\tfrac{A}{\lambda_0} = b_{1,m}^{(2)},
\end{aligned}
\tag{4.94}
$$

and

$$
B_m^{(1)} = (s_{1,m}^{(1)} + s_{2,m}^{(1)})\tfrac{A}{\lambda_0} = (s_{2,m}^{(2)} + s_{1,m}^{(2)})\tfrac{A}{\lambda_0} = B_m^{(2)}, \tag{4.95}
$$

for $m = 1, \cdots, M$, so that by (4.7), (4.8),

$$
\begin{aligned}
\rho_1(\mathbf{s}^{(1)}) &= \rho_2(\mathbf{s}^{(2)}), \\
\rho_1(\mathbf{s}^{(2)}) &= \rho_2(\mathbf{s}^{(1)}), \\
\bar{\rho}(\mathbf{s}^{(1)}) &= \bar{\rho}(\mathbf{s}^{(2)}).
\end{aligned}
\tag{4.96}
$$

By (4.9), (4.94)–(4.96), it follows that

$$
\begin{aligned}
K_1(\mathbf{s}^{(1)}) &= \lambda_0 \textstyle\sum_{m=1}^{M} \left[B_m^{(1)} \log\left(1 + \bar{\rho}(\mathbf{s}^{(1)}) B_m^{(1)}\right) - 2b_{2,m}^{(1)} \log\left(1 + \rho_2(\mathbf{s}^{(1)}) b_{2,m}^{(1)}\right) \right] \\
&= \lambda_0 \textstyle\sum_{m=1}^{M} \left[B_m^{(2)} \log\left(1 + \bar{\rho}(\mathbf{s}^{(2)}) B_m^{(2)}\right) - 2b_{1,m}^{(2)} \log\left(1 + \rho_1(\mathbf{s}^{(2)}) b_{1,m}^{(2)}\right) \right] \\
&= -K_2(\mathbf{s}^{(2)}).
\end{aligned}
$$

Property 3: Without loss of generality, let $a \geq b \geq 0$. By (3.8), with $N = 2$,

$$
\begin{aligned}
I(a,\, b,\, \mathbf{s}^{(1)}) = {} & (a - b) \sum_{m=1}^{M} \zeta\left(s_{1,m}^{(1)} A,\, \lambda_0\right) + b \sum_{m=1}^{M} \zeta\left((s_{1,m}^{(1)} + s_{2,m}^{(1)}) A,\, \lambda_0\right) \\
& - \sum_{m=1}^{M} \zeta\left(a s_{1,m}^{(1)} A + b s_{2,m}^{(1)} A,\, \lambda_0\right)
\end{aligned}
$$

$$
\begin{aligned}
&= (a-b)\sum_{m=1}^{M}\zeta\left(s_{2,m}^{(2)}A,\,\lambda_0\right)+b\sum_{m=1}^{M}\zeta\left((s_{2,m}^{(2)}+s_{1,m}^{(2)})A,\,\lambda_0\right)\\
&\quad-\sum_{m=1}^{M}\zeta\left(as_{2,m}^{(2)}A+bs_{1,m}^{(2)}A,\,\lambda_0\right)\\
&= I(b,\,a,\,\mathbf{s}^{(2)}).
\end{aligned}
$$

Property 4: By property 3, we get that $I(a_2,\,b_2,\,\mathbf{s}^{(2)}) = I(b_2,\,a_2,\,\mathbf{s}^{(1)})$. Therefore, it suffices to prove that

$$
\begin{aligned}
&D(a_1,\,a_2,\,b_1,\,b_2,\,\mathbf{s}^{(1)})\\
&\triangleq 2I\left(\frac{a_1+b_2}{2},\,\frac{a_2+b_1}{2},\,\mathbf{s}^{(1)}\right)-I(a_1,\,b_1,\,\mathbf{s}^{(1)})-I(b_2,\,a_2,\,\mathbf{s}^{(1)})\\
&\geq 0.
\end{aligned}
$$

From (3.8), after rearrangement, we get

$$
D(a_1,\,a_2,\,b_1,\,b_2,\,\mathbf{s}^{(1)}) = Z(a_1,\,a_2,\,b_1,\,b_2,\,\mathbf{s}^{(1)})+L(a_1,\,a_2,\,b_1,\,b_2,\,\mathbf{s}^{(1)}),
$$

where

$$
\begin{aligned}
&Z(a_1,\,a_2,\,b_1,\,b_2,\,\mathbf{s}^{(1)})\\
&= \sum_{m=1}^{M}\Big[\zeta\left(a_1s_{1,m}^{(1)}A+b_1s_{2,m}^{(1)}A,\,\lambda_0\right)+\zeta\left(b_2s_{1,m}^{(1)}A+a_2s_{2,m}^{(1)}A,\,\lambda_0\right)\\
&\quad-2\zeta\left(\frac{a_1+b_2}{2}s_{1,m}^{(1)}A+\frac{b_1+a_2}{2}s_{2,m}^{(1)}A,\,\lambda_0\right)\Big],
\end{aligned}
\tag{4.97}
$$

and

$$
\begin{aligned}
&L(a_1,\,a_2,\,b_1,\,b_2,\,\mathbf{s}^{(1)})\\
&= \left([a_1-b_1]^+ + [b_2-a_2]^+ - [a_1-b_1+b_2-a_2]^+\right)\times\\
&\quad\sum_{m=1}^{M}\left(\zeta(s_{1,m}^{(1)}A+s_{2,m}^{(1)}A,\,\lambda_0)-\zeta(s_{1,m}^{(1)}A,\,\lambda_0)-\zeta(s_{2,m}^{(1)}A,\,\lambda_0)\right),
\end{aligned}
\tag{4.98}
$$

where we use the notation $[x]^+ = \max\{x, 0\}$. By Jensen's inequality applied to the convex function $\zeta(\cdot, \lambda_0)$, from (4.97) we get $Z(a_1,\, a_2,\, b_1,\, b_2,\, \mathbf{s}^{(1)}) \geq 0$. Finally, noting that $[x]^+ + [y]^+ \geq [x+y]^+$, $x, y \in \mathbb{R}$, by (3.6) and (4.98), we get $L(a_1,\, a_2,\, b_1,\, b_2,\, \mathbf{s}^{(1)}) \geq 0$, so that $D(a_1,\, a_2,\, b_1,\, b_2,\, \mathbf{s}^{(1)}) \geq 0$. $\square$

Proof of Lemma 2:

Suppose $p_1 : (\mathbb{R}_0^+)^{2\times M} \to [0,\, 1]$ and $p_2 : (\mathbb{R}_0^+)^{2\times M} \to [0,\, 1]$ constitute a feasible power control law, i.e., $\mathbb{E}[p_n(\mathbf{S})] \leq \sigma$, $n = 1,\, 2$. For every $\mathbf{s} \in (\mathbb{R}_0^+)^{2\times M}$, define $p_1^{'}(\mathbf{s}) = \frac{1}{2}(p_1(\mathbf{s}) + p_2(\mathbf{s}^{mir}))$, and $p_2^{'}(\mathbf{s}) = \frac{1}{2}(p_1(\mathbf{s}^{mir}) + p_2(\mathbf{s}))$, where $\mathbf{s}^{mir}$ is the mirror state of $\mathbf{s}$. Clearly, $p_n^{'}(\mathbf{s}) \in [0,\, 1]$, $\mathbf{s} \in (\mathbb{R}_0^+)^{2\times M}$, $n = 1,\, 2$, and by property 1 of Lemma 1, it follows that

$$\mathbb{E}[p_1^{'}(\mathbf{S})] = \mathbb{E}[p_2^{'}(\mathbf{S})] = \frac{1}{2}\,\mathbb{E}[p_1(\mathbf{S}) + p_2(\mathbf{S})] \leq \sigma.$$

Therefore, $\{p_1^{'}(\cdot),\, p_2^{'}(\cdot)\}$ is a feasible power control law. Furthermore, by property 4 of Lemma 1, it follows that for any $\mathbf{s} \in (\mathbb{R}_0^+)^{2\times M}$ (and the corresponding mirror state $\mathbf{s}^{mir}$)

$$\begin{aligned} I(p_1^{'}(\mathbf{s}),\, p_2^{'}(\mathbf{s}),\, \mathbf{s}) &= I\left(\frac{p_1(\mathbf{s}) + p_2(\mathbf{s}^{mir})}{2},\, \frac{p_1(\mathbf{s}^{mir}) + p_2(\mathbf{s})}{2},\, \mathbf{s}\right) \\ &\geq \frac{1}{2}\left[I(p_1(\mathbf{s}),\, p_2(\mathbf{s}),\, \mathbf{s}) + I(p_1(\mathbf{s}^{mir}),\, p_2(\mathbf{s}^{mir}),\, \mathbf{s}^{mir})\right], \end{aligned}$$

so that

$$\mathbb{E}[I(p_1^{'}(\mathbf{S}),\, p_2^{'}(\mathbf{S}),\, \mathbf{S})] \geq \mathbb{E}[I(p_1(\mathbf{S}),\, p_2(\mathbf{S}),\, \mathbf{S})].$$

In other words, given any feasible power control law $\{p_1(\cdot),\, p_2(\cdot)\}$ satisfying the constraints $\mathbb{E}[p_n(\mathbf{S})] \leq \sigma$, $n = 1,\, 2$, we can obtain another feasible power control law $\{p_1^{'}(\cdot),\, p_2^{'}(\cdot)\}$ satisfying $\mathbb{E}[p_n^{'}(\mathbf{S})] \leq \sigma$, $n = 1,\, 2$, and the conditions $p_1^{'}(\mathbf{s}) =$

$p_2^{'}(\mathbf{s}^{mir})$, and $p_2^{'}(\mathbf{s}) = p_1^{'}(\mathbf{s}^{mir})$ for all pairs of mirror states $\mathbf{s}$ and $\mathbf{s}^{mir}$, such that $\mathbb{E}[I(p_1^{'}(\mathbf{S}), p_2^{'}(\mathbf{S}), \mathbf{S})] \geq \mathbb{E}[I(p_1(\mathbf{S}), p_2(\mathbf{S}), \mathbf{S})]$. Therefore, the optimal power control law $\{\mu_1^*(\cdot), \mu_2^*(\cdot)\}$ must satisfy $\mu_1^*(\mathbf{s}) = \mu_2^*(\mathbf{s}^{mir})$, for all pairs of mirror states $\mathbf{s}$ and $\mathbf{s}^{mir}$. This completes the proof of Lemma 2. □

Proof of Theorem 8:

Let $\mu_n^* : (\mathbb{R}_0^+)^{2\times M} \to [0, 1]$, $n = 1, 2$, constitute an optimal power control law in (4.5) with $N = 2$. At optimality,

$$\begin{aligned}
\mathbb{E}_{F_{\mathbf{S}}}[\mu_1^*(\mathbf{S})] &= \int_{\mathcal{S}_0} \mu_1^*(\mathbf{s}) dF_{\mathbf{S}}(\mathbf{s}) + \int_{\mathcal{S}_1} \mu_1^*(\mathbf{s}) dF_{\mathbf{S}}(\mathbf{s}) + \int_{\mathcal{S}_2} \mu_1^*(\mathbf{s}) dF_{\mathbf{S}}(\mathbf{s}) \\
&= \int_{\mathcal{S}_0} \frac{1}{2}(\mu_1^*(\mathbf{s}) + \mu_2^*(\mathbf{s})) dF_{\mathbf{S}}(\mathbf{s}) + \int_{\mathcal{S}_1} (\mu_1^*(\mathbf{s}) + \mu_2^*(\mathbf{s})) dF_{\mathbf{S}}(\mathbf{s}) \quad (4.99) \\
&= \frac{1}{2} \mathbb{E}_{\hat{F}_{\mathbf{S}}}[\mu_1^*(\mathbf{S}) + \mu_2^*(\mathbf{S})], \quad (4.100)
\end{aligned}$$

where (4.99) is by Lemma 1, property 1 and Lemma 2; and (4.100) is by (4.14). Similarly,

$$\mathbb{E}_{F_{\mathbf{S}}}[\mu_2^*(\mathbf{S})] = \frac{1}{2} \mathbb{E}_{\hat{F}_{\mathbf{S}}}[\mu_1^*(\mathbf{S}) + \mu_2^*(\mathbf{S})]. \quad (4.101)$$

Furthermore, for every pair of mirror states $\{\mathbf{s}, \mathbf{s}^{mir}\}$,

$$\begin{aligned}
I(\mu_1^*(\mathbf{s}), \mu_2^*(\mathbf{s}), \mathbf{s}) &= I(\mu_2^*(\mathbf{s}), \mu_1^*(\mathbf{s}), \mathbf{s}^{mir}) \quad (4.102) \\
&= I(\mu_1^*(\mathbf{s}^{mir}), \mu_2^*(\mathbf{s}^{mir}), \mathbf{s}^{mir}), \quad (4.103)
\end{aligned}$$

where (4.102) is by Lemma 1, property 3; and (4.103) is by Lemma 2. By property 1 of Lemma 1, we now get

$$\mathbb{E}_{F_{\mathbf{S}}}[I(\mu_1^*(\mathbf{S}), \mu_2^*(\mathbf{S}), \mathbf{S})] = \mathbb{E}_{\hat{F}_{\mathbf{S}}}[I(\mu_1^*(\mathbf{S}), \mu_2^*(\mathbf{S}), \mathbf{S})]. \quad (4.104)$$

Summarizing collectively (4.5), (4.100), (4.101), (4.104), we get

$$C \leq \max_{\substack{\mu_1:(\mathbb{R}_0^+)^{2\times M}\to[0,\,1]\\ \mu_2:(\mathbb{R}_0^+)^{2\times M}\to[0,\,1]\\ \mathbf{E}_{\hat{F}_{\mathbf{S}}}[\mu_1(\mathbf{S})+\mu_2(\mathbf{S})]\leq 2\sigma}} \mathbb{E}_{\hat{F}_{\mathbf{S}}}[I(\mu_1(\mathbf{S}),\, \mu_2(\mathbf{S}),\, \mathbf{S})]. \tag{4.105}$$

Note that the optimization problem in the right side of (4.105) bears a structural resemblance to the capacity formula of the MIMO Poisson channel with constant fade and an average sum power constraint (see (3.11)). Suppose $\mu_1^{\diamond} : (\mathbb{R}_0^+)^{2\times M} \to [0,\,1]$ and $\mu_2^{\diamond} : (\mathbb{R}_0^+)^{2\times M} \to [0,\,1]$ constitute an optimal solution that maximizes the right side of (4.105). Then, by Corollary 6,

$$\begin{aligned} \mu_1^{\diamond}(\mathbf{s}) &= \mu_2^{\diamond}(\mathbf{s}), \quad \mathbf{s} \in \mathcal{S}_0, \\ \mu_1^{\diamond}(\mathbf{s}) &\geq \mu_2^{\diamond}(\mathbf{s}), \quad \mathbf{s} \in \mathcal{S}_1. \end{aligned} \tag{4.106}$$

For $\mathbf{s} \in (\mathbb{R}_0^+)^{2\times M}$, define

$$\mu_1^{\dagger}(\mathbf{s}) = \begin{cases} \mu_1^{\diamond}(\mathbf{s}), & \text{if } \mathbf{s} \in \mathcal{S}_0 \cup \mathcal{S}_1, \\ \mu_2^{\diamond}(\mathbf{s}^{mir}), & \text{otherwise}, \end{cases} \tag{4.107}$$

and

$$\mu_2^{\dagger}(\mathbf{s}) = \begin{cases} \mu_2^{\diamond}(\mathbf{s}), & \text{if } \mathbf{s} \in \mathcal{S}_0 \cup \mathcal{S}_1, \\ \mu_1^{\diamond}(\mathbf{s}^{mir}), & \text{otherwise}, \end{cases} \tag{4.108}$$

where $\mathbf{s}^{mir}$ is the mirror state of $\mathbf{s}$. By (4.14), (4.107), (4.108), it follows that

$$\mathbb{E}_{\hat{F}_{\mathbf{S}}}[\mu_1^{\diamond}(\mathbf{S}) + \mu_2^{\diamond}(\mathbf{S})] = \mathbb{E}_{\hat{F}_{\mathbf{S}}}[\mu_1^{\dagger}(\mathbf{S}) + \mu_2^{\dagger}(\mathbf{S})],$$

and

$$\mathbb{E}_{\hat{F}_{\mathbf{S}}}[I(\mu_1^{\diamond}(\mathbf{S}),\, \mu_2^{\diamond}(\mathbf{S}),\, \mathbf{S})] = \mathbb{E}_{\hat{F}_{\mathbf{S}}}[I(\mu_1^{\dagger}(\mathbf{S}),\, \mu_2^{\dagger}(\mathbf{S}),\, \mathbf{S})].$$

Therefore, the pair $\{\mu_1^\dagger(\cdot),\, \mu_2^\dagger(\cdot)\}$ also constitutes an optimal solution that maximizes the right side of (4.105). Furthermore,

$$\begin{aligned}
\mathbb{E}_{F_\mathbf{S}}[\mu_1^\dagger(\mathbf{S})] &= \int_{\mathcal{S}_0} \mu_1^\diamond(\mathbf{s})dF_\mathbf{S}(\mathbf{s}) + \int_{\mathcal{S}_1} \mu_1^\diamond(\mathbf{s})dF_\mathbf{S}(\mathbf{s}) + \int_{\mathcal{S}_2} \mu_2^\diamond(\mathbf{s}^m)dF_\mathbf{S}(\mathbf{s}) \\
&= \int_{\mathcal{S}_0} \frac{1}{2}(\mu_1^\diamond(\mathbf{s}) + \mu_2^\diamond(\mathbf{s}))dF_\mathbf{S}(\mathbf{s}) + \int_{\mathcal{S}_1} (\mu_1^\diamond(\mathbf{s}) + \mu_2^\diamond(\mathbf{s}))dF_\mathbf{S}(\mathbf{s}) \quad (4.109) \\
&= \frac{1}{2}\,\mathbb{E}_{\hat{F}_\mathbf{S}}[\mu_1^\diamond(\mathbf{S}) + \mu_2^\diamond(\mathbf{S})] \quad (4.110) \\
&\le \sigma, \quad (4.111)
\end{aligned}$$

where (4.109) is by (4.106) and Lemma 1, property 1; (4.110) is by (4.14); and (4.111) is by (4.105). Similarly,

$$\mathbb{E}_{F_\mathbf{S}}[\mu_2^\dagger(\mathbf{S})] = \frac{1}{2}\,\mathbb{E}_{\hat{F}_\mathbf{S}}[\mu_1^\diamond(\mathbf{S}) + \mu_2^\diamond(\mathbf{S})] \le \sigma. \quad (4.112)$$

By (4.5), (4.111), (4.112), it thus follows that

$$C \ge \mathbb{E}_{F_\mathbf{S}}[I(\mu_1^\dagger(\mathbf{S}),\, \mu_2^\dagger(\mathbf{S}),\, \mathbf{S})]. \quad (4.113)$$

Proceeding further with the right side of (4.113), we get that

$$\begin{aligned}
&\mathbb{E}_{F_\mathbf{S}}[I(\mu_1^\dagger(\mathbf{S}),\, \mu_2^\dagger(\mathbf{S}),\, \mathbf{S})] \\
&= \int_{\mathcal{S}_0\cup\mathcal{S}_1\cup\mathcal{S}_2} I(\mu_1^\dagger(\mathbf{s}),\, \mu_2^\dagger(\mathbf{s}),\, \mathbf{s})dF_\mathbf{S}(\mathbf{s}) \\
&= \int_{\mathcal{S}_0\cup\mathcal{S}_1} I(\mu_1^\diamond(\mathbf{s}),\, \mu_2^\diamond(\mathbf{s}),\, \mathbf{s})dF_\mathbf{S}(\mathbf{s}) + \int_{\mathcal{S}_2} I(\mu_2^\diamond(\mathbf{s}^{mir}),\, \mu_1^\diamond(\mathbf{s}^{mir}),\, \mathbf{s})dF_\mathbf{S}(\mathbf{s}) \quad (4.114) \\
&= \int_{\mathcal{S}_0\cup\mathcal{S}_1} I(\mu_1^\diamond(\mathbf{s}),\, \mu_2^\diamond(\mathbf{s}),\, \mathbf{s})dF_\mathbf{S}(\mathbf{s}) + \int_{\mathcal{S}_1} I(\mu_1^\diamond(\mathbf{s}),\, \mu_2^\diamond(\mathbf{s}),\, \mathbf{s})dF_\mathbf{S}(\mathbf{s}) \quad (4.115) \\
&= \mathbb{E}_{\hat{F}_\mathbf{S}}[I(\mu_1^\diamond(\mathbf{S}),\, \mu_2^\diamond(\mathbf{S}),\, \mathbf{S})], \quad (4.116)
\end{aligned}$$

where (4.114) is by (4.107), (4.108); (4.115) is by Lemma 1, properties 1 and 3; and

(4.116) is by (4.14). Summarizing collectively (4.105), (4.113), (4.116), we get

$$C = \max_{\substack{\mu_1:(\mathbb{R}_0^+)^{2\times M}\to[0,\,1] \\ \mu_2:(\mathbb{R}_0^+)^{2\times M}\to[0,\,1] \\ \mathbf{E}_{\hat{F}_{\mathbf{S}}}[\mu_1(\mathbf{S})+\mu_2(\mathbf{S})]\le 2\sigma}} \mathbb{E}_{\hat{F}_{\mathbf{S}}}[I(\mu_1(\mathbf{S}),\, \mu_2(\mathbf{S}),\, \mathbf{S})]. \tag{4.117}$$

Recall that $\hat{F}_{\mathbf{S}}(\mathbf{s}) = 0$ for all $\mathbf{s} : K_1(\mathbf{s}) + K_2(\mathbf{s}) < 0$. Therefore, by Corollary 6, it suffices to consider $\mu_1(\mathbf{S}) \geq \mu_2(\mathbf{S})$ with probability 1 in the computation of the optimal solution in (4.117), whence by (3.8), we get

$$C = \max_{\substack{\mu_1:(\mathbb{R}_0^+)^{2\times M}\to[0,\,1] \\ \mu_2:(\mathbb{R}_0^+)^{2\times M}\to[0,\,1] \\ \mu_1(\mathbf{S})\ge\mu_2(\mathbf{S}) \\ \mathbf{E}_{\hat{F}_{\mathbf{S}}}[\mu_1(\mathbf{S})+\mu_2(\mathbf{S})]\le 2\sigma}} \mathbb{E}_{\hat{F}_{\mathbf{S}}}[I_1 > 2(\mu_1(\mathbf{S}) - \mu_2(\mathbf{S}),\, \mu_2(\mathbf{S}),\, \mathbf{S})], \tag{4.118}$$

where $I_{1>2}(\cdot, \cdot, \cdot)$ is as defined in (3.94). This concludes the proof of Theorem 8. $\square$

Proof of Theorem 9:

By (4.15), it follows that

$$C \leq \max_{\substack{\mu:(\mathbb{R}_0^+)^{2\times M}\to[0,\,1] \\ \mathbf{E}_{\hat{F}_{\mathbf{S}}}[\mu(\mathbf{S})]\le\sigma}} \mathbb{E}_{\hat{F}_{\mathbf{S}}}\left[I_{max}(\mu(\mathbf{S}),\, \mathbf{S})\right], \tag{4.119}$$

where

$$\begin{aligned} I_{max}(\mu_0,\, \mathbf{s}) &= \max_{\substack{0\le\mu_2\le\mu_1\le 1 \\ \mu_1+\mu_2\le 2\mu_0}} I_{1>2}(\mu_1 - \mu_2,\, \mu_2,\, \mathbf{s}), \\ &\qquad \mathbf{s} \in (\mathbb{R}_0^+)^{2\times M}, \quad 0 \leq \mu_0 \leq 1. \end{aligned} \tag{4.120}$$

Since $\tilde{F}_{\mathbf{S}}(\mathbf{s}) = 0$ for all $\mathbf{s} \in \mathcal{S}_2$, it suffices to limit our attention to the solution of the optimization problem in the right side of (4.120) for only the states $\mathbf{s} \in \mathcal{S}_0 \cup \mathcal{S}_1$. By Theorem 6, for $\mathbf{s} \in \mathcal{S}_0 \cup \mathcal{S}_1$, the optimal pair $\{\kappa_1 = \kappa_1(\mu_0, \mathbf{s}),\, \kappa_2 = \kappa_2(\mu_0, \mathbf{s})\}$ that maximizes the right side of (4.120) is given as follows[4]:

[4]By Corollary 6, for all $\mathbf{s} \in \mathcal{S}_0 \cup \mathcal{S}_1$, the constraint set in the right side of (4.120) can be expanded to include all pairs $\{\mu_1,\, \mu_2\}$ satisfying $0 \leq \mu_1 \leq 1$, $0 \leq \mu_2 \leq 1$, $\mu_1 + \mu_2 \leq 2\mu_0$, without changing the optimal solution.

1. if $\mu_0 \geq \bar{\rho}(\mathbf{s})$, then

$$\kappa_1 \;=\; \kappa_2 \;=\; \bar{\rho}(\mathbf{s});$$

2. if $\mu_0 < \bar{\rho}(\mathbf{s})$, then

$$\begin{aligned}\kappa_1 &= \begin{cases}\min\{2\gamma_2(\mu_0,\, \mathbf{s}),\, 2\mu_0\}, & \text{if } \; K_2(\mathbf{s}) \geq d(\mu_0,\, \mu_0,\, \mathbf{s}),\\ \mu_0, & \text{otherwise,}\end{cases}\\ \kappa_2 &= 2\mu_0 - \kappa_1.\end{aligned}$$

By (4.119), (4.120), we thus get that

$$C \;\leq\; \max_{\substack{\mu:(\mathrm{I\!R}_0^+)^{2\times M}\to[0,\,1]\\ \mathbf{E}_{\hat{F}_\mathbf{S}}[\mu(\mathbf{S})]\leq\sigma}} \mathrm{I\!E}_{\hat{F}_\mathbf{S}}[I_{1>2}(\kappa_1(\mathbf{S},\mu(\mathbf{S})) - \kappa_2(\mathbf{S},\mu(\mathbf{S})),\, \kappa_2(\mathbf{S},\mu(\mathbf{S})),\, \mathbf{S})].$$

It can be verified that $0 \leq \kappa_2(\mu_0,\, \mathbf{s}) \leq \kappa_1(\mu_0,\, \mathbf{s}) \leq 1$, $0 \leq \mu_0 \leq 1$, $\mathbf{s} \in \mathcal{S}_0 \cup \mathcal{S}_1$; and $\mathrm{I\!E}_{\hat{F}_\mathbf{S}}[\kappa_1(\mu(\mathbf{S}),\, \mathbf{S}) + \kappa_2(\mu(\mathbf{S}),\, \mathbf{S})] \leq \mathrm{I\!E}_{\hat{F}_\mathbf{S}}[2\mu(\mathbf{S})] \leq 2\sigma$. Therefore, by (4.15),

$$C \;\geq\; \max_{\substack{\mu:(\mathrm{I\!R}_0^+)^{2\times M}\to[0,\,1]\\ \mathbf{E}_{\hat{F}_\mathbf{S}}[\mu(\mathbf{S})]\leq\sigma}} \mathrm{I\!E}_{\hat{F}_\mathbf{S}}[I_{1>2}(\kappa_1(\mathbf{S},\mu(\mathbf{S})) - \kappa_2(\mathbf{S},\mu(\mathbf{S})),\, \kappa_2(\mathbf{S},\mu(\mathbf{S})),\, \mathbf{S})],$$

whence we get the desired result. □

Proof of Corollary 8:

Consider the class of "simultaneous ON-OFF keying" strategies, i.e., the power control laws that dictate all the transmit apertures to simultaneously remain in either the ON or the OFF state. It can be verified that the maximum rate achievable by any simultaneous ON-OFF keying strategy is given by (4.20), and the optimal power control law $\mu_L : (\mathrm{I\!R}_0^+)^{2\times M} \to [0,\, 1]$ that achieves the maximum in (4.20) has a closed-form structure, which we state next[5]. For $\rho \geq 0$, $\mathbf{s} \in (\mathrm{I\!R}_0^+)^{2\times M}$, let $\mu = \mu_\rho(\mathbf{s})$

[5] The proof is similar to the proofs of Theorems 1 and 2, with appropriate modifications to account for the complexity of the MIMO channel, and is omitted here.

be the solution of the equation

$$\sum_{m=1}^{M} B_m \log \frac{(1+\alpha(B_m)B_m)}{(1+\mu B_m)} = \frac{\rho}{\lambda_0}, \tag{4.121}$$

with $B_m = \sum_{n=1}^{2} s_{n,m} \frac{A}{\lambda_0}$, $m = 1, \cdots, M$. If $\sigma_0 \triangleq \mathbb{E}[\mu_0(\mathbf{S})] > \sigma$, let $\rho = \rho^* > 0$ be the solution of the equation

$$\mathbb{E}\left[[\mu_\rho(\mathbf{S})]^+\right] = \sigma. \tag{4.122}$$

Then μ_L is given by

$$\mu_L(\mathbf{s}) = \begin{cases} \mu_0(\mathbf{s}), & \sigma_0 \leq \sigma, \\ [\mu_{\rho^*}(\mathbf{s})]^+, & \sigma_0 > \sigma, \quad \mathbf{s} \in (\mathbb{R}_0^+)^{2\times M}. \end{cases} \tag{4.123}$$

Clearly, if $\mathbf{s}^{(1)}$ and $\mathbf{s}^{(2)}$ are mirror states, then by definition, $B_m^{(1)} = B_m^{(2)}$, $m = 1, \cdots, M$, so that by (4.121)–(4.123),

$$\mu_L(\mathbf{s}^{(1)}) = \mu_L(\mathbf{s}^{(2)}). \tag{4.124}$$

By (4.20), we now get

$$\begin{aligned} C_{SOOK} &= \mathbb{E}[I(\mu_L(\mathbf{S}),\, \mu_L(\mathbf{S}),\, \mathbf{S})] \\ &= \mathbb{E}_{\hat{F}_{\mathbf{S}}}[I(\mu_L(\mathbf{S}),\, \mu_L(\mathbf{S}),\, \mathbf{S})] && (4.125) \\ &= \mathbb{E}_{\hat{F}_{\mathbf{S}}}[I_{1>2}(0,\, \mu_L(\mathbf{S}),\, \mathbf{S})] && (4.126) \\ &\leq \mathbb{E}_{\hat{F}_{\mathbf{S}}}[I_{max}(\mu_L(\mathbf{S}),\, \mathbf{S})] && (4.127) \\ &= \mathbb{E}_{\hat{F}_{\mathbf{S}}}[I_{1>2}(\kappa_1(\mathbf{S},\, \mu_L(\mathbf{S})) - \kappa_2(\mathbf{S},\, \mu_L(\mathbf{S})),\, \kappa_2(\mathbf{S},\, \mu_L(\mathbf{S})),\, \mathbf{S})] \\ &\leq C, && (4.128) \end{aligned}$$

where (4.125) is by (4.124) and property 1, Lemma 1; (4.126) is by (3.8) and (3.94); (4.127) is by (4.120); (4.128) is by (4.16). This concludes the proof. □

4.5 Numerical examples

We now discuss a few illustrative examples. For simplicity, we consider two-valued state sets. In the first example, we show that for certain fade realizations, one transmit aperture enjoys significantly better channel conditions and hence should be assigned higher average power levels than the other transmit aperture; the new lower bound (C_{LB}) is shown to yield higher information rates than the simultaneous ON-OFF keying lower bound (C_{SOOK}) for some values of σ. In the second example, we demonstrate that simultaneous ON-OFF keying is optimal for all values of $0 \leq \sigma \leq 1$, and achieves channel capacity.

Example 4.1: Consider a symmetric MIMO Poisson channel with $N = 2$, $M = 2$, $A_1 = A_2 = 1.0$, $\lambda_{0,1} = \lambda_{0,2} = \lambda_0 = 1.0$. Let $S_{n,m}$, $n = 1, 2$, $m = 1, 2$ be independent and uniformly distributed on $\mathcal{S} = \{0.1, 1.0\}$. A pictorial representation of the state set is provided in Figure 4.4. It is clear that for a state $\mathbf{s} \in \mathcal{S}^{2\times 2}$ which satisfies the conditions $K_1(\mathbf{s}) \geq 0$, $K_2(\mathbf{s}) \leq 0$, and which is represented by a circle in Figure 4.4, the optimal average conditional duty cycles satisfy the condition $\mu_1^*(\mathbf{s}) = \mu_2^*(\mathbf{s})$ for any $0 \leq \sigma \leq 1$. However, for the states which are represented by squares in Figure 4.4, the optimal average conditional duty cycles need not be equal. In Figure 4.5, we plot two lower bounds on channel capacity, viz., C_{SOOK} and C_{LB} (see (4.20) and (4.21)). It is clear from this figure that the new lower bound (C_{LB}) on capacity yields higher rates than the simultaneous ON-OFF lower bound (C_{SOOK}).

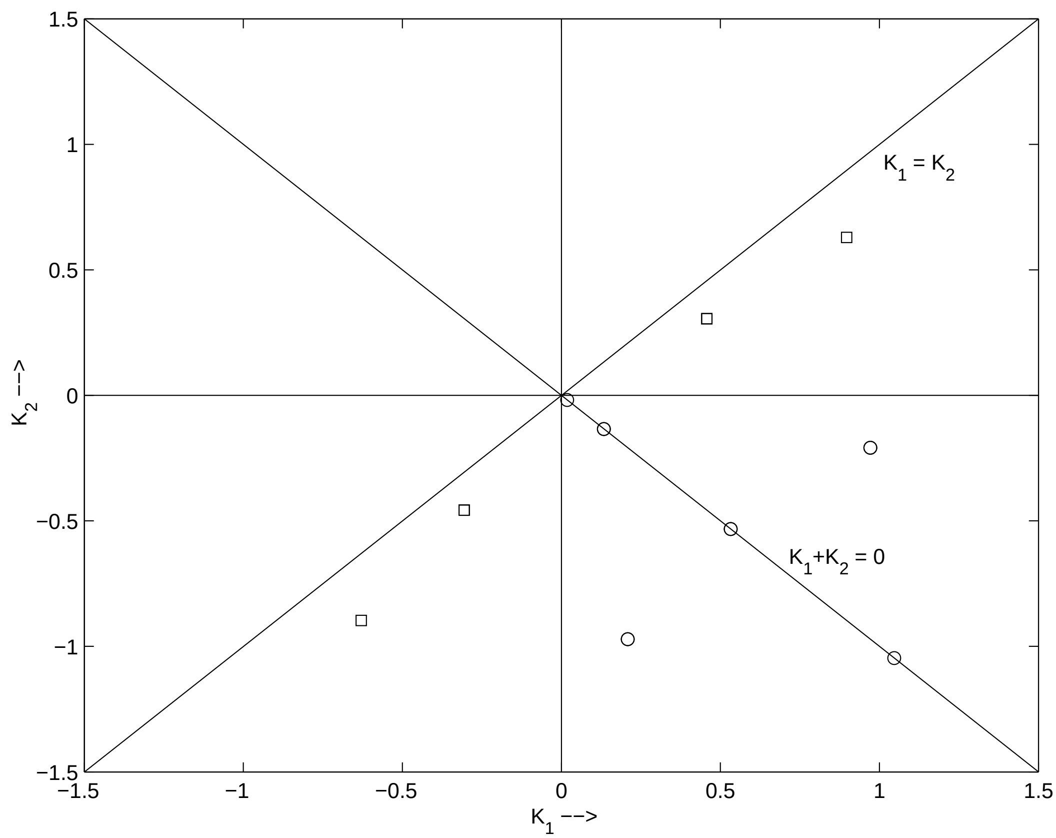

Figure 4.4: State set diagram for Example 4.1.

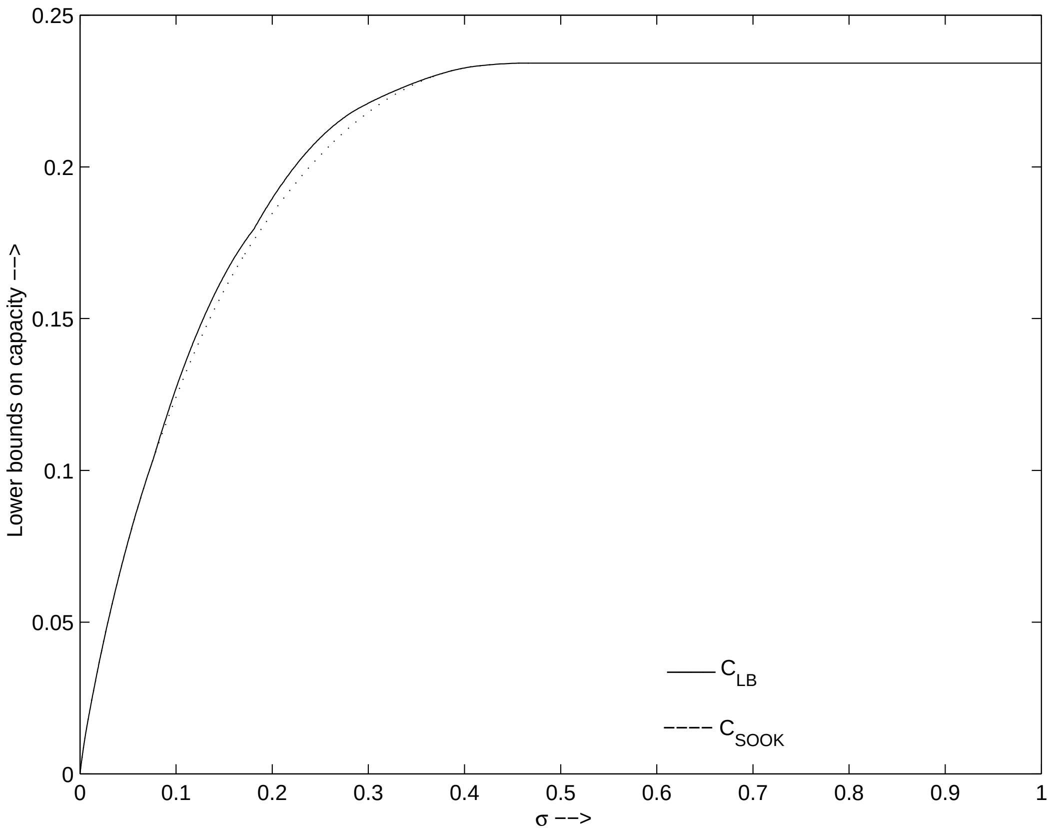

Figure 4.5: Lower bounds on capacity for Example 4.1.

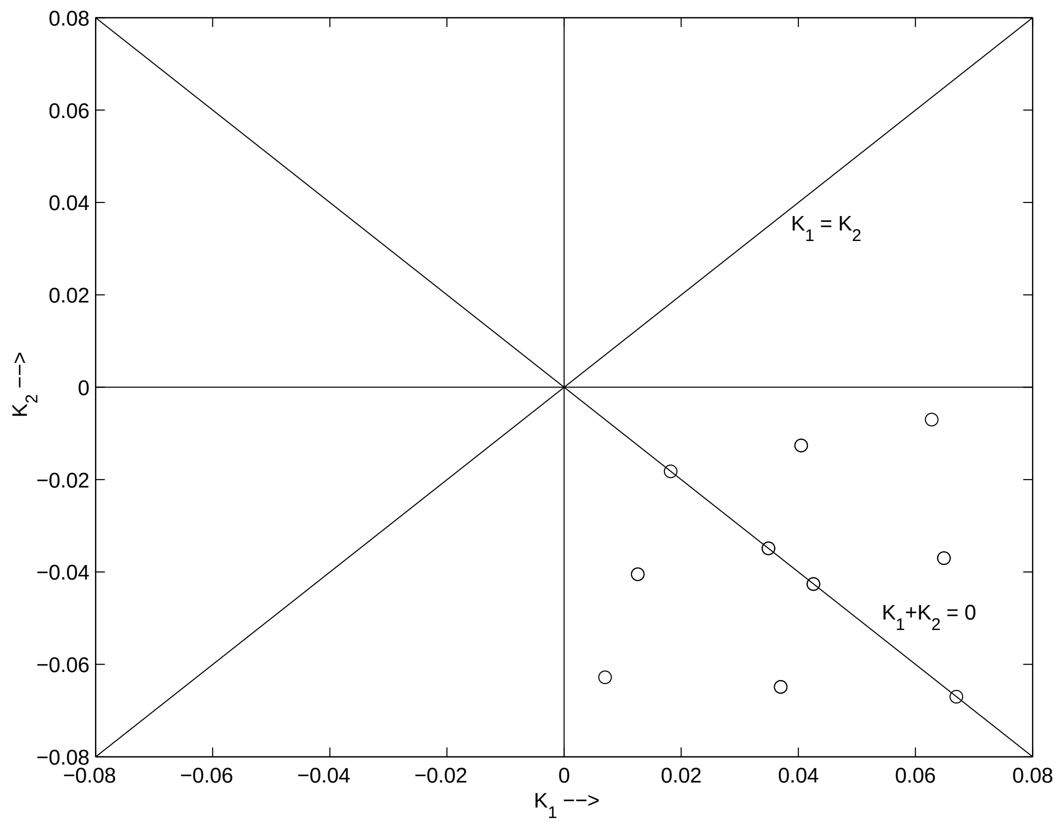

Figure 4.6: State set diagram for Example 4.2.

Example 4.2: Consider the 2×2 MIMO channel from the previous example, with $A = 1.0$, $\lambda_0 = 1.0$ as before, but now $S_{n,m}$, $n = 1, 2, m = 1, 2$ are independent and uniformly distributed on $\mathcal{S} = \{0.1, 0.2\}$. From Figure 4.6, it follows that for all $\mathbf{s} \in \mathcal{S}^{2\times 2}$, the conditions $K_1(\mathbf{s}) \geq 0$, $K_2(\mathbf{s}) \leq 0$ are satisfied, and hence the optimal conditional duty cycles satisfy $\mu_1^*(\mathbf{s}) = \mu_2^*(\mathbf{s})$. For this special case, both lower bounds, C_{LB} and C_{SOOK} are tight for all $0 \leq \sigma \leq 1$, as shown in Figure 4.7.

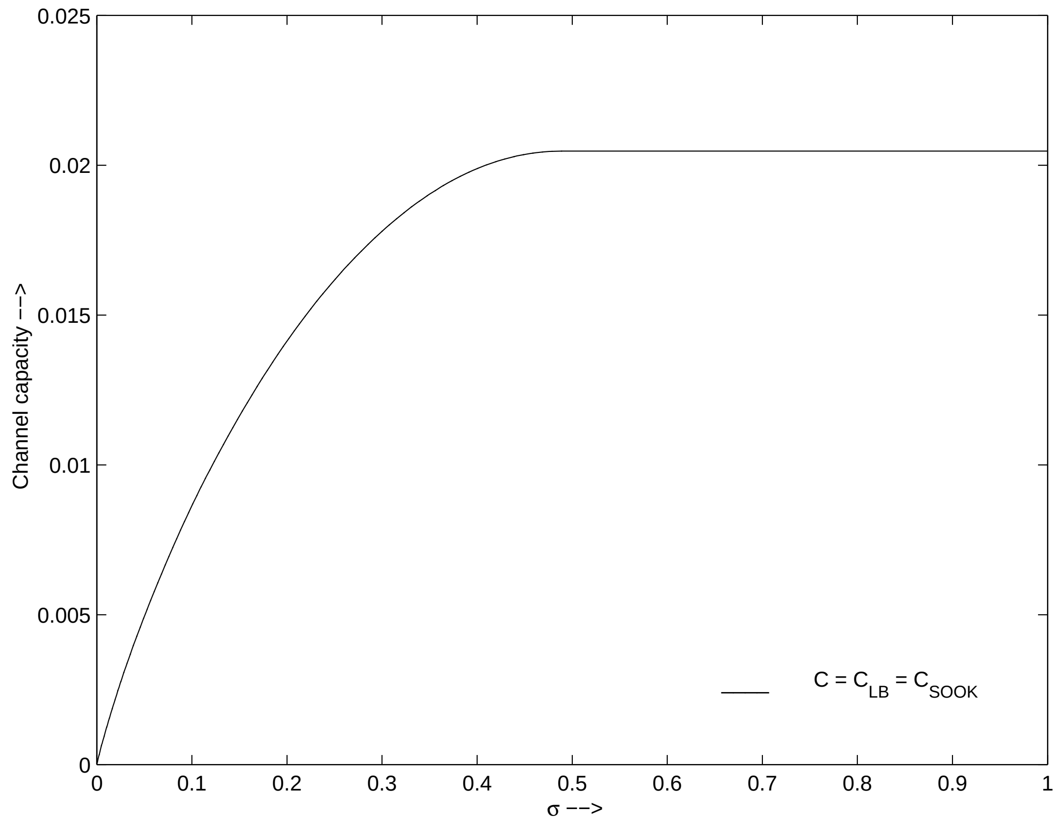

Figure 4.7: Channel capacity for Example 4.2.

4.6 Discussion

We have studied the shot-noise limited $N \times M$ MIMO Poisson fading channel with perfect receiver channel state information. At each receive aperture, the optical fields received from different transmit apertures are assumed to be sufficiently separated in frequency or angle of arrival such that the received total power is the sum of powers from individual transmit apertures, scaled by the respective channel fades. The transmit apertures are subject to asymmetric peak and average power constraints. A block fading channel model has been proposed to account for the slowly varying nature of fading in the free-space optical channel. The capacity of the MIMO Poisson channel with random channel fade has been explicitly characterized and several properties of optimum strategies for transmission over this channel have been discussed.

It has been shown that a two-level signaling scheme (ON-OFF keying) with arbitrarily fast intertransition times through each transmit aperture is capacity-achieving. The transmitted signals through the N apertures are, in general, correlated across apertures and i.i.d. in time. Furthermore, the optimum set of transmission events for every realization of random channel gains has at most $N + 1$ values (out of a possible 2^N values). Each of these values corresponds to an event in which exactly k transmit apertures are ON and the remaining $N - k$ are OFF, where $k = 0, 1, \cdots, N$.

While the structure of the optimal transmission strategy is yet to be characterized in full generality, several interesting properties have been identified. For the special case of a symmetric MIMO channel with isotropically varying channel fade, a simpler characterization of channel capacity has been obtained based on the notion of "mirror states." In the mirror states, the channel characteristics experienced by the transmit apertures are interchanged. By the symmetry of the channel and input parameters, it follows that the optimal power control laws for the respective transmit apertures are also interchanged in the mirror states. This observation leads to a more tractable expression for channel capacity. A new easily computable lower bound on channel capacity based on a simple suboptimal power control law has been proposed, which improves upon the previously known "simultaneous ON-OFF keying" lower bound [21].

Chapter 5

Conclusions

We now present some of the unresolved issues in this dissertation, and discuss possible directions for future research. We conclude by addressing some of the information theoretic issues associated with reliable communication over optical and RF wireless channels.

5.1 Directions for future research

We have addressed the general capacity problem of the block fading MIMO Poisson channel with peak and average transmitter power constraints. With perfect CSI at the receiver, it follows that that i.i.d. two-level signaling ("ON-OFF keying") with arbitrarily fast intertransition times through each transmit aperture, as a function of current transmitter CSI, can achieve channel capacity; a single-letter characterization of capacity, which does not depend on the channel coherence time, has been obtained. For this general setup, an exact characterization of the optimal power control law, which specifies the optimal average conditional duty cycles (conditioned on current transmitter CSI) of the transmit apertures, is yet to be determined.

It is natural to ask what happens when the assumption of perfect receiver CSI is relaxed. In RF communication, it is well-known (cf. e.g., [40]) that the conditional

Gaussianity of the channel transition pdf, and the optimality of i.i.d. Gaussian input signals, break down when the receiver CSI is imperfect. Consider now the SISO Poisson fading channel, for which the received signal $\{Y(t),\, 0 \le t \le T\}$ is a doubly stochastic Poisson process with rate

$$\Lambda(t) = S[\lceil t/T_c \rceil]x(t) + \lambda_0, \quad 0 \le t \le T.$$

Suppose that the receiver CSI is given by $\{D[k] = g(S[k]),\, k = 1, \cdots, K\}$, where $K = T/T_c$, and $g : \mathbb{R}_0^+ \to \mathcal{D}$ is a given mapping, with $\mathcal{D} \subset \mathbb{R}_0^+$ arbitrary. Then the channel output (represented by the sufficient statistics $(N_T,\, \mathbf{T}^{N_T})$) "sample function density," conditioned on the receiver CSI $\mathbf{D}^K$ and channel input X^T, is given by

$$\begin{aligned} & f_{N_T, \mathbf{T}_{N_T} | X^T, \mathbf{D}^K}(n_T,\, \mathbf{t}^{n_T} | x^T,\, \mathbf{d}^K) \\ &= \int_{\mathbf{s}^K \in (\mathbb{R}_0^+)^K} f_{N_T, \mathbf{T}_{N_T} | X^T, \mathbf{S}^K}(n_T,\, \mathbf{t}^{n_T} | x^T,\, \mathbf{s}^K) \cdot \prod_{k=1}^{K} 1(d[k] = g(s[k])) d\mathbf{s}^K, \end{aligned}$$

where $f_{N_T, \mathbf{T}_{N_T} | X^T, \mathbf{S}^K}(\cdot,\, \cdot | \cdot,\, \cdot)$ is as defined in (2.23). Thus, the channel transition pdf is now a mixture of Poisson counting processes; thenm conditioned on the transmitted signal x^T and the receiver CSI $\mathbf{D}^K = \mathbf{d}^K$, the received signal $\{Y_t,\, 0 \le t \le T\}$ need no longer be independent, unlike in Remark (ii), Section 2.3. Also, the optimum input distribution for the resulting channel with memory need no longer be i.i.d. Hence the capacity may, in general, depend on the coherence time T_c.

In practice, the transmitter and receiver devices are limited in bandwidth. The effects of various forms of bandwidth constraints on the capacity of Poisson channels without any fading have been studied in [41, 42, 43, 31], and bounds on capacity have been obtained. It is worthwhile to address these issues for the

Poisson fading channel. Another assumption recurrent in this work is the shot-noise limited operation regime of the direct detection receiver. In free space optical communication systems, the ambient background radiation is the primary source of receiver noise, whose intensity can often far exceed the intensity of the transmitted signal; in such scenarios, a Gaussian model for background noise is often used [26]. The capacity problem for such channels has been studied in [23, 53], but several issues remain unresolved.

A simple block fading channel model has been introduced in which the channel fade is assumed to remain unchanged for a fixed time interval of duration T_c, and varies in an i.i.d. manner across successive such intervals. It has been noted that given the slowly varying nature of optical fade and the existing high-data rate laser transmitters, millions of consecutive transmission bits experience identical fade. In this setting, the notion of *ergodic* capacity, which has been addressed in this dissertation, relies heavily on the ergodicity assumption $T \gg T_c$; for delay-limited applications, when this ergodicity condition does not hold, other notions of capacity, e.g., capacity versus outage may be more relevant (cf. e.g., [40]). Some of these issues for the MIMO Poisson channel have been addressed in [21], which can be further strengthened with the results discussed in this dissertation.

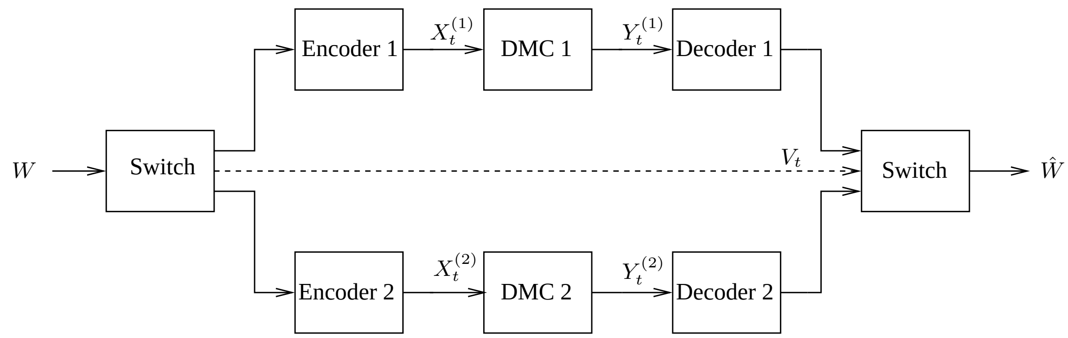

Figure 5.1: Block schematic of the DMC sum channel.

5.2 RF/optical wireless sum channel

5.2.1 Motivation

With the explosive growth in demand for RF bandwidth, and the increasing popularity of optical wireless systems, recent proposals have recommended the use of free space optics (FSO) as a viable alternative for RF communication [10, 38, 30]. There are several issues associated with the design and analysis of such heterogeneous communication systems. In the following, we introduce a simple "sum channel" model for communication over RF and optical wireless channels; at each time instant, information is transmitted using only one of the two channels, and additional information is conveyed by switching between the channels.

We are motivated by [46], in which Shannon demonstrated that by randomly switching between two DMCs, information can be conveyed. Consider a sum channel comprising two DMCs with disjoint input and output alphabets (see Figure 5.1). At each time instant, the switch at the transmitter chooses one of the two DMCs for transmission; by virtue of disjointness of the output alphabets, the switching information is losslessly conveyed to the receiver. The capacity of this sum channel is given by [46]

$$C_{sum} = \max_{0\leq\alpha\leq 1} \left[h_b(\alpha) + \alpha C_1 + (1-\alpha)C_2\right], \tag{5.1}$$

where α is the probability of using channel 1, and C_i is the capacity of channel i, $i = 1, 2$. The optimal α^* that achieves the maximum on the right side of (5.1) is given by

$$\alpha^* = \frac{\exp(C_1)}{\exp(C_1) + \exp(C_2)}, \tag{5.2}$$

and the corresponding maximum value is given by

$$C_{sum} = \log\left(\exp(C_1) + \exp(C_2)\right) \text{ nats/ channel use.} \tag{5.3}$$

We propose a new sum channel model for communication over RF and optical wireless channels. In RF fading channels, it is well known (cf. e.g., [40]) that under severe fading conditions, (i.e., when the multiplicative fade coefficient is very small), then the optimal power control strategy for the transmitter dictates it to remain silent. If an optical wireless link is available, then the optical channel can be used under such circumstances as a backup for the RF channel. On the other hand, the RF channel can also be used as a backup for the optical channel under certain fade conditions. Furthermore, additional information can be conveyed by the means of switching between the two channels, although its amount and whether it would make a significant difference, remains to be seen. In practice, this channel model may be useful for conveying critical information over unreliable wireless links, in which information is transmitted using the channel which experiences "better"

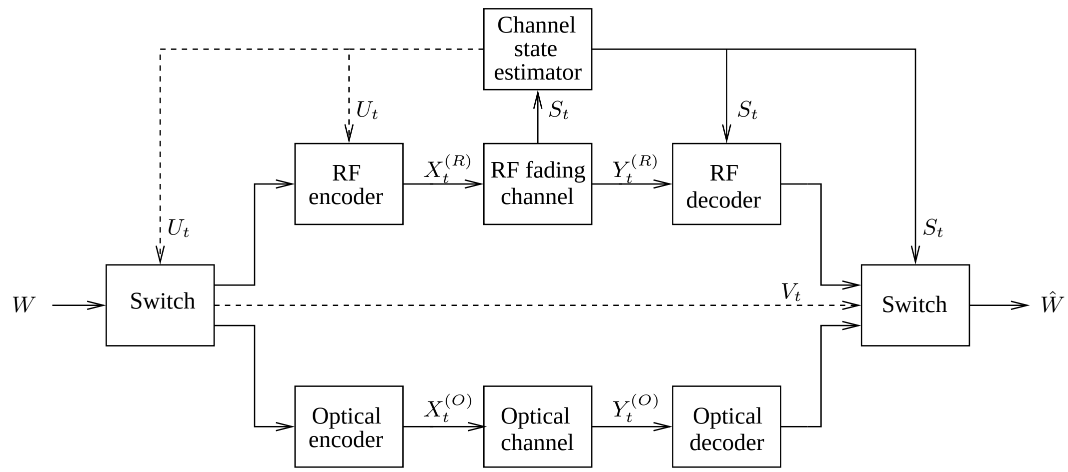

Figure 5.2: Block schematic of the combined RF/optical sum channel.

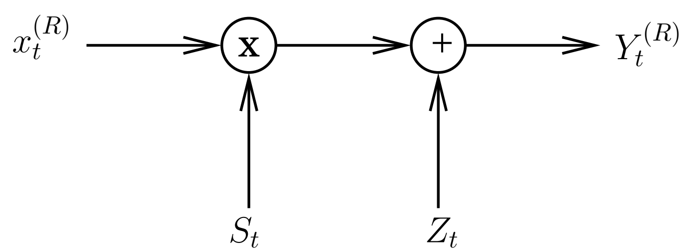

Figure 5.3: RF fading channel model.

instantaneous channel conditions; and in delay-limited applications, where using the sum channel to convey time-sensitive information may lead to higher communication throughput while incurring lower delay.

5.2.2 Problem formulation

We consider a discrete-time memoryless sum channel model for communication over RF and optical wireless channels. A block schematic diagram of the channel model is given in Figure 5.2. For each use of the channel, the switch at the transmitter picks one of two channels, viz., a RF fading channel (see Figure 5.3) and an

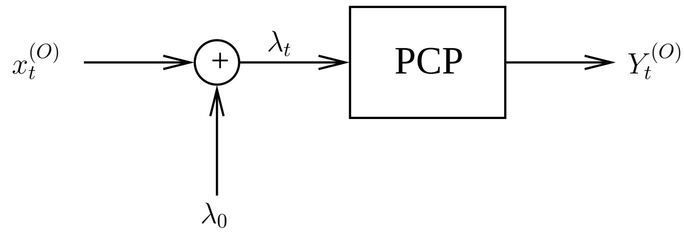

Figure 5.4: Optical channel model.

optical channel (see Figure 5.4).

The $\mathbb{R}$-valued received RF signal is given by

$$Y_t^{(R)} \;=\; S_t x_t^{(R)} + Z_t, \quad t \geq 1, \tag{5.4}$$

where $\{x_t^{(R)}\}_{t=1}^{\infty}$ is the $\mathbb{R}$-valued RF transmitted signal, which is proportional to the instantaneous *amplitude* of the transmitted RF waveform; $\{S_t\}_{t=1}^{\infty}$ is the $\mathbb{R}_0^+$-valued i.i.d. RF channel fade with $\mathbb{E}[S_t^2] < \infty$; and $\{Z_t\}_{t=1}^{\infty}$ is the $\mathbb{R}$-valued i.i.d. Gaussian noise rv with $\mathbb{E}[Z_t] = 0$, and $\mathbb{E}\left[(Z_t - \mathbb{E}[Z_t])^2\right] = \sigma^2$, with $\sigma \geq 0$ fixed. The rvs $\{S_t\}_{t=1}^{\infty}$ and $\{Z_t\}_{t=1}^{\infty}$ are assumed to be mutually independent. We assume that the receiver possesses perfect channel state information (CSI) $\{S_t\}_{t=1}^{\infty}$, while the transmitter CSI is given by $\{U_t = h(S_t)\}_{t=1}^{\infty}$, with the mapping $h : \mathbb{R}_0^+ \to \mathcal{U}$, where $\mathcal{U} \subseteq \mathbb{R}_0^+$ is arbitrary.

The $\mathbb{Z}_0^+$-valued received optical signal $\{Y_t^{(O)}, t \geq 1\}$ is a Poisson rv with intensity (or rate)[1]

$$\lambda_t \;=\; x_t^{(O)} + \lambda_0, \quad t \geq 1, \tag{5.5}$$

[1]For simplicity, in this dissertation we focus on an optical channel model without fading; however, the ideas can be extended (with appropriate modifications) to the case of an optical channel with an i.i.d. block fading structure as discussed in the previous chapters.

where $\{x_t^{(O)}\}_{t=1}^{\infty}$ is the $\mathbb{R}_0^+$-valued transmitted optical signal, which is proportional to the instantaneous *intensity* of the transmitted optical waveform; and $\lambda_0 \geq 0$ is the (constant) background noise ("dark current") rate.

The switching device at the transmitter decides which channel to use at each time instant, based on the available (causal) transmitter CSI. The $\{0, 1\}$-valued switching rv $\{V_t = V_t(\mathbf{U}^t)\}_{t=1}^{\infty}$ is given as follows: if the RF channel is picked for use at time instant t, then the switch sets $V_t = 1$; otherwise, i.e., if the optical channel is picked for use at time instant t, then the switch sets $V_t = 0$, $t \geq 1$. We assume that the switches at the transmitter and the receiver are perfectly coordinated, i.e., the switching rv $\{V_t\}_{t=1}^{\infty}$ is available to the receiver. Physically, the optical and RF modulation and demodulation schemes are fundamentally different from one another, which justifies the perfectly coordinated switching assumption, and also allows us to separate the respective input and output alphabets in our abstract model. Note that by allowing the switching to depend on transmitter CSI, we have introduced here an additional form of coding, which may lead to higher achievable information rates than the DMC sum channel discussed earlier, since now the switching probability ($\Pr\{V_t = 1\} = 1 - \Pr\{V_t = 0\}$) will, in general, depend on transmitter CSI $\mathbf{U}^t$, $t \geq 1$.

The channel transition pdf for n uses of the combined RF/optical sum channel is given by

$$
\begin{aligned}
&W_{\mathbf{Y}^n|\mathbf{X}^n,\mathbf{V}^n,\mathbf{S}^n}(\mathbf{y}^n|\mathbf{x}^n,\, \mathbf{v}^n,\, \mathbf{s}^n) \\
&\quad = \prod_{t=1}^{n} \left\{ W^{(R)}_{Y_t^{(R)}|X_t^{(R)},S_t}(y_t|x_t,\, s_t) \cdot 1(v_t = 1) + W^{(O)}_{Y_t^{(O)}|X_t^{(O)}}(y_t|x_t) \cdot 1(v_t = 0) \right\},
\end{aligned}
$$

$$\mathbf{x}^n,\, \mathbf{y}^n \in \mathbb{R}^n, \quad \mathbf{s}^n \in (\mathbb{R}_0^+)^n, \quad \mathbf{v}^n \in \{0,\, 1\}^n, \tag{5.6}$$

where the RF channel transition pdf is given by

$$W^{(R)}_{Y_t^{(R)}|X_t^{(R)},S_t}(y|x,\, s) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y-sx)^2}{2\sigma^2}\right),$$
$$x,\, y \in \mathbb{R}, \quad s \in \mathbb{R}_0^+, \quad t = 1,\, \cdots,\, n, \tag{5.7}$$

and the optical channel transition pmf is given by

$$W^{(O)}_{Y_t^{(O)}|X_t^{(O)}}(y|x) \;=\; \begin{cases} \frac{(x+\lambda_0)^y \exp(-(x+\lambda_0))}{y!}, & \text{if } x \in \mathbb{R}_0^+, \;\; y \in \mathbb{Z}_0^+, \\ 0, & \text{otherwise}, \;\; t = 1,\, \cdots,\, n. \end{cases} \tag{5.8}$$

The inputs to the sum channel are a $\{0,\, 1\}^n$-valued switching signal $\{v_t\}_{t=1}^n$, and a $\mathbb{R}^n$-valued transmitted signal $\{x_t\}_{t=1}^n$, where $\{x_t\}_{t=1}^n$ satisfies the following peak and average power constraints[2]:

$$0 \;\le\; x_t \;\le\; A, \quad \text{if } v_t = 0, \quad t = 1,\, \cdots,\, n, \tag{5.9}$$

and

$$\frac{1}{n}\sum_{t=1}^{n}[x_t^2 \cdot 1(v_t = 1) + x_t \cdot 1(v_t = 0)] \;\le\; P, \tag{5.10}$$

where $0 \le A < \infty$ and $P \ge 0$ are fixed. Note that x_t is proportional to the amplitude of the transmitted RF signal if $v_t = 1$, whereas if $v_t = 0$, then x_t is proportional to

[2]In (5.10), we have considered an average power constraint involving both the RF and optical transmitted signals; alternatively, individual average power constraints on the transmitted signals from both the channels can be considered as follows: (i) $\frac{1}{n_1}\sum_{t:v_t=1} x_t^2 \le P^{(R)}$, and (ii) $\frac{1}{n-n_1}\sum_{t:v_t=0} x_t \le P^{(O)}$, where $P^{(R)} \ge 0$ and $P^{(O)} \ge 0$ are fixed, and $n_1 = |\{t \in \{1,\, \cdots,\, n\} : v_t = 1\}|$.

the intensity of the transmitted optical signal. This accounts for the difference in the exponents of x_t in the expression of average power constraint in (5.10).

Let the set of messages be denoted by $\mathcal{W} = \{1, \cdots, W\}$. A length-$n$ block code for the channel under consideration is a pair (f, ϕ) defined as follows.

1. For each $\mathbf{u}^n \in \mathcal{U}^n$, the codebook comprises a set of W codewords $f(w, \mathbf{u}^n) = \{v_t(w, \mathbf{u}^t), x_t(w, \mathbf{u}^t, v_t(w, \mathbf{u}^t))\}_{t=1}^n$, $w \in \mathcal{W}$, where

$$x_t(w, \mathbf{u}^t, v_t(w, \mathbf{u}^t)) \quad = \quad \begin{cases} x_t^{(R)}(w, \mathbf{u}^t), & \text{if} \quad v_t(w, \mathbf{u}^t) = 1, \\ x_t^{(O)}(w), & \text{otherwise}, \end{cases} \tag{5.11}$$

satisfies the following peak and average power constraints which follow from (5.10):

$$0 \quad \leq \quad x_t^{(O)}(w) \quad \leq \quad A, \qquad t = 1, \cdots, n, \quad w \in \mathcal{W}, \tag{5.12}$$

and

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^{n} \Bigg[\left(x_t^{(R)}(w, \mathbf{u}^t) \right)^2 \cdot 1(v_t(w, \mathbf{u}^t) = 1) \\ + x_t^{(O)}(w) \cdot 1(v_t(w, \mathbf{u}^t) = 0) \Bigg] \quad \leq \quad P, \quad w \in \mathcal{W}. \end{aligned} \tag{5.13}$$

2. The decoder is a mapping[3] $\phi : (\mathbb{R} \uplus \mathbb{Z}_0^+)^n \times (\mathbb{R}_0^+)^n \to \mathcal{W}$.

For each message $w \in \mathcal{W}$ and transmitter CSI $\mathbf{u}^n \in \mathcal{U}^n$ corresponding to the fade vector $\mathbf{s}^n \in (\mathbb{R}_0^+)^n$, the transmitter generates a switching vector $\{v_t(w, \mathbf{u}^t)\}_{t=1}^n$; a RF signal vector $\{x_t^{(R)}(w, \mathbf{u}^t)\}_{t=1}^n$; and an optical signal vector $\{x_t^{(O)}(w)\}_{t=1}^n$. At time instant $t = 1, \cdots, n$, the transmitter sends $x_t^{(R)}(w, \mathbf{u}^t)$ (resp. $x_t^{(O)}(w)$) if

[3]The notation $\mathbb{R} \uplus \mathbb{Z}_0^+$ is used to denote the physical separability of the output sets of the RF and the optical detector signals.

$v_t(w, \mathbf{u}^t) = 1$ (resp. $v_t(w, \mathbf{u}^t) = 0$) over the RF (resp. optical) channel. The receiver, upon observing $\{y_t^{(R)}, t : v_t = 1\}$ at the RF detector, $\{y_t^{(O)}, t : v_t = 0\}$ at the optical detector, and being provided with $\mathbf{s}^n$, produces an output $\hat{w} = \phi(\{y_t^{(R)}, t : v_t = 1\}, \{y_t^{(O)}, t : v_t = 0\}, \mathbf{s}^n)$. The rate of this length-n block code (f, ϕ) is given by $R = \frac{1}{n} \log W$ nats/channel use, and the average probability of decoding error is given by

$$\begin{aligned} P_e(f, \phi) \;\; = \;\; & \frac{1}{W} \sum_{w=1}^{W} \mathbb{E}[\Pr\{\phi(\{Y_t^{(R)}, t : v_t(w, \mathbf{U}^t) = 1\}, \{Y_t^{(O)}, t : v_t(w, \mathbf{U}^t) = 0\}, \mathbf{S}^n)| \\ & \mathbf{v}^n(w, \mathbf{U}^n), \mathbf{x}^n(w, \mathbf{U}^n, \mathbf{v}^n(w, \mathbf{U}^n)), \mathbf{S}^n\}], \end{aligned} \tag{5.14}$$

where we have used the shorthand notation

$$\mathbf{v}^n(w, \mathbf{U}^n) = \{v_t(w, \mathbf{U}^t), t = 1, \cdots, n\}, \quad w \in \mathcal{W},$$

and

$$\mathbf{x}^n(w, \mathbf{U}^n, \mathbf{v}^n(w, \mathbf{U}^n)) = \{x_t(w, \mathbf{U}^t, v_t(w, \mathbf{U}^t)), t = 1, \cdots, n\}, \quad w \in \mathcal{W}.$$

We are interested in obtaining a "single-letter characterization" of the channel capacity C_{sum} of the RF/optical sum channel in terms of the signal and channel parameters.

5.2.3 Channel capacity

We provide below an expression of the channel capacity of the RF/optical sum channel in terms of the capacities of the individual RF and optical channels.

Theorem 10 *The capacity of the RF/optical sum channel is given by*

$$C_{sum} = \max_{\substack{\alpha:\mathcal{U}\to[0,\,1] \\ \mu_1:\mathcal{U}\to\mathbb{R}_0^+ \\ 0\le\mu_2\le A \\ \mathbf{E}[\alpha(U)\mu_1(U)+(1-\alpha(U))\mu_2]\le P}} \mathbb{E}[g(S,\, \alpha(U),\, \mu_1(U),\, \mu_2)], \tag{5.15}$$

with $U = h(S)$, *where*

$$\begin{aligned} g(s,\, a,\, m_1,\, m_2) &= h_b(a) + aC_{RF}(s,\, m_1) + (1-a)C_O(m_2), \\ &\qquad s \in \mathbb{R}_0^+, \quad a \in [0,\, 1], \quad m_1 \in \mathbb{R}_0^+, \quad 0 \le m_2 \le A, \end{aligned} \tag{5.16}$$

and $C_{RF}(\cdot,\, \cdot)$ *and* $C_O(\cdot)$ *are obtained as follows:*

$$\begin{aligned} C_{RF}(s,\, m_1) &= \max_{P_{X^{(R)}|U=h(s)}:\mathbf{E}[(X^{(R)})^2|U=h(s)]=m_1} I(X^{(R)} \wedge Y^{(R)}|S=s), \\ &\qquad s,\, m_1 \in \mathbb{R}_0^+, \end{aligned} \tag{5.17}$$

where $X^{(R)}$, $Y^{(R)}$, S *are related by*

$$\begin{aligned} W_{Y^{(R)}|X^{(R)},S}(y|x,\, s) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-sx)^2}{2\sigma^2}\right), \\ &\qquad x,\, y \in \mathbb{R}, \quad s \in \mathbb{R}_0^+, \end{aligned} \tag{5.18}$$

and

$$C_O(m_2) = \max_{\substack{P_{X^{(O)}}:0\le X^{(O)}\le A \\ \mathbf{E}[X^{(O)}]=m_2}} I(X^{(O)} \wedge Y^{(O)}), \quad 0 \le m_2 \le A, \tag{5.19}$$

where $X^{(O)}$, $Y^{(O)}$ *are related by*

$$W_{Y^{(O)}|X^{(O)}}(y|x) = \begin{cases} \frac{(x+\lambda_0)^y \exp(-(x+\lambda_0))}{y!}, & \textit{if} \quad x \in \mathbb{R}_0^+, \quad y \in \mathbb{Z}_0^+, \\ 0, & \textit{otherwise.} \end{cases} \tag{5.20}$$

Proof: See Appendix D. $\square$

Remark: Note the structural similarity of the capacity formula (5.15) with the capacity formula for the DMC sum channel (5.1). The additional complexity in (5.15) is due to (a) the dependence of the switching rv on transmitter CSI; and (b) the power constraints on the transmitted signals.

The capacity problem for the discrete-time Gaussian channel is well documented (cf. e.g., [8], Chapter 10). It is easy to see from ([8], Theorem 10.1.1) that

$$C_{RF}(s, m_1) = \frac{1}{2} \log \left(1 + \frac{s^2 m_1}{\sigma^2} \right), \quad s, m_1 \in \mathbb{R}_0^+.$$

On the other hand, to the best of our knowledge, an exact expression for $C_O(\cdot)$ is not available; inner and outer bounds on the capacity of the discrete-time Poisson channel have been obtained (cf. e.g., [31], and the references therein).

It remains to determine the optimal switching and transmitter power control strategies $\alpha^* : \mathcal{U} \to [0, 1]$, $\mu_1^* : \mathcal{U} \to \mathbb{R}_0^+$, and $0 \le \mu_2^* \le A$ that achieve the maximum in (5.15). One major difficulty is the lack of availability of an exact expression for $C_O(\cdot)$.

Appendix A

A.1 Proof of (2.37)

We begin with the following proposition:

Proposition 1 *If* $\{Y(\tau),\, 0 \le \tau \le T\}$ *is a PCP with (deterministic) rate function* $\{\lambda(\tau),\, 0 \le \tau \le T\}$*, then for any function* $g(\cdot)$ *integrable on* $[0,\, T]$*,*

$$\mathbb{E}\left[\sum_{i=1}^{N_T} g(T_i)\right] = \int_0^T \lambda(\tau) g(\tau) d\tau. \tag{A.1}$$

Proof: A proof of this proposition can be found in [19, 18]. Here, we provide a simpler proof. Observe that

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{N_T} g(T_i)\right] &= \sum_{n=1}^{\infty} \Pr\{N_T = n\}\, \mathbb{E}\left[\sum_{i=1}^{n} g(T_i) \middle| N_T = n\right] \\ &= \sum_{n=1}^{\infty} \Pr\{N_T = n\}\, \mathbb{E}\left[\sum_{i=1}^{n} g(U_i) \middle| N_T = n\right],\end{aligned} \tag{A.2}$$

where we denote the *unordered* arrival times on $[0,\, T]$ by U_i, $i = 1, \cdots, N_T$. It can be verified (cf. e.g., [47], pp. 62–63) that given $N_T = n$, U_i are i.i.d. $\sim U$ with pdf

$$f_U(u) = \begin{cases} \frac{\lambda(u)}{\int_0^T \lambda(\tau) d\tau}, & \text{if } 0 \le u \le T, \\ 0, & \text{otherwise.} \end{cases} \tag{A.3}$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{n} g(U_i) \middle| N_T = n\right] &= n\, \mathbb{E}[g(U)] \\ &= n \frac{\int_0^T g(\tau)\lambda(\tau) d\tau}{\int_0^T \lambda(\tau) d\tau},\end{aligned} \tag{A.4}$$

so that by (A.2), it follows that

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{N_T} g(T_i)\right] &= \frac{\int_0^T g(\tau)\lambda(\tau)d\tau}{\int_0^T \lambda(\tau)d\tau}\,\mathbb{E}[N_T] \\
&= \int_0^T g(\tau)\lambda(\tau)d\tau, \qquad \text{(A.5)}
\end{aligned}$$

since N_T is a Poisson rv with mean $\int_0^T \lambda(\tau)d\tau$. $\square$

Note that $(N_T, \mathbf{T}^{N_T})$ is a doubly stochastic PCP with rate process $\{\Lambda(t),\, 0 \leq t \leq T\}$. Therefore, by (A.1), for any $\{\lambda(t),\, 0 \leq t \leq T\}$ integrable on $[0, T]$,

$$\mathbb{E}\left[\sum_{i=1}^{N_T} \log \Lambda(T_i)\,\middle|\,\Lambda^T = \lambda^T\right] = \int_0^T \lambda(\tau)\log\lambda(\tau)d\tau,$$

so that

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{N_T} \log \Lambda(T_i)\right] &= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{N_T} \log \Lambda(T_i)\,\middle|\,\Lambda^T\right]\right] \\
&= \mathbb{E}\left[\int_0^T \Lambda(\tau)\log\Lambda(\tau)d\tau\right]. \qquad \text{(A.6)}
\end{aligned}$$

Similarly, by noting that $(N_T, \mathbf{T}^{N_T})$ conditioned on $\mathbf{S}^K$ is a self-exciting PCP with rate process $\{\hat{\Lambda}(t),\, 0 \leq t \leq T\}$, we get

$$\mathbb{E}\left[\sum_{i=1}^{N_T} \log \hat{\Lambda}(T_i)\right] = \mathbb{E}\left[\int_0^T \hat{\Lambda}(\tau)\log\hat{\Lambda}(\tau)d\tau\right]. \qquad \text{(A.7)}$$

Combining (A.6), (A.7), we get

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i=1}^{N_T}\left(\log\Lambda(T_i) - \log\hat{\Lambda}(T_i)\right)\right] \\
&= \mathbb{E}\left[\int_0^T \left(\Lambda(\tau)\log\Lambda(\tau) - \hat{\Lambda}(\tau)\log\hat{\Lambda}(\tau)\right)d\tau\right] \\
&= \mathbb{E}\left[\int_0^T \left(\zeta(S_{\lceil\tau/T_c\rceil}X(\tau),\, \lambda_0) - \zeta(S_{\lceil\tau/T_c\rceil}\hat{X}(\tau),\, \lambda_0)\right)d\tau\right] \qquad \text{(A.8)} \\
&= \int_0^T \left\{\mathbb{E}\left[\zeta(S_{\lceil\tau/T_c\rceil}X(\tau),\, \lambda_0)\right] - \mathbb{E}\left[\zeta(S_{\lceil\tau/T_c\rceil}\hat{X}(\tau),\, \lambda_0)\right]\right\}d\tau, \qquad \text{(A.9)}
\end{aligned}$$

where (A.8) is by (2.3), (2.25), (2.26); and (A.9) holds by interchanging the order of operations[1] in (A.8).

A.2 Proof of (2.42)

Consider rvs A, B, C, X with (arbitrary) alphabets $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{X}$, where

$$A \text{ is independent of } B, \quad C = k(B), \quad X = l(A, C), \tag{A.10}$$

with k, l being given mappings. Then,

$$\begin{aligned} 0 \leq I(X \wedge B|C) &\leq I(X, A \wedge B|C) \\ &= I(A \wedge B|C) + I(X \wedge B|A, C) \\ &\leq I(A \wedge B, C) + I(X \wedge B|A, C) \\ &= 0, \end{aligned} \tag{A.11}$$

since by (A.10), A is independent of (B, C) and $X = l(A, C)$. Set $A = (W, \mathbf{U}^{\lceil \tau/T_c \rceil - 1})$, $B = S_{\lceil \tau/T_c \rceil}$, $C = U_{\lceil \tau/T_c \rceil}$ and $X = X(\tau)$. Clearly (A.10) is satisfied, so that (2.42) follows from (A.11).

[1]The interchange is permissible as the assumed condition $\mathbb{E}[|\zeta(SA, \lambda_0)|] < \infty$ ensures the integrability of $\{\zeta(S_{\lceil \tau/T_c \rceil} X(\tau), \lambda_0), 0 \leq \tau \leq T\}$ and $\{\zeta(S_{\lceil \tau/T_c \rceil} \hat{X}(\tau), \lambda_0), 0 \leq \tau \leq T\}$.

A.3 Proof of (2.74)

In (2.64), for $L \gg 1$, we use the approximation $\exp(a/L) = 1 + a/L + O(L^{-2})$, $a \in \mathbb{R}$, to get

$$\begin{aligned} w_{10}(L) &= \lambda_0 T_c/L + O\left(L^{-2}\right), \\ w_{11}(s,\, L) &= (\lambda_0 + sA)T_c/L + O\left(L^{-2}\right), \end{aligned} \tag{A.12}$$

and in (2.71), for $x \ll 1$, we use the approximations [52]

$$\begin{aligned} h_b(x) &= -x \log x + x + O(x^2), \\ h_b(x + O(x^2)) &= h_b(x) + O(x^2 \log x). \end{aligned} \tag{A.13}$$

Then, for $L \gg 1$, we have from (2.71) and (A.12) that for $s \in \mathbb{R}_0^+$,

$$\begin{aligned} \beta_L(s) &= h_b\left((\lambda_0 + s\mu(h(s))A)T_c/L + O\left(L^{-2}\right)\right) \\ &\quad -\mu(h(s))h_b\left((\lambda_0 + sA)T_c/L + O\left(L^{-2}\right)\right) \\ &\quad -(1 - \mu(h(s)))h_b\left(\lambda_0 T_c/L + O\left(L^{-2}\right)\right) \\ &= \frac{T_c}{L}\left[\mu(h(s))(\lambda_0 + sA)\log(\lambda_0 + sA) + (1 - \mu(h(s)))\lambda_0 \log \lambda_0\right. \\ &\quad \left. -(\lambda_0 + s\mu(h(s))A)\log(\lambda_0 + s\mu(h(s))A)\right] + O\left(L^{-2}\log L\right) \qquad \text{(A.14)} \\ &= \frac{T_c}{L}\left[\mu(h(s))\zeta(sA,\, \lambda_0) - \zeta(s\mu(h(s))A,\, \lambda_0)\right] + O\left(L^{-2}\log L\right) \qquad \text{(A.15)} \end{aligned}$$

where (A.14) is by (A.13), and (A.15) is by (2.3). Hence,

$$\lim_{L\to\infty} \frac{\beta_L(s)}{T_c/L} = \mu(h(s))\zeta(sA,\, \lambda_0) - \zeta(s\mu(h(s))A,\, \lambda_0), \quad s \in \mathbb{R}_0^+,$$

whereby (2.74) follows.

Appendix B

B.1 Proof of (3.34)

By Proposition 1, noting that Y_m^T is a doubly stochastic PCP with rate process $\{\Lambda_m(t),\, 0 \le t \le T\}$, we get that

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{N_m(T)} \log(\Lambda_m(T_{m,\,i}))\right] &= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{N_m(T)} \log(\Lambda_m(T_{m,\,i}))\,\middle|\, \Lambda_m^T\right]\right] \\
&= \mathbb{E}\left[\int_0^T \Lambda_m(\tau)\log(\Lambda_m(\tau))d\tau\right]. \qquad \text{(B.1)}
\end{aligned}$$

Similarly, by noting that Y_m^T is a self-exciting PCP with rate process $\{\hat{\Lambda}_m(t),\, 0 \le t \le T\}$, by (A.1), we get

$$\mathbb{E}\left[\sum_{i=1}^{N_m(T)} \log(\hat{\Lambda}_m(T_{m,\,i}))\right] = \mathbb{E}\left[\int_0^T \hat{\Lambda}_m(\tau)\log(\hat{\Lambda}_m(\tau))d\tau\right]. \qquad \text{(B.2)}$$

From (3.33), we thus get that

$$\begin{aligned}
&\sum_{m=1}^{M}\mathbb{E}\left[\sum_{i=1}^{N_m(T)}\left(\log\Lambda_m(T_{m,\,i}) - \log\hat{\Lambda}_m(T_{m,\,i})\right)\right] \\
&= \sum_{m=1}^{M}\mathbb{E}\left[\int_0^T \Lambda_m(\tau)\log(\Lambda_m(\tau))d\tau\right] - \mathbb{E}\left[\int_0^T \hat{\Lambda}_m(\tau)\log(\hat{\Lambda}_m(\tau))d\tau\right] \qquad \text{(B.3)} \\
&= \sum_{m=1}^{M}\mathbb{E}\left[\int_0^T\left\{\zeta\left(\sum_{n=1}^{N} s_{n,\,m}X_n(\tau),\ \lambda_{0,\,m}\right)\right.\right. \\
&\qquad \left.\left. -\zeta\left(\sum_{n=1}^{N} s_{n,\,m}\hat{X}_n(\tau),\ \lambda_{0,\,m}\right)\right\}d\tau\right] \qquad \text{(B.4)} \\
&= \sum_{m=1}^{M}\int_0^T\left\{\mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} s_{n,\,m}X_n(\tau),\ \lambda_{0,\,m}\right)\right.\right. \\
&\qquad \left.\left. -\zeta\left(\sum_{n=1}^{N} s_{n,\,m}\hat{X}_n(\tau),\ \lambda_{0,\,m}\right)\right]\right\}d\tau, \qquad \text{(B.5)}
\end{aligned}$$

where (B.3) is by (B.1), (B.2); (B.4) is by (2.3), (3.24), (3.25); and (B.5) holds by interchanging the order of operations[1] in (B.4).

B.2 Proof of (3.43)

From (3.41) and (3.42), after rearrangement we get

$$h(\tau,\, \boldsymbol{\chi}^N(\tau)) \;=\; \sum_{n=1}^{N} \chi_n(\tau) \sum_{m=1}^{M} \zeta(s_{n,\,m}A_n,\, \lambda_{0,\,m}) + \max_{\mathbf{q}(\tau)} \sum_{i=0}^{2^N-1} c_i q_i(\tau), \tag{B.6}$$

where

$$c_i \;=\; \sum_{m=1}^{M} \left\{ \zeta\left(\sum_{n=1}^{N} s_{n,\,m} b_n(i) A_n,\, \lambda_{0,\,m} \right) - \sum_{n=1}^{N} b_n(i) \zeta\left(s_{n,\,m}A_n,\, \lambda_{0,\,m}\right) \right\} \tag{B.7}$$

for $i = 0, \cdots, 2^N - 1$. The following property of $\mathbf{c} = \{c_i\}_{i=0}^{2^N-1}$ can be derived from (3.6):

$$I_i \subseteq I_j \;\Rightarrow\; 0 \le c_i \le c_j, \quad i,\, j \in \{0, \cdots, 2^N - 1\}. \tag{B.8}$$

The linearity of the summand on the right side of (B.6) with respect to the variables $\{q_i(\tau)\}_{i=0}^{2^N-1}$, and the special property (B.8) of $\mathbf{c}$ suggest the following N-step algorithm to compute the optimal pmf vector $\mathbf{q}^*(\tau)$:

- *Step 1*: By (B.8), it follows that $c_{2^N-1} \ge c_i$ for all $i = 0, \cdots, 2^N - 1$. Therefore, $q^*_{2^N-1}(\tau)$ should be assigned the highest allowable value. By (3.41), it follows that $q_{2^N-1}(\tau) \le \chi_n(\tau)$, $n = 1, \cdots, N$. Therefore,

$$q^*_{2^N-1}(\tau) \;=\; \min_{n=1,\,\cdots,\,N} \chi_n(\tau) \;=\; \chi_{P_\tau(N)}(\tau), \tag{B.9}$$

[1]The interchange is permissible as the assumed conditions $0 \le X_n(\tau) \le 1$, $0 \le \tau \le T$, $n = 1, \cdots, N$, ensures the integrability of the rvs $\left\{\zeta\left(\sum_{n=1}^{N} s_{n,\,m} X_n(\tau),\, \lambda_{0,\,m}\right),\, 0 \le \tau \le T\right\}$ and $\left\{\zeta\left(\sum_{n=1}^{N} s_{n,\,m} \hat{X}_n(\tau),\, \lambda_{0,\,m}\right),\, 0 \le \tau \le T\right\}$, $m = 1, \cdots, M$.

where P_τ is as defined in (3.44). By (3.41), it follows that

$$q_i^*(\tau) \;=\; 0, \quad i : b_{P_\tau(N)}(i) = 1, \quad i \neq 2^N - 1. \tag{B.10}$$

- *Step 2*: For all $i \in \{0, \cdots, 2^N - 1\} : b_{P_\tau(N)}(i) = 0$, by (B.8), it follows that $c_j \geq c_i$, where $j = \sum_{n=1}^{N-1} 2^{P_\tau(n)-1}$. Furthermore, for all $i : b_{P_\tau(N)}(i) = 0$, by (3.41), it follows that $q_i(\tau) \leq \chi_n(\tau) - \chi_{P_\tau(N)}(\tau)$, $n = P_\tau(1), \cdots, P_\tau(N-1)$. Summarizing these facts, we have

$$q^*_{\sum_{n=1}^{N-1} 2^{P_\tau(n)-1}}(\tau) \;=\; \chi_{P_\tau(N-1)}(\tau) - \chi_{P_\tau(N)}(\tau), \tag{B.11}$$

whence by (3.41), we get

$$q_i^*(\tau) \;=\; 0, \quad i : b_{P_\tau(N)}(i) = 0, \quad b_{P_\tau(N-1)}(i) = 1, \quad i \neq \textstyle\sum_{n=1}^{N-1} 2^{P_\tau(n)-1}. \tag{B.12}$$

- *Step k*, $2 < k < N$: In this step, repeating the arguments of Step 2, we get

$$q^*_{\sum_{n=1}^{N-k+1} 2^{P_\tau(n)-1}}(\tau) \;=\; \chi_{P_\tau(N-k+1)}(\tau) - \chi_{P_\tau(N-k+2)}(\tau), \tag{B.13}$$

and

$$\begin{aligned} q_i^*(\tau) \;=\; 0, \quad & i : b_{P_\tau(j)}(i) = 0, \qquad j = N-k+2, \cdots, N, \\ & b_{P_\tau(N-k+1)}(i) = 1, \quad i \neq \textstyle\sum_{n=1}^{N-k+1} 2^{P_\tau(n)-1}. \end{aligned} \tag{B.14}$$

- *Step N*: The only remaining pmf is $q_0^*(\tau)$. By (3.41) and (B.9)–(B.14), it follows that

$$q_0^*(\tau) \;=\; 1 - \sum_{i=1}^{2^N-1} q_i^*(\tau) \;=\; 1 - \chi_{P_\tau(1)}(\tau). \tag{B.15}$$

B.3 Proof of (3.52)

It suffices to prove that

$$\sum_{n=1}^{N} \frac{1}{T} \int_0^T \psi_{n,\,P_\tau} \chi_{P_\tau(n)}(\tau) d\tau \;\; \leq \;\; \sum_{n=1}^{N} \psi_{n,\,\Pi} \mu_{\Pi(n)}. \tag{B.16}$$

Without loss of generality, we assume that $\Pi(n) = n$ for all $n = 1, \cdots, N$, i.e., $\mu_n \geq \mu_{n+1}$, $n = 1, \cdots, N-1$. Let $\mathcal{P}$ denote the set of all possible permutations of $\{1, \cdots, N\}$; clearly, $|\mathcal{P}| = N!$. We fix an indexing[2] $\{P_i,\, i = 1, \cdots, N!\}$ of the elements of $\mathcal{P}$ with $P_1 = \Pi$. Consider the following partition of $[0,\, T]$:

$$\mathcal{T}_i \;\; = \;\; \{\tau \in [0,\, T] \backslash \cup_{j<i} \mathcal{T}_j : \chi_{P_i(1)}(\tau) \geq \cdots \geq \chi_{P_i(N)}(\tau)\}, \tag{B.17}$$

and let

$$l_i \;\; = \;\; \frac{1}{T} \int_{\mathcal{T}_i} d\tau, \quad i = 1, \cdots, N!. \tag{B.18}$$

In other words, $\mathcal{T}_i$ is the set of time instants $\tau \in [0,\, T]$ when the elements of $\boldsymbol{\chi}^N(\tau)$ satisfy the permutation P_i but no other permutation P_j, $j < i$, and l_i is the proportion of $\mathcal{T}_i$ relative to the transmission interval $[0,\, T]$. Clearly, $l_i \geq 0$, $i = 1, \cdots, N!$, and $\sum_{i=1}^{N!} l_i = 1$. Define

$$\Theta_{n,\,i} \;\; = \;\; \frac{1}{\int_{\mathcal{T}_i} d\tau} \int_{\mathcal{T}_i} \chi_n(\tau) d\tau, \quad n = 1, \cdots, N, \quad i = 1, \cdots, N!. \tag{B.19}$$

By (3.50), (B.17) and (B.18), we have

$$\begin{aligned} \textstyle\sum_{i=1}^{N!} l_i \Theta_{n,\,i} \;\; &= \;\; \mu_n, && n = 1, \cdots, N, \\ \Theta_{P_i(n),\,i} \;\; &\geq \;\; \Theta_{P_i(n+1),\,i}, && n = 1, \cdots, N-1, \quad i = 1, \cdots, N!. \end{aligned} \tag{B.20}$$

[2]For the sake of brevity, we use similar notation to denote the permutations P_i, $i \in \{1, \cdots, N!\}$ and P_τ, $\tau \in [0,\, T]$. The distinction should be clear from the context.

In order to prove (B.16), we now establish the nonnegativity of the difference

$$\begin{aligned}
&\sum_{n=1}^{N} \psi_{n,\,\Pi}\mu_{\Pi(n)} - \sum_{n=1}^{N} \frac{1}{T}\int_0^T \psi_{n,\,P_\tau}\chi_{P_\tau(n)}(\tau)d\tau \\
&\quad= \sum_{n=1}^{N} \psi_{n,\,P_1}\sum_{i=1}^{N!} l_i\Theta_{n,\,i} - \sum_{n=1}^{N}\sum_{i=1}^{N!} \frac{l_i}{\int_{\mathcal{T}_i} d\tau}\int_{\mathcal{T}_i} \psi_{n,\,P_i}\chi_{P_i(n)}(\tau)d\tau && \text{(B.21)}\\
&\quad= \sum_{i=1}^{N!} l_i \sum_{n=1}^{N} \left(\psi_{n,\,P_1}\Theta_{n,\,i} - \psi_{n,\,P_i}\Theta_{P_i(n),\,i}\right) && \text{(B.22)}\\
&\quad= \sum_{i=1}^{N!} l_i \sum_{n=1}^{N} \left(\psi_{P_i(n),\,P_1} - \psi_{n,\,P_i}\right)\Theta_{P_i(n),\,i} \\
&\quad= \sum_{i=1}^{N!} l_i \left[\sum_{n=1}^{N-1}\left(\Theta_{P_i(n),\,i} - \Theta_{P_i(n+1),\,i}\right)\sum_{k=1}^{n}\left(\psi_{P_i(k),\,P_1} - \psi_{k,\,P_i}\right)\right. \\
&\qquad\qquad \left. +\Theta_{P_i(N),\,i}\sum_{k=1}^{N}\left(\psi_{P_i(k),\,P_1} - \psi_{k,\,P_i}\right)\right], && \text{(B.23)}
\end{aligned}$$

where (B.21) and (B.22) follow by (B.17)–(B.20) with $\Pi = P_1$; and (B.23) follows by (3.54) with the substitution $x_n = \Theta_{P_i(n),\,i}$ and $y_n = \psi_{P_i(n),\,P_1} - \psi_{n,\,P_i}$, $n = 1, \cdots, N$. By definition (see (3.46)), for every $n = 1, \cdots, N$ and $i = 1, \cdots, N!$,

$$\begin{aligned}
\sum_{k=1}^{n} \psi_{P_i(k),\,P_1} &= \sum_{k=1}^{n}\sum_{m=1}^{M}\left[\zeta\left(\sum_{j=1}^{P_i(k)} s_{j,\,m}A_j,\ \lambda_{0,\,m}\right) - \zeta\left(\sum_{j=1}^{P_i(k)-1} s_{j,\,m}A_j,\ \lambda_{0,\,m}\right)\right] \\
&= \sum_{m=1}^{M}\sum_{k=1}^{n}\zeta\left(s_{P_i(k),\,m}A_{P_i(k)},\ \lambda_{0,\,m} + \sum_{j=1}^{P_i(k)-1} s_{j,\,m}A_j\right), && \text{(B.24)}
\end{aligned}$$

where (B.24) follows by (3.6). On the other hand, by (3.46),

$$\begin{aligned}
&\sum_{k=1}^{n} \psi_{k,\,P_i} \\
&\quad= \sum_{k=1}^{n}\sum_{m=1}^{M}\left[\zeta\left(\sum_{j=1}^{k} s_{P_i(j),\,m}A_{P_i(j)},\ \lambda_{0,\,m}\right) - \zeta\left(\sum_{j=1}^{k-1} s_{P_i(j),\,m}A_{P_i(j)},\ \lambda_{0,\,m}\right)\right] \\
&\quad= \sum_{m=1}^{M}\zeta\left(\sum_{j=1}^{n} s_{P_i(j),\,m}A_{P_i(j)},\ \lambda_{0,\,m}\right) \\
&\quad= \sum_{m=1}^{M}\sum_{k=1}^{n}\zeta\left(s_{P_i(k),\,m}A_{P_i(k)},\ \lambda_{0,\,m} + \sum_{j=1}^{P_i(k)-1} s_{j,m}A_j 1(j \in \{P_i(1), \cdots, P_i(n)\})\right), && \text{(B.25)}
\end{aligned}$$

where (B.25) is by (3.6). Using the fact that $\zeta(x,\, y+z) \geq \zeta(x,\, y)$ for all $x,\, y,\, z \geq 0$, by (B.24) and (B.25), we get

$$\sum_{k=1}^{n} \left(\psi_{P_i(k),\, P_1} - \psi_{k,\, P_i} \right) \;\; \geq \;\; 0, \quad n = 1,\, \cdots,\, N, \quad i = 1,\, \cdots,\, N!, \tag{B.26}$$

whence by (B.20), (B.23), it follows that

$$\sum_{n=1}^{N} \psi_{n,\, \Pi} \mu_{\Pi(n)} - \sum_{n=1}^{N} \frac{1}{T} \int_0^T \psi_{n,\, P_\tau} \chi_{P_\tau(n)}(\tau) d\tau \;\; \geq \;\; 0,$$

which is the desired result.

Remark: It can be seen from the preceding proof that (B.16) holds for the case of a discrete sum instead of an integral, i.e., the following inequality is true:

$$\sum_{n=1}^{N} \frac{1}{K} \sum_{k=0}^{K} \psi_{n,\, P_k} \chi_{P_k(n)}(k) \;\; \leq \;\; \sum_{n=1}^{N} \psi_{n,\, \Pi} \mu_{\Pi(n)}, \tag{B.27}$$

where P_k is a permutation of $\{1,\, \cdots,\, N\}$ such that

$$\chi_{P_k(n)}(k) \;\; \geq \;\; \chi_{P_k(n+1)}(k), \quad n = 1,\, \cdots,\, N-1, \quad k = 1,\, \cdots,\, K,$$

$\mu_n = \frac{1}{K} \sum_{k=1}^{K} \chi_n(k)$, $n = 1,\, \cdots,\, N$, and Π is a permutation of $\{1,\, \cdots,\, N\}$ such that

$$\mu_{\Pi(n)} \;\; \geq \;\; \mu_{\Pi(n+1)}, \quad n = 1,\, \cdots,\, N-1.$$

B.4 Proof of (3.75)

In (3.62), for $L \gg 1$, we use the approximation $\exp(a/L) = 1 + a/L + O(L^{-2})$, $a \in \mathbb{R}$, to get

$$\omega_m(\tilde{\mathbf{x}}^N,\, L) \;\; = \;\; w_m(\tilde{\mathbf{x}}^N) \frac{T}{L} + O(L^{-2}), \quad \tilde{\mathbf{x}}^N \in \{0,\, 1\}^N, \tag{B.28}$$

for $m = 1, \cdots, M$. For $L \gg 1$, we have from (3.73) and (B.28) that

$$\begin{aligned}
\beta_L(\mathbf{p}) &= \sum_{m=1}^{M} \left[h_b \left(\sum_{i=0}^{2^N-1} p_i w_m(\mathbf{b}^N(i)) \frac{T}{L} + O(L^{-2}) \right) \right. \\
&\quad \left. - \sum_{i=0}^{2^N-1} p_i h_b \left(w_m(\mathbf{b}^N(i)) \frac{T}{L} + O(L^{-2}) \right) \right] \\
&= \frac{T}{L} \sum_{m=1}^{M} \sum_{i=0}^{2^N-1} p_i w_m(\mathbf{b}^N(i)) \log \left(\frac{w_m(\mathbf{b}^N(i))}{\sum_{i=0}^{2^N-1} p_i w_m(\mathbf{b}^N(i))} \right) + O(L^{-2} \log L) \quad \text{(B.29)} \\
&= \frac{T}{L} \sum_{m=1}^{M} \sum_{i=0}^{2^N-1} p_i \left(\sum_{n=1}^{N} s_{n,m} b_n(i) A_n + \lambda_{0,m} \right) \times \\
&\quad \log \left(\frac{\left(\sum_{n=1}^{N} s_{n,m} b_n(i) A_n + \lambda_{0,m} \right)}{\sum_{i=0}^{2^N-1} p_i \left(\sum_{n=1}^{N} s_{n,m} b_n(i) A_n + \lambda_{0,m} \right)} \right) + O(L^{-2} \log L) \quad \text{(B.30)} \\
&= \frac{T}{L} \sum_{m=1}^{M} \left[\sum_{i=0}^{2^N-1} p_i \zeta_m \left(\sum_{n=1}^{N} s_{n,m} b_n(i) A_n,\ \lambda_{0,m} \right) \right. \\
&\quad \left. - \zeta_m \left(\sum_{i=0}^{2^N-1} p_i \sum_{n=1}^{N} s_{n,m} b_n(i) A_n,\ \lambda_{0,m} \right) \right] + O(L^{-2} \log L), \quad \text{(B.31)}
\end{aligned}$$

where (B.29) is by (A.13); (B.30) is by (3.63); and (B.31) holds by (2.3). Hence

$$\begin{aligned}
\lim_{L \to \infty} \frac{\beta_L(\mathbf{p})}{T/L} &= \sum_{m=1}^{M} \left[\sum_{i=0}^{2^N-1} p_i \zeta_m \left(\sum_{n=1}^{N} s_{n,m} b_n(i) A_n,\ \lambda_{0,m} \right) \right. \\
&\quad \left. - \zeta_m \left(\sum_{i=0}^{2^N-1} p_i \sum_{n=1}^{N} s_{n,m} b_n(i) A_n,\ \lambda_{0,m} \right) \right], \quad \text{(B.32)}
\end{aligned}$$

whence (3.75) follows.

Appendix C

C.1 Proof of (4.37)

By Proposition 1, noting that Y_m^T is a doubly stochastic PCP with rate process $\{\Lambda_m(t),\, 0 \le t \le T\}$, by (A.1), we get that

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{N_m(T)} \log(\Lambda_m(T_{m,i}))\right] &= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{N_m(T)} \log(\Lambda_m(T_{m,i}))\middle|\Lambda_m^T\right]\right] \\
&= \mathbb{E}\left[\int_0^T \Lambda_m(\tau)\log(\Lambda_m(\tau))d\tau\right]. \qquad \text{(C.1)}
\end{aligned}$$

Similarly, by noting that Y_m^T is a self-exciting PCP with rate process $\{\hat{\Lambda}_m(t),\, 0 \le t \le T\}$, by (A.1), we get that

$$\mathbb{E}\left[\sum_{i=1}^{N_m(T)} \log(\hat{\Lambda}_m(T_{m,i}))\right] = \mathbb{E}\left[\int_0^T \hat{\Lambda}_m(\tau)\log(\hat{\Lambda}_m(\tau))d\tau\right]. \qquad \text{(C.2)}$$

By (C.1), (C.2), we then get

$$\begin{aligned}
&\sum_{m=1}^{M}\mathbb{E}\left[\sum_{i=1}^{N_m(T)}\left(\log\Lambda_m(T_{m,i}) - \log\hat{\Lambda}_m(T_{m,i})\right)\right] \\
&= \sum_{m=1}^{M}\mathbb{E}\left[\int_0^T \Lambda_m(\tau)\log(\Lambda_m(\tau))d\tau\right] - \mathbb{E}\left[\int_0^T \hat{\Lambda}_m(\tau)\log(\hat{\Lambda}_m(\tau))d\tau\right] \\
&= \sum_{m=1}^{M}\mathbb{E}\left[\int_0^T \left\{\zeta\left(\sum_{n=1}^{N} S_{n,m}[\lceil\tau/T_c\rceil]X_n(\tau),\, \lambda_{0,m}\right)\right.\right. \\
&\qquad \left.\left. -\zeta\left(\sum_{n=1}^{N} S_{n,m}[\lceil\tau/T_c\rceil]\hat{X}_n(\tau),\, \lambda_{0,m}\right)\right\} d\tau\right] \qquad \text{(C.3)} \\
&= \sum_{m=1}^{M}\int_0^T \left\{\mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} S_{n,m}[\lceil\tau/T_c\rceil]X_n(\tau),\, \lambda_{0,m}\right)\right]\right. \\
&\qquad \left. -\,\mathbb{E}\left[\zeta\left(\sum_{n=1}^{N} S_{n,m}[\lceil\tau/T_c\rceil]\hat{X}_n(\tau),\, \lambda_{0,m}\right)\right]\right\} d\tau, \qquad \text{(C.4)}
\end{aligned}$$

where (C.3) is by (2.3), (4.26), (4.27); and (C.4) holds by interchanging the order of operations[1] in (C.3).

C.2 Proof of (4.60)

Let Q_k be a permutation of $\{1, \cdots, N\}$ such that

$$\eta_{Q_k(n)}(k, \mathbf{u}) \ \ \geq \ \ \eta_{Q_k(n+1)}(k, \mathbf{u}), \quad n = 1, \cdots, N-1, \tag{C.5}$$

for $\mathbf{u} \in \mathcal{U}$, $k = 1, \cdots, K$, where we have suppressed the dependence of Q_k on $\mathbf{u}$ for notational brevity. By B.16, it follows that

$$\sum_{n=1}^{N} \frac{1}{T_c} \int_{(k-1)T_c}^{kT_c} \psi_{n, P_\tau}(\mathbf{s}) \chi_{P_\tau(n)}(\tau, \mathbf{u}) d\tau \ \ \leq \ \ \sum_{n=1}^{N} \psi_{n, Q_k}(\mathbf{s}) \eta_{Q_k(n)}(k, \mathbf{u}), \tag{C.6}$$

for $k = 1, \cdots, K$, $\mathbf{s} \in (\mathbb{R}_0^+)^{N \times M}$, $\mathbf{u} \in \mathcal{U}$. Therefore, by (C.6), we now get

$$\begin{aligned}
&\frac{1}{K} \sum_{k=1}^{K} \frac{1}{T_c} \int_{(k-1)T_c}^{kT_c} \mathbb{E}\left[\sum_{n=1}^{N} \psi_{n, P_\tau}(\mathbf{S}[k]) \chi_{P_\tau(n)}(\tau, \mathbf{U}[k]) \right] d\tau \\
&\quad \leq \ \ \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}\left[\sum_{n=1}^{N} \psi_{n, Q_k}(\mathbf{S}[k]) \eta_{Q_k(n)}(k, \mathbf{U}[k]) \right] \\
&\quad = \ \ \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}\left[\sum_{n=1}^{N} \psi_{n, Q_k}(\mathbf{S}) \eta_{Q_k(n)}(k, \mathbf{U}) \right] && \text{(C.7)} \\
&\quad \leq \ \ \mathbb{E}\left[\sum_{n=1}^{N} \psi_{n, \Pi}(\mathbf{S}) \mu_{\Pi(n)}(\mathbf{U}) \right] && \text{(C.8)} \\
&\quad = \ \ \mathbb{E}\left[\sum_{n=1}^{N} \nu_n(\mathbf{U}) \sum_{m=1}^{M} \zeta\left(\sum_{j=1}^{n} S_{\Pi(j), m} A_{\Pi(j)}, \ \lambda_{0, m} \right) \right], && \text{(C.9)}
\end{aligned}$$

[1]The interchange is permissible as the assumed conditions $\mathbb{E}\left[\left| \zeta\left(\sum_{n=1}^{N} S_{n, m} A_n, \lambda_{0, m} \right) \right| \right] < \infty$, $m = 1, \cdots, M$, ensure the integrability of $\left\{ \zeta\left(\sum_{n=1}^{N} S_{n, m}[\lceil \tau/T_c \rceil] X_n(\tau), \ \lambda_{0, m} \right), \ 0 \leq \tau \leq T \right\}$ and $\left\{ \zeta\left(\sum_{n=1}^{N} S_{n, m}[\lceil \tau/T_c \rceil] \hat{X}_n(\tau), \ \lambda_{0, m} \right), \ 0 \leq \tau \leq T \right\}$, $m = 1, \cdots, M$.

where (C.7) is by the i.i.d. nature of $\{\mathbf{S}[k]\}_{k=1}^{\infty}$; (C.8) follows by (4.61), (4.54), (B.27); and (C.9) is by the application of (3.54) with the substitution $x_n = \mu_{\Pi(n)}(\mathbf{U})$, $y_n = \psi_{n,\,\Pi}(\mathbf{S})$.

C.3 Proof of (4.85)

In (4.72), for $L \gg 1$, we use the approximation $\exp(a/L) = 1 + a/L + O(L^{-2})$, $a \in \mathbb{R}$, to get

$$\begin{aligned}\omega_m(\tilde{\mathbf{x}}^N,\, \mathbf{s},\, L) \;&=\; w_m(\tilde{\mathbf{x}}^N,\, \mathbf{s})\frac{T_c}{L} + O(L^{-2}),\\ &\qquad\qquad \tilde{\mathbf{x}}^N \in \{0,\, 1\}^N, \quad \mathbf{s} \in (\mathbb{R}_0^+)^{N\times M},\end{aligned} \tag{C.10}$$

for $m = 1, \cdots, M$. For $L \gg 1$, we have from (4.83) and (C.10) that

$$\begin{aligned}\beta_L(\mathbf{p},\, \mathbf{s}) \;&=\; \sum_{m=1}^{M}\left[h_b\left(\sum_{i=0}^{2^N-1} p_i w_m(\mathbf{b}^N(i),\, \mathbf{s})\frac{T_c}{L} + O(L^{-2})\right)\right.\\ &\qquad\left. - \sum_{i=0}^{2^N-1} p_i h_b\left(w_m(\mathbf{b}^N(i),\, \mathbf{s})\frac{T_c}{L} + O(L^{-2})\right)\right]\\ &=\; \frac{T_c}{L}\sum_{m=1}^{M}\left[\sum_{i=0}^{2^N-1} p_i \zeta_m\left(\sum_{n=1}^{N} s_{n,\,m} b_n(i) A_n,\; \lambda_{0,\,m}\right)\right.\\ &\qquad\left. -\zeta_m\left(\sum_{i=0}^{2^N-1} p_i \sum_{n=1}^{N} s_{n,\,m} b_n(i) A_n,\; \lambda_{0,\,m}\right)\right] + O(L^{-2}\log L),\end{aligned} \tag{C.11}$$

where (C.11) holds by (B.31). Hence

$$\begin{aligned}\lim_{L\to\infty}\frac{\beta_L(\mathbf{p},\, \mathbf{s})}{T_c/L} \;&=\; \sum_{m=1}^{M}\left[\sum_{i=0}^{2^N-1} p_i \zeta_m\left(\sum_{n=1}^{N} s_{n,\,m} b_n(i) A_n,\; \lambda_{0,\,m}\right)\right.\\ &\qquad\left. -\zeta_m\left(\sum_{i=0}^{2^N-1} p_i \sum_{n=1}^{N} s_{n,\,m} b_n(i) A_n,\; \lambda_{0,\,m}\right)\right],\end{aligned} \tag{C.12}$$

whence (4.85) follows.

Appendix D

Proof of Theorem 10

Denote the input and the output alphabets of the RF/optical sum channel by $\mathcal{X} = \mathbb{R} \uplus \mathbb{R}_0^+$ and $\mathcal{Y} = \mathbb{R} \uplus \mathbb{Z}_0^+$ respectively, where we use the notation $A \uplus B$ to denote the physical separability of the sets A and B by virtue of the material differences between the respective modulation and demodulation schemes. It can be inferred from [5, 31, 9] that the capacity of the memoryless RF/optical sum channel is given by

$$C_{sum} = \max_{P_{X,\,V|U}:\mathbf{E}[X^2\cdot 1(V=1)+X\cdot 1(V=0)]\leq P} I(X,\, V \wedge Y|S), \tag{D.1}$$

where the rvs X, Y, V, S, U take values in the sets $\mathcal{X}$, $\mathcal{Y}$, $\{0,\, 1\}$, $\mathbb{R}_0^+$, $\mathcal{U}$ respectively, and the joint distribution of the rvs X, Y, V, S, U is determined by $U = h(S)$, F_S, $P_{V|U}$,

$$\begin{aligned} X &= X^{(R)}1(V=1) + X^{(O)}1(V=0), \\ Y &= Y^{(R)}1(V=1) + Y^{(O)}1(V=0), \end{aligned} \tag{D.2}$$

with the rvs $X^{(R)}$, $Y^{(R)}$, $X^{(O)}$, $Y^{(O)}$ taking values in the sets $\mathbb{R}$, $\mathbb{R}$, $[0,\, A]$, $\mathbb{Z}_0^+$ respectively; the distribution function[1]

$$\begin{aligned} &W_{Y|X,V,S}(y|x,\, v,\, s) \\ &\quad = W^{(R)}_{Y^{(R)}|X^{(R)},S}(y|x,\, s)\cdot 1(v=1) + W^{(O)}_{Y^{(O)}|X^{(O)}}(y|x)\cdot 1(v=0), \end{aligned}$$

[1]We denote by $W_{Y|X,V,S}$ the conditional distribution function of a collection of discrete and continuous rvs.

$$x,\, y \in \mathbb{R}, \quad s \in \mathbb{R}_0^+, \quad v \in \{0,\, 1\}, \tag{D.3}$$

where $W^{(R)}_{Y^{(R)}|X^{(R)},S}(\cdot|\cdot,\,\cdot)$ and $W^{(O)}_{Y^{(O)}|X^{(O)}}(\cdot|\cdot)$ are given by (5.18) and (5.20) respectively; and the Markov condition

$$X,\, V -\!\circ\!- U -\!\circ\!- S. \tag{D.4}$$

Note that

$$\begin{aligned} I(X,\, V \wedge Y|S) &= I(V \wedge Y|S) + I(X \wedge Y|V,\, S) \\ &= H(V|S) - H(V|Y,\, S) + I(X \wedge Y|V,\, S) \\ &= H(V|S) + I(X \wedge Y|V,\, S) \qquad \text{(D.5)} \\ &= \mathbb{E}[h_b(\alpha(U))] + I(X \wedge Y|V,\, S), \qquad \text{(D.6)} \end{aligned}$$

where (D.5) follows by the physical separability of the RF and optical signals at the detector, so that $H(V|Y,\, S) = 0$; and (D.6) holds with $\alpha(u) = \Pr\{V = 1|U = u\}$, $u \in \mathcal{U}$, since by (D.4),

$$H(V|S) \;=\; H(V|U) \;=\; \mathbb{E}[h_b(\alpha(U))].$$

Next, for any $s \in \mathbb{R}_0^+$, note that

$$\begin{aligned} I(X \wedge Y|V = 1,\, S = s) &= I(X^{(R)} \wedge Y^{(R)}|V = 1,\, S = s) \qquad \text{(D.7)} \\ &= I(X^{(R)} \wedge Y^{(R)}|S = s), \qquad \text{(D.8)} \end{aligned}$$

where (D.7) is by (D.2), and (D.8) follows from the fact that the RF encoding and decoding are independent of the switching operation, so that V is conditionally independent of $(X^{(R)},\, Y^{(R)})$ given S. Similarly, for $s \in \mathbb{R}_0^+$,

$$I(X \wedge Y|V = 0,\, S = s) \;=\; I(X^{(O)} \wedge Y^{(O)}|V = 0,\, S = s) \tag{D.9}$$

$$= \quad I(X^{(O)} \wedge Y^{(O)}), \tag{D.10}$$

where (D.9) is by (D.2), and (D.10) follows from the fact that optical encoding and decoding are independent of the switching operation as well as RF channel fade, so that $(X^{(O)}, Y^{(O)})$ is independent of (V, S). Combining (D.6), (D.8), (D.10), we get that

$$\begin{aligned} I(X, V \wedge Y|S) \quad &= \quad \mathbb{E}[h_b(\alpha(U))] + \mathbb{E}[\alpha(U) I(X^{(R)} \wedge Y^{(R)}|S)] \\ &\quad + \mathbb{E}[(1-\alpha(U)) I(X^{(O)} \wedge Y^{(O)})]. \end{aligned} \tag{D.11}$$

Next, note that

$$\begin{aligned} &\mathbb{E}[X^2 \cdot 1(V=1) + X \cdot 1(V=0)] \\ &\quad = \quad \mathbb{E}[{X^{(R)}}^2 \cdot 1(V=1) + X^{(O)} \cdot 1(V=0)] && \text{(D.12)} \\ &\quad = \quad \mathbb{E}[\mathbb{E}[{X^{(R)}}^2 \cdot 1(V=1) + X^{(O)} \cdot 1(V=0)|S]] \\ &\quad = \quad \mathbb{E}[\alpha(U)\, \mathbb{E}[{X^{(R)}}^2|U] + (1-\alpha(U))\, \mathbb{E}[X^{(O)}]], && \text{(D.13)} \end{aligned}$$

where (D.12) is by (D.2); and (D.13) holds by (D.4), the conditional independence of V and $X^{(R)}$ given S, and the independence of (V, S) and $X^{(O)}$. Summarizing collectively (D.1), (D.11), (D.13), we get that

$$C_{sum} \quad = \quad \max_{\substack{\alpha:\mathcal{U}\to[0,\,1] \\ 0\le X^{(O)}\le A \\ \mathbf{E}[\alpha(U)\,\mathbf{E}[{X^{(R)}}^2|U]+(1-\alpha(U))\,\mathbf{E}[X^{(O)}]]\le P}} \begin{aligned} &\mathbb{E}[h_b(\alpha(U)) + \alpha(U)\, I(X^{(R)} \wedge Y^{(R)}|S) \\ &+(1-\alpha(U))\, I(X^{(O)} \wedge Y^{(O)})]. \end{aligned} \tag{D.14}$$

Finally, setting $C_{RF}(\cdot, \cdot)$ and $C_O(\cdot)$ as in (5.17) and (5.19) respectively, by (D.14), we have

$$C_{sum} \quad = \quad \max_{\substack{\alpha:\mathcal{U}\to[0,\,1] \\ \mu_1:\mathcal{U}\to\mathbb{R}_0^+ \\ 0\le\mu_2\le A \\ \mathbf{E}[\alpha(U)\mu_1(U)+(1-\alpha(U))\mu_2]\le P}} \begin{aligned} &\mathbb{E}[h_b(\alpha(U)) + \alpha(U)\, C_{RF}(S,\, \mu_1(U)) \\ &+(1-\alpha(U))\, C_O(\mu_2)], \end{aligned}$$

which is the desired result (5.15). $\square$

BIBLIOGRAPHY

[1] P. Bremaud. *Point Processes and Queues: Martingale Dynamics.* Springer-Verlag, New York, 1981.

[2] S. I. Bross, M. V. Burnashev, and S. Shamai. Error exponents for the two-user Poisson multiple-access channel. *IEEE Trans. Inform. Theory*, 47(5): 1999–2016, July 2001.

[3] S. I. Bross and S. Shamai. Capacity and decoding rules for Poisson arbitrarily varying channel. *IEEE Trans. Inform. Theory*, 49(11): 3076–3093, Nov 2003.

[4] M. V. Burnashev and Y. A. Kutoyants. On the sphere-packing bound, capacity, and similar results for Poisson channels. *Problems of Information Transmission*, 35(2): 95–111, Apr–June 1999.

[5] G. Caire and S. Shamai. On the capacity of some channels with channel state information. *IEEE Trans. Inform. Theory*, 45(6): 2007–2019, Sept 1999.

[6] G. Caire, G. Taricco, and E. Biglieri. Optimum power control over fading channels. *IEEE Trans. Inform. Theory*, 45(5): 1468–1489, July 1999.

[7] F. H. Clarke. *Optimization & Nonsmooth Analysis.* John Wiley & Sons, New York, 1983.

[8] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley & Sons, New York, 1991.

[9] A. Das and P. Narayan. Capacities of time-varying multiple-access channels with side information. *IEEE Trans. Inform. Theory*, 48(1): 4–25, Jan 2002.

[10] C. C. Davis, I. I. Smolyaninov, and S. D. Milner. Flexible optical wireless links and networks. *IEEE Commun. Mag.*, 41(3): 51–57, Mar 2003.

[11] M. H. A. Davis. Capacity and cutoff rate for Poisson-type channels. *IEEE Trans. Inform. Theory*, 26(6): 710–715, Nov 1980.

[12] M. R. Frey. Information capacity of the Poisson channel. *IEEE Trans. Inform. Theory*, 37(2): 244–256, Mar 1991.

[13] R. M. Gagliardi and S. Karp. *Optical Communications*. John Wiley & Sons, New York, 2nd edition, 1995.

[14] R. G. Gallager. *Information Theory and Reliable Communication*. John Wiley & Sons, New York, 1968.

[15] I. M. Gelfand and S. V. Fomin. *Calculus of Variations*. Prentice-Hall, New Jersey, 1963.

[16] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath. Capacity limits of MIMO channels. *IEEE J. Select. Areas Commun.*, 21(5): 684–702, June 2003.

[17] J. Gowar. *Optical Communication Systems*. Prentice-Hall International, Inc., London, 1984.

[18] S. M. Haas. *Capacity and coding for multiple-aperture, wireless, optical communications*. Ph.D. thesis, Massachusetts Inst. Tech., Cambridge, MA, 2003.

[19] S. M. Haas and J. H. Shapiro. Capacity of the multiple-input, multiple-output Poisson channel. In B. Pasik-Duncan, editor, *Proceedings of the Kansas Workshop on Stochastic Theory and Control*, Lecture notes in Control and Information Sciences, Lawrence, KS, Oct 2001.

[20] S. M. Haas, J. H. Shapiro, and V. Tarokh. Space-time codes for wireless optical communications. *Eurasip Journal on Applied Signal Processing*, 2002(3): 211–220, Mar 2002.

[21] S. M. Haas and J. H. Shapiro. Capacity of wireless optical communications. *IEEE J. Select. Areas Commun.*, 21(8): 1346–1357, Oct 2003.

[22] J. Hecht. *City of Light: The Story of Fiber Optics*. Oxford University Press, Inc., 1999.

[23] S. Hranilovic and F. R. Kschischang. Capacity bounds for power- and band-limited optical intensity channels corrupted by Gaussian noise. *IEEE Trans. Inform. Theory*, 50(5): 784–795, May 2004.

[24] A. Ishimaru. *Wave Propagation and Scattering in Random Media*. IEEE Press and Oxford University Press, 1997.

[25] Y. M. Kabanov. Capacity of a channel of the Poisson type. *Theory of Probability and its Applications*, 23(1): 143–147, 1978.

[26] J. M. Kahn and J. R. Barry. Wireless infrared communications. *Proc. IEEE*, 85(2): 265–298, Feb 1997.

[27] S. Karp, R. M. Gagliardi, S. E. Moran, and L. B. Stotts. *Optical Channels: Fibers, Clouds, Water, and the Atmosphere*. Plenum Press, NY, 1988.

[28] D. Killinger. Free space optics for laser communication through the air. *Optics & Photonics News*, 13(10): 36–42, Oct 2002.

[29] I. I. Kim, H. Hakakha, P. Adhikari, E. Korevaar, and A. K. Mazumdar. Scintillation reduction using multiple transmitters. *Free Space Laser Communication Technologies IX, Proc. SPIE*, 2990: 102–113, 1997.

[30] I. I. Kim and E. Korevaar. Availability of Free Space Optics (FSO) and hybrid FSO/RF systems. *Proceedings of SPIE ITCOM 2001; Optical Wireless Communications IV*, Vol. 4530, August 21, 2001.

[31] A. Lapidoth and S. M. Moser. Bounds on the capacity of the discrete-time Poisson channel. In *Proceedings of 41st Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 1-3, 2003.

[32] A. Lapidoth and S. Shamai. The Poisson multiple-access channel. *IEEE Trans. Inform. Theory*, 44(2): 488–501, Mar 1998.

[33] A. Lapidoth, I. E. Telatar, and R. Urbanke. On wideband broadcast channels. *IEEE Trans. Inform. Theory*, 49(12): 3250–3258, Dec 2003.

[34] E. J. Lee and V. W. S. Chan. Part 1: optical communication over the clear turbulent atmospheric channel using diversity. *IEEE J. Select. Areas Commun.*, 22(9): 1896–1906, Nov 2004.

[35] D. G. Luenberger. *Introduction to Linear and Nonlinear Programming.* Addison Wesley, MA, 1972.

[36] S. Masud, S. Buckley, and T. McKenna. Ten hottest technologies. *Telecommunications Americas*, 35(5): 30–35, May 2001.

[37] R. J. McEliece and W. E. Stark. Channels with block interference. *IEEE Trans. Inform. Theory*, 30(1): 44–53, Jan 1984.

[38] The Optical & RF Combined Link Experiment (ORCLE). Available online at http://www.darpa.mil/ato/programs/orcle.htm.

[39] M. S. Pinsker. *Information and Information Stability of Random Variables and Processes.* Holden-Day, Inc., San Fransisco, 1964.

[40] J. Proakis, E. Biglieri, and S. Shamai. Fading channels: information-theoretic and communications aspects. *IEEE Trans. Inform. Theory*, 44(6): 2619–2692, Oct 1998.

[41] S. Shamai. Capacity of a pulse-amplitude modulated direct detection photon channel. *Proc. IEEE*, 137(6): 424–436, Dec 1990.

[42] S. Shamai. On the capacity of a direct-detection photon channel with intertransition-constrained binary input. *IEEE Trans. Inform. Theory*, 37(6): 1540–1550, Nov 1991.

[43] S. Shamai and A. Lapidoth. Bounds on the capacity of a spectrally constrained Poisson channel. *IEEE Trans. Inform. Theory*, 39(1): 19–29, Jan 1993.

[44] J. H. Shapiro. Imaging and optical communication through atmospheric turbulence. In J. W. Strohbehn, editor, *Laser Beam Propagation in the Atmosphere*, chapter 6. Springer, Berlin, 1978.

[45] J. H. Shapiro. Random medium propagation theory applied to communication and radar system analyses. In J. Carl Leader, editor, *Laser Beam Propagation in the Atmosphere, Proceedings of SPIE*, 410: 98–102, Apr 1983.

[46] C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27: 379–423, 1948.

[47] D. N. Snyder and M. I. Miller. *Random Point Processes in Time and Space.* Springer-Verlag, New York, 2nd edition, 1991.

[48] V. I. Tatarskii. *Wave Propagation in a Turbulent Medium.* Dover, New York, 1961.

[49] H. A. Willebrand and B. S. Ghuman. Fiber optics without fiber. *IEEE Spectr.*, 38(8): 41–45, Aug 2001.

[50] H. A. Willebrand and B. S. Ghuman. *Free Space Optics: Enabling Optical Connectivity in Today's Networks*. SAMS Publishing, Indianapolis, IN, 2002.

[51] S. G. Wilson, M. Brandt-Pierce, Q. Cao, and J. Leveque. Free-space optical MIMO communication with Q-ary PPM. *IEEE Trans. Commun.*, 53(8): 1402–1412, Aug 2005.

[52] A. D. Wyner. Capacity and error exponent for the direct detection photon channel, parts I & II. *IEEE Trans. Inform. Theory*, 34(6): 1449–1471, Nov 1988.

[53] R. You and J. M. Kahn. Upper-bounding the capacity of optical IM/DD channels with multiple subcarrier modulation and fixed bias using trigonometric moment space method. *IEEE Trans. Inform. Theory*, 48(2): 514–523, Feb 2002.

[54] X. Zhu and J. M. Kahn. Free-space optical communication through atmospheric turbulent channels. *IEEE Trans. Commun.*, 50(8): 1293–1300, Aug 2002.