Resource Allocation for 5G Technologies under Statistical Queueing Constraints

Von der Fakultät für Elektrotechnik und Informatik
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades

Doktor-Ingenieur

genehmigte

Dissertation

von

M.Sc. Marwan Hammouda
geboren am 20. Dezember 1985 in Jabalia, Palästina

2019

1. Referent : Prof. Dr. Jürgen Peissig
2. Referent : Prof. Dr.-Ing. Harald Haas

Tag der Promotion : 02. Juli 2019

ABSTRACT

As the launch of fifth generation (5G) wireless networks is approaching, recent years have witnessed comprehensive discussions about a possible 5G standard. Many transmission scenarios and technologies have been proposed and initial over-the-air experimental trials have been conducted. Most of the existing literature studies on 5G technologies have mainly focused on the physical layer parameters and quality of service (QoS) requirements, e.g., achievable data rates. However, the demand for delay-sensitive data traffic over wireless networks has increased exponentially in the recent years, and is expected to further increase by the time of 5G. Therefore, other constraints at the data-link layer concerning the buffer overflow and delay violation probabilities should also be regarded. It follows that evaluating the performance of the 5G technologies when such constraints are considered is a timely task.

Motivated by this fact, in this thesis we explore the performance of three promising 5G technologies when operating under certain QoS at the data-link layer. We follow a cross-layer approach to examine the interplay between the physical and data-link layers when statistical QoS constraints are inflicted in the form of limits on the delay violation and buffer overflow probabilities. Noting that wireless systems, generally, have limited physical resources, in this thesis we mainly target designing adaptive resource allocation schemes to maximize the system performance under such QoS constraints.

We initially investigate the throughput and energy efficiency of a general class of multiple-input multiple-output (MIMO) systems with arbitrary inputs. As a cross-layer evaluation tool, we employ the effective capacity as the main performance metric, which is the maximum constant data arrival rate at a buffer that can be sustained by the channel service process under specified QoS constraints. We obtain the optimal input covariance matrix that maximizes the effective capacity under a short-term average power budget. Then, we perform an asymptotic analysis of the effective capacity in the low signal-to-noise ratio and large-scale antenna (massive MIMO) regimes. Such analysis has a practical importance for 5G scenarios that necessitate low latency, low power consumption, and/or ability to simultaneously support massive number of users.

Non-orthogonal multiple access (NOMA) has attracted significant attention in the recent years as a promising multiple access technology for 5G. In this thesis, we consider a two-user power-domain NOMA scheme in which both transmitters employ superposition coding and the receiver applies successive interference cancellation (SIC) with a certain order. For practical concerns, we consider limited transmission power budgets at the transmitters, and assume that both transmitters have arbitrarily distributed input signals. We again exploit the effective capacity as the main cross-layer

performance measure. We provide a resource management scheme that can jointly obtain the optimal power allocation policies at the transmitters and the optimal decoding order at the receiver, with the goal of maximizing the effective capacity region that provides the maximum allowable sustainable arrival rate region at the transmitters' buffers under QoS guarantees.

In the recent years, visible light communication (VLC) has emerged as a potential transmission technology that can utilize the visible light spectrum for data transmission along with illumination. Different from the existing literature studies on VLC, in this thesis we consider a VLC system in which the access point (AP) is unaware of the channel conditions, thus the AP sends the data at a fixed rate. Under this assumption, and considering an ON-OFF data source, we provide a cross-layer study when the system is subject to statistical buffering constraints. To this end, we employ the maximum average data arrival rate at the AP buffer and the non-asymptotic bounds on buffering delay as the main performance measures. To facilitate our analysis, we adopt a two-state Markov process to model the fixed-rate transmission strategy, and we then formulate the steady-state probabilities of the channel being in the ON and OFF states.

The coexistence of radio frequency (RF) and VLC systems in typical indoor environments can be leveraged to support vast user QoS needs. In this thesis, we examine the benefits of employing both technologies when operating under statistical buffering limitations. Particularly, we consider a multi-mechanism scenario that utilizes RF and VLC links for data transmission in an indoor environment. As the transmission technology is the main physical resource to be concerned in this part, we propose a link selection process through which the transmitter sends data over the link that sustains the desired QoS guarantees the most. Considering an ON-OFF data source, we employ the maximum average data arrival rate at the transmitter buffer and the non-asymptotic bounds on data buffering delay as the main performance measures. We formulate the performance measures under the assumption that both links are subject to average and peak power constraints.

Keywords: Quality of service, queueing constraints, cross-layer design, resource allocation, 5G, multiple-antenna systems, visible light communications, non-orthogonal multiple access.

ZUSAMMENFASSUNG

Da der Start von drahtlosen Netzwerken der fünften Generation (5G) bevorsteht, wurden in den letzten Jahren umfassende Diskussionen über einen möglichen 5G-Standard geführt. Viele Übertragungsszenarien und -technologien wurden vorgeschlagen, und es wurden anfängliche Versuche durchgeführt. Die meisten der vorhandenen Literaturstudien über 5G-Technologien haben sich hauptsächlich auf die Parameter der physikalischen Schicht und die Dienstgüte fokussiert, z. B. erreichbare Datenraten. Die Nachfrage nach verzögerungsempfindlichem Datenverkehr über drahtlose Netzwerke ist jedoch in den letzten Jahren exponentiell gestiegen und wird voraussichtlich bis zum Zeitpunkt des 5G weiter zunehmen. Daher sollten andere Beschränkungen auf der Sicherungsschicht, die die Pufferüberlauf- und Verzögerungsverletzungswahrscheinlichkeiten betreffen, ebenfalls betrachtet werden. Daraus folgt, dass die Bewertung der Leistung der 5G-Technologien, wenn solche Beschränkungen in Betracht gezogen werden, eine rechtzeitige Aufgabe ist.

Motiviert durch diese Tatsache untersuchen wir in dieser Arbeit die Leistung von drei viel versprechenden 5G-Technologien, wenn sie unter bestimmten Dienstgüte auf der Sicherungsschicht arbeiten. Wir verfolgen einen Cross-Layer-Ansatz, um die Interaktion zwischen der physikalischen und der Sicherungsschicht zu untersuchen, wenn statistische Dienstgüte Einschränkungen in Form von Grenzen für die Verzögerungsverletzung und Pufferüberlaufwahrscheinlichkeiten auftreten. Unter Berücksichtigung der Tatsache, dass drahtlose Systeme im Allgemeinen über begrenzte physische Ressourcen verfügen, zielen wir in dieser Arbeit hauptsächlich darauf ab, adaptive Ressourcenzuweisungsstrategien zu entwerfen, um die Systemleistung unter solchen Dienstgüte Einschränkungen zu maximieren.

Wir untersuchen zunächst den Durchsatz und die Energieeffizienz einer allgemeinen Klasse von Systemen mit mehreren Sende- und Empfangsantennen (MIMO) mit beliebigen Eingangssignalen. Als Cross-Layer-Evaluierungswerkzeug verwenden wir die effektive Kapazität als Hauptleistungsmetrik, die die maximale konstante Datenankunftsrate bei einem Puffer ist, die durch den Kanaldienstprozess unter spezifizierten Dienstgüte Einschränkungen aufrechterhalten werden kann. Wir formulieren die optimale Kovarianzmatrix, die die effektive Kapazität bei einem kurzfristigen durchschnittlichen Leistungsbudget maximiert. Dann führen wir eine asymptotische Analyse der effektiven Kapazität in dem niedrigen Signal-zu-Rausch-Verhältnis und großen Antennen- (massiven MIMO) Regimes durch. Eine solche Analyse hat eine praktische Bedeutung für 5G-Szenarien, die eine niedrige Latenz, einen niedrigen Stromverbrauch und / oder die Fähigkeit zur gleichzeitigen Unterstützung einer großen Anzahl von Benutzern erfordern.

Nicht-orthogonaler Mehrfachzugriff (NOMA) hat in den letzten Jahren als vielversprechende Multiple-Access-Technologie für 5G große Beachtung gefunden. In dieser Arbeit betrachten wir ein Zwei-Benutzer-Power-Domain-NOMA-Schema, bei dem beide Sender eine Überlagerungscodierung verwenden und der Empfänger aufeinanderfolgende Interferenzlöschung (SIC) mit einer bestimmten Reihenfolge anwendet. Aus praktischen Erwägungen betrachten wir begrenzte Übertragungsleistungsbudgets bei den Sendern und nehmen an, dass beide Sender willkürlich verteilte Eingangssignale haben. Wir nutzen die effektive Kapazität als wichtigste übergreifende Leistungsbewertung. Wir formulieren ein Ressourcenverwaltungsschema bereit, das gemeinsam die optimalen Leistungszuweisungsrichtlinien bei den Sendern und die optimale Decodierreihenfolge beim Empfänger mit dem Ziel erhält, den effektiven Kapazitätsbereich zu maximieren.

In den letzten Jahren hat sich die Kommunikation mit sichtbarem Licht (VLC) als eine mögliche Übertragungstechnologie herausgebildet. Im Unterschied zu den bisherigen Literaturstudien zu VLC betrachten wir in dieser Arbeit ein VLC-System, bei dem der Access Point (AP) die Kanalbedingungen nicht kennt und der AP die Daten daher mit einer festen Rate sendet. Unter dieser Annahme und unter Berücksichtigung einer ON-OFF-Datenquelle bieten wir eine Cross-Layer-Studie an, wenn das System statistischen Pufferbedingungen unterliegt. Zu diesem Zweck verwenden wir die maximale durchschnittliche Datenankunftsrate bei dem AP-Puffer und die nicht-asymptotischen Grenzen bei der Pufferung Beschränkungen als die Hauptleistungsmaße. Um unsere Analyse zu erleichtern, verwenden wir einen Markov-Prozess mit zwei Zuständen. Um die Übertragungsstrategie mit fester Rate zu modellieren, formulieren wir dann die stationären Wahrscheinlichkeiten des Kanals, der sich in den Zuständen EIN und AUS befindet.

Die Koexistenz von Radiofrequenz (RF) - und VLC-Systemen in typischen Innenraumumgebungen kann genutzt werden, um umfangreiche Benutzer- Dienstgüte -Anforderungen zu unterstützen. In dieser Arbeit untersuchen wir die Vorteile der Verwendung beider Technologien, wenn sie unter statistischen Pufferung Beschränkungen arbeiten. Insbesondere betrachten wir ein Multimechanismus-Szenario, das die Kombination von RF- und VLC-Verbindungen für die Datenübertragung in einer Innenumgebung verwendet. Da die Übertragungstechnologie die wichtigste physische Ressource in diesem Teil ist, schlagen wir einen Linkauswahlprozess vor, bei dem der Sender Daten über die Verbindung sendet, die die gewünschten Dienstgüte Einschränkungen am meisten unterstützt. Betrachtet man eine ON-OFF-Datenquelle, verwenden wir die maximale durchschnittliche Datenankunftsrate im Senderpuffer und die nicht asymptotischen Grenzen für die Datenpufferverzögerung als Hauptleistungsmaß. Wir formulieren die Leistungsmaße unter der Annahme, dass beide Verbindungen mittleren und maximalen Leistungsbeschränkungen unterliegen.

Schlagwörter: Dienstgüte Einschränkungen, Pufferung Beschränkungen, Cross-Layer-Analyse, Ressourcenallokation, 5G, Mehrantennensysteme, Kommunikation mit sichtbarem Licht, Nicht-orthogonaler Mehrfachzugriff.

CONTENTS

LIST OF FIGURES

GLOSSARY OF ACRONYMS

5G	Fifth-generation cellular systems
AP	Access Point
bpcu	bits per channel use
BPSK	Binary Phase-Shift Keying
CDF	Cumulative Distribution Function
CDMA	Code-Division Multiple Access
FDMA	Frequency-Division Multiple Access
FOV	Field of View
IoT	Internet of Things
IR	Infrared
LED	Light Emitting Diode
LoS	Line-of-Sight
MGF	Moment Generating Function
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
MMSE	Minimum Mean-Square Error
NOMA	Non-orthogonal Multiple Access
OMA	Orthogonal Multiple Access
PD	Photodetector
PDF	Probability Density Function
QAM	Quadrature Amplitude Modulation
QoS	Quality-of-Service
RF	Radio Frequency
SIC	Successive Interference Cancellation
SIMO	Single-Input Multiple-Output
SNR	Signal-to-Noise Ratio
TDMA	Time-Division Multiple Access
VLC	Visible Light Communications
VNI	Virtual Network Index

Part I

DISSERTATION

1 INTRODUCTION

1.1 RESOURCE ALLOCATION IN WIRELESS SYSTEMS

The time-varying nature of wireless channels is one of the fundamental and unique challenges in wireless communications. Different effects such as multipath fading and shadowing, which occur due to mobility and physical changes in the surrounding environment, can result in both short-term and long-term variations in the channel strength. In addition to the randomly changing channel over time, having limited transmission resources is another key factor that has potential impacts on the design and performance of wireless systems. These resources may include transmission power, bit rate, space (in systems with multiple antenna), and access technology (in hybrid systems).

Given the harsh channel conditions and the scarce transmission resources, how to design a wireless system that can guarantee certain quality of service (QoS) requirements has been a key research theme in the last few decades. In this direction, adaptive transmission schemes, in which the transmission resources are adapted with respect to the channel conditions, have been considered as a potential solution. In principle, such techniques enable a more efficient utilization of the transmission resources, thus enhance the system performance towards satisfying the required QoS needs. This line of research has attracted much interest and huge effort has been expended to design several wireless systems while considering different transmission resources and QoS requirements. As for instance, power optimization has been researched in [1–10], rate adaptation has been regarded in [10–15], and optimal access technology selection in hybrid systems has been investigated in [16–24].

1.2 5G TECHNOLOGIES

Since the first successful demonstrations of wireless telephony by Macroni in 1894, wireless communication has witnessed revolutionary improvements in the supported QoS to the end users. Driven by these achievements, wireless networks have also experienced exponentially increasing demands for wireless data, which are expected to keep increasing in the future. For example, the most recent visual network index (VNI) report by Cisco showed that the global mobile data traffic reached 7.2 exabytes[1] per month at the end of 2016, while this amount is expected to increase 7-fold by 2021 to reach 49 exabytes per month [25]. In addition to this deluge of data, the number of mobile-connected devices continuous to increase exponentially and is

1 exa = 10^{18}

expected to reach 11.6 billion devices by 2021, which exceeds the world's projected population of 7.8 billion at that time[25]. This ever-increasing demand for wireless traffic triggered a quest for technical solutions that can support stringent and diverse QoS needs of the end users.

As the current generation of cellular technology, i.e. 4G, has been deployed for about a decade, intense discussions have been made in the recent years to provide a possible standard about the next wireless generation, commonly known as 5G. The first commercial version of 5G is foreseen to be released by 2020. Thus, a major part of research activities in the recent years has been spent to form a comprehensive framework that defines new transmission scenarios, design requirements, technical challenges, and possible solutions for 5G systems, see e.g., [26–28]. As a matter of fact, the key 5G requirements include supporting 10 to 100 times higher data transmission rate and providing 10 times longer battery life than the current 4G technology, while achieving a round-trip latency of about 1 millisecond [26, 27]. Among the different technological solutions that have been proposed in the literature to achieve these requirements, in this thesis we mainly focus on three technologies, namely massive multiple-input multiple-output (MIMO), non-orthogonal multiple access (NOMA), and visible light communications (VLC). In the following three subsections, we briefly describe these technologies.

1.2.1 *Massive MIMO*

Multiple-input multiple-output (MIMO) wireless channels have been widely studied since the pioneering works by Foschini [29] and Telatar [9]. It has been recognized that using multiple transmit and receive antennas can remarkably enhance the system performance in terms of reliability and/or spectral efficiency [30]. However, maximizing the spectral efficiency is normally a conflicting objective with minimizing the transmission energy, which is also a critical requirement for 5G. Thus, system designers traditionally need to consider spectral and energy efficiency trade-off [31, 32]. To overcome this problem, the authors in [33] showed that increasing the transmit and/or receive antennas without bounds can substantially improve the system spectral efficiency while making the transmit power arbitrarily small. Commonly referred to as massive MIMO, this technology has been considered as a disruptive technology for 5G [27, 28, 34].

One of the key features of massive MIMO systems is the so-called channel hardening phenomena, which implies reducing the small-scale randomness in the channel and, thus, the channel behaves almost as a non-fading channel [35, 36]. Physically, increasing the number of antennas means increasing the number of channel observations, which results in a smoothed channel response as a consequence of the law of large numbers. For example, let us consider a simple scenario where we set the number of transmit

antennas to $M = 1$, and let $\mathbf{h}$ be the $N \times 1$ channel vector, where N is the number of receive antennas. The concept of channel hardening implies that

$$\frac{\|\mathbf{h}\|^2}{\mathbb{E}\{\|\mathbf{h}\|^2\}} \xrightarrow{P} 1, \qquad \text{as} \quad N \to \infty \tag{1.1}$$

where $\xrightarrow{P}$ refers to a convergence in probability. A fundamental advantage of channel hardening is that the resource allocation schemes can be adapted with respect to the large-scale variations in the channel instead of the small-scale fading, which obviously simplifies the system design and reduces the signaling overhead.

Massive MIMO systems have been investigated by many researchers from the information-theoretic perspectives [37–43]. For instant, the authors in [40] studied the energy and spectral efficiency on the uplink of multiuser massive MIMO systems. In addition, the authors in [43] derived the optimal input covariance matrices for multiple access channels with massive antennas at both sides to maximize the sum rate. Finally in this subsection, we refer to the first real-time testbed for massive MIMO [44, 45]. In this testbed, a MIMO system was implemented with a total number of 160 dual polarized and half-wavelength shorted patch antennas, and two racks of $0.8\text{m} \times 1.2\text{m} \times 1\text{m}$ were used to assemble all system components. The system was designed to operate at the center frequency of 3.7 GHz. It is worth remaking that more antennas can be mounted in smaller areas in the Millimeter Wave range, i.e., 30-300 GHz, since antenna sizes shrink with frequency [27]. However, operating in this frequency range requires developing new channel models that can capture the propagation characteristics [27].

1.2.2 *Non-Orthogonal Multiple Access*

Unlike wired systems where data transmission is normally performed over point-to-point links, wireless systems allow multiple transmitters and/or receivers to share the same transmission media and resources. Thus, developing multiple-access schemes that can efficiently utilize the shared transmission media and resources, while minimizing (or ideally eliminating) the resulting interference, is one of the major challenges in wireless systems. In principle, the conventional orthogonal multiple access (OMA) schemes of time-division multiple access (TDMA), frequency-division multiple access (FDMA), and code-division multiple access (CDMA) are commonly employed. In these techniques, different users are allocated orthogonal resources in either time, frequency, or code, respectively.

In principle, applying OMA schemes has a fundamental advantage that simple single-user detection methods can be employed since the inter-cell interference is, ideally, eliminated due to the orthogonal resource allocation. However, these schemes have several limitations, two of which we highlight in the following, making such methods not to be the proper solutions for 5G networks. First, the orthogonal schemes are in general sub-optimal in terms of the achievable rates and they cannot always achieve the capacity region

of multi-user systems [46, Chapter 6]. Recall that increasing the spectral efficiency is one of the key requirements in 5G. Second, it is obvious that the number of users that a given orthogonal scheme can support is limited by the number of available orthogonal resources. Indeed, a scenario with a massive connectivity is highly expected in 5G as the demand for wireless traffic keeps increasing and due to the rapid growth of the Internet of Things (IoT). To overcome such limitations, many researchers have suggested employing non-orthogonal multiple access (NOMA) schemes, where the ability of such schemes to enhance the system throughput and support massive connectivity has been proved, see e.g., [47–49] and references therein.

The key idea of NOMA is to allow all users to access the entire frequency and time resources simultaneously, while exploiting either the power or code domain to distinguish different users. This concept has been gaining an increasing attention in the recent years and several NOMA schemes have been proposed in many studies, see e.g., [47] and references therein. In this thesis, we are mainly interested in the power-domain NOMA, which has been shown to be a capacity-achieving multiple access scheme in multi-user systems [46, Chapter 6]. To achieve such a performance, active users are allocated different power levels based on their channel conditions, whereas successive interference cancellation (SIC) is employed with a certain order to manage the inter-user interference. It follows that a joint optimization of the transmitted power levels and the SIC order is required in power-domain NOMA in order to get the required capacity gain. This indeed creates a trade-off between the achieved performance gain and the system design complexity. Other practical issues in using SIC in wireless systems, such as imperfect channel knowledge and propagation error, have been discussed in [46, Chapter 6]. Power-domain NOMA has been investigated in many studied from an information-theoretic perspective, see e.g., [3, 50–52] for multiple-access channels (uplink) and [4, 53–55] for broadcast channels (downlink).

1.2.3 *Visible Light Communications*

As the radio frequency (RF) spectrum is a scarce resource and that parts of it are licensed to be used by certain services, the increasing demand for wireless services has motivated researchers and regulation bodies to find new frequency bands. In this regard, the Millimeter Wave (mmWave) range of $30 - 300$ GHz [56, 57], and the optical spectra [58] are promising candidates due to wide bandwidths available in both ranges. However, moving to the mmWave range still needs further investigations and many issues still need to be considered and explored, such as the channel and propagation characteristics [27]. The optical spectra includes the infrared (IR) and the visible light regions.

Recently, the noticeable progress in white light emitting diodes (LEDs) technology motivated utilizing the visible light spectrum for data transmission along with the white LEDs main function of illumination. In addition to

providing the required bandwidth, using white LEDs for data transmission has advantages over the RF and IR technologies [59]. For example, LEDs are cheap, energy efficient and no extra infrastructure is needed since white LEDs would be already installed for lighting. For instant, the European commission has recently decided to replace the standard incandescent lamps with more efficient lighting technologies including LED technology, which is expected to take the significant share of the future lighting market [60]. Furthermore, using white LEDs for data transmission would inherently ensure data security since light does not penetrate through walls. Using white LEDs for data communication in indoor scenarios is commonly referred to as visible light communications (VLC), which will be one of the target technologies in this thesis. Notice that the choice of focusing only on indoor scenarios in this regard is due to the observation that the majority of the data traffic (around 70%) occurs in indoor environments [61].

While having the aforementioned advantages and benefits, VLC systems have also certain limitations and challenges, e.g., smaller coverage, strong dependence on line-of-sight components and achievable rates that vary with spatial fluctuations [21]. In order to overcome these constraints, hybrid networks that integrate RF and VLC technologies have been proposed intensively in the recent years in order to achieve the end-user demands of both capacity and coverage, which are difficult to meet when either technology is operating solely. In addition, such networks can be practically feasible with no extra infrastructure costs since both RF and VLC systems already coexist and operate in the same area in many indoor scenarios, like offices. In this line of research, the authors in [21, 22, 61, 62] explored the performance of hybrid RF/VLC systems, where the results showed substantial improvements over pure RF and pure VLC networks in terms of throughput and energy efficiency. Furthermore, various load balancing schemes in hybrid RF/VLC systems were addressed in [63, 64].

At this point, we highlight the reason for choosing the aforementioned 5G technologies to be further investigated in this thesis. Among the other requirements for 5G, increasing data rates is the one that gets the most attention in the literature [27]. As mentioned in [27], such a need can be mostly achieved by solutions that belong to one (or a combination) of the following categories: *i)* increasing the area spectral efficiency through extreme densification, *ii)* improving the spectral efficiency mainly through advances in MIMO, and/or *iii)* increasing the bandwidth by moving to new frequency bands. A better representation for such solutions can be as follows. If the target is to increase the total amount of data that a network can support, i.e., area capacity in *bits per second and unit area*, then we can relate between this goal and the above-described solutions as follows:

$$\underbrace{\frac{\text{bits/s}}{\text{unit area}}}_{\text{area capacity}} = \underbrace{\frac{\text{nodes}}{\text{unit area}}}_{\text{densification}} \times \underbrace{\frac{\text{bits/s}}{\text{node} \times \text{Hz}}}_{\text{spectral efficiency}} \times \underbrace{\text{Hz}}_{\text{bandwidth}}$$

So, we can easily observe that increasing any of the three terms in the right-hand side, or any combination of them, can improve the area spectral

efficiency. Back to the 5G technologies considered in this thesis, we observe that each of them belongs to at least one of the solution categories. Specifically, massive MIMO and NOMA belong to the second category as both technologies can substantially increase the system spectral efficiency. Moreover, the VLC technology is a good example for a solution that increases both the spectrum and the densification as more APs are expected to be located in smaller areas compared to RF systems.

1.3 CROSS-LAYER CONCEPTS

All of the studies mentioned above analyzed wireless systems from the physical layer perspectives only. They mainly focused on QoS requirements at the physical layer and used signal-to-noise ratio (SNR) as a performance measure. Note that different QoS requirementrs at the physical layer, such as bit error rate and transmission data rate, can be expressed in terms of SNR. However, a dramatic increase in the demand on delay-sensitive services has been observed in the recent years. For example, mobile video traffic accounted for 60% of the global data traffic in 2016, which is further expected to increase 9-fold and reach 78% by the end of 2021 as reported in the VNI Cisco report [25]. For such services SNR is not a sufficient metric, and other QoS measures at the data-link layer, such as limitations on delay violation and buffer overflow probabilities, should be also regarded. Therefore, there is an insisting need for analytic frameworks that take the QoS requirements at the data-link layer regarding the buffer dynamics into account.

In this direction, cross-layer analysis has been gaining an increasing attention as a powerful tool to study and assess different QoS mechanisms in wireless networks [65, 66]. As we target evaluating the performance of three 5G technologies, in this thesis we are mainly interested in the cross-layer analysis regarding the physical and data link layers. Particularly, we investigate the impacts of different parameters at the physical layer when certain QoS requirements are imposed at the data-link layer as limits on the buffer overflow and delay violation probabilities. Recall that unlike wired networks, which provide deterministic transmission rates, wireless links have a time-varying nature, thus it is not easy to sustain stable transmission rate over wireless channels. This implies that deterministic QoS requirements can never be guaranteed. Instead, applying statistical QoS requirements seem to be more realistic in such a case, where the desired bounds on the buffer dynamics can be satisfied with a small violation probability. In this context, the cross-layer analysis regarding the physical and data-link layers has been investigated in many different wireless scenarios [2, 67–79].

The first cross-layer analysis was performed during the early 90's in wired networks, in which service processes can be safely assumed to have constant transmission rates. Then, by only considering the stochastic nature of data arrival processes, the notion of effective bandwidth was introduced as a cross-layer probing tool [80]. The effective bandwidth of a time-varying data arrival process identifies the minimum service rate required to satisfy

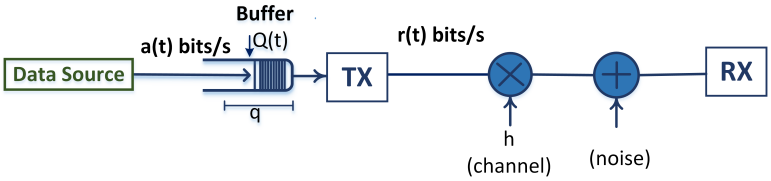


Figure 1.1: Simple channel model for cross-layer analysis.

certain QoS requirements, represented mathematically by an exponent θ (to be detailed later in this section). In contrast to the deterministic nature of wired networks, recall that wireless service links demonstrate generally a stochastic behavior. In this context, the authors in [81] introduced the notation of effective capacity by considering a stochastic service process while assuming a constant-rate arrival process. As the dual concept to the effective bandwidth, the effective capacity characterizes the maximum constant arrival rate that a time-varying service process can attain while satisfying given statistical QoS constraints defined in terms of the exponent θ.

In the remaining of this section, we present the mathematical framework of the cross-layer analysis to be investigated in this thesis. For this purpose, we present a simple channel model in Fig. 1.1. Here, the data initially arrives at the transmitter buffer from a source (or sources) with rate $a(t)$ bits/s for $t \in \{1, 2, \cdots\}$ and is stored in the buffer. Following the encoding and modulation processes, the transmitter sends the data to the receiver over the wireless channel with rate $r(t)$ bits/s. For buffer stability, notice that the average arrival rate should not exceed the average transmission rate, i.e., we should have $\mathbb{E}\{a(t)\} \leqslant \mathbb{E}\{r(t)\}$, where $\mathbb{E}\{\cdot\}$ denotes the expected value.

As the data is initially stored in the buffer before being transmitted, applying certain constraints on the buffer dynamics is indeed essential to control the data backlog and buffering delay. Some literature studies, see e.g., [82, 83] and reference therein, have already addressed the performance of wireless systems under delay requirements. However, these studies only regarded the average delay of the wireless transmission as the main performance metric. It is clear that such requirements are not sufficient for real-time services, such as the multimedia transmissions, where the bounded delay is the main concern. Subsequently, herein we assume that the system is operating under statistical constraints as limits on the buffer overflow (data backlog) and delay violation probabilities. In particular, let Q be the stationary queue length, $Q(t)$ (see Fig. 1.1), and q be the buffer overflow threshold. Then, we assume that the buffer overflow probability, i.e., $\Pr\{Q \geqslant q\}$, satisfies [80, Eq. (63)]:

$$\theta = -\lim_{q \to \infty} \frac{\log_e \Pr\{Q \geqslant q\}}{q}, \tag{1.2}$$

where $\theta > 0$ denotes the decay rate of the tail distribution of the data backlog, Q. Accordingly, we can approximate the buffer overflow probability for a large threshold, q_{max}, as

$$\Pr\{Q \geqslant q_{max}\} \approx e^{-\theta q_{max}}, \tag{1.3}$$

The approximation above implies that the buffer overflow probability should decay exponentially with a rate controlled by θ, which is called the QoS exponent. Basically, larger θ implies stricter QoS constraints, whereas smaller θ corresponds to looser constraints. In a similar way, an exponentially decaying approximation for the delay violation probability can be obtained as follows [84]:

$$\Pr\{D \geqslant d_{max}\} \approx e^{-\theta \vartheta d_{max}}, \tag{1.4}$$

for large threshold, d_{max}, where D is the steady-state delay experienced at the buffer and ϑ is jointly determined by both the arrival and service processes, see e.g., Fig. 3 in [84].

Now, let $\Lambda_a(\theta)$ and $\Lambda_c(\theta)$ be, respectively, the asymptotic log-moment generating functions (LMGFs) of the total amount of bits arriving at the transmitter buffer and the total service from the transmitter in the channel, i.e.,

$$\begin{aligned} \Lambda_a(\theta) &= \lim_{t\to\infty} \frac{1}{t} \log_e \mathbb{E}\left\{ e^{\theta \sum_{k=1}^{t} a(k)} \right\} \\ \text{and} \quad \Lambda_c(\theta) &= \lim_{t\to\infty} \frac{1}{t} \log_e \mathbb{E}\left\{ e^{\theta \sum_{k=1}^{t} r(k)} \right\}, \end{aligned} \tag{1.5}$$

where $a(t)$ and $r(t)$ are, respectively, the arrival and service processes, which are assumed to be stationary and ergodic. Let us further consider independent data arrival and work-conserving data service processes. For a given QoS exponent θ, it was shown in [85, Theorem 2.1] that the constraint in (1.2), or equivalently in (1.3) for a large threshold, is guaranteed only when the arrival and service processes satisfy

$$\Lambda_a(\theta) = -\Lambda_c(-\theta). \tag{1.6}$$

From the formulations above, we observe that characterizing the LMGFs is a main challenge in the cross-layer framework provided in this thesis. For more details about obtaining the LMGFs, we refer to [86, Example 7.2.7]. For example, let us consider the special case of a block-fading channel model. Specifically, we assume that the fading coefficient, h in Fig. 1.1, stays constant during one transmission block and changes independently from one block to another. We further assume that the fading process is identical in each block, and that the noise process n is also independent and identically distributed. In such a case, the log-moment generation function of the service process, i.e., $\Lambda_c(\theta)$ in (1.5), simplifies to

$$\Lambda_c(\theta) = \log_e \mathbb{E}\{e^{\theta r(k)}\}. \tag{1.7}$$

Proof: As the fading and noise follow independent and identically distributed processes in different blocks, then the service rate $\{r(k)\}$ is also an independent and identically distributed process. Subsequently, we have

$$\Lambda_c(\theta) = \lim_{t\to\infty} \frac{1}{t} \log_e \mathbb{E}\left\{ e^{\theta \sum_{k=1}^{t} r(k)} \right\} \tag{1.8}$$

$$= \lim_{t\to\infty} \frac{1}{t} \log_e \mathbb{E}\left\{ \prod_{k=1}^{t} e^{\theta r(k)} \right\} \tag{1.9}$$

$$= \lim_{t\to\infty} \frac{1}{t} \log_e \prod_{k=1}^{t} \mathbb{E}\{e^{\theta r(k)}\} \tag{1.10}$$

$$= \lim_{t\to\infty} \frac{1}{t} \log_e (\mathbb{E}\{e^{\theta r(k)}\})^t \tag{1.11}$$

$$= \lim_{t\to\infty} \frac{1}{t} t \log_e \mathbb{E}\{e^{\theta r(k)}\} \tag{1.12}$$

$$= \log_e \mathbb{E}\{e^{\theta r(k)}\} \tag{1.13}$$

Above, (1.10) follows from the independence of the service process, while we have (1.11) as the service process is identically distributed. □

Notice that when the service process has a fixed transmission rate, i.e., $r(t) = r$ for all t, then we simply have $\Lambda_c(\theta) = r\theta$. For such a setting, recall that the effective bandwidth of a given time-varying arrival process defines the minimum service rate, i.e., $r^\star \leqslant r$, required to fulfill the QoS needs defined by the exponent θ. Likewise, when the arrival process has a fixed rate, i.e., $a(t) = a$ for all t, then we have $\Lambda_a(\theta) = a\theta$. In this case, the effective capacity of a given time-varying service process defines the maximum arrival rate, i.e., $a^\star \geqslant a$, that can be sustained while achieving the required QoS needs. Consequently, based on the condition (1.6) we can immediately formulate the effective bandwidth of a time-varying arrival process and the effective capacity of a time-varying service process as a function of θ, respectively, as

$$r^\star = \frac{\Lambda_a(\theta)}{\theta} \quad \text{and} \quad a^\star = -\frac{\Lambda_c(-\theta)}{\theta}. \tag{1.14}$$

It follows that, the condition in (1.6) can be re-expressed as follows: the QoS constraints regarding the buffer overflow and delay violation probabilities, respectively expressed in (1.3) and (1.4), can be achieved in a given system only when the effective bandwidth of the arrival process is equal to the effective capacity of the service process. It is worth remarking that, when being plotted with respect to θ, the value of ϑ in (1.4) is equal to the intersection point of the effective bandwidth and the effective capacity curves, i.e., $\vartheta = r^\star = a^\star$, [84].

1.4 THESIS CONTRIBUTIONS

Regarding the physical and data-link layers, in this thesis we provide cross-layer studies for the three 5G technologies briefly described in Section 1.2. We are particularly interested in designing resource allocation schemes at

the physical layer that can optimize the system performance when certain QoS are imposed at the data-link layer as limits on the buffer overflow and delay violation probabilities. In the following we categorize the main contributions of this thesis based on the considered 5G technology.

MIMO Systems: In this part, we focus on a general MIMO scenario with arbitrary input signaling. We provide a mathematical toolbox that system designers can use in order to understand performance levels of spectrum and energy efficient systems under QoS constraints imposed as limits on the buffer overflow and delay violation probabilities. More specifically, we can list the main contributions of this part as follows:

- Assuming that the instantaneous channel fading gain estimate is available at both the transmitter and the receiver, we identify the optimal input covariance matrix that maximizes the effective capacity under a short-term average power constraint over the transmit antennas.
- We obtain the first and second derivatives of the effective capacity when the signal-to-noise ratio goes to zero. Using these derivatives, we obtain a linear approximation of the effective capacity in the low signal-to-noise ratio regime. We show that this approximation does not depend on the input distribution and covariance matrix.
- We further show that the minimum *energy-per-bit* is obtained when the signal-to-noise ratio goes to zero and that it is independent of the QoS constraints, the input distribution, and the covariance matrix.
- In the massive MIMO regime, we prove that the effective capacity approaches the average mutual information in the channel, i.e., the dependence of the effective capacity performance on the QoS constraints decreases with the increasing number of antennas.

NOMA Systems: Next, in this part we provide a cross-layer analysis of a power-domain NOMA system under statistical QoS needs. In particular, we focus on a two-user multiple access transmission scenario in which transmitters apply arbitrarily distributed input signaling under average power constraints. Recall that a joint optimization of the transmission powers at the transmitters and the decoding order at the receiver is an essential requirement in such a scenario. The main contributions of this part can be sorted as follows:

- Defining the effective capacity region by employing the effective capacity of each transmitter, we provide the optimal power allocation policies under an average transmission power constraint.
- We make use of the relationship between the mutual information and the minimum mean-square error (MMSE) in obtaining the power allocation policies.
- We attain the optimal decoding order that is administered at the receiver regarding the interplay between the channel fading coefficients.

VLC Systems: Then, we explore the performance of a VLC system that operates under statistical QoS constraints and considering an ON-OFF data source model. Different than the existing literature studies in VLC, we assume that the VLC access point has no knowledge about the user channel gain, thus the access point sends the data with a fixed rate. For the evaluation analysis, we employ the maximum average arrival rate at the transmitter buffer and the non-asymptotic bounds on buffering delay as the main performance metrics. To summarize, the main contributions of this part are as follows:

- Considering a practical fixed-rate transmission scenario, we initially model the VLC channel as a two-state Markov process, such that the channel is assumed to be in the ON state when a reliable transmission is guaranteed.
- Under the assumptions of line-of-sight transmission and random user distribution, we formulate the steady-state probabilities of the channel being in the ON and OFF states.
- For an ON-OFF data source model, we evaluate and simulate the maximum average arrival rate when the system is designed to satisfy certain QoS levels.

Hybrid RF/VLC Systems: Finally, in this thesis we provide a cross-layer study for a hybrid RF/VLC system in which the transmitter can use both RF and VLC channels, either separately or simultaneously, for data transmission. We further assume an ON-OFF modeled data arrival process at the transmitter buffer. We employ first the maximum average data arrival rate at the transmitter buffer considering the asymptotic buffer overflow probability approximation, and then non-asymptotic buffering delay violation probability as the main performance measures. We propose a mathematical toolbox to system designers for performance analysis in hybrid RF/VLC systems that work under low latency conditions. To summarize, the main contributions of this study are as the following:

- Assuming that both RF and VLC links are subject to average and peak power constraints, we express the maximum average data arrival rate at the transmitter that the data service process from the transmitter to the receiver can support under QoS constraints when either the RF or VLC link is used, or both links are simultaneously used for data transmission.
- We propose three different link usage strategies. We base two of the proposed strategies on the assumption that the receiver does not have a multihoming capability, thus data transmission is possible over only one link, either the RF or VLC link. In the third strategy, we assume that link aggregation is possible and data can be transmitted over both links simultaneously following a power sharing policy.

- We obtain the non-asymptotic data backlog and buffering delay violation probability bounds considering the proposed link usage strategies.

1.5 THESIS OUTLINE

As we explore the cross-layer performance in different 5G technologies, each of the following chapters particularly focuses on one of the technologies introduced in Section 1.2, with the main target of achieving the corresponding contributions, as summarized in Section 1.4. Each of these chapters begins with an introduction part, where we review related work on the performance evaluation of the corresponding technology, especially those studies with information-theoretic and cross-layer investigations. Then, we present the system model to be analyzed throughout that chapter before providing detailed performance analysis. At the end of each chapter, we present comprehensive numerical results to validate our theoretical findings.

The rest of this thesis is organized as follows:

- Chapter 2 provides the cross-layer analysis in MIMO systems. We initially provide the optimal input covariance matrix to maximize the effective capacity. Then, we perform asymptotic analyses in the low signal-to-noise ratio regime and in the massive-MIMO regime.
- Chapter 3 explores the cross-layer analysis in two-user power-domain NOMA channels. We mainly target an adaptive resource allocation scheme that can jointly optimize the transmitting power levels at the transmitters and the decoding order at the receiver to maximize the effective capacity.
- Chapter 4 investigates the cross-layer analysis in VLC systems. Considering fixed-rate transmission and an ON-OFF source model, we formulate the maximums average arrival rate that can be supported.
- Chapter 5 examines the cross-layer analysis in hybrid RF/VLC systems. We essentially focus on deriving different selection strategies, through which the transmitter sends data over the link that sustains the desired QoS guarantees the most.
- Chapter 6 contains the conclusions and an outlook on future work.
- We relegate the proofs to the Appendix part.

We should remark, that this thesis provides mainly a theoretical study to understand the performance of different wireless settings when statistical delay bounds are regarded. Therefore, the settings used in the simulation evaluations are not necessarily reflecting real systems, but to clearly show the main theoretical results. Such investigations can provide a good mathematical toolbox for system designers.

2

EFFECTIVE CAPACITY IN MIMO CHANNELS WITH ARBITRARY INPUTS

Recently, communication systems that are both spectrum and energy efficient have attracted significant attention. Different from the existing research, in this chapter we investigate the throughput and energy efficiency of a general class of multiple-input multiple-output (MIMO) systems with arbitrary inputs when they are subject to statistical quality of service (QoS) constraints, which are imposed as limits on the delay violation and buffer overflow probabilities. We provide a cross-layer study regarding the physical and data-link layers by employing the effective capacity as the main performance metric, which is the maximum constant data arrival rate at a buffer that can be sustained by the channel service process under specified QoS constraints.

We obtain the optimal input covariance matrix that maximizes the effective capacity under a short-term average power constraint. Following that, we perform an asymptotic analysis of the effective capacity in the low signal-to-noise ratio and large-scale antenna (massive MIMO) regimes. Such analysis has a practical importance for 5G scenarios that necessitate low latency, low power consumption, and/or ability to simultaneously support massive number of users. In addition, we put forward the non-asymptotic backlog and delay violation bounds by utilizing the effective capacity.

In the low signal-to-noise ratio regime analysis, in order to determine the minimum energy-per-bit and also the slope of the effective capacity versus energy-per-bit curve at the minimum energy-per-bit, we utilize the first and second derivatives of the effective capacity when the signal-to-noise ratio approaches zero. In the massive MIMO analysis, we benefit from the so-called self-averaging property to show that the effective capacity approaches the average transmission rate in the channel with the increasing number of transmit and/or receive antennas.

2.1 INTRODUCTION

Following the research of Foschini [29] and Telatar [9], multiple-input multiple-output (MIMO) transmission systems have been widely studied, and it was shown that employing multiple antennas at a transmitter and/or a receiver can remarkably enhance the system performance in terms of both reliability and spectral efficiency [30]. Herein, the information-theoretic analysis of MIMO systems formed the basis to understand the system dynamics [87–97]. For instance, the ergodic capacity of MIMO systems was explored, and analytical characterizations of spatial fading correlations and their effect on the ergodic capacity were provided in [88]. Moreover, regarding the available information about the channel statistics at the transmitter,

the optimal input covariance matrix that achieves the maximum ergodic capacity in a one-to-one MIMO system was investigated in [89]. Considering line-of-sight characterizations in a wireless medium, the structures of the capacity-achieving input covariance matrices were researched as well [95–97].

The efficient use of energy is a fundamental requirement in communication networks because most of the portable communication devices are battery-driven and environmental concerns are to be carefully mediated. Thus, energy efficiency along with spectral efficiency is in the focus of attention in prospective transmission system designs. For example, the next generation wireless communication technology, commonly known as 5G, targets to support 10 to 100 times higher data transmission rate and to provide 10 times longer battery life than the current mobile technology [26]. In this regard, the ergodic capacity of MIMO systems were primarily studied in low-power regimes [98–101]. These studies revealed that when the objective capacity function is concave, the minimum energy required to transmit one bit of information, i.e., energy-per-bit, is obtained when the signal-to-noise ratio approaches zero [98]. Subsequently, a more comprehensive energy efficiency analysis was conducted considering any power regime [32]. Particularly, MIMO scenarios with Rayleigh fading channel models were investigated, and a fairly accurate closed-form approximation for the energy-per-bit was obtained by engaging different power models. Similar investigations were conducted in distributed MIMO systems as well [102].

Another approach that maximizes the spectral efficiency while minimizing the energy-per-bit is to increase the spatial dimension by increasing the number of transmit and/or receive antennas. It was shown that the spectral efficiency improves substantially with the increasing number of antennas while making the transmit power arbitrarily small [33]. On this account, massive MIMO (or large-scale antenna [103]) systems have evolved as a candidate technology for 5G wireless communications [27, 34], and they have been investigated from information-theoretic perspectives [37–43]. Particularly, energy and spectral efficiency in the uplink channels of multi-user massive MIMO systems were studied with different information processing techniques such as maximum-ratio combining, zero forcing, and minimum mean-square error (MMSE) estimation [40]. Likewise, power allocation policies were also studied and optimal input covariance matrices in multi-access channels with massive number of antennas at both transmitters and receivers, which maximize the sum transmission rate, were derived [43].

Quality of service (QoS) constraints, which generally emerge in the form of delay and/or data buffering limitations, are generally disregarded when the ergodic capacity is set as the only performance metric. However, the increasing demand for delay-sensitive services, such as video streaming and online gaming over wireless networks, has brought up the need for a comprehensive investigation of delay-sensitive scenarios [25]. For wireless communications systems with such delay-sensitive services, the ergodic

capacity solely is not a sufficient metric. On the contrary, QoS constraints in the data-link layer that are attributed to delay violation and buffer overflow probabilities should be invoked as performance measures as well. Relying on this motivation, cross-layer design goals were acquired as new research grounds.

Recall that the effective capacity was proposed in [81] as a cross-layer performance measure for time-varying service processes. Particularly, the effective capacity provides the maximum constant data arrival rate at a transmission node that can be sustained by a given stochastic service process under statistical QoS constraints imposed as limits on the buffer overflow and delay violation probabilities. The concept of the effective capacity has gained a notable attention, and it has been investigated in several transmission scenarios, including MIMO systems [68, 104, 105]. Specifically, point-to-point MIMO scenarios were explored under QoS constraints by employing the effective capacity as the performance metric in the low and high signal-to-noise ratio regimes and the wide-band regime [68]. A comparable analysis was extended to cognitive MIMO systems, where the effects of channel uncertainty on the effective capacity performance of secondary users following channel sensing errors are studied [104]. Regarding the antenna beam-forming, optimal transmit strategies that maximize the effective capacity were derived in MIMO systems with doubly correlated channels and a covariance feedback [105].

Because Gaussian input signaling in certain cases is optimal in the sense of maximizing the mutual information between the input and output in a transmission channel, it has been invoked in many research scenarios. Even though Gaussian input signaling is not practical, it is preferred by many researchers since it typically simplifies the analytical presentations. On the other hand, it is of importance to understand the effects of signaling choice on the the system performance, because the type of input signaling may critically affect the tradeoff between the data arrival process to a node and the data service process from that node [106]. A general look at wireless systems employing finite and discrete input signaling methods can be found in [107–114]. However, QoS constraints are generally not included in these studies. Particularly, the optimal precoding matrix in a point-to-point MIMO system, which maximizes the mutual information in the low and high signal-to-noise ratio regimes, was proposed [107].

With the same objective, channel diagonalization was applied in order to obtain the optimal channel precoder [109, 111], i.e., parallel and non-interfering Gaussian channels are formed to reach the optimal input covariance matrix. In another study [8], the optimal power allocation policy that maximizes the mutual information, named as *mercury/water-filling*, was shown to be a generalization to the well-known *water-filling* algorithm. Multi-access systems were studied as well [113], where linear precoding matrices are obtained in order to maximize the weighted sum rate. An extension of the same analysis was performed in scenarios in which transmitters have only statistical information about the wireless channels [114]. Asymptotic analyses in the large-scale antenna regimes were also provided.

Here, the notion of mutual information was utilized as the performance metric, and the rudimentary relation between the mutual information and the MMSE, which was introduced in [115, 116], was exploited.

In this chapter, we focus on a more general MIMO scenario with input signaling is arbitrary and statistical QoS constraints. We investigate the system performance from cross-layer perspectives by exploiting the effective capacity. The contributions of this chapter have theoretical as well as practical implications. In this sense, we provide a mathematical toolbox[1] that system designers can use as guidelines in order to understand performance levels of spectrum and energy efficient systems under QoS constraints imposed as limits on the buffer overflow and delay violation probabilities, which are two of the main objectives in the 5G technology [26]. Furthermore, we can apply the analysis provided in this chapter in different practical scenarios that necessitate low latency, low power consumption or ability to simultaneously support massive number of users. Here, we refer to the vehicular-based communication scenarios defined by the well-known European project 'Mobile and wireless communications Enablers for the Twenty-twenty Information Society (METIS)' [26, 119]. For instance, we can perform our analysis in scenarios such as 'Best experience follows you' [26] and 'Traffic Jam' and 'Traffic Efficiency and Safety' [119].

The rest of this chapter is organized as follows: We describe the MIMO system in Section 2.2. Then, we discuss the instantaneous mutual information between the channel input and output, and then introduce the effective rate and capacity expressions in Section 2.3. We provide the optimal input covariance matrix. We perform asymptotic analyses in the low signal-to-noise ratio regime in Section 2.4.1 and in the large-scale antenna regime in Section 2.4.2. We investigate non-asymptotic backlog and delay bounds in Section 2.5. We present the numerical results in Section 2.6 and we relegate the proofs to the Appendix. Table 2.1 summarizes the symbols used in this chapter.

2.2 CHANNEL MODEL

As shown in Figure 2.1, we consider a point-to-point MIMO transmission system in which one transmitter and one receiver are equipped with M and N antennas, respectively. The data generated by a source (or sources) initially arrives at the transmitter buffer with rate *a(t) bits/channel use*[2] for $t \in \{1, 2, \cdots\}$ and is stored in the buffer. Following the encoding and modulation processes, the transmitter sends the data to the receiver over the wireless channel packet by packet in frames (blocks) of T *channel uses*. During the transmission of the data, the input-output relation in the flat-fading channel at time instant t is expressed as follows:

$$\mathbf{y}_t = \sqrt{P}\mathbf{H}_t\mathbf{x}_t + \mathbf{w}_t, \tag{2.1}$$

1 We refer interested readers to [44, 45, 117, 118] and references therein for practical massive MIMO settings.

2 Each *channel use* duration can be considered equal to the sampling duration of one symbol, i.e., bits/sec/Hz.

Table 2.1: Table of symbols in this chapter

Symbol	Description
d	Delay threshold
$\mathbb{E}\{\cdot\}$	Expected value
ε'	Delay violation probability
γ	Signal-to-noise ratio
Γ_t	Matrix with independent and identically distributed complex elements
$\widehat{\Gamma}_t$	Channel estimate
$\widetilde{\Gamma}_t$	Channel estimation error
$\mathbf{H}_t$	$N \times M$ random channel matrix
$\mathbf{I}_{N\times N}$	$N \times N$ identity matrix
$\mathcal{I}(\mathbf{x}_t;\mathbf{y}_t)$	Mutual information between $\mathbf{x}_t$ and $\mathbf{y}_t$
$\mathbf{K}_t$	Input covariance matrix
M	Number of transmit antennas
N	Number of receive antennas
P	Transmit average power
$\mathbf{R}_v$ ($\mathbf{R}_r$)	Transmit (Receive) correlation matrix
$\mathcal{S}_0$	Slope of the effective capacity versus the energy-per-bit at ζ_{min}
σ_e^2	Estimation error variance
σ_h^2	Channel variance
σ_w^2	Noise variance
T	Frame duration
θ	QoS exponent
$\mathbf{tr}\{\cdot\}$	Trace operator
$\widetilde{\mathbf{w}}_t$	Noise plus channel estimation error, i.e., $= \sqrt{P}\widetilde{\mathbf{H}}_t\mathbf{x}_t + \mathbf{w}_t$
$\mathbf{w}_t$	N-dimensional additive noise vector at time instance t
$\mathbf{x}_t$	M-dimensional input vector at time instance t
$\mathbf{y}_t$	N-dimensional output vector at time instance t
ζ	Energy-per-bit
ζ_{min}	Minimum energy-per-bit
$\{\cdot\}^\dagger$	Transpose operator

where $\mathbf{x}_t$ and $\mathbf{y}_t$ are the M-dimensional input and N-dimensional output vectors, respectively, and $\mathbf{w}_t$ represents the N-dimensional additive noise vector with independent and identically distributed elements. Each element of $\mathbf{w}_t$ is circularly symmetric, complex Gaussian distributed with zero-mean and variance σ_w^2. Hence, we have $\mathbb{E}\{\mathbf{w}_t\mathbf{w}_t^\dagger\} = \sigma_w^2\mathbf{I}_{N\times N}$, where $\mathbb{E}\{\cdot\}$ denotes

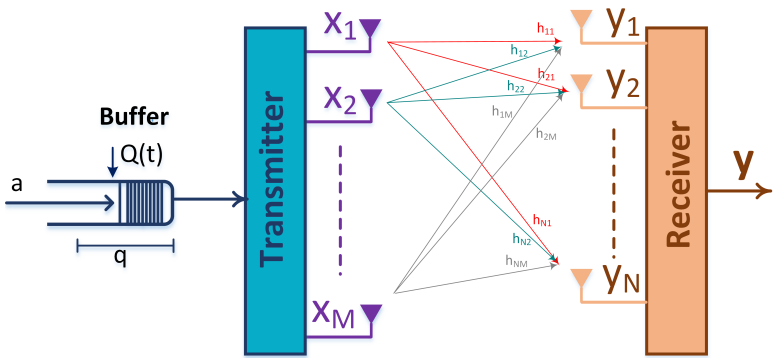


Figure 2.1: MIMO Channel model.

the expected value, $\{\cdot\}^\dagger$ is the transpose operator and $\mathbf{I}_{N\times N}$ is the $N \times N$ identity matrix. Furthermore, $\mathbf{H}_t = \{h_{nm}(t)\}$ is the $N \times M$ random channel matrix, where $h_{nm}(t)$ is the channel fading coefficient with an arbitrary distribution between the m^{th} transmit antenna and the n^{th} receive antenna.

We further assume that the channel matrix remains constant during one transmission frame (T *channel uses*) and changes independently from one frame to another. We also consider a short-term power constraint, i.e., P indicates the power allocated for the transmission of the data in one *channel use*. Then, we have $\mathbf{tr}\{\mathbb{E}\{\mathbf{x}_t\mathbf{x}_t^\dagger\}\} = \mathbf{tr}\{\mathbf{K}_t\} \leqslant 1$, where $\mathbf{tr}\{\cdot\}$ is the trace operator and $\mathbf{K}_t$ is a positive semi-definite Hermitian matrix.

We assume that the instantaneous channel realizations are available at both the transmitter and the receiver, and that the channel fading coefficients are correlated with each other. We invoke the Kronecker product model, which is widely used in modeling real channels because of its analytical tractability with a reasonable accuracy [120, Ch. 2], [121]. Hence, the channel matrix is expressed as

$$\mathbf{H}_t = \mathbf{R}_r^{\frac{1}{2}}\Gamma_t\mathbf{R}_v^{\frac{1}{2}}, \tag{2.2}$$

where Γ_t is an $N \times M$ matrix with independent and identically distributed complex elements. $\mathbf{R}_v$ and $\mathbf{R}_r$ are the transmit and receive correlation matrices, respectively, which are usually modeled with an exponential correlation structure [122, 123]. The transmit and receive correlation matrices depend on the array spacing at the transmitter and the receiver, and the characteristic distances proportional to the spatial coherence distances at the transmitter and the receiver, respectively. Particularly, the elements of $\mathbf{R}_v$ and $\mathbf{R}_r$ are expressed as

$$\{\mathbf{R}_v\}_{kl} = e^{\frac{d_v}{\Delta_v}|k-l|} \quad \text{for} \quad k, l \in \{1, \cdots, M\}$$

and

$$\{\mathbf{R}_r\}_{kl} = e^{\frac{d_r}{\Delta_r}|k-l|} \quad \text{for} \quad k, l \in \{1, \cdots, N\},$$

respectively, where d_v and d_r are the corresponding antenna spacings, and Δ_v and Δ_r are the corresponding characteristic distances. Therefore, the correlation matrix at one end can be locally estimated without any feedback from the other end. On the other hand, Γ_t is estimated by the receiver, and

then forwarded to the transmitter at the beginning of each transmission frame. Similar to the strategy in [124–126], we assume that the feedback channel is delay-free and error-free. Because we have a block-fading channel, the channel information is valid until the end of the transmission frame. Even if we consider a feedback delay, it will only reduce the time allocated for data transmission. In particular, when the channel feedback arrives after a certain portion of the time frame (T channel uses), i.e., αT for $0 < \alpha < 1$, the remaining $(1-\alpha)T$ will be the time duration for data transmission. Moreover, the reliable feedback can be sustained with strong channel codes.

We further know that in practical settings the channel estimation is obtained imperfectly. Therefore, we have

$$\mathbf{H}_t = \mathbf{R}_r^{\frac{1}{2}}(\widehat{\Gamma}_t + \widetilde{\Gamma}_t)\mathbf{R}_v^{\frac{1}{2}} = \widehat{\mathbf{H}}_t + \widetilde{\mathbf{H}}_t, \tag{2.3}$$

where $\widehat{\Gamma}_t$ is the channel estimate and $\widetilde{\Gamma}_t$ is the channel estimation error. Given that the receiver employs MMSE estimator in order to obtain the channel knowledge, we have $\widehat{\Gamma}_t$ and $\widetilde{\Gamma}_t$ uncorrelated with each other. Similar to [126], we further assume that $\widetilde{\Gamma}_t$ is a zero-mean process with a known variance at both the transmitter and the receiver. Above,

$$\widehat{\mathbf{H}}_t = \mathbf{R}_r^{\frac{1}{2}}\widehat{\Gamma}_t\mathbf{R}_v^{\frac{1}{2}} \quad \text{and} \quad \widetilde{\mathbf{H}}_t = \mathbf{R}_r^{\frac{1}{2}}\widetilde{\Gamma}_t\mathbf{R}_v^{\frac{1}{2}}$$

Hence, the input-output relation in (2.1) becomes

$$\mathbf{y}_t = \sqrt{P}\widehat{\mathbf{H}}_t\mathbf{x}_t + \sqrt{P}\widetilde{\mathbf{H}}_t\mathbf{x}_t + \mathbf{w}_t = \sqrt{P}\widehat{\mathbf{H}}_t\mathbf{x}_t + \widetilde{\mathbf{w}}_t. \tag{2.4}$$

2.3 EFFECTIVE CAPACITY

Recall that it is not very easy to sustain a stable transmission rate in wireless channels due to the time-varying nature of such links. In particular, reliable transmission may not be provided all the time. Therefore, depending on the type of data transmission, delay violation and buffer overflow concerns become critical at the transmitter. Respectively, given a statistical transmission (service) process, how to determine the maximum data arrival rate at the transmitter buffer so that the QoS requirements in the form of limits on delay violation and buffer overflow probabilities can be satisfied is one of the main research questions. In this regard, the effective capacity can be employed as a performance metric. Specifically, the effective capacity identifies the maximum constant data arrival rate at the transmitter buffer that the time-varying transmission process can support under desired QoS constraints [81].

In Fig. 2.1, $Q(t)$ is the number of bits in the data buffer at time instant t and q is the buffer threshold. As detailed in Section 1.3, in this thesis we impose certain QoS constraints as limits on the the buffer overflow probability in the steady-state, i.e., $\Pr\{Q(t \to \infty) \geqslant q\}$, as expressed in (1.2), or equivalently in (1.3) for large threshold q. In particular, the constraint in (1.3) implies that the buffer overflow probability should decay exponentially with a rate controlled by the QoS exponent $\theta > 0$. Recall that, for a given

θ a similar exponentially decaying approximation for the delay violation probability can be obtained, as expressed in (1.4). All in all, the effective capacity defines the maximum constant arrival rate that can be attained by the time-varying channel under the buffer overflow constraint expressed in (1.3) or the delay constraint expressed in (1.4). Noting that the average arrival rate is equal to the average departure rate in the steady-state [85], then the effective capacity can also be explained as the maximum throughput under such constraints.

Mathematically, for a given discrete-time, ergodic and stationary stochastic service process, $r(t)$, the effective capacity as a function of the decay rate parameter, θ, is given by [81, Eq. (11)]

$$C_E(\theta) = -\lim_{\tau\to\infty} \frac{1}{\theta\tau T} \log_e \mathbb{E}\{e^{-\theta \sum_{t=1}^{\tau T} r(t)}\},$$

where $r(t)$ is the service rate in the wireless channel at time instant t, $\sum_{t=1}^{\tau T} r(t)$ is the time-accumulated service process, i.e., the total number of bits served from the transmitter in τT *channel uses*, and $\tau \in \{1, 2, \cdots, \}$ is the time frame index. Recall that the encoding and modulation of data and its transmission are performed in frames of T *channel uses*.

Given the channel estimate, $\widehat{\mathbf{H}}_t$, the service rate in one frame can be set to the mutual information between $\mathbf{x}_t$ and $\mathbf{y}_t$, i.e., $r(t) = \mathcal{I}(\mathbf{x}_t; \mathbf{y}_t|\widehat{\mathbf{H}}_t)$. However, considering the input-output relation (2.4), it is difficult to evaluate the mutual information in closed-form. Therefore, the service rate in the channel is set to a lower bound on the mutual information by considering the worst-case noise and modeling the estimation error as an additional Gaussian noise vector with zero-mean, independent and identically distributed samples [127, 128], i.e.,

$$r(t) = \mathcal{I}_L(\mathbf{x}_t; \mathbf{y}_t|\widehat{\mathbf{H}}_t) \leqslant \mathcal{I}(\mathbf{x}_t; \mathbf{y}_t|\widehat{\mathbf{H}}_t) \quad \text{and} \quad \mathbb{E}\{\widetilde{\mathbf{w}}_t \widetilde{\mathbf{w}}_t^\dagger\} = \sigma_{\widetilde{w}}^2 \mathbf{I}_{N\times N},$$

where

$$\sigma_{\widetilde{w}}^2 = \sigma_w^2 + \frac{P}{NM} \mathbf{tr}\left\{\mathbb{E}\left\{\widetilde{\mathbf{H}}_t \mathbf{x}_t \mathbf{x}_t^\dagger \widetilde{\mathbf{H}}_t^\dagger\right\}\right\}.$$

Since the service rate in the channel is smaller than or equal to the mutual information, the reliable transmission is guaranteed. Hence, the service rate is expressed as

$$r(t) = \mathcal{I}_L(\mathbf{x}_t; \mathbf{y}_t|\widehat{\mathbf{H}}_t) = \mathbb{E}_{\mathbf{x}_t,\mathbf{y}_t}\left\{\log_2 \frac{f_{\mathbf{y}_t|\mathbf{x}_t}(\mathbf{y}_t|\mathbf{x}_t)}{f_{\mathbf{y}_t}(\mathbf{y}_t)}\right\}, \tag{2.5}$$

where

$$f_{\mathbf{y}_t}(\mathbf{y}_t) = \sum_{\mathbf{x}_t} p(\mathbf{x}_t) f_{\mathbf{y}_t|\mathbf{x}_t}(\mathbf{y}_t|\mathbf{x}_t)$$

is the probability density function of $\mathbf{y}_t$ and

$$f_{\mathbf{y}_t|\mathbf{x}_t}(\mathbf{y}_t|\mathbf{x}_t) = (\pi\sigma_{\widetilde{w}}^2)^{-N} e^{-\frac{1}{\sigma_{\widetilde{w}}^2}||\mathbf{y}_t - \sqrt{P}\,\widehat{\mathbf{H}}_t \mathbf{x}_t||^2}$$

is the conditional probability density function of $\mathbf{y}_t$ given $\mathbf{x}_t$. For notational convenience in this chapter, we use $\mathcal{I}(\mathbf{x}_t; \mathbf{y}_t)$ to refer to the lower bound, $\mathcal{I}_L(\mathbf{x}_t; \mathbf{y}_t|\widehat{\mathbf{H}}_t)$.

Because the channel matrix stays constant during one transmission frame and changes independently from one frame to another, and that the encoding and modulation of the data packets are performed in T *channel uses*, we can express the normalized effective rate in *bits/channel use/receive dimension* as

$$R_E(\theta) = -\frac{1}{\theta NT} \log_e \mathbb{E}_{\widehat{\mathbf{H}}_t} \left\{ e^{-\theta T \mathcal{I}(\mathbf{x}_t;\mathbf{y}_t)} \right\}. \tag{2.6}$$

Above, while the receiver has the instantaneous channel estimate, the transmitter has no information regarding the channel matrix. If the transmitter is aware of the channel statistics but not the actual value of $\widehat{\mathbf{H}}_t$, then the transmitter sets the input covariance matrix to a value, i.e., $\mathbf{K}_t = \mathbf{K}$, in order to maximize the effective rate in (2.6) by considering the QoS constraints and the channel statistics, i.e.,

$$R_E(\theta) = \max_{\substack{\mathbf{K} \succeq 0 \\ \mathbf{tr}\{\mathbf{K}\} \leqslant 1}} -\frac{1}{\theta NT} \log_e \mathbb{E}_{\widehat{\mathbf{H}}_t} \left\{ e^{-\theta T \mathcal{I}(\mathbf{x}_t;\mathbf{y}_t)} \right\} \tag{2.7}$$

in *bits/channel use/receive dimension*. In (2.7), the covariance matrix, $\mathbf{K}$, depends on the statistics of $\widehat{\mathbf{H}}_t$ and the worst-case noise, and is independent of its actual realization. On the other hand, if the instantaneous knowledge of $\widehat{\mathbf{H}}_t$ is available at the transmitter and the receiver, the transmitter can adaptively set the input covariance matrix by considering both the QoS constraints and the instantaneous realization of the channel matrix[3]. Hence, the maximum effective rate, which we call as the effective capacity, in *bits/channel use/receive dimension* is given as follows:

$$C_E(\theta) = \max_{\substack{\mathbf{K}_t \succeq 0 \\ \mathbf{tr}\{\mathbf{K}_t\} \leqslant 1}} -\frac{1}{\theta NT} \log_e \mathbb{E}_{\widehat{\mathbf{H}}_t} \left\{ e^{-\theta T \mathcal{I}(\mathbf{x}_t;\mathbf{y}_t)} \right\}. \tag{2.8}$$

Above, $\mathbf{K}_t$ is time-varying unlike $\mathbf{K}$ in (2.7), because it is a function of $\widehat{\mathbf{H}}_t$.

Here, a key research problem is the optimal selection of the power allocation policy (or input covariance matrix) given the channel matrix and the QoS requirements. In particular, the central question is the following: What is the instantaneous input covariance matrix, $\mathbf{K}_t$, that solves (2.8) given that the channel matrix, $\widehat{\mathbf{H}}_t$, is available at the transmitter and the receiver, and that there are certain QoS constraints? In the following theorem, we identify the optimal policy that the transmitter should employ to obtain (2.8).

Theorem 1 *The input covariance matrix, $\mathbf{K}_t \succeq 0$, that maximizes the effective capacity given in (2.8) is the solution of the following equality:*

$$\mathbf{K}_t = \frac{\theta T \gamma e^{-\theta T \mathcal{I}(\mathbf{x}_t;\mathbf{y}_t)}}{\lambda} \widehat{\mathbf{H}}_t^{\dagger} \widehat{\mathbf{H}}_t \mathbf{mmse}_t, \tag{2.9}$$

3 In case there is a delay in the feedback channel, and the delay is smaller than the block duration (T *channel uses*), the effective capacity can be reformulated as $C_E(\theta) = -\frac{1}{\theta NT} \log_e \mathbb{E}_{\widehat{\mathbf{H}}_t} \left\{ e^{-\theta T(1-\alpha)\mathcal{I}(\mathbf{x}_t;\mathbf{y}_t)} \right\}$, where αT is the delay and $0 < \alpha < 1$.

where $\gamma = \frac{P}{\sigma_{\widetilde{w}}^2}$ *is the average signal-to-noise ratio at the receiver,* λ *is the Lagrange multiplier of the constraint* $\mathbf{tr}\{\mathbf{K}_t\} \leqslant 1$*, and*

$$\mathbf{mmse}_t = \mathbb{E}\left\{\left(\mathbb{E}\{\mathbf{x}_t|\mathbf{y}_t\} - \mathbf{x}_t\right)\left(\mathbb{E}\{\mathbf{x}_t|\mathbf{y}_t\} - \mathbf{x}_t\right)^\dagger\right\}$$

is the MMSE matrix.

Proof: See Appendix A. □

In (2.9), both the mutual information and $\mathbf{mmse}_t$ are functions of the input covariance matrix, $\mathbf{K}_t$, and (2.9) is non-concave over the space spanned by $\mathbf{K}_t$ [108, 110, 111]. Therefore, the solution obtained from (2.9) is not necessarily unique. On the other hand, we follow a different strategy and start with the singular value decomposition of the channel matrix, i.e.,

$$\widehat{\mathbf{H}}_t = \mathbf{U}_t \mathbf{D}_t \mathbf{V}_t^\dagger,$$

where $\mathbf{U}_t$ and $\mathbf{V}_t$ are $N \times N$ and $M \times M$ unitary matrices, respectively, and $\mathbf{D}_t$ is an $N \times M$ matrix with non-negative real numbers on the diagonal, which are the square roots of the non-zero eigenvalues of $\widehat{\mathbf{H}}_t\widehat{\mathbf{H}}_t^\dagger$ and $\widehat{\mathbf{H}}_t^\dagger\widehat{\mathbf{H}}_t$. Then, we re-express the input-output model in (2.4) as follows:

$$\widetilde{\mathbf{y}}_t = \sqrt{P}\mathbf{D}_t\widetilde{\mathbf{x}}_t + \widetilde{\mathbf{n}}_t, \tag{2.10}$$

where $\widetilde{\mathbf{y}}_t = \mathbf{U}_t^\dagger \mathbf{y}_t$ and $\widetilde{\mathbf{x}}_t = \mathbf{V}_t^\dagger \mathbf{x}_t$. The new noise vector is denoted by $\widetilde{\mathbf{n}}_t = \mathbf{U}_t^\dagger \widetilde{\mathbf{w}}_t$, which is a zero-mean, Gaussian, complex vector with independent and identically distributed elements [9]. We further know that $\mathcal{I}(\mathbf{x}_t; \mathbf{y}_t) = \mathcal{I}(\widetilde{\mathbf{x}}_t; \widetilde{\mathbf{y}}_t)$, because the information regarding $\widehat{\mathbf{H}}_t$ is available at both the transmitter and the receiver. Now, let $\widetilde{\mathbf{K}}_t$ be the covariance matrix of $\widetilde{\mathbf{x}}_t$, i.e.,

$$\widetilde{\mathbf{K}}_t = \mathbb{E}\{\widetilde{\mathbf{x}}_t\widetilde{\mathbf{x}}_t^\dagger\} = \mathbb{E}\{\mathbf{V}_t^\dagger \mathbf{x}_t \mathbf{x}_t^\dagger \mathbf{V}_t\} = \mathbf{V}_t^\dagger \mathbf{K}_t \mathbf{V}_t.$$

In particular, if we can find the optimal $\widetilde{\mathbf{K}}_t$, we can also determine the optimal input covariance matrix, $\mathbf{K}_t$. Therefore, we provide the optimal input covariance matrix in the following theorem and show that this is the global solution in its proof.

Theorem 2 *The input covariance matrix,* $\mathbf{K}_t \succeq 0$*, that provides (2.8) is*

$$\mathbf{K}_t = \mathbf{V}_t \mathbf{\Sigma}_t \mathbf{V}_t^\dagger, \tag{2.11}$$

where $\mathbf{V}_t$ *is the* $M \times M$ *unitary matrix, columns of which are the left-singular vectors of* $\widehat{\mathbf{H}}_t$*.* $\widetilde{\mathbf{K}}_t = \mathbf{\Sigma}_t = diag\{\sigma_t(1), \cdots, \sigma_t(M)\}$ *is an* $M \times M$ *diagonal matrix that satisfies*

$$\sigma_t(i) = \frac{\theta T \gamma d_t(i)}{\lambda} e^{-\theta T \mathcal{I}(\widetilde{\mathbf{x}}_t; \widetilde{\mathbf{y}}_t)} \mathbf{mmse}_t(i), \text{ if } \sigma_t(i) \geqslant 0,$$
$$\sigma_t(i) = 0, \text{ otherwise,}$$
$$\sigma_t(i) = 0, \text{ for } \min\{M, N\} < i \leqslant M,$$

given that λ *is the Lagrange multiplier associated with the constraint* $\sum_{i=1}^{M} \sigma_t(i) \leqslant 1$, *and*

$$\mathbf{mmse}_t(i) = \mathbb{E}\left\{|\mathbb{E}\{\widetilde{x}_t(i)|\widetilde{y}_t(i)\} - \widetilde{x}_t(i)|^2\right\}$$

is the MMSE function. Furthermore, $d_t(i)$ *is the* i^{th} *eigenvalue of* $\widehat{\mathbf{H}}_t\widehat{\mathbf{H}}_t^{\dagger}$ *and* $\widehat{\mathbf{H}}_t^{\dagger}\widehat{\mathbf{H}}_t$.

Proof: See Appendix B. □

Remark 1 *The input covariance matrix,* $\mathbf{K}_t$, *is set according to the channel estimate. However, the constraint* $tr\{\mathbf{K}_t\} \leqslant 1$ *(or* $\sum_{i=1}^{M} \sigma_i \leqslant 1$ *in Theorem 2) is independent of the channel estimate. Therefore, the worst-case noise variance,* $\sigma^2_{\widetilde{w}}$, *and hence the signal-to-noise ratio,* γ, *do not depend on the actual channel estimate.*

2.4 EFFECTIVE CAPACITY IN ASYMPTOTIC REGIMES

Having obtained the effective capacity and rate expressions, and having characterized the optimal input covariance matrices that maximize the effective capacity performance, we note that due to the complexity in the analytical formulations, it becomes difficult to gain insight on the system performance in general scenarios. On the other hand, asymptotic approaches can help us set the design criteria in certain asymptotic regimes. Therefore, we investigate the effective capacity of MIMO systems in the low signal-to-noise ratio and large-scale antenna regimes. We also note that we drop the time index in the sequel unless otherwise it becomes necessary.

2.4.1 *Effective Capacity in Low Signal-to-Noise Ratio Regime*

In this section, we explore the effective capacity performance of the aforementioned MIMO system with an arbitrary input distribution in the low signal-to-noise ratio regime. In this direction, we determine the minimum energy-per-bit and the slope of the effective capacity versus the energy-per-bit at the minimum energy-per-bit, which are denoted by $\zeta_{\min}$ and $\mathcal{S}_0$, respectively. The benefit of the low signal-to-noise ratio analysis is that many battery-driven applications require operations at low energy costs and energy efficiency generally increases with decreasing transmission power when the transmission throughput is a concave[4] function of the transmission power. For this purpose, we start the low signal-to-noise ratio analysis with the following second-order expansion[5] of the effective capacity with respect to the transmission power, P, at $P = 0$:

$$C_E(\theta, P) = \dot{C}_E(\theta, 0)P + \ddot{C}_E(\theta, 0)\frac{P^2}{2} + o(P^2), \tag{2.12}$$

4 It is known that the minimum energy-per-bit is obtained as the signal-to-noise ratio goes to zero [98]. In our model, the signal-to-noise ratio, $\gamma = \frac{P}{\sigma^2_{\widetilde{w}}}$, goes to zero with the transmission power going to zero.

5 We utilize the Taylor series representation of the effective capacity with respect to P at $P = 0$.

where $\dot{C}_E(\theta,0)$ and $\ddot{C}_E(\theta,0)$ are, respectively, the first and second derivatives of the effective capacity with respect to P at $P=0$. Note that we express the effective capacity as a function of θ and P.

Now, let $\zeta = \frac{P}{C_E(\theta,P)}$ denote the energy-per-bit required for given θ and P. Following [129, Proposition 1], we can show that the effective capacity is concave in the space spanned by P. Notice that it is sufficient to prove the concavity of the lower bound on the mutual information over the space spanned by the transmission power P, because the signal-to-noise ratio is an increasing function of the transmission power. The concavity of the same lower bound on the mutual information is also shown in [124, Eq. 16] when the channel input is Gaussian distributed. Thus, we can obtain the minimum energy-per-bit when the transmission power goes to zero, i.e., $P \to 0$, as follows:

$$\zeta_{\min} = \lim_{P\to 0} \frac{P}{C_E(\theta,P)} = \frac{1}{\dot{C}_E(\theta,0)}. \tag{2.13}$$

Moreover, considering the result in [98, Eq. (29)], we can show the slope of the effective capacity versus ζ (in dB) curve at $\zeta_{\min}$ as

$$S_0 = \lim_{\zeta\downarrow\zeta_{\min}} \frac{C_E(\zeta)}{10\log_{10}\zeta - 10\log_{10}\zeta_{\min}} 10\log_{10} 2, \tag{2.14}$$

where $C_E(\zeta)$ is the effective capacity as a function of the energy-per-bit, ζ, and $\zeta_{\min}$ is the minimum energy-per-bit and obtained when the transmission power goes to zero, i.e., $P \to 0$. Above, $\zeta \downarrow \zeta_{\min}$ indicates the limit when the value of ζ is reduced and approaches $\zeta_{\min}$. Using the first and second derivatives [98, Th. 9], we can express the slope in *bits/channel use/(3 dB)/receive antenna* as

$$S_0 = \frac{2\left[\dot{C}_E(\theta,0)\right]^2}{-\ddot{C}_E(\theta,0)} \log_e 2. \tag{2.15}$$

Accordingly, having $\zeta_{\min}$ and S_0, we can form a linear approximation of $C_E(\zeta)$ in the low signal-to-noise ratio regime.

In order to better understand the effective capacity performance in the low signal-to-noise ratio regime, we provide the following theorem.

Theorem 3 *The first derivative of the effective capacity in (2.8) with respect to* P *at* $P=0$ *is given as*

$$\dot{C}_E(\theta,0) = \frac{1}{N\log_e 2}\mathbb{E}_{\widehat{\mathbf{H}}}\{\lambda_{max}(\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}})\}, \tag{2.16}$$

and the second derivative of the effective capacity with respect to P *at* $P=0$ *is given as*

$$\begin{aligned}\ddot{C}_E(\theta,0) =& \frac{\theta T}{N\log_e^2 2}\left[\mathbb{E}^2_{\widehat{\mathbf{H}}}\{\lambda_{max}(\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}})\} - \mathbb{E}_{\widehat{\mathbf{H}}}\{\lambda^2_{max}(\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}})\}\right] \\ &- \frac{\mathbb{E}_{\widehat{\mathbf{H}}}\{\lambda^2_{max}(\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}})\}}{Nl\log_e 2} - \frac{2\sigma_e^2}{N\log_e 2}\mathbb{E}_{\widehat{\mathbf{H}}}\{\lambda_{max}(\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}})\},\end{aligned} \tag{2.17}$$

where $\lambda_{\max}(\widehat{\mathbf{H}}^{\dagger}\widehat{\mathbf{H}})$ *in (2.16) and (2.17) is the maximum eigenvalue of* $\widehat{\mathbf{H}}^{\dagger}\widehat{\mathbf{H}}$ *and* l *in (2.17) is the multiplicity of* $\lambda_{max}(\widehat{\mathbf{H}}^{\dagger}\widehat{\mathbf{H}})$. *Above,*

$$\sigma_e^2 = \frac{P}{NM}\mathbf{tr}\left\{\mathbb{E}\left\{\widetilde{\mathbf{H}}_t\mathbf{x}_t\mathbf{x}_t^{\dagger}\widetilde{\mathbf{H}}_t^{\dagger}\right\}\right\}.$$

Proof: See Appendix C. □

Remark 2 *The first and second derivatives of the effective capacity,* $\dot{C}_E(\theta,0)$ *and* $\ddot{C}_E(\theta,0)$, *respectively, are independent of the input distribution. Particularly, the minimum energy-per-bit,* $\zeta_{\min}$, *and the slope of the effective capacity versus* ζ *(in dB) curve at* ζ_{min}, $\mathcal{S}_0$, *are not functions of* $\mathbf{x}$ *and/or its probability density function. Additionally, our results also confirm the findings in [68], where the effective capacity of MIMO systems are investigated when the input is Gaussian distributed and the channel is perfectly known at both the transmitter and receiver.*

Remark 3 *As also detailed in the proof in Appendix C, the minimum energy-per-bit is achieved by allocating data power in the direction of the eigenspace of the maximum eigenvalue of* $\widehat{\mathbf{H}}^{\dagger}\widehat{\mathbf{H}}$.

Remark 4 *The minimum energy-per-bit,* $\zeta_{\min}$, *does not change with increasing or decreasing QoS constraints or the channel estimation error, while the slope of the effective capacity at* $\zeta_{\min}$, $\mathcal{S}_0$, *is a function of both the exponential decay rate parameter,* θ, *and the estimation error variance,* σ_e^2. *With increasing* σ_e^2, *the slope decreases.*

Remark 5 *The aforementioned minimum energy-per-bit and slope are acquired given the fact that the input vector,* $\mathbf{x}$, *is complex. On the other hand, when the modulation is performed over the real axis of the constellation only, e.g., binary phase-shift keying (BPSK) and* M*-pulse-amplitude-modulation, the minimum energy-per-bit stays the same because the first derivative does not change, but the slope becomes half of the slope achieved with a complex modulation because the second derivative is the double of the second derivative in the case of a complex modulation [8].*

2.4.2 *Effective Capacity in Large-Scale Antenna Regime*

With the increasing number of antennas the transmitters and the receivers are equipped with, there are more communication pathways and increased transmission link reliability. One more advantage of employing many antennas is the energy efficiency, due to the fundamental principle that with a large number of antennas, energy can be focused with extreme sharpness onto small regions in space [34]. Therefore, in this section, we turn our attention to analyzing the system performance in the large-scale antenna regime. Principally, we obtain the effective capacity while the number of transmit or/and receive antennas goes to infinity.

In particular, we are interested in the effective capacity given in (2.8) when both M and N approach, or either M or N approaches, infinity, i.e.,

$$\lim_{M \text{ and/or } N\to\infty} C_E(\theta,P) = C_E^{\infty}(\theta,P). \tag{2.18}$$

The following theorem provides a significant property of $C_E^{\infty}(\theta, P)$, which follows from the increase in the number of antennas at the transmitter and/or the receiver.

Theorem 4 *For the MIMO system described in (2.4), the effective capacity, $C_E^{\infty}(\theta, P)$ defined in (2.18), is independent of the QoS exponent, θ, and approaches the average transmission rate, i.e.,*

$$C_E^{\infty}(\theta, P) = \lim_{M \text{ and/or } N \to \infty} \frac{1}{N} \mathbb{E}_{\widehat{\mathbf{H}}}\{r\} \tag{2.19}$$

where r is the service rate defined in (2.5).

Proof: See Appendix D. □

Remark 6 *Note that $N \times C_E(\theta, \gamma)$ indicates the throughput level the wireless channel can support under given QoS and transmission power constraints, and that $\mathbb{E}\{r\}$ is the average service rate in the wireless channel in one channel use. Since $N \times C_E(\theta, \gamma) \leqslant \mathbb{E}\{r\}$ for any θ, the transmitter cannot accept data to its buffer at a rate more than the effective capacity, $N \times C_E(\theta, \gamma)$, due to the delay violation and buffer overflow constraints even though the average service rate in the channel is higher. Therefore, the link utilization, which is defined to be the ratio of the data flow rate to a link to the link capacity [130, Ch. 5] and [66, Ch. 16], decreases with increasing QoS constraints. We can consider the effective capacity as the maximum data flow rate and the channel throughput as the link capacity. Herein, Theorem 4 states that the maximum link utilization can be achieved under QoS constraints by increasing the number of antennas.*

Remark 7 *As made clear in the proof of Theorem 4, the knowledge of the channel realizations is not necessary at the transmitter side to achieve the transmission rate given in (2.19) when the number of transmit and/or receive antennas becomes larger. Indeed, the statistical information regarding the channel matrix, $\mathbf{H}$, is sufficient.*

Example 1 *Let us assume that the channel is perfectly known and that the channel coefficients $\{h_{nm}(t)\}$ are zero-mean, independent and identically distributed with finite variance σ_h^2, i.e., $\mathbb{E}\{|h_{nm}|^2\} = \sigma_h^2$. When the number of antennas is going to infinity, the minimum energy-per-bit defined in (2.13), ζ_{min}, and the slope of the effective capacity versus ζ curve at ζ_{min} defined in (2.14), S_0, are*

$$\zeta_{min}^{\infty} = \lim_{M \text{ and/or } N \to \infty} \zeta_{min} = \lim_{M \text{ and/or } N \to \infty} \frac{1}{\dot{C}_E(\theta, 0)} \tag{2.20}$$

$$= \lim_{M \text{ and/or } N \to \infty} \frac{N \log_e 2}{\mathbb{E}_{\mathbf{H}}\{\lambda_{max}(\mathbf{H}^{\dagger}\mathbf{H})\}} \tag{2.21}$$

$$= \lim_{M \text{ and/or } N \to \infty} \frac{\min\{M, N\} N \log_e 2}{MN\sigma_h^2} \tag{2.22}$$

$$= \begin{cases} 0, & \text{if } M \to \infty, \\ \frac{\log_e 2}{\rho\sigma_h^2}, & \text{if } M, N \to \infty, \frac{M}{N} = \rho > 1 \\ \frac{\log_e 2}{\sigma_h^2}, & \text{if } \frac{M}{N} \leqslant 1 \end{cases} \tag{2.23}$$

and

$$\begin{aligned}
\mathcal{S}_0^{\infty} &= \lim_{M \text{ and/or } N\to\infty} \mathcal{S}_0 \\
&= \lim_{M \text{ and/or } N\to\infty} \frac{2\left[\dot{C}_E(\theta,0)\right]^2}{-\ddot{C}_E(\theta,0)} \log_e 2 && (2.24)\\
&= \lim_{M \text{ and/or } N\to\infty} \frac{2\log_e 2 \left[\frac{\mathbb{E}_{\mathbf{H}}\{\lambda_{max}(\mathbf{H}^{\dagger}\mathbf{H})\}}{N\log_e 2}\right]^2}{\frac{\mathbb{E}_{\mathbf{H}}\{\lambda^2_{max}(\mathbf{H}^{\dagger}\mathbf{H})\}}{N l \log_e 2}} && (2.25)\\
&= \lim_{M \text{ and/or } N\to\infty} 2\frac{\min\{M,N\}}{N} && (2.26)\\
&= \begin{cases} 0, & \text{if } N\to\infty, \\ 2\rho, & \text{if } M,N\to\infty, \frac{M}{N}=\rho\leqslant 1 \\ 2, & \text{if } \frac{M}{N}>1, \end{cases} && (2.27)
\end{aligned}$$

respectively.

2.5 NON-ASYMPTOTIC PERFORMANCE ANALYSIS

So far, we have investigated the throughput and energy efficiency of the aforementioned MIMO systems in two different asymptotic regimes by employing the effective capacity, which is also an asymptotic measure in time. Nevertheless, non-asymptotic performance bounds regarding the statistical characterizations of buffer overflow and queueing delay are of importance for practical research agendas. Therefore, we benefit from the tools of the stochastic network calculus [131–133], and provide a statistical bound on the buffer overflow and queueing delay probabilities by utilizing the effective capacity.

Recall that the transmission of a packet is performed over a block duration of T *channel uses* and the transmission rate in the channel during one transmission block is constant. Now, let us define $s(i)$ as the total number of bits transmitted (served) in the i^{th} transmission block. Subsequently, considering the normalized effective rate for the input covariance matrix given in (2.6), $R_E(\theta)$, and following the setting in [133, Definition 7.2.1], we define a statistical affine bound for the aforementioned channel model for any decay rate value, θ, as follows:

$$\mathbb{E}\left\{e^{-\theta S(i,j)}\right\} \leqslant e^{-\theta[(j-i)NTR_E(\theta)-\sigma_R(\theta)]}, \tag{2.28}$$

where $S(i,j)=\sum_{l=i+1}^{j} s(l)$, and $\sigma_R(\theta)$ is a slack term that defines an initial transmission delay. Due to $-\theta$, the expression in (2.28) is in fact a lower bound on the expected amount of the transmitted data in the channel. Subsequently, noting Chernoff's lower bound $\Pr\{X\leqslant x\}\leqslant e^{\theta x}\mathbb{E}\{e^{-\theta X}\}$ for $\theta\geqslant 0$, we have the exponentially bounded fluctuation model described in [134] with parameters $R_E(\theta)>0$ and $b\geqslant 0$ as

$$\Pr\{S(i,j) < (j-i)\mathrm{NTR}_E(\theta) - b\} \leqslant \varepsilon(b),$$

where $\varepsilon(b) = e^{\theta\sigma_R(\theta)}e^{-\theta b}$ is a specific exponentially decaying deficit profile of the amount of the transmitted data in the channel. Now, using the union bound, we express the sample path guarantee as follows:

$$\Pr\{\exists i \in [0,j] : S(i,j) < (j-i)\mathrm{NTR}^{\star}_E(\theta) - b\} \leqslant \varepsilon^{'}(b),$$

where

$$\varepsilon^{'}(b) = \frac{e^{\theta\sigma_R(\theta)}}{1-e^{-\theta\delta}}e^{-\theta b} \tag{2.29}$$

and $\mathrm{NTR}^{\star}_E(\theta) = \mathrm{NTR}_E(\theta) - \delta$ with a free parameter $0 < \delta \leqslant \mathrm{NTR}_E(\theta) - Ta$ for a constant data arrival rate at the transmitter buffer, i.e., a *bits/channel use*. For a more detailed derivation, we refer to [133]. We also refer to [135], where capacity-delay-error boundaries are provisioned as performance models for networked sources and systems. Exclusively, the backlog at the transmitter buffer with the constant data arrival rate a, i.e.,

$$Q(j) = \max_{i\in[0,j]}\{(j-i)Ta - S(i,j)\},$$

has a statistical bound

$$q = \max_{i\in[0,j]}\left\{(j-i)Ta - [(j-i)\mathrm{NTR}^{\star}_E(\theta) - b]_+\right\}$$

and may fail with probability $\Pr\{Q(j) > q\} \leqslant \varepsilon^{'}(b)$, where $[x]_+ = 0$ if $x < 0$ and $[x]_+ = x$ otherwise, which accounts for $S(i,j) \geqslant 0$. In this place, if $a \leqslant \mathrm{NR}^{\star}_E(\theta)$ for stability,

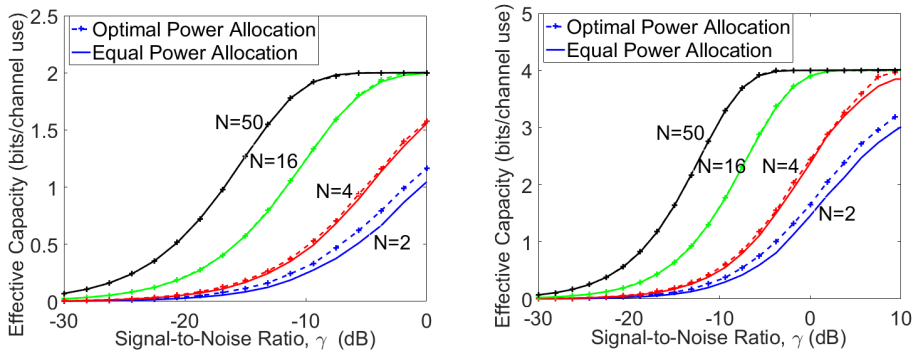
$$q = Ta\frac{b}{\mathrm{NTR}_E(\theta)-\delta} \tag{2.30}$$

is valid for all j. Accordingly, we can express the delay bound $\Pr\{D(j) > d\}$ with $d = \frac{q}{a}$, which is expressed in *channel use*. In other words, $\frac{Tb}{\mathrm{NTR}_E(\theta)-\delta}$ in (2.30) provides us the initial latency caused by the variations in the transmission. Finally, we can express b by inversion of (2.29) for any given $\varepsilon^{'}$ as

$$b = \sigma_R(\theta) - \frac{1}{\theta}\left[\log_e(\varepsilon^{'}) + \log_e\left(1-e^{-\theta\delta}\right)\right]. \tag{2.31}$$

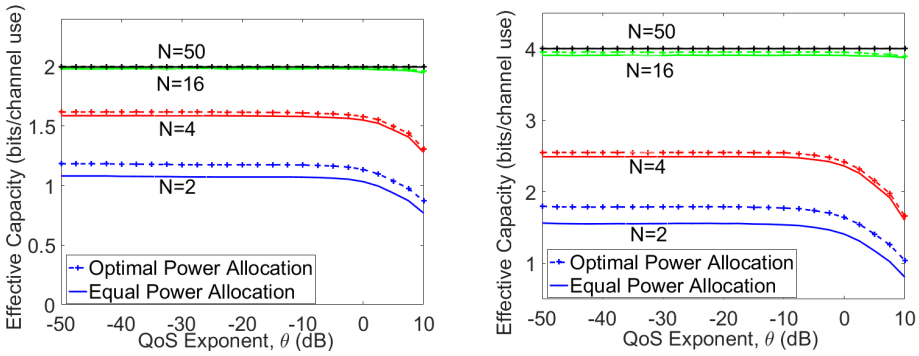
As for the existence of the slack term in (2.31), we refer to the following Lemma.

Lemma 1 *If $S(i,j)$ has an envelope rate $\mathrm{NTR}_E(\theta) < \infty$ for every $\epsilon > 0$, there exists $\sigma_R(\theta) < \infty$ such that $S(i,j)$ is $(\sigma_R(\theta), \mathrm{NTR}_E(\theta) - \epsilon)$-upper constrained [70, Lemma 1].*



(a) BPSK (b) 4-QAM

Figure 2.2: Effective capacity as a function of the signal-to-noise ratio, γ, when $M = 2$ and $\theta = 1$ with different number of receive antennas, i.e., $N = \in \{2, 4, 16, 50\}$.



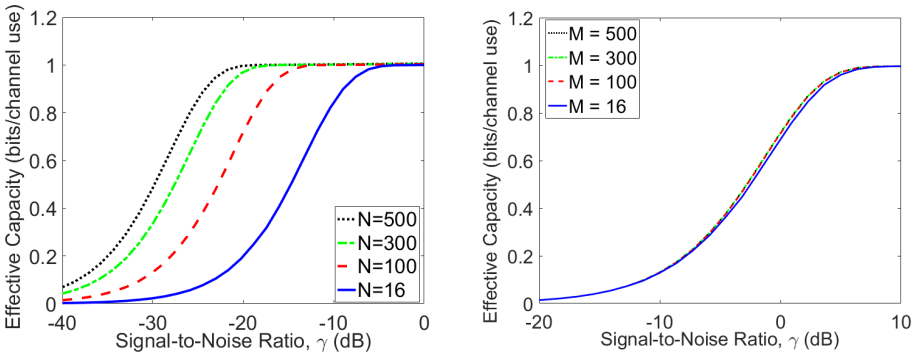
(a) BPSK (b) 4-QAM

Figure 2.3: Effective capacity vs. the QoS exponent, θ, when $M = 2$ and $\gamma = 0$ dB with different number of receive antennas. The input is BPSK-modulated.

2.6 NUMERICAL RESULTS

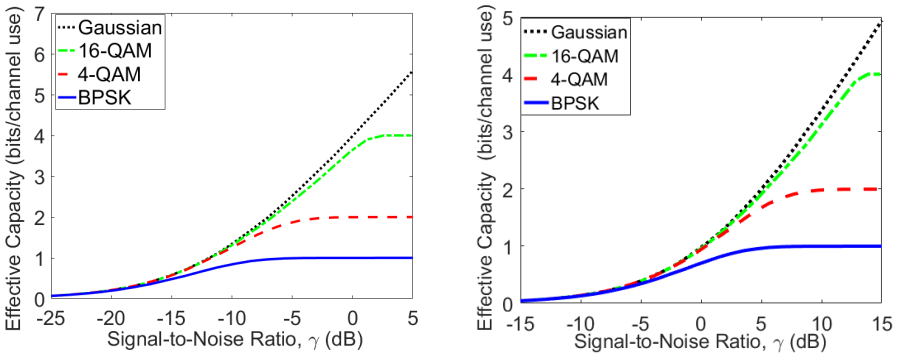
In this section, we substantiate our analytical results through numerical analysis. We initially assume that the channel is perfectly known at both the transmitter and receiver, and that the channel coefficients are uncorrelated, i.e., $\mathbf{R}_r = \mathbf{I}_N$ and $\mathbf{R}_t = \mathbf{I}_M$. In addition, we consider a Rayleigh fading channel model, where the components of the channel matrix, $\mathbf{H}$, are independent and identically distributed, zero-mean, unit variance ($\sigma_h^2 = 1$), circularly symmetric Gaussian random variables, i.e., $\{h_{nm}\} \sim \mathcal{CN}(0, 1)$ for $n \in \{1, \cdots, N\}$ and $m \in \{1, \cdots, M\}$. In addition, we set the noise power to $\sigma_w^2 = 1$. Thus, the signal-to-noise ratio is same with the transmission power, i.e., $\gamma = P$. Moreover, for the sake of simplicity, we set the number of *channel uses* in one transmission frame to 1, i.e., $T = 1$.

Initially, we plot the effective capacity of the MIMO system as a function of the signal-to-noise ratio, γ, for different numbers of receive antennas, N, in Fig. 2.2 when the number of transmit antennas and the queue decay rate



(a) M = 1 (b) N = 1

Figure 2.4: Effective capacity of different transmission scenarios as a function of signal-to-noise ratio γ for BPSK and $\theta = 5$.

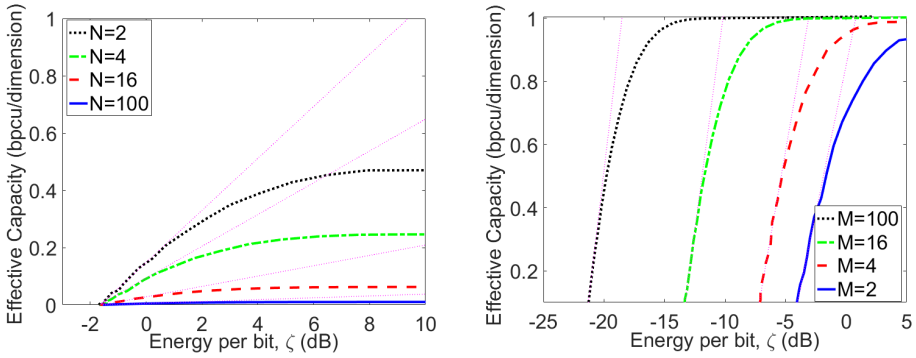


(a) M = 1 and N = 16 (b) M = 16 and N = 1

Figure 2.5: Effective capacity of different transmission scenarios vs. signal-to-noise ratio γ for different input signaling and $\theta = 1$.

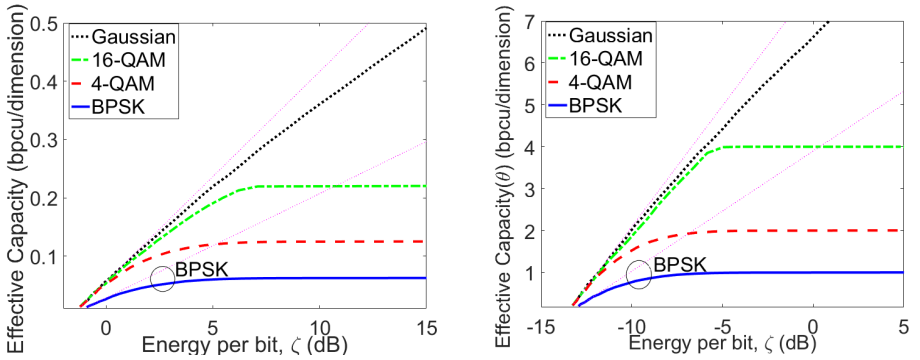
are set to 2 and 1, i.e., $M = 2$ and $\theta = 1$, respectively. We employ BPSK in Fig. 2.2a and 4-quadrature amplitude modulation (4-QAM) in Fig. 2.2b. This transmission scenario with 2 transmit antennas and many receive antennas can be considered as an uplink communication channel. We obtain the optimal input covariance matrix (i.e., optimal power allocation across the transmit antennas) and compare the effective capacity performance with the ones obtained when the input covariance matrix is diagonal (i.e., equal power allocation across the transmit antennas, where $\mathbf{K} = \frac{1}{M}\mathbf{I}$).

We clearly observe that the performance gap decreases with the increasing number of the receive antennas. In particular, given that BPSK and 4-QAM are employed, it is not very necessary to perform power optimization across the transmit antennas when the delay concerns are of importance. With the increasing number of antennas, the channel behaves almost deterministic and non-fading. In other words, the statistical dispersion index (Fano factor) of the channel service rates, i.e., a normalized measure of the dispersion of a probability distribution [136], approaches zero. The key point behind this behavior is the *self-averaging* property that we use to prove Theorem 4, and it shows that the so-called free energy converges in probability to its



(a) M = 1 (b) N = 1

Figure 2.6: Effective capacity of different transmission scenarios as a function of energy-per-bit ζ for BPSK and $\theta = 1$. bpcu: *bits/channel use*.



(a) M = 1 and N = 16 (b) M = 16 and N = 1

Figure 2.7: Effective capacity of different transmission scenarios as a function of energy-per-bit ζ for different input signaling and $\theta = 1$. bpcu: *bits/channel use*.

expectation over the distribution of the channel matrix in the large-antenna regime. Moreover, we see that because the number of bits that can be transmitted in one modulated symbol is limited (i.e., 1 bit with BPSK and 2 bits with 4-QAM, and hence 2 and 4 bits in total with 2 transmit antennas), when γ is higher we can send the data by employing equal power allocation across the transmit antennas.

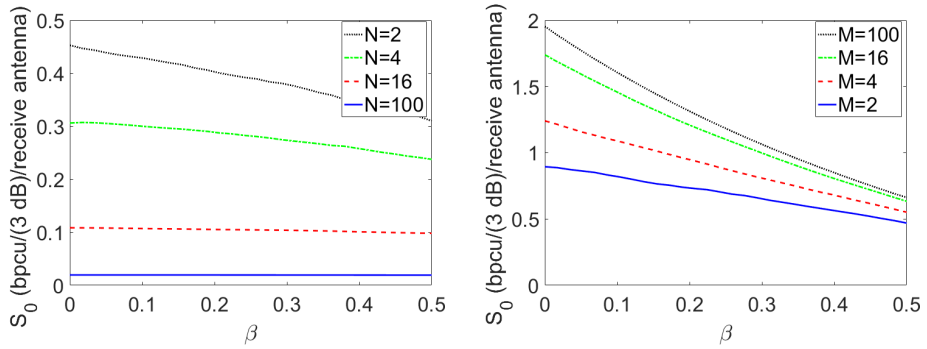
Regarding the system performance when the QoS metrics are of importance, we plot the effective capacity as a function of θ in Fig. 2.3 by employing BPSK and setting $\gamma = 0$ dB. With increasing θ, the effective capacity performance decreases and approaches zero. The effective capacity goes to the average transmission rate in the channel with decreasing θ. Moreover, the performance gain by employing power optimization is again not significant when the number of receive antennas is higher.

Employing the equal power allocation policy, we plot the effective capacity as a function of γ when the number of transmit antennas is fixed to 1 for different number of receive antennas in Fig. 2.4a and when the number of receive antennas is fixed to 1 for different number of transmit antennas in

Fig. 2.4b. The input data is BPSK-modulated. In order to understand the system behavior under strict QoS constraints, we set $\theta = 5$. Again, we can refer to the scenario in Fig. 2.4a as an uplink scenario and the scenario in Fig. 2.4b as a down-link scenario.

We observe that increasing the number of the receive antennas while keeping the number of transmit antennas constant boosts the effective capacity performance when the signal-to-noise ratio is small as seen in Fig. 2.4a. On the other hand, increasing the number of transmit antennas while keeping the number of receive antennas fixed does not provide a performance increase when the delay violation and buffer overflow concerns are present as seen in Fig. 2.4b. The reason behind this is the fact that increasing the number of receive antennas provides more power gain[6] [46, Chapter 8].

Subsequently, regarding the system performance with different modulation techniques, we again plot the effective capacity as a function of γ in Fig. 2.5 when we have BPSK, 4-QAM, 16-QAM and Gaussian signaling for $\theta = 1$. We set the number of transmit and receive antennas as $M = 1$ and $N = 16$ in Fig. 2.5a and $M = 16$ and $N = 1$ in Fig. 2.5b. Likewise, the former scenario can be considered as an uplink transmission and the latter can be considered as a down-link transmission. Regardless of the modulation technique, increasing the number of receive antennas helps improve the system performance more than increasing the number of transmit antennas does under the same conditions.



(a) $M = 1$ (b) $N = 1$

Figure 2.8: Effective capacity slope $\mathcal{S}_0$ as a function of the error ratio, β, and $\theta = -20$ dB, where $\beta = \frac{\sigma_e^2}{\sigma_h^2}$. bpcu: *bits/channel use*.

As for the system performance in the low signal-to-noise ratio regime, we plot the effective capacity as a function of the energy-per-bit, ζ, for different numbers of transmit and receive antennas in Fig. 2.6 by employing optimal power allocation policy when $\theta = 1$. We have the results for different number of receive antennas when the number of transmit antennas is set to

6 In [46, Chapter 8], comparing multi-input-single-output (MISO) and single-input-multi-output (SIMO) channel models, the author showed that SIMO systems outperforms MISO systems having the same number of receive and transmit antennas, respectively, which is also valid for the effective capacity performance.

1, i.e., $M = 1$, in Fig. 2.6a, and for different number of transmit antennas when the number of receive antennas is set to 1, i.e., $N = 1$, in Fig. 2.6b. We plot the effective capacity in *bits/channel use/dimension*. The minimum energy-per-bit, $\zeta_{\min}$, decreases with the increasing number of transmit antennas, whereas it is independent of the number of receive antennas given that the number of transmit antennas is fixed.

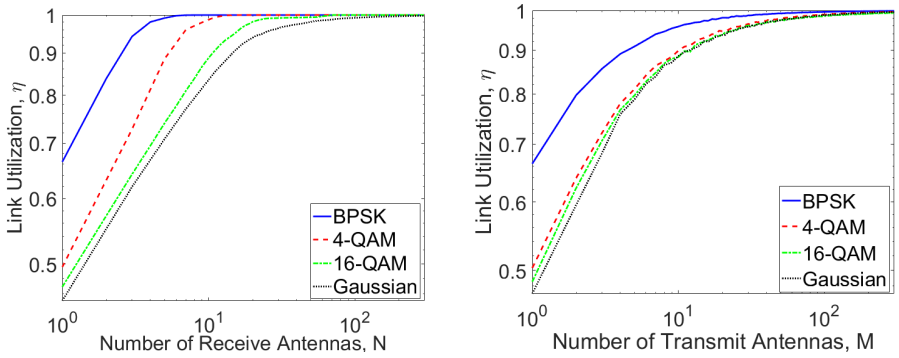
This observation verifies our analytical derivation in (2.23), which provides us the minimum energy-per-bit when either the number of transmit antennas or the number of receive antennas goes to infinity, or both go to infinity. In addition, we again plot the effective capacity as a function of ζ and compare the system performance when different modulation techniques are employed.

In Fig. 2.7a and Fig. 2.7b, we set the number of antennas as follows: $M = 1$ and $N = 16$, and $M = 16$ and $N = 1$, respectively. In both figures, the minimum energy-per-bit, $\zeta_{\min}$, is independent of the input modulation. We also note that the slope of the effective capacity versus ζ curve at $\zeta_{\min}$, $\mathcal{S}_0$, when BPSK is employed is half of the slope when the other modulation techniques are employed, which are formed in the complex domain.

Theorem 3 shows that the slope of the effective capacity versus ζ (in dB) curve at $\zeta_{\min}$, i.e., $\mathcal{S}_0$, decreases with the decreasing channel estimation quality. Hence, we plot $\mathcal{S}_0$ as a function of the additive Gaussian noise variance, σ_e^2, for different number of receive antennas in Fig. 2.8a and for different number of transmit antennas in Fig. 2.8b. The results confirm that the slope decreases with the decreasing estimation quality. In other words, the effective capacity increases slowly with the increasing transmission power in the low signal-to-noise ratio regime. Moreover, the decreasing estimation quality affects the effective capacity with the increasing number of receive antennas less than the increasing number of transmit antennas.

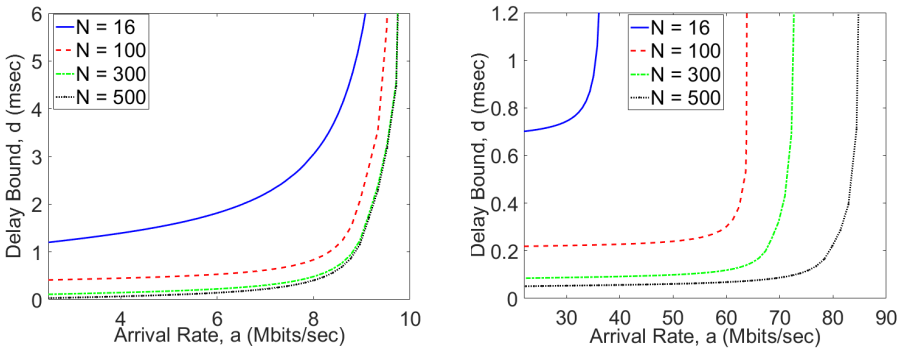
In addition, we display the system performance in the large-scale antenna regime. Hence, by setting $\theta = 5$ and $\gamma = 0$ dB and by employing the equal power allocation policy, we plot the link utilization as a function of the number of receive antennas in Fig. 2.9a and the number of transmit antennas in Fig. 2.9b. And then, we compare the system performance by having different modulation techniques. Recall that we define the link utilization as the ratio of the effective capacity to the average transmission rate. The fact that the link utilization approaches one with the increasing number of receive or transmit antennas justifies the result in Theorem 4. The link utilization reaches 1 faster with the increasing number of receive antennas than it does with the increasing number of transmit antennas. In addition, the link utilization is higher when BPSK is employed than it is when the others are employed, while it is lower when Gaussian distributed input is employed than it is when the others are employed. This is because the scattering of the probability of the achievable transmission rates in the channel is reduced when BPSK is employed and the scattering increases with the complexity of the modulation technique [106].

Finally, we display the non-asymptotic performance of an uplink MIMO scenario when the number of receive antennas is $N = 16$ and the number of



(a) M = 1 (b) N = 1

Figure 2.9: Link Utilization of different transmission scenarios for different input signaling and $\gamma = 0$ dB.



(a) BPSK (b) Gaussian

Figure 2.10: Delay bound of an uplink MIMO scenario as a function of the data arrival rate when M = 1 and N = 16 for $\gamma = 0$ dB and $\varepsilon' = 10^{-6}$.

transmit antennas is $M = 1$, where we employ the equal power allocation policy. Here, we set the delay violation probability to $\varepsilon' = 10^{-6}$ when $\gamma = 0$ dB and $T = 10^{-7}$ seconds. We plot the delay bound as a function of the data arrival rate when the transmitted data is modulated with BPSK and Gaussian input signaling in Fig. 2.10a and Fig. 2.10b, respectively. We observe that Gaussian distributed input provides lower delay bounds for a given delay violation probability than BPSK-modulated input does. We further see that the delay bound goes to infinity when the data arrival rate approaches the average transmission rate in the channel. In addition, the number of receive antennas affects the transmission performance by decreasing the delay bound for a given delay violation probability. However, after a certain value, increasing the number of receive antennas does not contribute to the delay performance.

3

EFFECTIVE CAPACITY IN NON-ORTHOGONAL MULTIPLE ACCESS CHANNELS

In this chapter, we target an non-orthogonal multiple access (NOMA) fading channel with two transmitters and one common receiver under quality of service (QoS) constraints. In particular, we consider a power-domain NOMA, in which the two transmitters are distinguished in the power domain while the receiver employs successive interference cancellation with a certain order to suppress inter-user interference. Recall that, in such a scenario a resource management scheme that can jointly optimize the transmission power levels at both transmitters and the decoding order at the receiver is a critical requirement. We initially formulate the transmission rates for both transmitters, assuming that they have arbitrarily distributed input signals.

Then, we establish the effective capacity region that provides the maximum allowable sustainable arrival rate region at the transmitters' buffers under QoS guarantees. Assuming limited transmission power budgets at the transmitters, we attain the power allocation policies that maximize the effective capacity region. As for the decoding order at the receiver, we characterize the optimal decoding order regions in the space spanned by channel fading power parameters for given power allocation policies. In order to accomplish the aforementioned objectives, we make use of the relationship between the minimum mean square error and the first derivative of the mutual information with respect to the power allocation policies.

3.1 INTRODUCTION

With the growth in wireless networks, recent years witnessed a large body of research on cooperative transmissions [137]. The researchers in some of these studies concentrated on multiple access transmission scenarios and investigated these scenarios from an information-theoretic perspective. Conventionally, wireless systems employ orthogonal multiple access (OMA) schemes, in which different users are allocated orthogonal resources in terms of time, frequency, or code. While these schemes have a fundamental advantage of detection simplicity due to the orthogonality, orthogonal schemes are known to be sub-optimal in terms of the achievable transmission rates and cannot always achieve the capacity region of multi-user systems [46].

Alternatively, power-domain non-orthogonal multiple access (NOMA) scheme has been shown to achieve the capacity region by allowing all users to share the available time and bandwidth simultaneously. In power-domain NOMA, different users are distinguished in the power domain while the receiver employs successive interference cancellation with a certain order to mitigate inter-user interference. As the need for radically higher data

rates is one of the key requirements in 5G [27], power-domain NOMA has been gaining an increasing attention as an emerging technology for 5G networks [49, 138]. In this context, different studies have explored the power-domain NOMA from an information-theoretic perspective [3, 50, 139–142]. For instance, the authors in [3] defined the ergodic capacity region for multiple access fading channels and derived the optimal resource allocation policies that maximize this region. Similarly, addressing the optimal power allocation policies that achieve any point on the capacity region boundary subject to a sum-power constraint, Gupta *et al.* studied Gaussian parallel (non-interacting) multiple access channels [140]. Moreover, taking the vector fading multiple access channels, the authors examined the dynamic resource allocation policies as an important means to increase the sum capacity in uplink synchronous code-division multiple-access systems [50].

It is very well known that the use of discrete and finite constellation diagrams is required for input signaling in many practical systems. Different than the above studies where the authors consider Gaussian input signaling, the authors in [52] researched two-user Gaussian multiple access channels with finite input constellations. Equivalently, the authors in [8] considered parallel Gaussian channels with arbitrary inputs as well. They investigated the optimal power allocation that maximizes the mutual information subject to an average power constraint by exploiting the relationship between the mutual information and the minimum mean-square error (MMSE), which was established in [115]. Furthermore, the authors in [143] explored the optimal power policies that minimize the outage probability over block-fading channels with arbitrary input distributions that were subject to both peak and average power constraints. Power allocation policies for a two-way relay channel with arbitrary inputs were studied in low and high signal-to-noise ratio regimes. In another line of research, the author studied the multiple access multiple-input multiple-output channels, and showed the relationship between the input-output mutual information and the MMSE [51].

In the meantime, recalling that current wireless systems require data transmission with strict constraints on delay performance, cross-layer design concerns have become of interest to many system designers. Therefore, quality of service (QoS) requirements regarding buffer overflow and delay have been addressed in wireless communications studies regarding the data-link and physical layers. As mentioned early, effective capacity was established as a measure to indicate the maximum sustainable rate at a transmitter queue by a given service (channel) process [81], and it has been investigated in several different transmission scenarios, e.g., [2, 68, 104]. More recently, Ozcan *et al.* [144] studied the effective capacity of point-to-point channels and derived the optimal power allocation policies to maximize the system throughput by employing arbitrary input distributions under average power constraints. Similarly, a cross-layer study was also conducted in [145] considering different data traffic models.

In this chapter we focus on a two-user power-domain NOMA scenario. Different from the aforementioned studies, we assume that both trans-

Table 3.1: Table of symbols in this chapter

Symbol	Description
a_i	Constant arrival rate at i, $i \in \{1,2\}$
α_i	Normalized power policy for transmitter i, $i \in \{1,2\}$, i.e., $\alpha_i = P_i/\overline{P}$
B	Transmission bandwidth
$h_i(t)$	Channel coeffiencnt between receiver and transmitter i, $i \in \{1,2\}$
$\mathcal{I}(a;b)$	Mutual information between a and b
K	Rician channel factor
λ_i	Weight of transmitter i in the optimization problem, $i \in \{1,2\}$
$\overline{P}$	Average transmission power limit
$P_i(t)$	Instantaneous power policy for transmitter i, $i \in \{1,2\}$
Ψ_i	$\mathbb{E}_{(z1,z2)}\{e^{-\theta_i TBr_i(z)}\}$
$r_i(z_1,z_2)$	Transmission rate of transmitter i, $i \in \{1,2\}$
T	Frame duration
θ_i	QoS exponent for transmitter i, $i \in \{1,2\}$
ε	Lagrange multiplier of the average power constraint
$x_i(t)$	Channel input at transmitter i, $i \in \{1,2\}$
$y(t)$	Channel output at time instance t
y_i	Channel output due to transmitter i only, $i \in \{1,2\}$
$\mathcal{Z}$	(z_1,z_2)-space where decoding order is $(2,1)$
$\mathcal{Z}^c$	(z_1,z_2)-space where decoding order is $(1,2)$
z_i	Channel fading power, $z_i = \lvert h_i\rvert^2$, between receiver and i, $i \in \{1,2\}$

mitters apply arbitrarily distributed input signaling under average power constraints and QoS requirements which are inflicted as limits on the delay violation and buffer overflow probabilities. We emphasize that our analysis can be easily expanded to multiple access scenarios with more than two transmitters. Our attention in this chapter is mainly focused on a resource management scheme that can jointly optimize the transmission power at the transmitters and the decoding order at the receiving to with the goal of maximizing the effective capacity region. In our approach, we benefit from the fundamental relation between the mutual information and the MMSE, which was initially formulated in [115]. Table 3.1 summarizes the symbols used in this chapter.

3.2 SYSTEM DESCRIPTION

In this section, we provide a comprehensive description regarding the system under investigation and the main performance criteria. Specifically, the

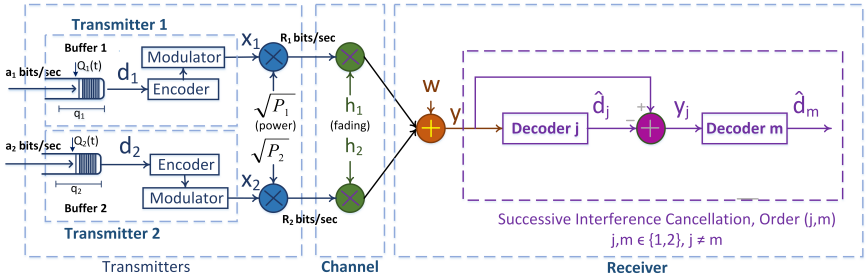


Figure 3.1: NOMA Channel model with two transmitters and one receiver. Each transmitter has its own data buffer, and the receiver performs successive interference cancellation with a certain order.

channel model is introduced in Section 3.2.1, while the superposition coding and successive interference cancellation process and the corresponding achievable rates are detailed in Section 3.2.2. For the performance criteria to be maximized, the effective capacity region of the considered system is defined in Section 3.2.3.

3.2.1 *Channel Model*

We consider a multiple access channel scenario in which two transmitters send data to one common receiver as seen in Figure 3.1. We initially assume that the data arrive at both transmitters from a source (or sources), and they are stored in the transmitters' data buffers before being conveyed into the wireless channel. Then, each transmitter divides the available data into data packets and performs the encoding, modulation and transmission of each packet in frames of T seconds. Thus, we impose certain QoS requirements in each transmitter buffer in order to control the buffer violation probabilities.

During the transmission in the channel, the input-output relation at time instant t is given as

$$y(t) = \sqrt{P_1(t)}h_1(t)x_1(t) + \sqrt{P_2(t)}h_2(t)x_2(t) + w(t),$$

for $t = 1, 2, \cdots$. Above, $x_1(t)$ and $x_2(t)$ are the channel inputs at the corresponding transmitters (i.e., Transmitter 1 and 2, respectively, in Fig. 1), and $y(t)$ is the channel output at the receiver. $P_1(t)$ and $P_2(t)$ are the instantaneous power allocation policies employed by Transmitter 1 and 2, respectively, with the following average power constraint:

$$\mathbb{E}\{P_1(t)\} + \mathbb{E}\{P_2(t)\} \leqslant \overline{P}, \tag{3.1}$$

where $\overline{P}$ is finite. Moreover, $w(t)$ denotes the zero-mean, circularly symmetric, complex Gaussian random variable with a unit variance, i.e., $\mathrm{E}\{|w|^2\} = 1$. The noise samples $\{w(t)\}$ are independent and identically distributed. Meanwhile, $h_1(t)$ and $h_2(t)$ represent the fading coefficients between Transmitter 1 and the receiver, and Transmitter 2 and the receiver, respectively. The magnitude squares of the fading coefficients are denoted by $z_1(t) = |h_1(t)|^2$ and $z_2(t) = |h_2(t)|^2$ with finite averages, i.e., $\mathrm{E}\{z_1\} < \infty$ and $\mathrm{E}\{z_2\} < \infty$. We

consider a block-fading channel, and assume that the fading coefficients stay constant for a frame duration of T seconds and change independently from one frame to another. The channel coefficients, h_1 and h_2, are perfectly known to the receiver and both transmitters, and hence, each transmitter can adapt its transmission power policy accordingly. We finally note that the available transmission bandwidth is B Hz. In the rest of the chapter, we omit the time index t unless otherwise needed for clarity.

3.2.2 *Achievable Rates*

We can express the instantaneous achievable rate between the transmitters and the receiver by invoking the mutual information between the inputs at the transmitters, i.e., x_1, x_2, and the output at the receiver, i.e., y. Hence, given that the instantaneous channel fading values, h_1 and h_2, are available at the transmitters and the receiver, the instantaneous achievable rate can be given as [146]

$$\mathcal{I}(x_1, x_2; y) = \mathbb{E}\left\{\log_2 \frac{f_{y|x_1,x_2}(y|x_1, x_2)}{f_y(y)}\right\}, \tag{3.2}$$

where $f_y(y) = \sum_{x_1,x_2} p(x_1, x_2) f_{y|x_1,x_2}(y|x_1, x_2)$ is the marginal probability density function (pdf) of the received signal y and

$$f_{y|x_1,x_2}(y|x_1, x_2) = \frac{1}{\pi} e^{-|y-\sqrt{\alpha_1 \overline{P}} h_1 x_1 - \sqrt{\alpha_2 \overline{P}} h_2 x_2|^2}.$$

Above, we consider the normalized power allocation policies: $\alpha_1 = \frac{P_1}{\overline{P}}$ and $\alpha_2 = \frac{P_2}{\overline{P}}$.

We assume that the receiver performs successive interference cancellation with a certain order (j, m) for $j, m \in \{1, 2\}$ and $j \neq m$. The decoding order depends on the channel conditions, i.e., the magnitude squares of channel fading coefficients, z_1 and z_2. In particular, the receiver initially decodes x_j while treating x_m as noise, and then subtracts x_j from the received signal y and decodes x_m. Let $\mathcal{Z}$ be the region of the (z_1, z_2)-space where the decoding order is (2,1). Then, $\mathcal{Z}^c$, which is the complement of $\mathcal{Z}$, is the region where the decoding order is (1,2). Now, we can express the instantaneous transmission rates for each transmitter as follows:

$$r_1(z_1, z_2) = \begin{cases} \mathcal{I}(x_1; y_1), & \mathcal{Z}, \\ \mathcal{I}(x_1; y), & \mathcal{Z}^c, \end{cases} \tag{3.3}$$

and

$$r_2(z_1, z_2) = \begin{cases} \mathcal{I}(x_2; y), & \mathcal{Z}, \\ \mathcal{I}(x_2; y_2), & \mathcal{Z}^c, \end{cases} \tag{3.4}$$

where

$$y_1 = \sqrt{\alpha_1 \overline{P}} h_1 x_1 + w, \quad \text{and} \quad y_2 = \sqrt{\alpha_2 \overline{P}} h_2 x_2 + w. \tag{3.5}$$

The decoding regions can be determined in such a way to maximize the objective throughput. Furthermore, we have

$$\mathcal{I}(x_j; y_j) = \mathbb{E}\left\{\log_2 \frac{f_{y_j|x_j}(y_j|x_j)}{f_{y_j}(y_j)}\right\},$$

where $f_{y_j}(y_j) = \sum_{x_j} p(x_j) f_{y_j|x_j}(y_j|x_j)$ is the marginal pdf of y_j and

$$f_{y_j|x_j}(y_j|x_j) = \frac{1}{\pi} e^{-|y_j - \sqrt{\alpha_j \overline{P}} h_j x_j|^2}.$$

3.2.3 *Effective Capacity Region*

Recall that the data packets are stored in the buffers of the transmitters until they are reliably decoded by the receiver. Thus, the delay and buffer overflow concerns are of interest for system designers. Therefore, we concentrate on the data arrival processes, i.e., a_1 and a_2 in Fig. 3.1, and we propose the effective capacity that provides us the maximum constant arrival rate that a given service (channel) process can support in order to guarantee a desired statistical QoS specified with the QoS exponent θ [81].

In the aforementioned multiple access transmission scenario, each transmitter has its own buffer to store the data, and it has its own QoS requirements. Therefore, we denote the decay rate of Transmitter 1 and Transmitter 2 by θ_1 and θ_2, respectively. Noting that the transmission bandwidth is B Hz, the block duration is T seconds, and the channel fading coefficients change independently from one transmission frame to another, we can express the effective capacity of each transmitter, i.e., the maximum sustainable data arrival rate at Transmitter j, in bits/sec/Hz as

$$-\frac{1}{\theta_j TB} \log_e \mathbb{E}\left\{e^{-\theta_j TB r_j(z_1, z_2)}\right\} \quad j \in \{1, 2\}, \tag{3.6}$$

where the expectation is taken over the (z_1, z_2)-space. Now, invoking the definition given in [147], we express the effective capacity region of the given multiple access transmission scenario as follows:

$$\mathcal{C}_E(\Theta) = \bigcup_{r_1, r_2} \left\{C(\Theta) \geqslant \mathbf{0} : C_j(\theta_j) \leqslant -\frac{1}{\theta_j TB} \log_e \mathbb{E}\left\{e^{-\theta_j TB r_j(z_1, z_2)}\right\}\right\}, \tag{3.7}$$

where $\Theta = [\theta_1, \theta_2]$, and $C(\Theta) = [C_1(\theta_1), C_2(\theta_2)]$ is the vector of the effective capacity values.

3.3 PERFORMANCE ANALYSIS

In this section, we focus on maximizing the effective capacity region defined in (3.7) under the QoS guarantees required at each transmitter and the average total power constraint defined in (3.1). Noting that the effective capacity region is convex [148], our objective turns out to be maximizing the

boundary surface of the region, which can be characterized by the following optimization problem [3]:

$$\max_{\substack{\mathcal{Z},\mathcal{Z}^c \\ \mathbb{E}\{P_1\}+\mathbb{E}\{P_2\}\leqslant\overline{P}}} \quad \lambda_1 C_1(\theta_1)+\lambda_2 C_2(\theta_2), \tag{3.8}$$

for $\lambda_1,\lambda_2 \in [0,1]$ such that $\lambda_1+\lambda_2=1$. In order to solve this optimization problem, we first obtain the power allocation policies in defined decoding regions $\mathcal{Z}$ and $\mathcal{Z}^c$, and then we provide the optimal decoding regions.

3.3.1 *Optimal Power Allocation*

Here, we study the optimal power allocation policies that solve the optimization problem in (3.8) in given decoding regions $\mathcal{Z}$ and $\mathcal{Z}^c$. In the subsequent result, we provide the following proposition that gives us the optimal power allocation policies:

Proposition 1 *The optimal normalized power allocation policies, α_1 and α_2, that solve the optimization problem in (3.8) are the solutions of the following equalities:*

$$\frac{\lambda_1}{\psi_1}e^{-\theta_1 TBr_1(z)}\frac{dr_1(z)}{d\alpha_1}+\frac{\lambda_2}{\psi_2}e^{-\theta_2 TBr_2(z)}\frac{dr_2(z)}{d\alpha_1}=\varepsilon, \tag{3.9}$$

$$\frac{\lambda_2}{\psi_2}e^{-\theta_2 TBr_2(z)}\frac{dr_2(z)}{d\alpha_2}=\varepsilon, \tag{3.10}$$

for $z=(z_1,z_2)\in\mathcal{Z}$, and

$$\frac{\lambda_1}{\psi_1}e^{-\theta_1 TBr_1(z)}\frac{dr_1(z)}{d\alpha_1}=\varepsilon, \tag{3.11}$$

$$\frac{\lambda_1}{\psi_1}e^{-\theta_1 TBr_1(z)}\frac{dr_1(z)}{d\alpha_2}+\frac{\lambda_2}{\psi_2}e^{-\theta_2 TBr_2(z)}\frac{dr_2(z)}{d\alpha_2}=\varepsilon, \tag{3.12}$$

for $z\in\mathcal{Z}^c$. Above, $\psi_1=\mathbb{E}_z\{e^{-\theta_1 TBr_1(z)}\}$, $\psi_2=\mathbb{E}_z\{e^{-\theta_2 TBr_2(z)}\}$, and ε is the Lagrange multiplier of the average power constraint in (3.1).

Proof: Let us rewrite (3.6) for Transmitter 1 as

$$\begin{aligned} C_1(\theta_1) &= \frac{-1}{\theta_1 TB}\log_e\left\{\mathbb{E}_{\mathcal{Z}}\{e^{-\theta_1 TBJ(x_1;y_1)}\}+\mathbb{E}_{\mathcal{Z}^c}\{e^{-\theta_1 TBJ(x_1;y)}\}\right\} \\ &= \frac{-1}{\theta_1 TB}\log_e\psi_1, \end{aligned} \tag{3.13}$$

and for Transmitter 2 as

$$\begin{aligned} C_2(\theta_2) &= \frac{-1}{\theta_2 TB}\log_e\left\{\mathbb{E}_{\mathcal{Z}}\{e^{-\theta_2 TBJ(x_2;y)}\}+\mathbb{E}_{\mathcal{Z}^c}\{e^{-\theta_2 TBJ(x_2;y_2)}\}\right\} \\ &= \frac{-1}{\theta_2 TB}\log_e\psi_2. \end{aligned} \tag{3.14}$$

Since the objective function in (3.7) is convex and the constraint (3.1) is linear with respect to α_1 and α_2, we can use the Lagrangian method to solve the optimization problem (3.8). We can form the Lagrangian as

$$\mathcal{B}=\lambda_1 C_1(\theta_1)+\lambda_2 C_2(\theta_2)-\varepsilon\{\mathbb{E}_{z\in\mathcal{Z}}\{\alpha_1+\alpha_2\}+\mathbb{E}_{z\in\mathcal{Z}^c}\{\alpha_1+\alpha_2\}-1\},$$

where ε is the Lagrangian multiplier. Now, taking the derivatives of $\mathcal{B}$ with respect to α_1 and α_2 and setting them to zero, we obtain (3.9) and (3.12), respectively, when $z \in \mathcal{Z}$, and (3.11) and (3.10), respectively, when $z \in \mathcal{Z}^c$. □

Above, the derivatives of the transmission rates with respect to the corresponding normalized power allocation policies are given as

$$\frac{dr_1(z)}{d\alpha_1} = \begin{cases} \frac{d\mathcal{I}(x_1;y_1)}{d\alpha_1}, & \mathcal{Z}, \\ \frac{d\mathcal{I}(x_1;y)}{d\alpha_1}, & \mathcal{Z}^c, \end{cases}$$

$$\frac{dr_2(z)}{d\alpha_2} = \begin{cases} \frac{d\mathcal{I}(x_2;y)}{d\alpha_2}, & \mathcal{Z}, \\ \frac{d\mathcal{I}(x_2;y_2)}{d\alpha_2}, & \mathcal{Z}^c, \end{cases}$$

and

$$\frac{dr_m(z)}{d\alpha_j} = \frac{d\mathcal{I}(x_j;y)}{d\alpha_j} - \frac{d\mathcal{I}(x_j;y_j)}{d\alpha_j},$$

for $m, j \in \{1,2\}$ and $m \neq j$.

In the following theorem, we provide the derivatives of the mutual information expressions with respect to the normalized power allocation policies:

Theorem 5 *Let, h_1, h_2, and $\overline{P}$ be given. In the multiple access transmission scenario described in Section 3.2, the first derivative of the mutual information between x_j and y with respect to the power allocation policy, α_j, is given by*

$$\frac{d\mathcal{I}(x_j;y)}{d\alpha_j} = \overline{P} z_j MMSE(x_j;y) + \overline{P}\sqrt{\frac{\alpha_m}{\alpha_j}} Re\left(h_j h_m^* \mathbb{E}\left\{x_j x_m^* - \hat{x}_j(y)\hat{x}_m^*(y)\right\}\right), \tag{3.15}$$

and similarly, the derivative of the mutual information between x_j and y_j with respect to α_j is given by

$$\frac{d\mathcal{I}(x_j;y_j)}{d\alpha_j} = \overline{P} z_j MMSE(x_j;y_j), \tag{3.16}$$

for $j, m \in \{1,2\}$, $j \neq m$, and $(\cdot)^*$ is the complex conjugate operation. In (3.15), the MMSE expression is given as

$$MMSE(x_j;y) = 1 - \frac{1}{\pi}\int \frac{\left|\sum_x x_j p(x_j) f_{y|x}(y|x)\right|^2}{f_y(y)} dy,$$

and the MMSE estimates of the channel inputs are

$$\hat{x}_j(y) = \frac{\sum_x x_j p(x) f_{y|x}(y|x)}{f_y(y)}.$$

Similarly, the MMSE expression in (3.16) is obtained by

$$MMSE(x_j;y_j) = 1 - \frac{1}{\pi}\int \frac{\left|\sum_{x_j} x_j p(x_j) f_{y_j|x_j}(y_j|x_j)\right|^2}{f_{y_j}(y_j)} dy_j,$$

where y_1 and y_2 are as given in (3.5).

Proof: See Appendix E. □

As seen in (3.9)-(3.12), closed-form solutions for α_1 and α_2 cannot be obtained easily which is mainly due to the cross-relation between α_1 and α_2. For instance, α_1 is a function of α_2 as observed in (3.9) for $z \in \mathcal{Z}$, whereas α_2 is a function of α_1 as seen in (3.12) for $z \in \mathcal{Z}^c$. Therefore, we need to employ numerical techniques which consist of iterative solutions.

In the following, we wrap up the above steps into an iterative algorithm that can be used to obtain the optimal power policies in given decoding regions. In Algorithm 1, we obtain the optimal normalized power allocation policies α_1 and α_2.

Algorithm 1

Given λ_1, λ_2, $\mathcal{Z}$ and $\mathcal{Z}^c$;
Initialize ψ_1, ψ_2;
while True **do**
    Initialize ε;
    Initialize α_1,α_2;
    while True **do**
        if $z \in \mathcal{Z}$ **then**
            For given α_1, compute the optimal α_2 by solving (3.10) ;
            For computed α_2, compute the optimal $\alpha_1^\star$ by solving (3.9) ;
            Calculate the absolute difference, i.e., $\kappa = |\alpha_1 - \alpha_1^\star|$;
        else
            For given α_2, compute the optimal α_1 by solving (3.11) ;
            For computed α_1, compute the optimal $\alpha_2^\star$ by solving (3.12) ;
            Calculate the absolute difference, i.e., $\kappa = |\alpha_2 - \alpha_2^\star|$;
        end if
        if $\kappa \leqslant \epsilon$ for small $\epsilon > 0$ **then**
            break;
        else
            if $z \in \mathcal{Z}$ **then**
                Set $\alpha_1 = \alpha_1^\star$;
            else
                Set $\alpha_2 = \alpha_2^\star$;
            end if
        end if
    end while
    Check if the average power constraint in (3.1) is satisfied with quality;
    If not, update ε and return to Step 5
    Compute $\psi_1^\star = \mathbb{E}_z\{e^{-\theta_1 n r_1(z)}\}$ and $\psi_2^\star = \mathbb{E}_z\{e^{-\theta_2 n r_2(z)}\}$
    if $|\psi_1 - \psi_1^\star| \leqslant \epsilon$ and $|\psi_2 - \psi_2^\star| \leqslant \epsilon$ **then**
        break;
    else
        Set $\psi_1 = \psi_1^\star$ and $\psi_2 = \psi_2^\star$;
    end if
end while

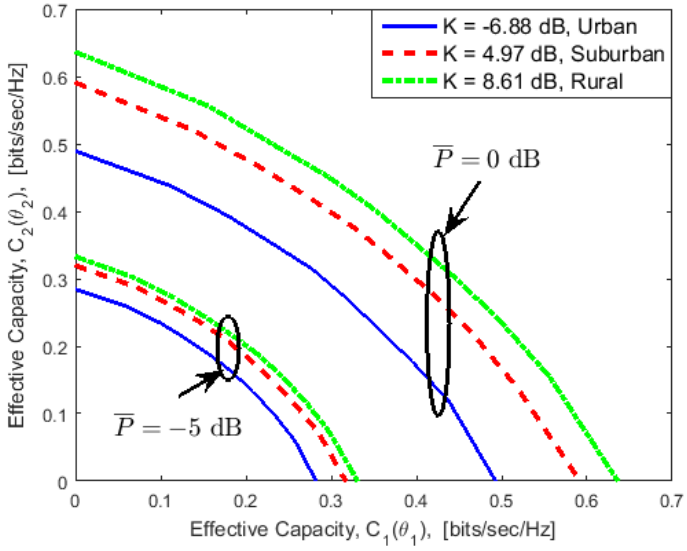


Figure 3.2: Effective capacity region, $C_1(\theta_1)$ vs. $C_2(\theta_2)$, when BPSK input signaling is employed for different values of $\overline{P}$ and K.

Given λ_j and ψ_j for $j \in \{1, 2\}$, it is shown in [149] that both (3.10) and (3.11) has at most one solution. We can further show that (3.9) has at most one solution for α_1 when α_2 is given, and that (3.12) has at most one solution for α_2 when α_1 is given. Then, we can guarantee that Steps 8, 9, 13 and 12 in Algorithm 1 will converge to a single unique solution. It is also clear that (3.9) and (3.11) are monotonically decreasing functions of α_1, and (3.10) and (3.12) are monotonically decreasing functions of α_2. Hence, in region $\mathcal{Z}$, we first obtain α_2 by solving (3.10), and then we find α_1 by solving (3.9) after inserting α_2 into (3.9). Similarly, in region $\mathcal{Z}^c$, we first obtain α_1 by solving (3.11), and then we find α_2 by solving (3.12) after inserting α_1 into (3.12). We can employ bisection search methods to obtain α_1 and α_2. In the above approach, when either α_1 or α_2 becomes negative, we set it to zero.

3.3.2 *Optimal Decoding Order*

Following the optimal power allocation policies, we identify the optimal decoding order regions. We initially note that when there are no QoS requirements, i.e., $\theta_1 = \theta_2 = 0$, the effective capacity region is reduced to be the ergodic capacity region. The authors in [55] showed that the ergodic capacity region is maximized when the symbol of the transmitter with the strongest channel is decoded first. Principally, when $z_j > z_m$, the symbol of Transmitter j is decoded first, and then the symbol of Transmitter m is decoded. Furthermore, the authors in [147] considered a special case and set $\theta_1 = \theta_2 = \theta$ for $\theta > 0$. Then, they derived the optimal decoding order that maximizes the effective capacity region. However, their result is based on the assumption of Gaussian input signaling. Nevertheless, obtaining the optimal decoding order regions is a difficult task when $\theta_1 \neq \theta_2$ and

arbitrary input distribution is employed. In the following, we provide the optimal decoding order regions given that the transmitters have the equal queue decay rates, i.e., $\theta_1 = \theta_2$, and they employ arbitrary input distributions.

Theorem 6 *Let h_1, h_2, and $\overline{P}$ be given. Define $z_2^\star$ for any given $z_1 \geqslant 0$, such that the decoding order is (2,1) when $z_2 > z_2^\star$, and it is (1,2) otherwise for the given z_1. In the multiple access transmission scenario described in Section 3.2, with arbitrary input distributions and the normalized power allocation policies at the transmitters, the optimal $z_2^\star$ for any given z_1 value is the solution of the following equality:*

$$\mathcal{I}(x;y|z_1,z_2^\star) = \mathcal{I}(x_1;y_1|z_1) + \mathcal{I}(x_2;y_2|z_2^\star).$$

Proof: See Appendix F. □

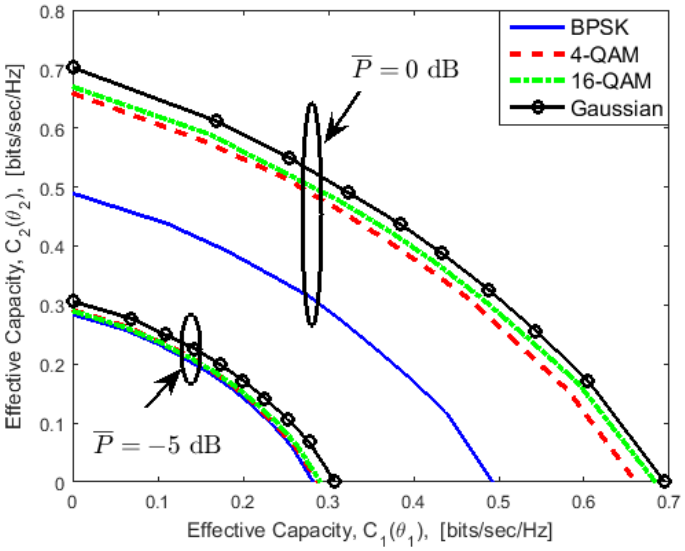


Figure 3.3: Effective capacity region, $C_1(\theta_1)$ vs. $C_2(\theta_2)$, considering different input signaling for $K = -6.88$ dB and different values of $\overline{P}$.

3.4 NUMERICAL RESULTS

In this section, we present the numerical results. Throughout the chapter, we set the available channel bandwidth to $B = 100$ Hz and the transmission duration block to $T = 1$ sec.. We further assume that h_1 and h_2 are independent of each other and set $\mathbb{E}\{|h_1|^2\} = \mathbb{E}\{|h_2|^2\} = 1$. Unless indicated otherwise, we set the QoS exponents $\theta_1 = \theta_2 = 0.01$. We define the signal-to-noise ratio with $\frac{\overline{P}}{\mathbb{E}\{|w|^2\}} = \overline{P}$ where $\mathbb{E}\{|w|^2\} = 1$.

We initially consider binary phase shift keying (BPSK) at both transmitters, and we plot the effective capacity region in Fig. 3.2. We have the results for different values of the signal-to-noise ratio, $\overline{P}$, and K. Recall that when $K = 0$, the channel fading has a Rayleigh distribution, i.e., there is not a strong line-of-sight propagation path between the transmitters and the receiver. On the other hand, when $K > 0$, there is a line-of-sight path between

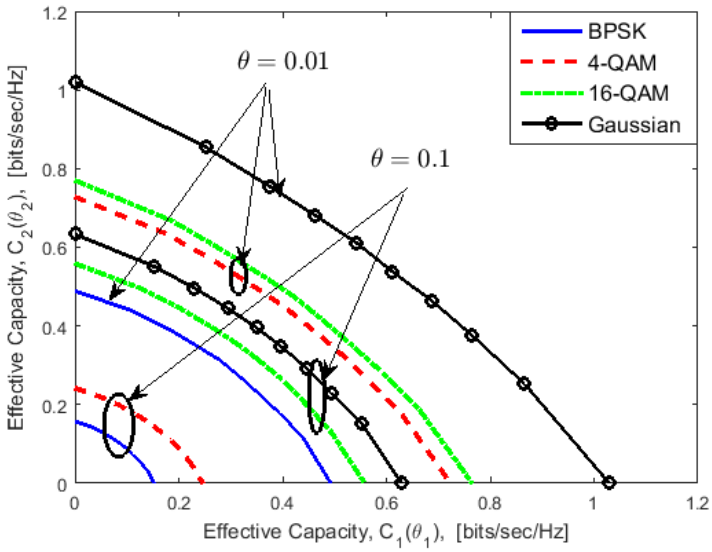


Figure 3.4: Effective capacity region, $C_1(\theta_1)$ vs. $C_2(\theta_2)$, considering different input signaling for $K = -6.88$ dB, $\overline{P} = 5$ dB and different values of $\theta = \theta_1 = \theta_2$.

the transmitters and the receiver, and the line-of-sight propagation path becomes dominant with increasing K[1]. As expected, with increasing K, the effective capacity region broadens. Moreover, we observe the broadening of the effective capacity region with increasing $\overline{P}$ more clearly.

Setting $K = -6.88$ dB, we plot the effective capacity region for different $\overline{P}$ values and signal modulation methods such as BPSK, quadrature amplitude modulation (QAM) and Gaussian distributed signaling in Fig. 3.3. We can easily notice that Gaussian input signaling has the best performance for both $\overline{P} = -5$ dB and $\overline{P} = 0$ dB, while BPSK has the lowest performance. However, the performance gap is reduced with decreasing $\overline{P}$. Furthermore, we investigate the effect of the QoS exponent, θ, on the effective capacity region in Fig. 3.4. Here, we set $\overline{P} = 5$ dB and $K = -6.88$ dB, and compare the effective capacity region for different modulation techniques. As clearly seen, increasing θ results in a decrease in the effective capacity region since the system is subject to stricter QoS constraints. We can further observe that the performance gaps among the modulation techniques are smaller with increasing θ. We finally display the effective capacity region for transmitters having different modulation methods than each other in Fig. 3.5. We can clearly notice that the transmitter with an input signal of higher modulation order can sustain higher effective capacity.

1 K is the ratio of the power in the line-of-sight component to the total power in the non-line-of-sight components in a channel. Therefore, the ratio of the power in the line-of-sight component to the total channel power is defined as $\nu = \frac{K}{K+1}$. It is shown in [150] that the empirical means of K are -6.88 dB, 8.61 dB and 4.97 dB for urban, rural and suburban environments, respectively, at 781 MHz.

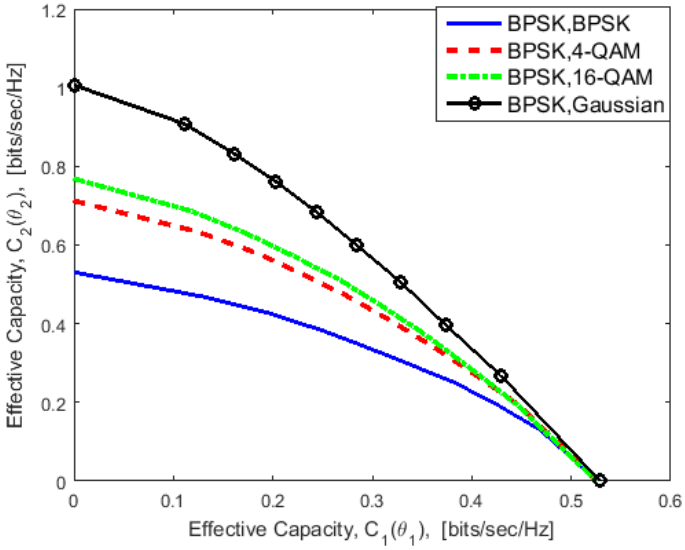


Figure 3.5: Effective capacity region, $C_1(\theta_1)$ vs. $C_2(\theta_2)$, considering mixed input signaling for $K = -6.88$ dB, $\overline{P} = 0$ dB and $\theta_1 = \theta_2 = 0.01$.

4

STATISTICAL QOS PROVISIONINGS FOR VLC SYSTEMS WITH FIXED-RATE TRANSMISSIONS

Having applied the cross-layer concepts in different radio frequency (RF) settings in the previous two chapters, in this chapter we extend our analysis to visible light communication (VLC) systems. In this context, one of the fundamental differences between RF and VLC systems is that VLC channels do not suffer from the small-scale multi-path variations (i.e., fading). In other words, for a given user location VLC channels are known as time-invariant. This "favorable" property makes VLC systems suitable for 5G networks, where strict delay requirements are to be satisfied. Recall that multi-path fading (channel randomness) is a key limiting factor that restricts the ability of RF channels to support such strict QoS needs. However, VLC systems may have other sources of randomness (uncertainty), rather than the channel, e.g., random data traffic and/or uncertain (unknown) user location.

Having this background, in this chapter we consider a VLC system in which the VLC access point (AP) is unaware of the user location and considering an ON-OFF data source. For such a scenario, we provide a cross-layer study when statistical QoS needs, in the form of limits on the delay violation and buffer overflow probabilities, should be fulfilled. To this end, in this chapter we employ the maximum average data arrival rate at the AP buffer and the non-asymptotic bounds on buffering delay as the main performance measures. Through numerical results, we illustrate the impacts of different physical characteristics of the VLC system on the performance levels.

4.1 INTRODUCTION

Recently, visible light communication (VLC) has emerged as a potential transmission technology, thanks to the remarkable advances in white light emitting diodes (LEDs) technology that enabled utilizing the visible light spectrum for data transmission along with illumination. In addition to providing the bandwidth required to meet the increasing demand on wireless services, using LEDs for data transmission has many advantages over the radio frequency (RF) technology [58]. For example, LEDs are cheap, energy efficient, and installed almost everywhere in indoor scenarios for lighting.

Most of the existing literature studies on VLC, see e.g., [151–154] and references therein, focused mainly on the physical layer aspects of the VLC systems. However, the increasing demand on delay-sensitive applications, as reported by Cisco[25], requires involving additional constraints on the buffer dynamics at the data-link layer. In this regard, cross-layer analysis has been gaining an increasing attention as a powerful tool to study and assess

different quality-of-service (QoS) mechanisms in wireless networks. Regarding the physical and data-link layers, cross-layer analyses were addressed by many researchers in the RF literature, see e.g., [79, 80, 145, 155] and references therein. On the other hand, to the best of our knowledge, there are only very few studies that recently investigated cross-layer concepts in VLC systems[71, 72].

Nevertheless, these studies are based on the assumption that the VLC access point (AP) has a full and instant knowledge about the channel conditions. Indeed, this is not a practical assumption in many indoor scenarios, e.g., as in airports and exhibition halls, where the AP is expected to serve a large number of users, thus learning the channel of each user can be difficult and a resource-consuming task. Furthermore, notice that the authors in[71, 72] focused on the case of constant data arrival rates at the transmitter buffer, which may not be realistic in many practical settings.

In this chapter, we focus on a VLC system that operates under statistical QoS constraints, which are applied as limits on the buffer overflow and delay violation probabilities, and considering an ON-OFF data source. Different than the aforementioned cross-layer studies in VLC, we assume that the VLC AP has no knowledge about the user channel gain, thus the AP sends the data with a fixed rate. For such a system, we provide a cross-layer study regarding the physical and data-link layers by employing the maximum average arrival rates at the transmitter buffer and the non-asymptotic bounds on buffering delay as the main performance metrics. To summarize, the main contributions of this chapter are *i*) VLC systems with a practical assumption such that no knowledge about the channel quality is required at the access points, *ii*) cross-layer analyses when the system is operating under statistical quality of services constraints, imposed as limits on the buffer dynamics, and considering an ON-OFF data source, and *ii*) non-asymptotic bounds on the buffering delay. We emphasize that, to the best of our knowledge, the analytical framework provided in this chapter has not been addressed by other studies in the literature yet. Table 4.1 summarizes the symbols used in this chapter.

4.2 SYSTEM MODEL

We target a point-to-point VLC network in which one LED-based AP[1] provides communication services to one user within the indoor environment. Throughout the chapter, it will become clear that the analytical framework provided here can be easily extended to a more general multi-AP and multi-user scenario. We consider a downlink scenario and assume that the AP is equipped with a data buffer. Particularly, the data initially arrives at the AP from a source (or sources) and is stored in the data buffer before being divided into packets and transmitted to the user in frames of T seconds. Throughout this chapter, we assume that the user is randomly located within the AP coverage area.

1 In this chapter, we use the terms "AP" and "transmitter" interchangeably.

Table 4.1: Table of symbols in this chapter

Symbol	Description
α_s	Source transmission probability from OFF state to ON state
β_s	Source transmission probability from ON state to OFF state
$\delta(\theta)$	Maximum average arrival rate of the considered system
$\delta_{ref}(\theta)$	Maximum average arrival rate of the reference system
ε	Delay violation probability
h	Channel gain
λ	Constant arrival rate during the source ON state
P	Average transmission power
$P_{o,c}$	Steady-state probability of the channel being in the ON state
$P_{o,s}$	Steady-state probability of the source being in the ON state
$P^{\star}_{o,c}$	Channel ON probability corresponding to the maximum average arrival rate
$P^{\star}_{o,s}$	Source ON probability corresponding to the maximum average arrival rate
R	Achievable rate in the channel
ρ	Fixed transmission rate
ρ_{min}	Minimum achievable transmission rate, i.e., at cell edge
ρ_{max}	Minimum achievable transmission rate, i.e., at cell center
$\rho^{\star}$	Optimal fixed-transmission rate
τ	Delay threshold
θ	QoS exponent
θ_t	Target QoS exponent

In this section, we provide a detailed description for the system under examination. Specifically, the VLC channel model is defined in Section 4.2.1, while the fixed-rate transmission process and its mathematical representation as a discrete-time Markov process in detailed in Section 4.2.2. The ON-OFF source model regarded in this chapter is introduced in Section 4.2.3.

4.2.1 *VLC Channel Model*

VLC channels normally contain both line-of-sight (LoS) and non-LOS parts. Nevertheless, it was observed in [156] that, in typical indoor environments the main received energy at the photodetector (more than 95%) comes from the LoS component. Subsequently, and without any loss of generality in this chapter we only consider the LoS path, as depicted in Figure 4.1. We further assume that the LED-based AP follows the Lambertian radiation pattern, and that the VLC AP is directed downwards and the user photodetector (PD)

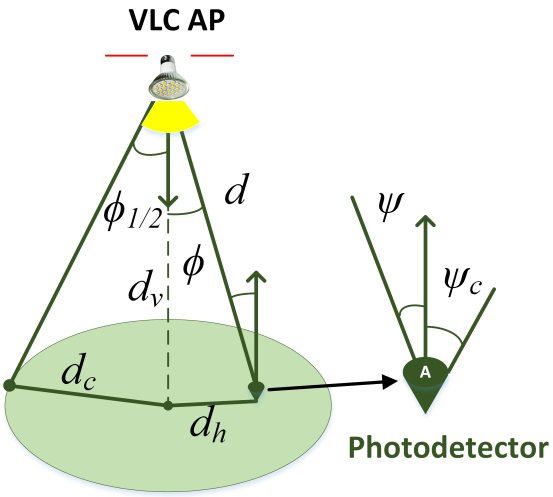


Figure 4.1: VLC channel via LoS link.

is directed upwards. Then, the LoS channel gain at a horizontal distance d_h from the cell center is expressed as [157]

$$h = \frac{(m+1)AL(\psi)g(\psi)d_v^{m+1}}{2\pi(d_v^2+d_h^2)^{\frac{m+3}{2}}}\text{rect}(\psi/\psi_C). \tag{4.1}$$

where d_v, A, and ψ are, respectively, the vertical distance, the PD physical area and the angle of incidence with respect to the normal axis to the receiver plane. In addition, $L(\psi)$ is the gain of the optical filter and

$$g(\psi)\frac{n^2}{\sin^2(\psi_C)}$$

is the optical concentrator gain at the receiver, where n is the refractive index and ψ_C is the field of view (FOV) angle of the receiver. In (4.1),

$$m = \frac{-1}{\log_2(\cos(\phi_{1/2}))}$$

is the Lambertian index, where $\phi_{1/2}$ is the LED half intensity viewing angle. Finally, rect(z) is an indicator function such that rect$(z) = 1$ if $z \leqslant 1$ and rect$(z) = 0$ otherwise.

Throughout this chapter, we consider the achievable rate for the VLC link of the form [158]

$$R = \frac{TB}{2}\log_2\left(1+\frac{(\mu\Omega Ph)^2}{\varkappa^2\sigma_n^2}\right) \quad \text{bits per frame} \tag{4.2}$$

for some constant μ. For example, setting $\mu = \sqrt{e/2\pi}$ defines the achievable rate when the transmitted light intensity is exponentially distributed [159]. Above, B is the available bandwidth of the VLC channel and Ω is the optical-to-electrical conversion efficiency of the PD. Further, P is the average

transmission power limit[2], σ_n^2 is the power of the zero-mean AWGN noise at the receiver, and $\varkappa$ is the ratio between the average optical power and the average electrical power of the transmitted signal. Setting $\varkappa = 3$ can guarantee a neglected clipping noise, thus the LED can be assumed to be operating in its linear region [24].

4.2.2 *Fixed-Rate Transmission*

In this chapter, we assume that the VLC AP has no prior information about the channel gain h, or equivalently the exact user location. Thus, the AP sends the data at the fixed rate of ρ bits/frame. Based on the known fact that the decoding error is negligible when transmitting at rates less than or equal to the instantaneous channel rate [160, Eq. (10)] [161], we assume that reliable communication is achieved when $\rho \leqslant R$ and thus the transmitted data is correctly decoded. On the other hand, we assume that an outage occurs when $\rho > R$, thus no reliable communication is guaranteed and data re-transmission is needed.

For an analytical presentation, we adopt a two-state ON-OFF Markov process to model the aforementioned VLC channel. Notice that a similar approach was also considered in many research studies to model several RF systems under different conditions and settings, see e.g., [67, 162]. Such an ON-OFF model can also be used to describe different physical characteristics of the channel, such as LoS blockage (see e.g., [163]) and inter-symbol interference.

Here, we assume that the channel is in the ON state if $\rho \leqslant R$, whereas the channel is in the OFF state if $\rho > R$. Notice that a number of ρ bits will be transmitted and successfully received, hence removed from the AP buffer, in one frame when the channel is in the ON state, while the transmission rate is effectively zero if the channel is on the OFF state. Here, the transition probability from the ON state to the OFF state is denoted by α_c, and the transition from the OFF state to the ON state is denoted by β_c. Thus, the state-transition matrix for this two-state Markov process is

$$J = \begin{bmatrix} 1-\alpha_c & \alpha_c \\ \beta_c & 1-\beta_c \end{bmatrix}.$$

Furthermore, let $p_{o,c}$ and $p_{f,c}$ be the steady-state probabilities of the channel being in the ON and OFF states, respectively, where $p_{o,c} + p_{f,c} = 1$. Then, the steady-state probabilities can be obtained by solving the following equation: $[p_{o,c}\ p_{f,c}] = [p_{o,c}\ p_{f,c}]J$. Subsequently, we can easily show that

$$p_{o,c} = \frac{\beta_c}{\alpha_c + \beta_c} = \Pr\{\rho \leqslant R\} = \Pr\{d_h \leqslant \Delta\} \tag{4.3}$$

and

$$p_{f,c} = \frac{\alpha_c}{\alpha_c + \beta_c} = \Pr\{\rho > R\} = 1 - \Pr\{d_h \leqslant \Delta\}, \tag{4.4}$$

2 Notice that a peak intensity constraint can also be imposed for practical and safety concerns. However, we ignore such a limit in this chapter for the sake of simplicity.

where

$$\Delta = \left(\left(\frac{(\mu\alpha P(m+1) A d_v^{m+1} g)^2}{(2\pi\sigma_n \varkappa)^2 (2^{2\rho/(TB)} - 1)} \right)^{\frac{1}{m+3}} - d_v^2 \right)^{\frac{1}{2}} \tag{4.5}$$

Notice that the value of Δ should satisfy $0 \leqslant \Delta \leqslant d_c$. For simplicity, and due to the random user distribution, we consider a block-channel model and assume that the channel status remains the same during the frame duration of T s, whereas the status changes independently from one frame to another. Formally, this yields that $\beta_c + \alpha_c = 1$ and hence we have $p_{o,c} = \beta_c$ and $p_{f,c} = \alpha_c$.

4.2.3 *Source Model*

Herein, we consider a two-state discrete-time Markov process with ON and OFF states to model the data source. Notice that we can utilize the ON-OFF source to describe certain practical data arrival processes. As for instance, the ON-OFF discrete model has been widely used to describe voice sources[164, 165]. Nevertheless, we emphasize that other source models can be easily integrated into our framework. We consider a constant data arrival rate of λ bits/frame in the ON state, whereas no bits arriving at the transmitter buffer in the OFF state. Herein, the transition probability from the ON state to the OFF state is denoted by α_s and the transition probability from the OFF state to the ON state is denoted by β_s. If $p_{o,s}$ denotes the steady-state probability of the data arrival process being in the ON state, then we have $p_{o,s} = \frac{\beta_s}{\alpha_s + \beta_s}$, and the average data arrival rate at the transmitter buffer is

$$r_{avg} = \lambda p_{o,s} = \lambda \frac{\beta_s}{\alpha_s + \beta_s}. \tag{4.6}$$

4.3 SYSTEM ANALYSIS

In this section, we explore the performance levels that the aforementioned system can achieve. Since the data is initially stored in the transmitter buffer before being transmitted, then applying certain constraints on the buffer length is required in order to control the buffering delay and overflow probabilities. Particularly, we target applying the statistical QoS constraints expressed in (1.3). Recall that (1.3) implies that the buffer violation probability should decay exponentially with a rate controlled by θ, which is also denoted as the QoS exponent. Here, larger θ implies stricter QoS constraints, whereas smaller θ corresponds to looser constraints.

4.3.1 *Maximum Average Arrival Rate*

Considering the ON-OFF source model described in Section 4.2.3, we initially aim to determine the maximum average arrival rate the can be supported by the VLC channel with the fixed-rate transmission policy, as

described in Section 4.2.2, while satisfying the QoS requirement in (1.3) for a given θ. To this end, we benefit from the fundamental condition expressed in (1.6), which states that the constraint in (1.3) is satisfied when we have $\Lambda_a(\theta) = -\Lambda_c(-\theta)$, where $\Lambda_a(\theta)$, and $\Lambda_c(\theta)$ are, respectively, the asymptotic log-moment generating functions of the total amount of bits arriving at the AP buffer and the total amount of bits served from the transmitter.

Recall that we assume a block-channel model such that the channel status changes independently from one transmission frame to another. Then, we can readily express the log-moment generating function of the service process for a given rate ρ as follows[3]:

$$\Lambda_c(-\theta) = \log_e\{p_{o,c}e^{-\theta\rho} + p_{f,c}\}, \tag{4.7}$$

where $p_{o,c}$ and $p_{f,c}$ are given in (4.3) and (4.4), respectively. Furthermore, we have

$$\Lambda_a(\theta) = \log_e\left\{\frac{1-\beta_s+(1-\alpha_s)e^{\theta\lambda}}{2} + \frac{\sqrt{[1-\beta_s+(1-\alpha_s)e^{\theta\lambda}]^2 - 4(1-\alpha_s-\beta_s)e^{\theta\lambda}}}{2}\right\}. \tag{4.8}$$

Now, using the condition $\Lambda_a(\theta) = -\Lambda_c(-\theta)$, we can formulate the maximum average arrival rate that the aforementioned VLC system can support as

$$\delta(\theta) = \frac{p_{o,s}}{\theta}\log_e\left\{\frac{1-(1-\beta_s)D}{(1-\alpha_s)D-(1-\alpha_s-\beta_s)D^2}\right\}, \tag{4.9}$$

where $D = p_{o,c}e^{-\theta\rho} + p_{f,c}$ and $p_{o,s} = \frac{\beta_s}{\alpha_s+\beta_s}$. The expression of $\delta(\theta)$ is obtained as follows. Given the log-moment generation functions in (4.7) and (4.8) we solve the equation $\Lambda_a(\theta) = -\Lambda_c(-\theta)$ with respect to λ as follows:

$$\lambda = \frac{1}{\theta}\log_e\left\{\frac{1-(1-\beta_s)D}{(1-\alpha_s)D-(1-\alpha_s-\beta_s)D^2}\right\}. \tag{4.10}$$

We then obtain $\delta(\theta)$ by substitution (4.10) in (4.6).

Remark 8 *In practice, the VLC system might be designed to support certain data transmission settings with a pre-defined QoS level, which defines the type and/or the quality of the supported service. As for example, in some places like airports and exhibition halls, users might be allowed to only download text files and/or audio files with a low quality. In such scenarios, the transmission fixed rate can be optimized to maximize the channel performance given the target QoS needs, which we denote as θ_t. It follows that, the channel log-moment generating function in (4.7) can be updated as*

$$\begin{aligned}\Lambda_c^\star(-\theta_t) &= \min_{\rho\geqslant 0}\log_e\{p_{o,c}e^{-\theta_t\rho} + p_{f,c}\}\\ &= \log_e\{p_{o,c}^\star e^{-\theta_t\rho^\star} + p_{f,c}^\star\}\end{aligned} \tag{4.11}$$

3 For more details about obtaining the log-moment generating functions, we refer to [86, Example 7.2.7].

where $\rho^\star$ is the optimal fixed-transmission rate and $p^\star_{o,c}$ and $p^\star_{f,c}$ are the corresponding channel ON and OFF probabilities, respectively. Notice that $p^\star_{o,c}$ and $p^\star_{f,c}$ also depend on the transmission rate ρ. Intuitively, we expect that the system designed to support a given θ_t will provide lower performance levels for any other $\theta > \theta_t$ since the system is subject to stricter QoS needs. To show this mathematically, we will consider the special case of a fixed-rate arrival process, i.e., $\beta_s = 1$ and $\alpha_s = 0$. In such a case, $\delta(\theta)$ in (4.9) reduces to

$$\delta(\theta) = \delta_{EC}(\theta) = -\frac{1}{\theta}\log_e\{p^\star_{o,c}e^{-\theta\rho^\star} + p^\star_{f,c}\}, \tag{4.12}$$

which is normally referred to as the effective capacity of the service process, and it defines the maximum constant arrival rate that the process can support for a given QoS exponent θ. It was shown in [155, Prop. 2] that the effective capacity is a non-increasing function of θ, which confirms the initial claim that the performance level either degrades or remains the same, depending on θ_t and $\rho^\star$, for $\theta > \theta_t$. Equivalently, designing the system for θ_t guarantees that services with QoS needs $\theta \leqslant \theta_t$ will be supported.

4.3.2 *Non-asymptotic Bounds*

Notice that the aforementioned analysis provides an asymptotic performance measure which assumes that the number of time frames is very large. Nevertheless, non-asymptotic bounds on the buffer overflow and delay violation probabilities at the transmitter are of interest from practical perspectives. Recall that Q and q define, respectively, the buffer length and threshold. For a given buffer overflow probability ε, i.e., $\Pr\{Q > q\} \leqslant \varepsilon$, [135, Theorem 2] states that a minimal bound on the queue length can be found as follows:

$$q = \inf_{c>0}\{q_c + q_a\}, \tag{4.13}$$

where

$$\begin{aligned} q_c &= -\sup_\theta\left\{\frac{\log_e\{-\varepsilon_c\left[\Lambda_c(-\theta)+\theta c\right]\}}{\theta}\right\} \\ &\text{for } \max\left\{0, -\frac{1}{c\varepsilon_c} - \frac{\Lambda_c(-\theta)}{c}\right\} < \theta, \end{aligned} \tag{4.14}$$

and

$$\begin{aligned} q_a &= -\sup_\theta\left\{\frac{\log_e\left\{\varepsilon_a\left[\theta c - \sup_{t>0}\{\Lambda_a(\theta,t)\}\right]\right\}}{\theta}\right\} \\ &\text{for } 0 < \theta < \frac{1}{c\varepsilon_a} + \frac{\sup_{t>0}\{\Lambda_a(\theta,t)\}}{c}, \end{aligned} \tag{4.15}$$

Above, the buffer violation probability is $\varepsilon = \varepsilon_c + \varepsilon_a$, and θ and c are free parameters. In (4.15), the time-variant log-moment generating function of the arrival process, $\Lambda_a(\theta,t)$, is given by [135, Eq. (21)]

$$\Lambda_a(\theta,t) = \frac{1}{t}\log_e\left\{\begin{bmatrix} p_{o,s} & p_{f,s}\end{bmatrix}\right.$$

Table 4.2: Simulation Parameters

LED half intensity viewing angle, $\phi_{1/2}$	60°
PD field of view (FOV), ψ_C	90°
PD physical area, A	$1\,cm^2$
Modulation bandwidth, B	40 MHz
PD opt.-to-elect. conversion efficiency, Ω	0.53 A/W
Refractive index, n	1.5
Optical filter gain, $L(\psi)$	1
Noise power spectral density, N_0	10^{-21} A^2/ Hz
Vertical distance, d_v	3 m
Transmission frame, T	1 ms

$$\times \left(\begin{bmatrix} (1-\alpha_s)e^{\theta\lambda} & \alpha_s e^{\theta\lambda} \\ \beta_s & 1-\beta_s \end{bmatrix} \right)^{(t-1)} \begin{bmatrix} e^{\theta\lambda} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Bigg\} \tag{4.16}$$

Next, we target the statistical bound regarding the queuing delay in the form $\Pr\{\mathcal{D} > \tau\} \leqslant \varepsilon$, where $\mathcal{D}$ is the buffering delay and τ is the delay threshold. To this end, let us assume that first-come first-served protocol exists at the transmitter buffer. Thus, the minimal bound on the buffering delay can be expressed as follows [135, Theorem 1]:

$$\tau = \inf_{c>0} \left\{ \frac{q_c + q_a}{c} \right\} \tag{4.17}$$

4.4 NUMERICAL RESULTS

In this section, we present the numerical results. While the analytical results in the above sections are valid for any user distribution, herein we show the results assuming that the user is uniformly located within the VLC cell area. Formally, let us define the probability density function of the horizontal distance, d_h as $f_{d_h}(h) = \frac{2h}{d_c^2}$. Subsequently, the cumulative distribution function (CDF) of d_h can be obtained as

$$F_{d_h}(\Delta) = \Pr\{d_h \leqslant \Delta\} = \frac{\Delta^2}{d_c^2}, \tag{4.18}$$

where $\Delta \leqslant d_c$. Then, we can calculate the channel ON and OFF probabilities, i.e., $p_{o,c}$ in (4.3) and $p_{f,c}$ in (4.4), respectively. Throughout this chapter, we denote by ρ_{min} and ρ_{max} the minimum and the maximum achievable rate over the VLC cell, respectively, and we set the supported range of ρ as $\rho_{min} \leqslant \rho < \rho_{max}$[4]. Obviously, ρ_{min} is the rate achieved when the user is

4 Notice that we cannot set $\rho \geqslant \rho_{max}$ since the channel will always experience an outage, unless the user is located at the cell center and when $\rho = \rho_{max}$.

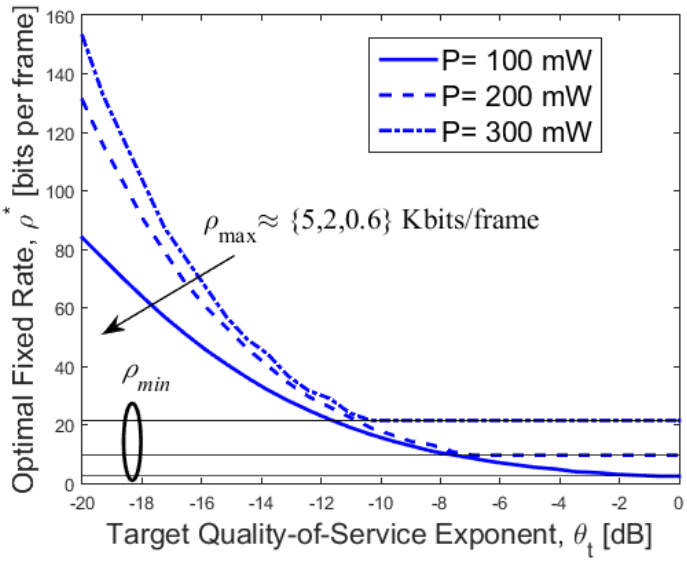


Figure 4.2: Optimal fixed-transmission rate as a function of the target QoS exponent, θ_t, and for different transmission power values.

located at the cell edge, while ρ_{max} is the achievable rate when the user is located at the cell center. Finally, the thermal noise power at the photodiode is $\sigma_n^2 = N_0 B$, where N_0 is the noise power spectral density. Table 4.2 summarizes the parameters used in this section.

In Fig. 4.2, we initially display the optimal fixed-transmission rate, following (4.11), as a function of the target QoS needs, θ_t, and for different transmission power levels, P. We clearly see that the optimal rate rapidly decreases with increasing θ_t to a certain value of θ_t, after which the optimal rate saturates at a level that equals to the minimum achievable rate, i.e., $\rho^\star = \rho_{min}$. This behavior can be explained since the system randomness, either in the source or in the channel, is a key factor that limits the system ability to satisfy strict QoS requirement. Thus, stricter QoS needs can be satisfied only by reducing the randomness in the channel activity. In the VLC scenario considered in this chapter, the channel randomness can be eliminated by setting $\rho^\star = \rho_{min}$, since the channel stays in the ON state regardless of the user location.

We further observe that increasing the transmission power results in higher possible fixed rates. To better understand Fig. 4.2, we remark the following. Noting that increasing the transmission rate, generally, results in a higher outage probability, thus reducing the system ability to satisfy stricter QoS needs, Fig. 4.2 reveals that increasing the transmission power allows transmission at higher rates, and yet satisfying the QoS needs. However, the allowed rates are still far lower than the maximum achievable rate over the cell, i.e., ρ_{max}. Notice that the influence of the transmission power on the system performance has a practical significance since the power can be related to the number of served users in multi-user scenarios, in which the transmission resources are divided among the users. Also, the transmission power shows the impact of light dimming.

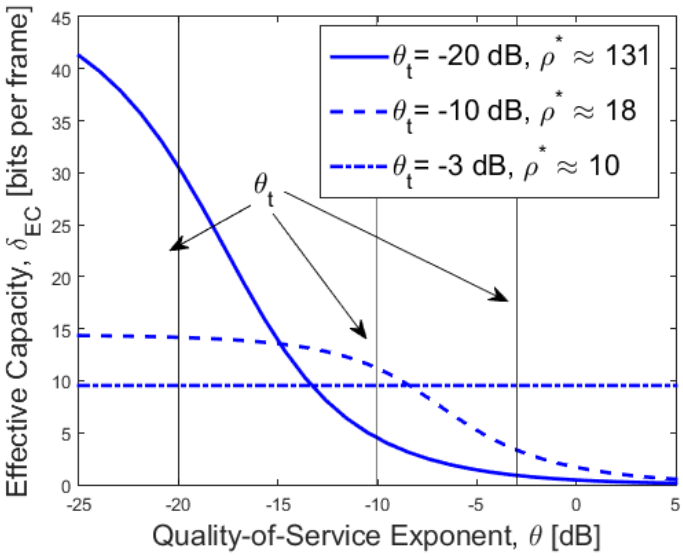


Figure 4.3: Effective capacity as a function of θ and for different target QoS needs, θ_t. Here, $P = 200$ mW.

To explore the impact of designing the system to support given QoS needs, in Fig. 4.3 we plot the effective capacity, given in (4.12), as a function of the QoS exponent, θ, and for different values of the target exponent θ_t. We immediately observe that the performance curves either degrade or remain the same with increasing θ, which confirm the conclusion explained in Remark 8. Specifically, we observe that the performance gain for $\theta < \theta_t$ decreases for larger values of θ_t. This is since the optimal transmission rate increases by decreasing θ_t, thus the exponential decay rate, controlled by the product $-\theta\rho^\star$, increases. As for instance, for $\theta_t = -3$ dB, the performance curve is flat with θ, since the optimal transmission rate is equal to the minimum rate, i.e., $\rho^\star = \rho_{min}$, and then the effective capacity reduces to $\delta_{EC}(\theta) = \rho_{min}$. This behavior also agrees with the results shown in Fig. 4.2.

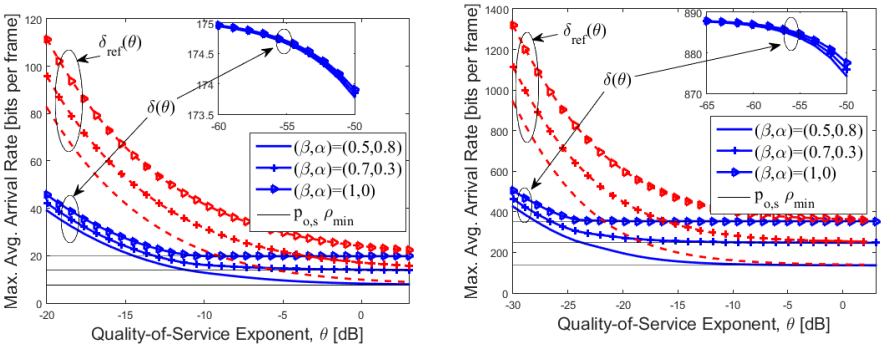
In Fig. 4.4, we plot the maximum average arrival rate at the transmitter buffer as a function of the QoS exponent[5], θ, and for different source statistics. We further compare the results with those obtained with a reference scenario in which the AP has a perfect knowledge of the channel gain. Thus, the AP can send the data with a rate equals to the channel achievable rate, i.e., $\rho = R$, where R is expressed in (4.2) for a given horizontal distance d_h. In such a scenario, and assuming a random user distribution within the cell coverage area, the channel log-moment generating function for a give θ can be expressed as follows:

$$\Lambda_c(-\theta) = \log_e \left\{ \mathbb{E}_{d_h}\{e^{-\theta R}\} \right\}, \tag{4.19}$$

and the maximum average arrival rate, denoted as $\delta_{ref}(\theta)$, has the same expression as in (4.9) with $D = \mathbb{E}_{d_h}\{e^{-\theta R}\}$.

In Fig. 4.4, we immediately observe that the supported average rate decreases with increasing θ. We further notice that learning the channel

5 Int his figure, we simply set $\theta_t = \theta$.



(a) $\phi_{1/2} = 60°$ (b) $\phi_{1/2} = 45°$

Figure 4.4: Maximum average arrival rate considering as a function of the QoS exponent, θ, and for different source statistics. Here, $P = 200$ mW, $\beta = \beta_s$ and $\alpha = \alpha_s$.

gain has a positive impact on the system performance at lower QoS levels, whereas the performance curves of both scenarios approach the same level, for given source statistics, as $\theta \to \infty$. This observation also means that, while the reference scenario outperforms the fixed-rate scenario at lower θ values, the reference scenario is affected more by increasing θ, as the decreasing rate of the performance curves is faster with increasing θ. This can be explained due to the higher level of the transmission randomness in the reference scenario. Then, we can conclude that transmitting with a fixed rate is preferred when stricter QoS needs are required since it is simpler to implement. We further notice the following two observations. First, the average arrival rate is independent of the source statistics as the QoS exponent diminishes, i.e., as $\theta \to 0$. This is sensible since the average arrival rate is expected to approach the average transmission rate in the channel as θ goes to zero.

Mathematically, we can easily prove that $\delta(\theta)$ in (4.9) simplifies to $\delta(\theta) = p^{\star}_{o,c}\rho^{\star}$ as $\theta \to 0$. Second, the performance curves saturation at certain levels as θ increases agrees with the results shown in Fig. 4.2, that the optimal transmission rate approaches ρ_{min} as $\theta \to \infty$. Subsequently, we can easily show that $\delta(\theta)$ in (4.9) reduces to $\delta(\theta) = p_{o,s}\rho^{\star} = p_{o,s}\rho_{min}$ as $\theta \to \infty$, where $p_{o,s} = \frac{\beta_s}{\beta_s+\alpha_s}$. From these two observations, we can conclude that the average arrival rate depends only on the service (channel) dynamics as $\theta \to 0$, whereas the average arrival rate depends mainly on the arrival (source) dynamics as $\theta \to \infty$. We finally notice that decreasing the cell area, i.e., by decreasing the transmission viewing angle $\phi_{1/2}$, improves the system performance as higher transmission rates can be supported and the channel randomness due to the user random location decreases.

Finally, in Fig. 4.5 we illustrate the delay bounds achieved by the VLC system as a function of the average arrival rates. We set $\alpha_s = 0.3$ and $\beta_s = 0.7$ and we consider different transmission power values. For each value of the transmission power, we find the value of θ that minimizes the channel-related buffer overflow probability, following (5.32). We then compute the

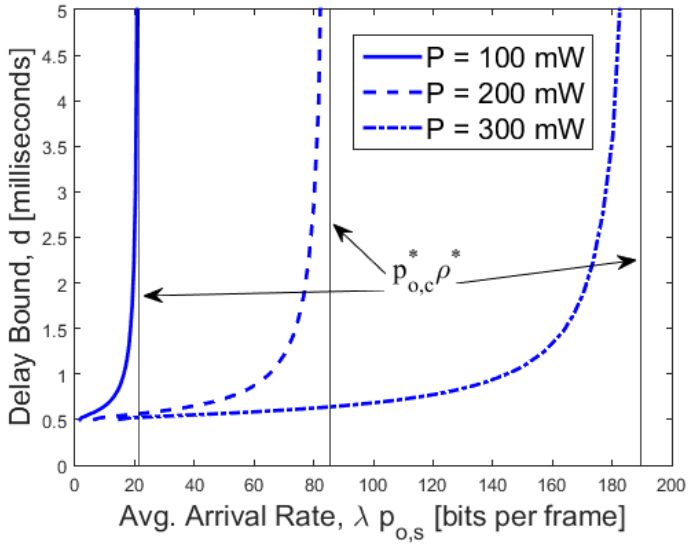


Figure 4.5: Delay bounds as a function of the average arrival rate and considering different power levels. Here, $\alpha_s = 0.3$ and $\beta_s = 0.7$.

average transmission rate that corresponds to the optimal channel log-moment generating function at that θ level, i.e., $p^{\star}_{o,c}\rho^{\star}$. The vertical lines in Fig. 4.5 show that the delay bounds increase asymptotically as the average arrival rate approaches the average transmission rate in the channel, since the system becomes unstable when the average arrival rate is greater than the average service rate in the channel, and long buffering periods are expected. We further observe the effect of increasing the transmission power on improving the delay performance.

5

HYBRID RF/VLC SYSTEMS UNDER STATISTICAL QUEUEING CONSTRAINTS

As detailed in the previous chapter, VLC systems have unique advantages over the counterpart RF systems. However, VLC systems have several challenges and limitations that should be taken into account, such as smaller coverage areas and strong dependence on line-of-sight. Therefore, VLC can be seen as a complement, rather than an alternative, technology to RF. In other words, integrating RF and VLC technologies can form strong complementary hybrid networks that can achieve diverse end-user demands, e.g., capacity and coverage, which are difficult to obtain when either technology is operating solely. Therefore, the deployment of such hybrid networks has been proposed and intensively investigated in the recent years to enhance network performances and to address specific quality of service (QoS) constraints. Notice that such networks can be employed in practice with negligible installation costs as RF and VLC systems already coexist in most of the indoor environment, such as offices and factories.

In this chapter, we explore the benefits of employing both technologies when the QoS requirements are imposed as limits on the buffer overflow and delay violation probabilities, which are important metrics in designing low latency wireless networks. As we mainly focus on developing resource management schemes, in this chapter we regard the access technology as the primary resource at the physical layer. Particularly, we consider a multi-mechanism scenario that utilizes RF and VLC links for data transmission in an indoor environment, and then propose a link selection process through which the transmitter sends data over the link that sustains the desired QoS guarantees the most.

Considering an ON-OFF data source, we employ the maximum average data arrival rate at the transmitter buffer and the non-asymptotic bounds on data buffering delay as the main performance measures. We formulate the performance measures under the assumption that both links are subject to average and peak power constraints. Furthermore, we investigate the performance levels when either one of the two links is used for data transmission, or when both are used simultaneously. Finally, we show the impacts of different physical layer parameters on the system performance through numerical analysis.

5.1 INTRODUCTION

The ever-growing demand for mobile communications triggered a quest for technical solutions that will support stringent quality of service (QoS) constraints. Thanks to the significant advances in white light emitting diodes (LEDs) research, and the availability of an extensive unregulated spectrum,

visible light communication (VLC) has emerged as a promising technology. We can utilize LEDs simultaneously for data transmission and illumination, since they have many unique aspects compared to the other communication technologies [62]. Moreover, we can improve data security because light does not penetrate the surrounding walls. We can also sustain an all-important *green* agenda and minimize the installation costs because we do not require an extensive infrastructure. Nevertheless, attention must be paid to certain limitations and challenges in VLC systems, e.g., smaller coverage, strong dependence on line-of-sight components and achievable rates that vary with spatial fluctuations [21]. In order to overcome these constraints, researchers proposed hybrid RF/VLC systems [22, 24, 62, 64, 163, 166–172], where end users can benefit from the wide coverage area that RF systems support and the stable rates that VLC systems provide. Such networks are practically feasible as RF and VLC systems can coexist without causing interference on each other and operate in the same environment, such as offices and rooms.

Comparing hybrid RF/VLC systems with systems that employ either RF or VLC only, the authors in [22, 62, 166, 167, 173] demonstrated a remarkable increase in data transmission throughput, energy efficiency and delay performance in hybrid RF/VLC systems. Moreover, the authors in [168, 169] projected a hybrid system in which they use VLC links for down-link communication and RF links for up-link communication. In such a system, the authors in [163, 170, 171] and the ones in [24, 63, 64] investigated handover and load balancing mechanisms, respectively. Alternatively, considering an outdoor environment, the authors in [23] studied a point-to-point transmission scenario in which the system can switch between RF links and VLC links after comparing the signal-to-noise ratio levels in each link. Regarding the same system setting, the authors in [174] assumed that both RF and VLC links have the same transmission rates, and then proposed a diversity-based transmission scheme such that the transmitter sends data by employing both links simultaneously.

The aforementioned studies analyzed the hybrid RF/VLC systems mostly from the physical layer perspective, i.e., they did not concentrate on the data link layer metrics, such as limits on the buffer overflow and buffering delay probabilities, as much as needed. However, recall that we also need QoS metrics that can be a cross-layer analysis tool between the physical layer features and the performance levels in data-link layer due to the dramatic increase in the demand for reliable delay-sensitive services in recent years. To the best of our knowledge, there are only a few studies that investigated cross-layer performance levels in VLC systems. For example, the authors employed effective capacity as a performance measure in resource allocation schemes in VLC systems [71] and heterogeneous networks composed of VLC and RF links [72]. Here, we note that the authors in [72] concentrated on the case of constant data arrival rates at the transmitter buffer, which is not realistic in certain practical settings. For more details in effective capacity, we refer to [81].

In this chapter, we perform a cross-layer analysis of a hybrid RF/VLC system in which the transmitter can use both RF and VLC channels for

data transmission. We investigate the performance gains achieved by a hybrid RF/VLC system when it operates under statistical QoS constraints, which are inflicted as limits on the buffer overflow and delay violation probabilities. Assuming an ON-OFF modeled data arrival process at the transmitter buffer, we employ first the maximum average data arrival rate at the transmitter buffer considering the asymptotic buffer overflow probability approximation, and then non-asymptotic buffering delay violation probability as the main performance measures.

In this chapter, we propose a mathematical toolbox to system designers for performance analysis in hybrid RF/VLC systems that work under low latency conditions. Particularly, we provide a rudimentary model for multi-mechanisms in communication systems that operate under QoS constraints. We employ RF and VLC links as two different mechanisms. Our model can be easily invoked in settings with more than two different mechanisms as well. The reason behind multi-mechanisms in communications is to boost performance levels through the increased degree of freedom. Therefore, in order to introduce our model smoothly and make it easier for readers to understand our objective, we also benefit from the existing literature in RF and VLC studies. However, to the best of our knowledge, the analytical framework provided in this chapter, in which we investigate the QoS performance, is not addressed in other studies. One aspect of this hybrid system is that VLC links provide time-invariant transmission rates, while RF links provide rates that vary over time. A communication setting that depends solely on an RF link may suffer low transmission rates, and have longer data backlogs in the transmitter buffer. However, a communication setting that can utilize both RF and VLC links, for instance, can take advantage of the constant transmission rate in the VLC link when the transmission rate in the RF link falls below a certain level. Table 5.1 summarizes the symbols used in this chapter.

5.2 SYSTEM MODEL

We consider a network access controller that provides a connection to a user[1], through either an RF access point or a VLC access point, which are positioned at different locations in an indoor environment as seen in Figure 5.1. Herein, we assume a down-link scenario, i.e., the network access controller acts as a transmitter and the user acts as a receiver. In the sequel, we use *transmitter* and *receiver* instead of *network access controller* and *user*, respectively. Initially, the transmitter receives data from a source (or sources) and stores it as packets in its buffer. Subsequently, it sends the data packets in frames of T seconds to the receiver following a given transmission strategy. The receiver is considered to be equipped with an RF front-end and a photo-diode. We also note that the VLC coverage area is generally smaller than the RF coverage area, and that they overlap[2]. Finally,

1 The analytical framework provided in this chapter can easily be extended to a multi-user scenario. For more details, we refer to Remark 12 in Section III.

2 This model is also considered in [163].

Table 5.1: Table of symbols in this chapter

Symbol	Description
α	Source transition probability from the ON state to the OFF state
β	Source transition probability from the OFF state to the ON state
B_r	RF channel Bandwidth
B_v	VLC channel Bandwidth
d	Delay threshold
d_0	Distance between the user and the RF access point
d_1	Distance between the user and the VLC access point
g	VLC channel gain
h	Fading coefficient of the RF channel
λ	Fixed arrival rate during the source ON state
n	Number of sub-frames with respect to the handover time i.e., $T = n \times T_H$
ν	Average-to-peak power ratio
$P_{avg,r}$	Average power constraint in the RF system
$P_{avg,v}$	Average power constraint in the VLC system
P_{ON}	Steady-state probability of the source being in the ON state
P_{OFF}	Steady-state probability of the source being in the OFF state
$P_{peak,r}$	Peak power constraint in the RF system
$P_{peak,v}$	Peak power constraint in the VLC system
$\rho_r(\theta)$	Maximum average arrival rate in the RF channel
$\rho_v(\theta)$	Maximum average arrival rate in the VLC channel
$\rho_{rv}(\theta)$	Maximum average arrival rate due to the Hybrid-type I strategy
$\rho_{srv}(\theta)$	Maximum average arrival rate due to the Hybrid-type II strategy
R_l	Achievable rate of the RF link
σ_r^2	Noise variance in the RF system
σ_v^2	Noise variance in the VLC system
T	Frame duration
T_H	Handover time
θ	QoS exponent
V	Achievable rate of the VLC link
ε	Delay violation probability

we consider a power-limited system and assume that the network controller is constrained by a fixed average power budget, denoted as P_{avg}, for data

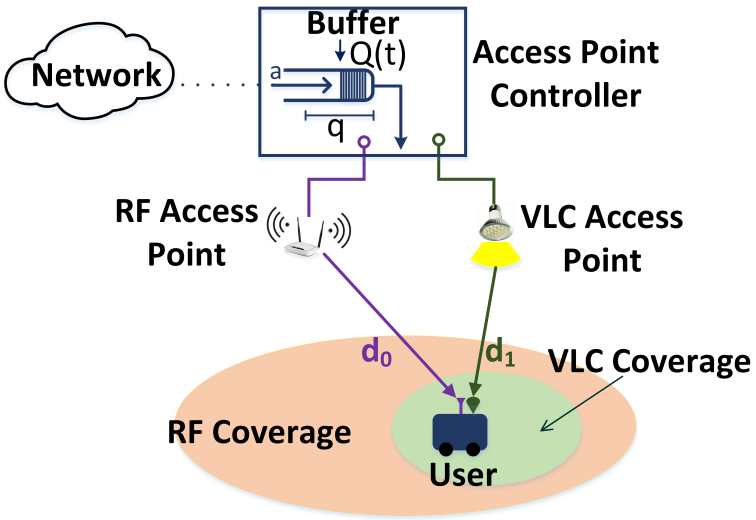


Figure 5.1: Hybrid RF/VLC system.

transmission. In the sequel, we initially introduce the RF and VLC channels, and then describe the data source model.

5.2.1 *RF Channel Model*

During the data transmission in the flat-fading RF channel, the input-output relation at time instant t is expressed as

$$y_r(t) = x_r(t)h(t) + w_r(t), \tag{5.1}$$

where $x_r(t)$ and $y_r(t)$ are the complex channel input and output at the RF access point of the transmitter and the RF front-end of the receiver, respectively. The complex channel input is subject to an average power constraint, $P_{avg,r}$, i.e., $\mathbb{E}\{|x_r(t)|^2\} \leqslant P_{avg,r}$, and a peak power constraint, $P_{peak,r}$, i.e., $|x_r(t)|^2 \leqslant P_{peak,r}$. Above, $h(t)$ is the complex channel fading gain with an arbitrary distribution having a finite average power, i.e., $\mathbb{E}\{|h(t)|^2\} < \infty$. Furthermore, we consider a block-fading channel and assume that the fading gain stays constant during one transmission frame (T seconds), i.e., $h(lT) = h(lT + \frac{1}{B_r}) = \cdots = h((l+1)T - \frac{1}{B_r}) = h_l$, where the available bandwidth is B_r Hz in the channel and h_l is the channel fading gain in the l^{th} time frame.

Note that there are TB_r symbols transmitted in one time frame. Moreover, the fading gain changes independently from one frame to another. Meanwhile, $w_r(t)$ is the additive noise at the RF front-end of the receiver, which is a zero-mean, circularly symmetric complex Gaussian random variable with variance σ_r^2, i.e., $\mathbb{E}\{|w_r(t)|^2\} = \sigma_r^2 < \infty$. The noise samples $\{w_r(t)\}$ are assumed to be independent and identically distributed.

We assume that a reliable data transmission exists as long as the transmission rate in the channel is lower than or equal to the instantaneous mutual

information between the channel input and output[3]. In particular, when the transmission rate in the l^{th} time frame (i.e., R_l bits per frame) is lower than or equal to the instantaneous mutual information (i.e., $C_{r,l}$ bits per frame) between the channel input $[x_r(lT), x_r(lT + \frac{1}{B_r}), \cdots, x_r((l+1)T - \frac{1}{B_r})]$ and the channel output $[y_r(lT), y_r(lT + \frac{1}{B_r}), \cdots, y_r((l+1)T - \frac{1}{B_r})]$, a reliable data transmission occurs and R_l bits are decoded correctly by the receiver. Here, we assume that TB_r is large enough so that the decoding error probability is negligible when $R_l \leqslant C_{r,l}$.

In [175], a lower bound on the maximum mutual information (or channel capacity) is provided, where the input has a two-dimensional circularly truncated Gaussian distribution. Therefore, we set the instantaneous data transmission rate to the lower bound and assume that the input has a two-dimensional circularly truncated Gaussian distribution. Specifically, we have a reliable transmission in the l^{th} time frame when

$$R_l = TB_r \log_2 \left\{ 1 + \frac{2|h_l|^2}{a\sigma_r^2} \exp\left\{ \frac{bP_{avg,r}}{2} - 1 \right\} \right\} \leqslant C_{r,l} \tag{5.2}$$

bits per frame, where a and b are the solutions of the following equations:

$$\frac{a}{b}\left[1 - \exp\left\{-\frac{bP_{peak,r}}{2}\right\}\right] = 1, \tag{5.2a}$$

and

$$2\frac{a}{b}(bP_{peak,r})^{-1}\left[1 - \exp\left\{-\frac{bP_{peak,r}}{2}\right\}\left(1 + \frac{bP_{peak,r}}{2}\right)\right] = \frac{P_{avg,r}}{P_{peak,r}}. \tag{5.2b}$$

Above, $C_{r,l}$ is time-dependent and changes from one time frame to another because the maximum instantaneous mutual information in each frame is a function of the channel fading gain.

5.2.2 *VLC Channel Model*

We assume that the transmitter employs *intensity modulation/direct detection*. Principally, the VLC access point of the transmitter is equipped with an LED and the data is modulated on the intensity of the emitted light. The receiver that collects light using a photo-diode generates an electrical current or voltage proportional to the intensity of the received light. Besides, we know that VLC channels are typically composed of line-of-sight as well as multi-path components.

However, the majority of the collected energy at photo-diodes (more than 95%) comes from the line-of-sight components in typical indoor scenarios [156]. Therefore, we can assume that the VLC channel is flat with a

3 This assumption is based on the known result in the literature that transmitting data at rates less than or equal to the instantaneous mutual information has a high reliability, thus the decoding error is negligible [160, Eq. (10)] [161].

dominant line-of-sight component [176–178], and the channel gain does not vary during the data transmission as long as the receiver is stationary[4]. Accordingly, the input-output relation in the VLC channel between the VLC access point of the transmitter and the photo-diode of the receiver at time instant t is given as follows:

$$y_\nu(t) = \Omega x_\nu(t) g + w_\nu(t), \tag{5.3}$$

where $x_\nu(t)$ and $y_\nu(t)$ are the real-valued channel input and output, respectively. Above, Ω is the optical-to-electrical conversion efficiency (or detector responsivity) of the photo-diode in amperes per watt and $w_\nu(t)$ is the additive noise at the photo-diode of the receiver, which is a zero-mean, real Gaussian random variable with variance σ_ν^2, i.e., $\mathbb{E}\{w_\nu^2\} = \sigma_\nu^2$. The noise samples $\{w_\nu\}$ are independent and identically distributed. Moreover, g is the time-invariant optical channel gain.

Recall that the data transmitted over the VLC link is modulated on the light that illuminates the environment. Hence, assuming that the operation range of the radiated optical power is limited between P_{min} and P_{max} when the light is on, we modulate the data between the power levels P_{min} and P_{max}, i.e., $P_{min} \leqslant x_\nu(t) \leqslant P_{max}$. As a result, the data bearing symbol, $\tilde{x}_\nu = x_\nu(t) - P_{min}$ is limited as follows:

$$\tilde{x}_\nu \leqslant P_{max} - P_{min} = P_{peak,\nu}.$$

Hence, we can re-express the input-output relation in (5.3) as

$$\tilde{y}_\nu(t) = y_\nu(t) - \Omega P_{min} g = \Omega \tilde{x}_\nu(t) g + w_\nu(t). \tag{5.4}$$

Now, assuming that the expected value of $\tilde{x}_\nu(t)$ is bounded by $P_{avg,\nu}$, i.e., $\mathbb{E}\{\tilde{x}_\nu(t)\} \leqslant P_{avg,\nu}$, and that the available bandwidth in the optical channel is B_ν Hz, we set the transmission rate in the channel in bits per frame to the lower bound on the channel capacity, which is defined as follows [159]:

$$V = \frac{TB_\nu}{2} \log_2 \left\{ 1 + P_{peak,\nu}^2 \frac{\Omega^2 g^2}{2\pi\sigma_\nu^2} \times \exp\left\{ 2\frac{P_{avg,\nu}}{P_{peak,\nu}} \mu^\star - 1 \right\} \left(\frac{1 - e^{-\mu^\star}}{\mu^\star} \right)^2 \right\} \leqslant C_\nu, \tag{5.5}$$

when $0 \leqslant \frac{P_{avg,\nu}}{P_{peak,\nu}} < \frac{1}{2}$, and

$$V = \frac{TB_\nu}{2} \log_2 \left\{ 1 + P_{peak,\nu}^2 \frac{\omega^2 g^2}{2\pi\sigma_\nu^2} e^{-1} \right\} \leqslant C_\nu, \tag{5.6}$$

when $\frac{1}{2} \leqslant \frac{P_{avg,\nu}}{P_{peak,\nu}} \leqslant 1$, where $\mu^\star$ is the unique solution to

$$\frac{P_{avg,\nu}}{P_{peak,\nu}} = \frac{1}{\mu} - \frac{e^{-\mu}}{1 - e^{-\mu}}. \tag{5.7}$$

4 Small-scale variations in VLC channels (i.e., fading) is mitigated since the area of a photo-diode is much larger than the light wavelength [179, Sec. 2.5]. Thus, VLC channels are known as time-invariant. This fact is almost true regardless of the frame duration or the user being stationary or mobile in indoor environments, because the users are either stationary or move very slowly.

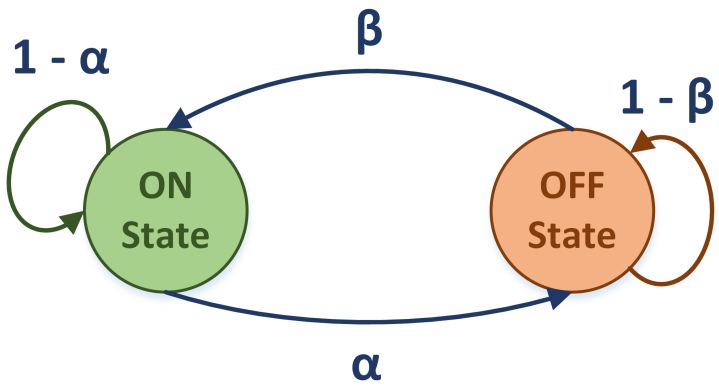


Figure 5.2: State transition model of the data arrival process.

Above, C_v and V are constant values, because the mutual information in the VLC link does not change by time, i.e., due to the strong line-of-sight channel component, the channel gain, g, does not change. As for the input distribution, we refer to [159, Eq. 42].

5.2.3 *Source Model*

Similar to Chapter 4, in this chapter we also consider a two-state discrete-time Markov process with ON and OFF states in each time frame to model the arrival process. Recall that we can project certain data arrival models on ON-OFF Markov processes. For instance, voice sources are generally modeled with ON and OFF states [164]. However, we stress that the analytical framework provided in this chapter can easily be extended to other source models.

Herein, when the source is in the ON state in one time frame, the data from a source (or sources) arrives at the transmitter buffer. In the ON state, we consider a constant data arrival rate, i.e., λ bits per frame. The number of bits arriving at the transmitter buffer is zero in the OFF state. As shown in Fig. 5.2, the transition probability from the ON state to the OFF state is denoted by α and the transition from the OFF state to the ON state is denoted by β. The probability of staying in the ON state is $1-\alpha$ and the probability of staying in the OFF state is $1-\beta$. Hence, the state transition matrix becomes

$$J = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}. \tag{5.8}$$

Now, let p_{ON} and p_{OFF} be the steady-state probabilities of the data arrival process being in the ON and OFF states, respectively, where $p_{ON} + p_{OFF} = 1$. Then, we have the following equality: $[p_{ON}\ p_{OFF}] = [p_{ON}\ p_{OFF}]J$. Subsequently, we have $p_{ON} = \frac{\beta}{\alpha+\beta}$ and $p_{OFF} = \frac{\alpha}{\alpha+\beta}$, and hence, the average data arrival rate at the transmitter buffer is $\frac{\beta\lambda}{\alpha+\beta}$ bits per frame.

5.3 PERFORMANCE ANALYSIS

In this section, we investigate the performance levels that the aforementioned system achieves by opportunistically exploiting the RF and VLC channels for data transmission. Herein, because the data is initially stored in the transmitter buffer before transmission, we assume that certain constraints are applied on the amount of the data in the buffer and the buffering delays. Therefore, we examine the system under QoS constraints that are associated with buffer overflow (data backlog) and buffering delay as explained in Section 1.3. However, for the presentation completeness, in the following we repeat the main expressions and notations of Section 1.3. Herein, we express the decay rate of the tail distribution of the queue length as [80, Eq. (63)]

$$\theta = -\lim_{q\to\infty} \frac{\log_e \Pr\{Q \geqslant q\}}{q}, \tag{5.9}$$

where Q is the stationary queue length, $Q(t)$ (see Fig. 5.1), and q is the buffer overflow threshold. Above, $\theta > 0$ denotes the decay rate of the tail distribution of the data backlog, Q. Accordingly, we can approximate the buffer overflow probability for a large threshold, q_{max}, as $\Pr\{Q \geqslant q_{max}\} \approx e^{-\theta q_{max}}$. Notice that the buffer overflow probability decays exponentially with a rate controlled by θ, which is also defined as the QoS exponent. Basically, larger θ implies stricter QoS constraints, whereas smaller θ corresponds to looser constraints.

Recall that the outgoing service from the transmitter queue is R_l, given in (5.2), when the data is sent through the RF channel and it is V, given in (5.5) and (5.6), when the data is sent through the VLC channel, while the data arrival is ON-OFF Markov process with λ bits per frame in the ON state and zero bits in the OFF state. Hence, assuming that the buffer size is infinite and considering independent data arrival and work-conserving data service processes, the constraint in (5.9) can be satisfied only when $\Lambda_a(\theta) = -\Lambda_r(-\theta)$ if the RF link is utilized and when $\Lambda_a(\theta) = -\Lambda_v(-\theta)$ if the VLC link is used, respectively, where $\Lambda_a(\theta)$, $\Lambda_r(\theta)$ and $\Lambda_v(\theta)$ are the asymptotic log-moment generating functions of the total amount of bits arriving at the transmitter buffer, the total service from the transmitter in the RF channel and the total service from the transmitter in the VLC channel, respectively [85, Theorem 2.1]. In particular, the asymptotic log-moment generating functions for any θ are

$$\Lambda_a(\theta) = \log_e \left\{ \frac{1-\beta+(1-\alpha)e^{\theta\lambda}}{2} + \frac{\sqrt{[1-\beta+(1-\alpha)e^{\theta\lambda}]^2 - 4(1-\alpha-\beta)e^{\theta\lambda}}}{2} \right\}, \tag{5.10}$$

$$\Lambda_r(\theta) = \log_e \left\{ \mathbb{E}_h \left\{ e^{\theta R_l} \right\} \right\} \text{ and } \Lambda_v(\theta) = \theta V. \tag{5.11}$$

We refer to [86, Example 7.2.7] for obtaining the log-moment generating functions. Using $\Lambda_a(\theta) = -\Lambda_r(-\theta)$, we can express the maximum average

data arrival rate at the transmitter buffer that the service process in the RF channel can sustain for any $\theta > 0$ as

$$\begin{aligned}\rho_r(\theta) &= \frac{\beta}{(\alpha+\beta)\theta}\log_e\left\{\frac{e^{-2\Lambda_r(-\theta)}-(1-\beta)e^{-\Lambda_r(-\theta)}}{(1-\alpha)e^{-\Lambda_r(-\theta)}-(1-\alpha-\beta)}\right\}\\ &= \frac{\beta}{(\alpha+\beta)\theta}\log_e\left\{\frac{1-(1-\beta)D}{(1-\alpha)D-(1-\alpha-\beta)D^2}\right\},\end{aligned} \tag{5.12}$$

where $D = \mathbb{E}_h\left\{e^{-\theta R_l}\right\}$ and R_l is given in (5.2). For the derivation of $\rho_r(\theta)$, we refer to Appendix G. Likewise, using $\Lambda_a(\theta) = -\Lambda_v(-\theta)$ and following the steps in Appendix G, we can also express the maximum average data arrival rate at the transmitter buffer that the service process in the VLC channel can sustain for any $\theta > 0$ as

$$\begin{aligned}\rho_v(\theta) &= \frac{\beta}{(\alpha+\beta)\theta}\log_e\left\{\frac{e^{-2\Lambda_v(-\theta)}-(1-\beta)e^{-\Lambda_v(-\theta)}}{(1-\alpha)e^{-\Lambda_v(-\theta)}-(1-\alpha-\beta)}\right\}\\ &= \frac{\beta}{(\alpha+\beta)\theta}\log_e\left\{\frac{e^{2\theta V}-(1-\beta)e^{\theta V}}{(1-\alpha)e^{\theta V}-(1-\alpha-\beta)}\right\},\end{aligned} \tag{5.13}$$

where V is given in (5.5) and (5.6) accordingly with the relation between P_{avg} and P_{peak}. Moreover, in the special case where $\alpha = 0$ and $\beta = 1$, the expressions in (5.12) and (5.13) provide the effective capacity, which is the maximum sustainable constant data arrival rate by the channel process given the QoS constraints [81]. In another special case where $\alpha + \beta = 1$, i.e., the state-transitions are independent of the past and current states, we have

$$\begin{aligned}&\rho_r(\theta) = \frac{\beta}{\theta}\log_e\left\{\frac{1-\alpha\mathbb{E}_h\left\{e^{-\theta R_l}\right\}}{\beta\mathbb{E}_h\left\{e^{-\theta R_l}\right\}}\right\}\\ &\text{and}\quad \rho_v(\theta) = \frac{\beta}{\theta}\log_e\left\{\frac{e^{2\theta V_l}-\alpha e^{\theta V_l}}{\beta e^{\theta V_l}}\right\}.\end{aligned} \tag{5.14}$$

5.3.1 *Link Selection Policy*

In this section, we focus on the channel selection process that the transmitter employs. We set the maximum average data arrival rate under QoS constraints as the objective in the channel selection process. In particular, the transmitter chooses the channel in which the service process maximizes the average data arrival rate at the transmitter buffer. Notice that the transmission rate in the VLC channel is constant, whereas the transmission rate in the RF channel varies due to the changes in the channel fading gain. Now, due to the fact that the channel fading gains are known by the receiver as well as the transmitter, we provide the following proposition.

Proposition 2 *In the aforementioned RF/VLC system, the transmitter sends data to the receiver over the VLC link when the following condition for a given QoS exponent, θ, holds:*

$$V \geqslant \frac{1}{\theta}\log_e\left\{\frac{1-\beta+(1-\alpha)\xi}{2} + \frac{\sqrt{[1-\beta+(1-\alpha)\xi]^2-4(1-\alpha-\beta)\xi}}{2}\right\}, \tag{5.15}$$

where

$$\xi = \frac{1-(1-\beta)\mathbb{E}_h\left\{e^{-\theta R_l}\right\}}{(1-\alpha)\mathbb{E}_h\left\{e^{-\theta R_l}\right\}-(1-\alpha-\beta)\mathbb{E}_h^2\left\{e^{-\theta R_l}\right\}}. \tag{5.16}$$

Proof: See Appendix H. □

Proposition 2 states that if the maximum attainable transmission rate in the VLC channel is greater than the right-hand side of (5.15), the transmitter should perform transmission over the VLC link because the statistical variations in the RF channel deteriorates the buffer stability. Meanwhile, in the special case when $\alpha+\beta=1$, we re-express (5.15) as

$$V \geqslant -\frac{1}{\theta}\log_e\left\{\mathbb{E}_h\left\{e^{-\theta R_l}\right\}\right\}. \tag{5.17}$$

Specifically, the constant rate in the VLC channel should be greater than the effective capacity of the RF channel such that the VLC channel is chosen for data transmission. In the following, we present two transmission strategies, such that (*i*) the data is transmitted over the link with the highest instantaneous transmission rate in one time frame, and (*ii*) the data is transmitted over both links simultaneously.

HYBRID-TYPE I TRANSMISSION STRATEGY In the above analysis, we obtain the performance levels when the transmitter chooses either of these two channels for data transmission following a link selection process based on the maximum average data arrival rates that the service processes in the channel can support. On the other hand, if there exists a fast and stable handover mechanism between the transmitter and the receiver, the transmitter will forward the data to the receiver over the link that provides the maximum lower bound on the instantaneous mutual information in the corresponding channel. For instance, when the lower bound on the instantaneous mutual information in the RF channel in the l^{th} time frame, R_l, is greater than the lower bound on the instantaneous mutual information in the VLC channel, V, i.e., $R_l > V$, the transmitter sends the data over the RF link in the corresponding time frame only. Otherwise, it prefers sending the data over the VLC link. Respectively, we can establish the channel selection criterion as follows: The transmitter sends the data over the RF link when

$$|h_l|^2 > \frac{a\sigma_r^2}{2}\left(2^{\frac{V}{TB_r}}-1\right)\exp\left\{1-\frac{bP_{\text{avg},r}}{2}\right\} = \kappa. \tag{5.18}$$

Otherwise, it sends the data over the VLC link. The aforementioned selection test can also be considered as the outage condition in the RF channel, i.e., the RF link is in outage when $|h_l|^2 \leqslant \kappa$. Noting that the channel fading gain changes independently from one time frame to another in the RF channel, the log-moment generating function of the service process becomes

$$\Lambda_{rv}(\theta) = \log_e \left\{ \mathbb{E}_{|h|^2>\kappa} \left\{ e^{\theta R_l} \right\} + \Pr\{|h_l|^2 \leqslant \kappa\} e^{\theta V} \right\}, \tag{5.19}$$

where $\mathbb{E}_{|h|^2>\kappa}\{e^{\theta R_l}\}$ is the conditional expectation given that $|h|^2 > \kappa$, i.e.,

$$\mathbb{E}_{|h|^2>\kappa}\{e^{\theta R_l}\} = \frac{1}{\Pr\{|h_l|^2 > \kappa\}} \int_{\kappa}^{\infty} e^{\theta R_l} f_{|h|^2}(|h|^2) d|h|^2,$$

where $f_{|h|^2}(|h|^2)$ is the probability density function of $|h|^2$. Hence, the maximum average data arrival rate at the transmitter buffer that the hybrid service process can sustain under the QoS constraints specified by $\theta > 0$ becomes

$$\rho_{rv}(\theta) = \frac{\beta}{(\alpha+\beta)\theta} \log_e \left\{ \frac{1-(1-\beta)D}{(1-\alpha)D-(1-\alpha-\beta)D^2} \right\}, \tag{5.20}$$

where $D = \mathbb{E}_{|h|^2>\kappa} \left\{ e^{-\theta R_l} \right\} + \Pr\{|h_l|^2 \leqslant \kappa\} e^{-\theta V}$.

HYBRID-TYPE II TRANSMISSION STRATEGY Different than the aforementioned protocols, we consider a transmitter that sends data over both links simultaneously in each frame. We assume that the receiver has a multihoming capability. We further assume a multiplexing-based transmission scheme such that the data streams transmitted over the RF and VLC links are different and independent from each other. Indeed, this scenario is feasible since the light and RF waves do not cause interference on each other. We further assume a power allocation policy between the two links, i.e., the average power constraint in the RF link is set to $P_{avg,r} = \gamma_l P_{avg}$, and the one in the VLC link is set to $P_{avg,v} = (1-\gamma_l)P_{avg}$, where P_{avg} is the total average power constraint, and $0 < \gamma_l < 1$. Then, the instantaneous transmission rate in the RF channel in the l^{th} time frame in bits per frame becomes

$$R_l = TB_r \log_2 \left\{ 1 + \frac{2|h_l|^2}{a\sigma_r^2} \exp\left\{ \frac{b\gamma_l P_{avg}}{2} - 1 \right\} \right\}, \tag{5.21}$$

where a and b are the solutions of (5.2a) and (5.2b). Similarly following (5.5)–(5.6), the transmission rate in the VLC channel becomes

$$V = \frac{TB_v}{2} \log_2 \left\{ 1 + P_{peak,v}^2 \frac{\Omega^2 g^2}{2\pi\sigma_v^2} \times \exp\left\{ 2\frac{(1-\gamma_l)P_{avg}}{P_{peak,v}} \mu^\star - 1 \right\} \left(\frac{1-e^{-\mu^\star}}{\mu^\star} \right)^2 \right\}, \tag{5.22}$$

when $0 \leqslant \frac{(1-\gamma_l)P_{avg}}{P_{peak,v}} < \frac{1}{2}$, and

$$V = \frac{TB_v}{2} \log_2 \left\{ 1 + P_{peak,v}^2 \frac{\Omega^2 g^2}{2\pi\sigma_v^2} e^{-1} \right\}, \tag{5.23}$$

when $\frac{1}{2} \leqslant \frac{(1-\gamma_l)P_{avg}}{P_{peak,v}} \leqslant 1$, where $\mu^\star$ is the unique solution of (5.7). It follows that the total transmission rate in each frame is the sum of the transmission rates in both links, i.e., $R_l + V$. Then, the log-moment generating function can be readily expressed as

$$\Lambda_{srv}(\theta) = \log_e \mathbb{E}\{e^{\theta(R_l+V)}\} = \theta V + \log_e \mathbb{E}\{e^{\theta R_l}\}, \tag{5.24}$$

and the maximum average data arrival rate is equal to

$$\rho_{srv}(\theta) = \frac{\beta}{(\alpha+\beta)\theta} \log_e \left\{ \frac{1-(1-\beta)D}{(1-\alpha)D-(1-\alpha-\beta)D^2} \right\}, \tag{5.25}$$

where $D = e^{-\theta V}\mathbb{E}_h\{e^{-\theta R_l}\}$. We finally remark that the optimal value of γ_l that maximizes the sum transmission rate, i.e., $\max_{\gamma_l}\{R_l + V\}$, can be obtained numerically.

Remark 9 *Employing the aforementioned strategies requires the perfect knowledge of both RF and VLC channels at the transmitter side in each frame. We assume that the channel estimation is performed at the receiver, and forwarded to the transmitter in a delay-free and error-free feedback channel. This increases the signaling overhead. Moreover, applying* Hybrid-Type II Transmission Strategy *increases the implementation complexity because power sharing should be performed in each transmission frame. From the implementation perspective, exploiting* Hybrid-Type II Transmission Strategy *is limited because the receiver should have a multi-homing capability that enables it to perform link aggregation and receive data from different transmission technologies simultaneously. On the other hand, the maximum average data arrival rate is a steady-state performance measure. Thus, the selection process explained in Proposition 2 is considered as a large-timescale operation, which can be performed over periods of multiple transmission frames, and therefore, the implementation complexity decreases. Such a large-timescale operation is also proposed in [180].*

Remark 10 *The selection process can also be employed in cases the access point controller has to choose between more than two transmission links. For example, let there be* N *transmission links, and let* $\{\rho_1(\theta), \ldots, \rho_N(\theta)\}$ *be the maximum average data arrival rates at the transmitter buffer sustained by these links. Then, Proposition 1 in the chapter will be updated with the solution of the following maximization problem:*

$$\textit{Transmission Link} = \max\{i : \rho_i(\theta)\}. \tag{5.26}$$

Similarly, let $\{R_1(l), \cdots, R_N(l)\}$ *be the instantaneous transmission rates provided by each link in the* l^{th} *time frame. Then, the link selection criterion in* Hybrid-Type I Transmission Strategy *will be updated with the solution of the following maximization problem:*

$$\textit{Transmission Link} = \max\{i : R_i\}. \tag{5.27}$$

Subsequently, the log-moment generating function of the service process will be

$$\Lambda_{1,\ldots,N}(\theta) = \log_e \left\{ Pr_i \mathbb{E} e^{\theta R_i(l)} \right\}, \tag{5.28}$$

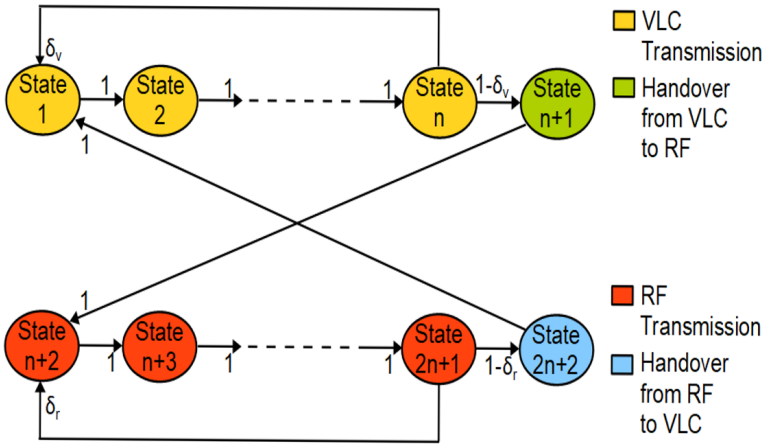


Figure 5.3: State transition model of the hybrid scenario with handover.

where $\Pr_i$ *is the probability that the transmitter chooses link* i *for data transmission. Finally, if the receiver has a multihoming capability, Hybrid-Type II Transmission Strategy can be applied with power sharing to maximize the total transmission rate in each frame.*

5.3.2 *Impacts of Handover Delay*

Handover delay occurs when the transmitter moves from one link to the other, which is the case in *Hybrid-Type I Transmission Strategy*. In this strategy, data transmission in one time frame is performed over the link that provides the maximum instantaneous transmission rate in that frame. Particularly, the transmitter switches from one link to the other or stay in the same link at the end of any time frame after comparing the instantaneous transmission rates in both links.

Now, given that T_H denotes the duration of one single handover phase, let us initially assume that the frame duration is larger than the handover phase, i.e., $T > T_H$. For the sake of simplicity, we divide the frame duration into n sub-frames that are equal to T_H, i.e., $T = n \times T_H$, where $n \in \mathbb{N}$ and $n > 1$. Particularly, a series of n sub-frames of data transmission phase is followed by one sub-frame of handover process if the transmitter changes the transmission link, or by another series of n sub-frames of data transmission phase if the transmitter stays in the same transmission link. For an analytical representation, we model the buffer activity at the end of each sub-frame as a discrete-time Markov process.

As shown in Fig. 5.3, we have $2n+2$ states. The first n states, i.e., {State $1,\cdots$,State n}, represent the data transmission sub-frames in the VLC link, and State $(n+1)$ represents the handover process from the VLC link to the RF link. Similarly, the subsequent n states, i.e., {State $(n+2),\cdots$,State $(2n+1)$}, represent the data transmission sub-frames in the RF link, and

State $(2n+2)$ represents the handover process from the RF link to the VLC link.

Notice that the state transition probability from State i to State $i+1$ is 1 for $i \in \{1, \cdots, n-1\}$ and for $i \in \{n+2, \cdots, 2n\}$ because the data transmission in each link is completed in n sub-frames and the link change may occur at the end of the n^{th} and $(2n+1)^{th}$ sub-frames. On the other hand, in State n, either the transmitter changes the link and the system enters State $n+1$ with probability $1-\delta_v$, or the transmitter stays in the same link and the system enters State 1 with probability δ_v. Similarly, in State $(2n+1)$, either the transmitter changes the link and the system enters State $(2n+2)$ with probability $1-\delta_r$, or the transmitter stays in the same link and the system enters State $(n+2)$ with probability δ_r. Finally, the system moves from State $(n+1)$ to State $(n+2)$ and from State $(2n+2)$ to State 1 with probability 1 because at the end of one handover phase, the transmitter starts data transmission.

As seen in (5.18), if $|h_l|^2 > \kappa$, the transmitter sends the data over the RF link. Otherwise, it sends the data over the VLC link. Therefore, for this specific case, we have $\delta_r = \Pr\{|h_l|^2 > \kappa\} = \delta$ and $\delta_v = 1-\delta$. Then, we can express the $(2n+2) \times (2n+2)$ transition matrix as follows:

$$\Gamma = [p_{ji}], \text{where}$$

$$p_{ji} = \begin{cases} 1, & \text{for } j = i+1 \quad \text{and } 1 \leqslant i \leqslant n-1 \text{ or } n+2 \leqslant i \leqslant 2n, \\ 1, & \text{for } (i,j) = (n+1, n+2) \text{ or } (i,j) = (2n+2, 1), \\ \delta, & \text{for } (i,j) = (n, n+1) \text{ or } (i,j) = (2n+1, n+2), \\ 1-\delta, & \text{for } (i,j) = (n, 1) \text{ or } (i,j) = (2n+1, 2n+2), \\ 0, & \text{otherwise,} \end{cases} \tag{5.29}$$

is the state transition probability from State i to State j. Then, we can re-characterize the log-moment generating function of the hybrid system for any $\theta > 0$, which is provided in (5.19), as follows [86, Chap. 7, Example 7.2.7]:

$$\Lambda_{rv}(\theta) = \log_e \mathrm{sp}(\Phi(\theta)\Gamma), \tag{5.30}$$

where $\mathrm{sp}(\Phi(\theta)\Gamma)$ is the spectral radius of the matrix $\Phi(\theta)\Gamma$, and $\Phi(\theta)$ is a diagonal matrix whose components are the moment generating functions of the processes in $2n+2$ states. Notice that the transmitted bits are removed from the transmitter buffer only at the ends of the n^{th} and $(2n+1)^{th}$ frames. Therefore, the moment generating functions are $e^{\theta V}$ and $\mathbb{E}_{|h|^2>\kappa}\{e^{\theta R_l}\}$ in the n^{th} and $(2n+1)^{th}$ frames, respectively. However, there are no bits removed in the other states, i.e., the service rates in the other states are effectively zero. Hence, the moment generating functions are 1 in other states. Moreover, the unique QoS exponent, $\theta^\star$, is obtained when $\Lambda_a(\theta^\star) = -n\Lambda_{rv}(-\theta^\star)$.

5.3.3 *Non-asymptotic Bounds*

Recall that the aforementioned results provide the performance analysis in the steady-state. Particularly, the analysis is obtained when the number of time frames is very large. On the other hand, non-asymptotic bounds regarding the statistical queueing and delay characterizations at the transmitter buffer are of interest for system designers as well. Therefore, we address the framework of Network Calculus [86, 133, 181], and consider [135, Theorem 2], which states that a minimal bound on the queue length can be found for a given buffer overflow probability. Particularly, given the RF-based data service process in Section 5.2.1 and the two-state Markov modeled data arrival process in Section 5.2.3, the buffer threshold, q, is expressed as

$$q = \inf_{c>0}\{q_r + q_a\}, \tag{5.31}$$

for a given buffer overflow probability, $\Pr\{Q \geqslant q\} = \varepsilon$, where

$$\begin{aligned} q_r &= -\sup_{\theta}\left\{\frac{\log_e\{-\varepsilon_r\left[\Lambda_r(-\theta)+\theta c\right]\}}{\theta}\right\} \\ &\text{for } \max\left\{0, -\frac{1}{c\varepsilon_r} - \frac{\Lambda_r(-\theta)}{c}\right\} < \theta, \end{aligned} \tag{5.32}$$

and

$$\begin{aligned} q_a &= -\sup_{\theta}\left\{\frac{\log_e\left\{\varepsilon_a\left[\theta c - \sup_{t>0}\{\Lambda_a(\theta,t)\}\right]\right\}}{\theta}\right\} \\ &\text{for } 0 < \theta < \frac{1}{c\varepsilon_a} + \frac{\sup_{t>0}\{\Lambda_a(\theta,t)\}}{c}, \end{aligned} \tag{5.33}$$

with

$$\begin{aligned} \Lambda_a(\theta,t) = \frac{1}{t}\log_e\Bigg\{&\begin{bmatrix} p_{ON} & p_{OFF}\end{bmatrix}\left(\begin{bmatrix} e^{\theta\lambda} & 0 \\ 0 & 1\end{bmatrix}\begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta\end{bmatrix}\right)^{(t-1)} \\ &\times\begin{bmatrix} e^{\theta\lambda} & 0 \\ 0 & 1\end{bmatrix}\begin{bmatrix} 1 \\ 1\end{bmatrix}\Bigg\}. \end{aligned} \tag{5.34}$$

Above, the buffer violation probability is $\varepsilon = \varepsilon_r + \varepsilon_a$, and θ and c are free parameters. Notice also that the log-moment generating function in (5.32), i.e., $\Lambda_r(-\theta)$, is time-invariant because the service process is memory-less, whereas the log-moment generating function in (5.33), i.e., $\Lambda_a(\theta,t)$, is time-variant because the arrival process depends on its current state. Moreover, we remark that when t goes to infinity, we have $\lim_{t\to\infty}\Lambda_a(\theta,t) = \Lambda_a(\theta)$, where $\Lambda_a(\theta)$ is expressed in (5.10). Herein, we refer to [135] for further calculation details in (5.32) and (5.33).

Likewise, when the VLC-based data service process in Section 5.2.2 is employed, we have

$$q = \inf_{c>0}\{q_v + q_a\}, \tag{5.35}$$

and

$$\begin{aligned} q_v &= -\sup_{\theta} \frac{\log_e\{-\varepsilon_r\left[\Lambda_v(-\theta)+\theta c\right]\}}{\theta} \\ &= -\sup_{\theta} \frac{\log_e\{-\varepsilon_r\left[-\theta V+\theta c\right]\}}{\theta}. \end{aligned} \tag{5.36}$$

Notice that q in (5.35) is minimized when $c = V$. Moreover, because q_v cannot be smaller than zero, we have $q = q_a$ and $c = V$. Therefore, when the VLC-based service channel is chosen as the service process, we have

$$\begin{aligned} q = q_a &= -\sup_{\theta}\left\{\frac{\log_e\left\{\varepsilon_a\left[\theta V-\sup_{t>0}\{\Lambda_a(\theta,t)\}\right]\right\}}{\theta}\right\} \\ &\text{and } \varepsilon = \varepsilon_a. \end{aligned} \tag{5.37}$$

Now, assuming a fast and stable handover mechanism between the transmitter and the receiver, and a service channel selection process as described in *Hybrid-Type I Transmission Strategy*, we can characterize the delay bound as follows: $q = \inf_{c>0}\{q_a + q_{rv}\}$ where

$$\begin{aligned} q_{rv} &= -\sup_{\theta}\left\{\frac{\log_e\{-\varepsilon_r\left[\Lambda_{rv}(-\theta)+\theta c\right]\}}{\theta}\right\} \\ &\text{for } \max\left\{0, -\frac{1}{c\varepsilon_r} - \frac{\Lambda_{rv}(-\theta)}{c}\right\} < \theta, \end{aligned} \tag{5.38}$$

and $\Lambda_{rv}(\theta)$ is given in (5.19).

Remark 11 *Let us assume a first-come first-served protocol exists at the transmitter buffer. Then, the minimal bound on the buffering delay is expressed as follows [135, Theorem 1]:*

$$\begin{aligned} &d = \inf_{c>0}\left\{\frac{q_r+q_a}{c}\right\}, \textit{ or } d = \inf_{c>0}\left\{\frac{q_v+q_a}{c}\right\}, \\ &\textit{or } d = \inf_{c>0}\left\{\frac{q_{rv}+q_a}{c}\right\}, \end{aligned} \tag{5.39}$$

when the RF-based service process, or the VLC-based service process, or Hybrid-Type I Transmission Strategy *is employed.*

Remark 12 *We consider a user-related performance measure, i.e., the maximum average data arrival rate at the data buffer, and formulate the link selection by employing the transmission rates provided by the RF and VLC links. Our analytical framework can easily be extended to a more general multi-user scenario by regarding the rate allocations for each user in the transmission links and the* receiver-oriented *data arrival processes at the transmitter buffer. In this regard, we refer to Fig. 5.6 in Section 5.4, where we employ frequency-division multiple access (FDMA) and time-division multiple access (TDMA) protocols for numerical illustrations. Moreover, our chapter is different than [24, 72, 180, 182]. The system sum throughput is maximized in [24, 72], and the system average power consumption is minimized in [180, 182]. In these studies, a framework in which a joint resource allocation*

Table 5.2: Simulation Parameters

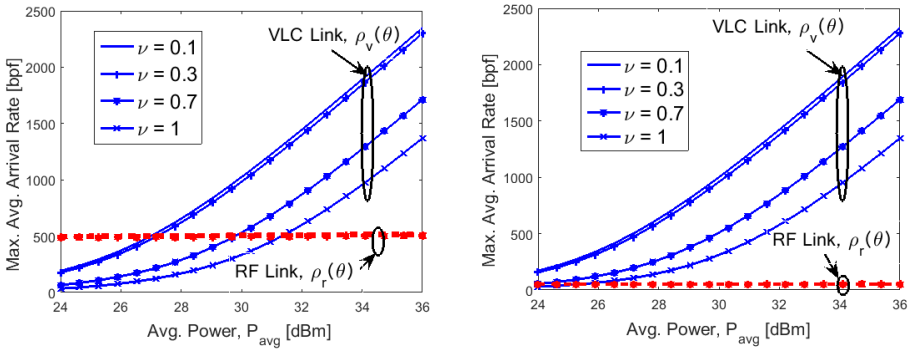
VLC System	
LED half intensity viewing angle, $\phi_{1/2}$	$\{30°, 45°, 60°\}$
PD field of view (FOV), ψ_C	$90°$
PD physical area, A	$1\,\text{cm}^2$
Channel bandwidth, B_v	10 MHz
PD opt.-to-elect. conversion efficiency, Ω	0.53 A/W
PD optical concentrator gain, $u(\psi)$	1
Vertical distance, d_v	2.5 m
Noise power spectral density, N_v	10^{-21} A^2/ Hz
RF System	
Channel bandwidth, B_r	10 MHz
Path loss exponent, q	1.8
Rician factor, K	10 dB
Log-normal standard deviation, σ	3.6 dB
Ambient Temperature, T_t	$280°$ K

and link assignment process is employed is not provided, and the optimization problems are in principle mixed integer and non-linear programming problems, which are mathematically intractable. Therefore, the main optimization problems are decomposed into solvable sub-problems, and iterative algorithms are provided.

Remark 13 *The link selection process can easily be adopted into scenarios where the receiver is mobile. All that one needs is to consider the log-moment generating function in the VLC link for changing transmission rates, i.e.,* $\Lambda_v(\theta) = \log_e\left\{\mathbb{E}_v\{e^{\theta V_i}\}\right\}$*, and base the channel selection process on the link that increases the maximum average data arrival rate at the transmitter buffer. In* Hybrid-Type I Transmission Strategy *and* Hybrid-Type II Transmission Strategy*, because the channel fading gains in the RF and VLC links are instantaneously known at the transmitter in each time frame, the aforementioned analysis will not be different than what we currently have in the chapter even if the transmission rates vary due to mobility. We also note that the user mobility is normally low in indoor scenarios. In this regard, we refer to [183].*

5.4 NUMERICAL RESULTS

In this section, we present numerical results that substantiate our theoretical findings. Unless otherwise specified, we set the transmission time frame to 0.1 milliseconds, i.e., $T = 10^{-4}$. We assume that the LED at the VLC access point of the transmitter has a Lambertian radiation pattern, and that the transmitter and receiver planes are parallel to each other. We further assume



(a) $\theta = 0.01$ (b) $\theta = 0.1$

Figure 5.4: Maximum average arrival rates of VLC and RF links as a function of the average power limit, P_{avg}, for different values of the average-to-peak power ratio ν and the QoS exponent, θ. Here, $d_0 = 15$ m, $d_1 = 3$ m, $\phi_{1/2} = 60°$, $\alpha = 0.3$, and $\beta = 0.7$. {bpf : bits per frame}.

that the transmitter is directed downwards, while the receiver is directed upwards, as depicted in Figure 4.1. However, our theoretical results can easily be adopted into different positional settings. Herein, the line-of-sight channel gain, i.e., g, is given as follows [157]:

$$g = \frac{(m+1)AL(\psi)u(\psi)d_v^{m+1}}{2\pi d_1^{m+3}}\text{rect}(\psi/\psi_C), \tag{5.40}$$

where d_1 is the distance between the LED at the transmitter and the photodiode at the receiver, and other parameters are similar to those in (4.1).

Regarding the RF channel, we consider a Rician fading distribution with the Rician factor, K, where the channel realizations, $\{h_l\}$, are independent and identically distributed, circularly symmetric complex Gaussian random variables with mean and variance

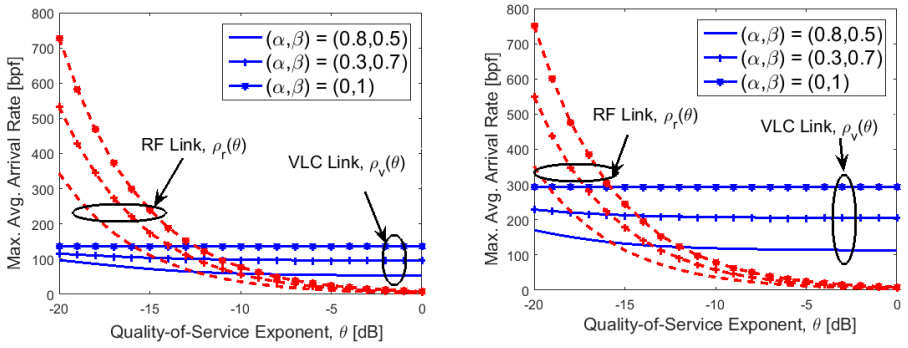
$$\mu = \sqrt{\frac{e^{-L(d_0)/10}K}{K+1}} \quad \text{and} \quad \sigma_h^2 = \frac{e^{-L(d_0)/10}}{K+1},$$

respectively. Setting K to a reasonable value, we can reflect the channel characteristics in millimeter wave range communications as well [21]. Here, $L(d_0)$ is the large-scale path loss in decibels as a function of the distance between the RF access point at the transmitter and the RF front-end at the receiver, d_0, and it is given by [184]

$$L(d_0) = L(d_{ref}) + 10q\log_e\left(\frac{d_0}{d_{ref}}\right) + X_\sigma, \tag{5.41}$$

where $L(d_{ref}) = 40$ dB is the path loss at a reference distance, $d_{ref} = 1$ m, and an operating frequency of 2.4 GHz. In addition, q is the path loss exponent and X_σ represents the shadowing effect, which is assumed to be a zero-mean Gaussian random variable with a standard deviation σ expressed in decibels[5]. Finally, the thermal noise power at the RF front-end of the receiver

5 Empirical values of K, q, and σ^2 in different indoor scenarios were provided in [184–186]. For instance, the value of q ranges from 1.2 to 1.7, while σ varies between 3.6 dB and 4.0 dB inside buildings [184].



(a) $P_{avg} = 24$ dBm (b) $P_{avg} = 30$ dBm

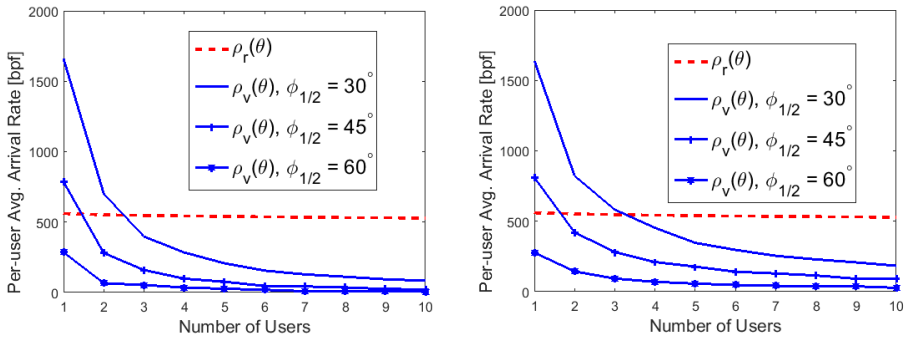
Figure 5.5: Maximum average arrival rates of VLC and RF links as a function of the QoS exponent, θ, for different values of the average power limit P_{avg} and the source statistics, α and β. Here, $d_0 = 15$ m, $d_1 = 3$ m, $\alpha = 0.3$, $\phi_{1/2} = 60°$, and $\nu = 0.7$. {bpf : bits per frame}.

is $\sigma_r^2 = \kappa_B T_t B_r$, where κ_B is the Boltzmann constant and T_t is the ambient temperature [173]. Table 5.2 summarizes the simulation parameters, unless otherwise stated. Finally, setting the average transmission power constraint to P_{avg} in all the transmission strategies, we define ν as the average-to-peak power ratio in the RF and VLC links, i.e., $\nu = \frac{P_{avg}}{P_{peak,r}} = \frac{P_{avg}}{P_{peak,v}}$.

5.4.1 *Transmission Strategies*

We consider the scenario in which a transmitter has a VLC access point and an RF access point, as shown in Figure 5.1. The receiver is located at a distance $d_1 = 3$ m from the VLC access point, i.e., the user is located at a horizontal distance $d_h \approx 1.6$ m from the cell center, where $\phi_{1/2} = 60°$ and $d_v = 2.5$ m, and at a distance $d_0 = 15$ m from the RF access point. In Figure 5.4, we plot the maximum average data arrival rates at the transmitter buffer, $\rho_v(\theta)$ and $\rho_r(\theta)$, as functions of the average power constraint, P_{avg}, with different average-to-peak power ratios, i.e, $\nu \in \{0.1, 0.3, 0.7, 1\}$, when the VLC and RF links are employed, respectively. We have the results for $\theta = 0.01$ in Figure 5.4a and for $\theta = 0.1$ in Figure 5.4b. We observe that the maximum average data arrival rates increase faster with the increasing average power constraint in the VLC link than they do in the RF link. For instance, when $\nu = 0.3$ and $\theta = 0.01$, the maximum average data arrival rate increases more than 145 bits per frame with P_{avg} increasing from 27 dBm to 28 dBm in the VLC link, whereas it increases 2 bits per frame in the RF link.

We observe the same behavior when the peak power constraint increases. We can explain this result with the stochastic nature of the transmission rates in the RF channel. Particularly, when the instantaneous transmission rate in the RF channel becomes very low, more data packets are accumulated in the transmitter buffer. Therefore, in order to sustain the QoS constraints, the transmitter buffer should accept data at lower arrival rates. On the



(a) FDMA (i.e., power and bandwidth)

(b) TDMA (i.e., time)

Figure 5.6: Per-user maximum average arrival rates of VLC and RF links as a function of the number of served receivers (or equivalently the receiver allocated resources) and for different values of the LED viewing angle, $\phi_{1/2}$. Here, $\alpha = 0.3$, $\beta = 0.7$, $\nu = 1$, and $P_{avg} = 24$ dBm. {bpf : bits per frame}.

other hand, because the transmission rate in the VLC channel is constant, the maximum average data arrival rate increases almost linearly with the increasing transmission rate in the VLC channel. Moreover, the performance in the RF link is better than the VLC link when the average power constraint is lower, and the performance in the VLC link is better than the RF link when the average power constraint is higher.

In Figure 5.5, we plot $\rho_r(\theta)$ and $\rho_v(\theta)$ as functions of the QoS exponent, θ, for different values of α and β when $\nu = 0.7$. We have $P_{avg} = 24$ dBm in Figure 5.5a and $P_{avg} = 30$ dBm in Figure 5.5b. The performance in the RF link is better than the VLC link when θ is low, whereas the performance in the RF link decreases faster with increasing θ and becomes less than the performance in the VLC link. In other words, the stochastic nature of the RF channel prevents the RF link from supporting data arrival rates at the transmitter buffer when the QoS constraints are stringent. Indeed, the maximum average data arrival rate that the RF link supports approaches zero exponentially with increasing θ regardless of the average power constraint and the source statistics, i.e., α and β.

However, the RF link can support higher data arrival rates if the QoS constraints are looser. We further observe that increasing β (or decreasing α), results in better performance values in the VLC and RF links because the steady-state probability of the ON state, i.e., $P_{ON} = \frac{\beta}{\alpha+\beta}$, and the average arrival rate at the transmitter buffer increase. However, the effect of the source statistics on the performance values is much less than that of the QoS constraints, especially in the RF link. In other words, the randomness in the service process has a higher impact on the system performance than the randomness in the arrival process has.

In typical indoor scenarios, VLC and RF access points serve multiple receivers. This is applied by sharing available resources (i.e., power, time, and bandwidth) among the served receivers. Herein, we assume that the transmitter employs the commonly known FDMA and TDMA schemes in

both links. In Figure 5.6, we plot the maximum average data arrival rate per user given that all the users are uniformly positioned within the coverage area of the VLC access point.

We observe that the performance per user in the RF link is generally much higher than the performance per user in the VLC link when the number of receivers is above 4. Basically, the system can serve more users in the RF link than the VLC link when the QoS constraints are of interest. In addition, the results in Figure 5.6 agree with the results in Figure 5.4, where the performance of the VLC link is highly affected by the decreasing average power constraint. We finally see that with the decreasing LED viewing angle, the performance in the VLC link becomes better because the transmission power is concentrated in smaller areas. Notice also that the VLC channel gain is affected by $\phi_{1/2}$ through the Lambertian index. Herein, we show the performance sensitivity of the RF and VLC links to the allocated transmission resources such as power, time, and bandwidth, given that the available resources are equally shared among the users. We also show that our framework can easily be invoked in a multi-user scenario.

We further explore the system performance with respect to the receiver location. We set $(x_v, y_v, z_v) = (0, 0, 0)$ as the Cartesian coordinates of the VLC access point, $(x_r, y_r, z_r) = (10, 0, 0)$ as the coordinates of the RF access point, and (x_u, y_u, z_u) as the coordinates of the receiver, where $z_u = -d_v$. Particularly, we consider the following strategies:

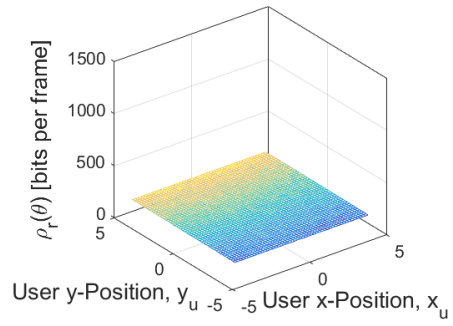
1. *RF-only Strategy*: The transmitter sends data over the RF link only, and the maximum average arrival rate, $\rho_r(\theta)$, is expressed in (5.12).

2. *VLC-only Strategy*: The transmitter sends data over the VLC link only, and the maximum average arrival rate, $\rho_v(\theta)$, is expressed in (5.13).

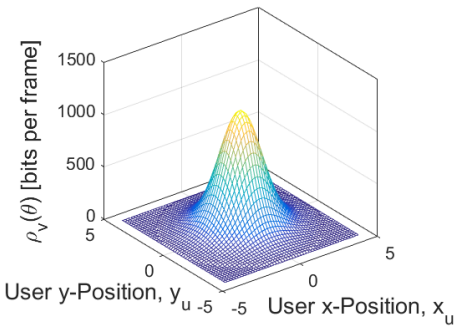
3. *Hybrid-type I Transmission Strategy*: The transmitter sends data over the link that provides the highest transmission rate, and the maximum average arrival rate, $\rho_{rv}(\theta)$, is expressed in (5.20).

4. *Hybrid-type II Transmission Strategy*: The transmitter sends data over both links simultaneously following a power allocation policy to maximize the transmission rate, and the maximum rate, $\rho_{srv}(\theta)$, is expressed in (5.25).

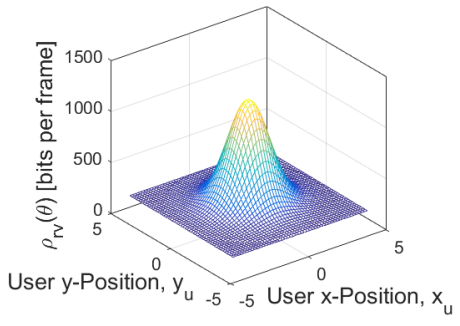
In Figure 5.7, we plot the maximum average arrival rate as a function of the x_u and y_u coordinates of the receiver location for different QoS constraints, where $x_u \in [-5, 5]$ and $y_u \in [-5, 5]$. As seen in Figure 5.7a and Figure 5.7e, the position of the receiver does not impact the performance levels in the RF link necessarily, i.e., the performance level stays almost constant when the receiver stays in the defined range. However, the performance levels in the other strategies are affected by the position of the receiver, and the maximum average data arrival rate increases as the receiver gets closer to the point $(x_v, y_v, z_v) = (0, 0, -d_v)$ because the constant transmission rate from the VLC access point to the receiver increases and the stochastic nature of the RF link is mitigated more with the increasing rate in the VLC link.



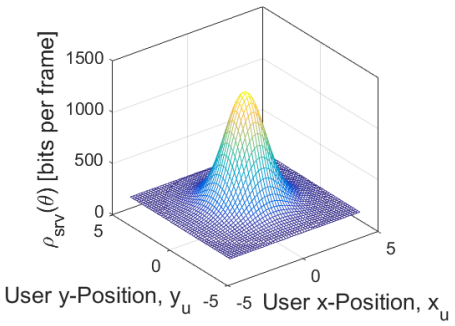
(a) RF-only, $\theta = 0.01$



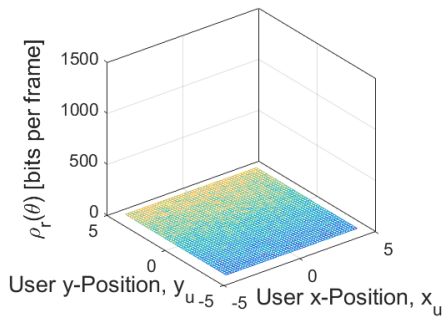
(b) VLC-only, $\theta = 0.01$



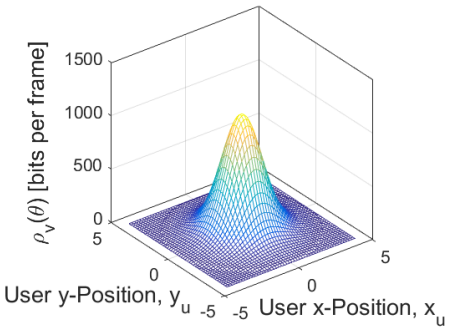
(c) Hybrid-type I, $\theta = 0.01$



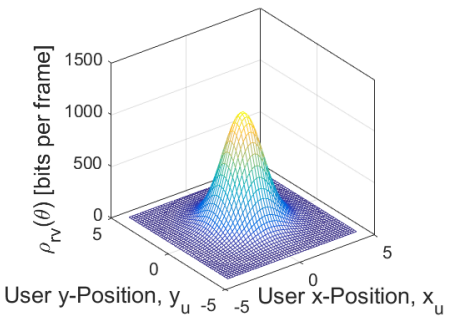
(d) Hybrid-type II, $\theta = 0.01$



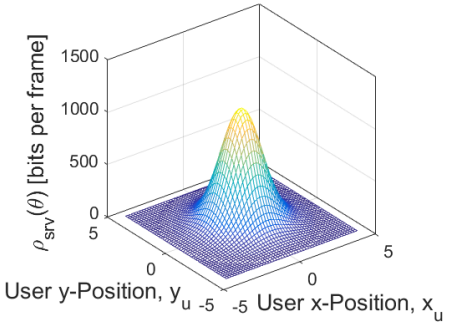
(e) RF-only,, $\theta = 0.1$



(f) VLC-only, $\theta = 0.1$

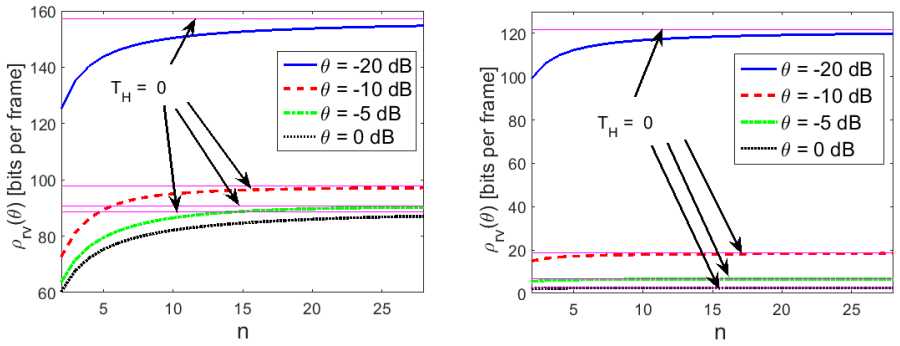


(g) Hybrid-type I, $\theta = 0.1$



(h) Hybrid-type II, $\theta = 0.1$

Figure 5.7: Maximum average arrival rates for different selection strategies as a function of the receiver position in terms of x_u and y_u and for different values of θ. Here, $P_{avg} = 24$ dBm, $\nu = 0.7$, $\alpha = 0.3$, $\beta = 0.7$ and $\phi_{1/2} = 60°$. {bpf : bits per frame}.



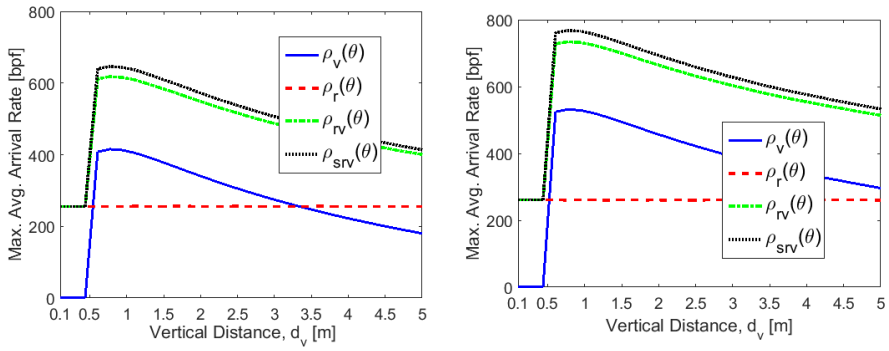
(a) $(x_u, y_u, z_u) = (0, 0, -d_v)$ i.e., cell center (b) $(x_u, y_u, z_u) = (d_c, 0, -d_v)$ i.e., cell edge

Figure 5.8: Maximum average arrival rate as a function of n and for different values of the QoS exponent θ and user position in terms of x_u. Here, $\phi_{1/2} = 60°$, $\alpha = 0.3$, $\beta = 0.7$, $P_{avg} = 24$ dBm, and $\nu = 1$.

As seen in Figure 5.7b and Figure 5.7f, the maximum average data arrival rate goes to zero in *VLC-only Strategy* as the receiver goes out of the coverage area of the VLC access point. Similarly, as seen in Figure 5.7c, Figure 5.7d, Figure 5.7g and Figure 5.7h, the maximum average data arrival rate becomes equal to the one in *RF-only Strategy* as the receiver goes out of the the coverage area of the VLC access point. Furthermore, *Hybrid-type I Transmission Strategy* provides higher performance levels than *RF-only Strategy* and *VLC-only Strategy* do because the transmitter, employing *Hybrid-type I Transmission Strategy*, sends the data over the RF link when the instantaneous transmission rate in the RF link is higher than the rate in the VLC link, and mitigates the lower transmission rates in the RF link by sending the data in the VLC link. Finally, *Hybrid-type II Transmission Strategy* outperforms all the other strategies. However, the performance gap between *Hybrid-type I Transmission Strategy* and *Hybrid-type II Transmission Strategy* is not necessarily large. Hence, it is more advantageous to employ *Hybrid-type I Transmission Strategy* in order to avoid the hardware complexity that follows the addition of multihoming capability in *Hybrid-type II Transmission Strategy*.

In Figure 5.8, assuming that the handover process causes a transmission delay, where the handover process takes $T_H = \frac{1}{n}T$ seconds for $n \in \mathbb{N}$ and $n > 1$, we plot the maximum average data arrival rate in *Hybrid-type I Transmission Strategy*, $\rho_{rv}(\theta)$, as a function of n considering different user locations. Noting that smaller n means a longer handover period, we observe that the transmission performance is highly affected by the handover process. With increasing n, the performance levels approach the values that are obtained when there is no handover delay. Moreover, the maximum average data arrival rates are higher in Figure 5.8a than in Figure 5.8b, because the constant transmission rate in the VLC link is higher when the user is at the center.

Subsequently in Figure 5.9, we plot the maximum average data arrival rates as functions of the vertical distance between the VLC access point and the receiver. We set the position of the receiver to $(x_u, y_u, z_u) = (0.8, 0, -d_v)$,



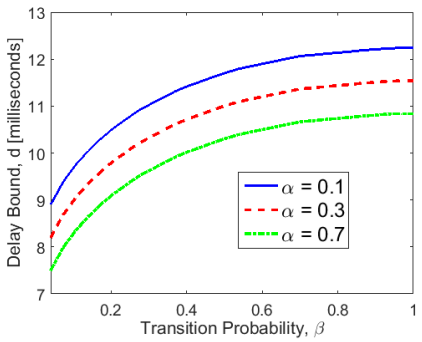
(a) $P_{avg} = 24$ dBm (b) $P_{avg} = 30$ dBm

Figure 5.9: Maximum average arrival rate of different transmission strategies as a function of the vertical distance and for different average power limit P_{avg}. Here, $\phi_{1/2} = 60^\circ$, $(x_u, y_u, z_u) = (0.8, 0, -d_v)$, $\alpha = 0.3$, $\beta = 0.7$, $\theta = 0.1$, and $\nu = 1$. {bpf : bits per frame}.

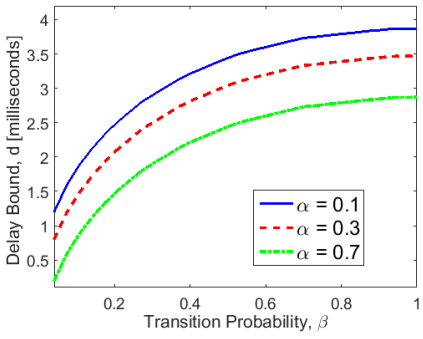
i.e., the horizontal distance to the VLC cell center is $d_h = 0.8$ m. When $d_v \leqslant 0.6$ m, the performance level in the VLC link is zero because the cell area is very small and does not cover the point where the user stands, i.e., $d_c = d_v \tan(\theta_{1/2}) < d_h = 0.8$ m, and d_c is the cell radius. The performance levels in all strategies except *RF-only Strategy* increase up to a value, and then decrease with increasing d_v. This is because the increase in the LED viewing angle is relatively less than the increase in d_v at the beginning. Therefore, the user is effectively getting closer to the cell center and having more rate in the VLC link. In other words, the gain achieved by getting closer to the cell center is higher than the expected degradation due to the increasing cell radius. However, with d_v increasing beyond a certain value, the user gets far away from the VLC access point, and hence, the radiated power spreads over more area, which eventually leads to decreased transmission rates in the VLC link. Therefore, the gain in the VLC link vanishes as the distance to the VLC access point becomes larger.

5.4.2 *Non-asymptotic Delay Bounds*

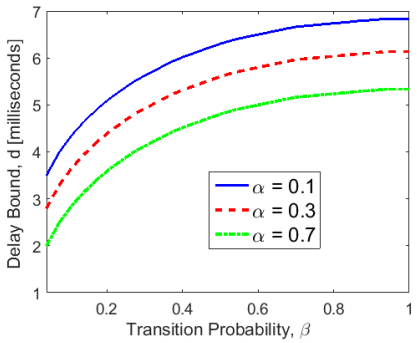
In the aforementioned results, we analyze the system performance in the steady-state. In the following, we provide results regarding the non-asymptotic bounds, i.e., the bounds on the buffering delay experienced by the data in the transmitter buffer. Particularly, we plot the delay bound as a function of the state transition probability from the OFF state to the ON state in the data arrival process, β, for different α values, where α is the state transition probability from the ON state to the OFF state in the data arrival process. We set the average data arrival rate, λp_{ON}, to a value very close to the average data service (transmission) rate in the transmission channel. We note that the average data service rate in the transmission channel depends on the chosen transmission strategy.



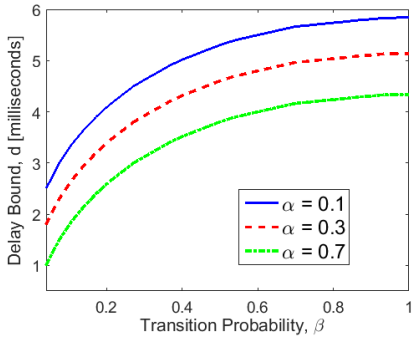
(a) RF-only Transmission, λ = 21 kbpf



(b) VLC-only Transmission, λ = 7 kbpf



(c) Hybrid-Type I Transmission, λ = 24 kbpf

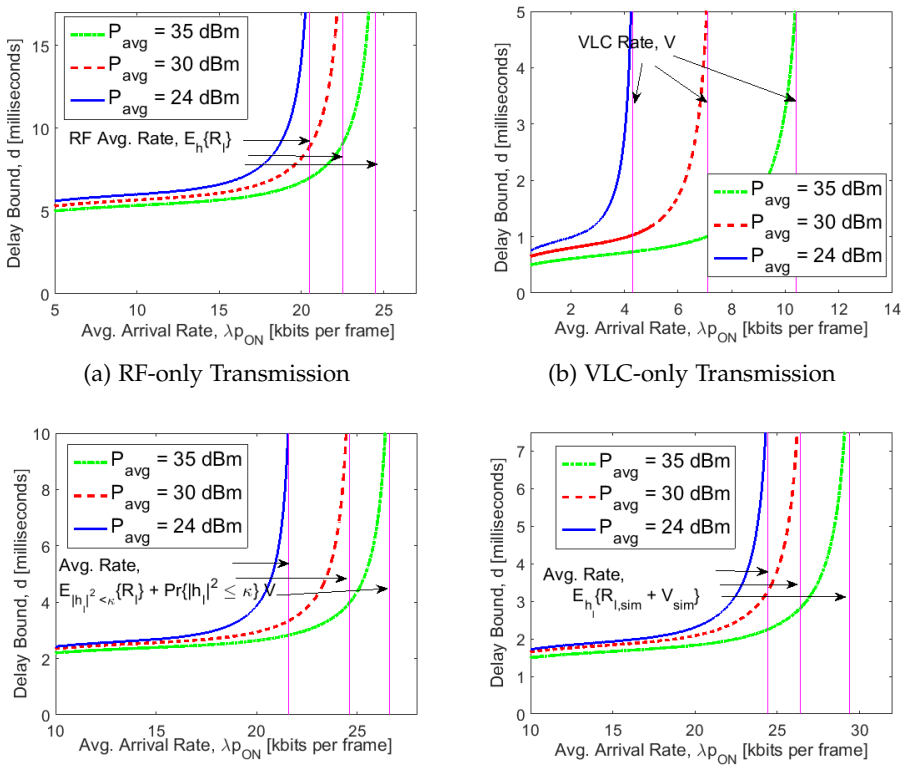


(d) Hybrid-Type II Transmission, λ = 26 kbpf

Figure 5.10: Delay Bounds for different transmission strategies as a function of the transition probability β and for different values of α. Here, $d_0 = 10$ m, $d_1 = 3$ m, $P_{avg} = 30$ dBm, $\nu = 0.7$, and $\phi_{1/2} = 60°$. {bpf : bits per frame}

The delay bound is the highest in *RF-only Strategy* as seen in Figure 5.10a, whereas it is the lowest in *VLC-only Strategy* as seen in Figure 5.10b. However, the arrival rate that *RF-only Strategy* supports is higher than the rate that *VLC-only Strategy* supports. More interestingly, the hybrid strategies can support higher arrival rates with less delay bounds, and *Hybrid-type II Transmission Strategy* outperforms all the others. Herein, the system takes advantage of the occasional higher rates in the RF links, and mitigates the lower rates in the RF link by the constant transmission rate in the VLC link. Moreover, increasing β and decreasing α cause the delay bound to increase.

Finally, we explore the effects of the data arrival rate, λ, on the delay bound performance in Figure 5.11. We set $\alpha = 0.3$ and $\beta = 0.7$, and consider different average power constraints, i.e., $P_{avg} = 24$, 30, and 35 dBm. The delay bounds increase asymptotically as the average arrival rate approaches the average data service rates in the channels, because when the average data arrival rate is greater than the average data service rate in one channel, the system becomes unstable, and long buffering periods are expected. Moreover, as seen in Figure 5.11b, the delay bounds are the minimum in *VLC-only Strategy* but the ranges of the average data arrival rates are smaller



(a) RF-only Transmission

(b) VLC-only Transmission

(c) Hybrid-Type I transmission

(d) Hybrid-Type II transmission

Figure 5.11: Delay Bounds for different transmission strategies as a function of the arrival rate λ and for different values of P_{avg}. Here, $\alpha = 0.3$, $\beta = 0.7$, $d_0 = 10$ m, $d_1 = 3$ m, $\nu = 0.7$, and $\phi_{1/2} = 60^\circ$.

than the other strategies, and as seen in Figure 5.11a, the delay bounds are the maximum in *RF-only Strategy*.

However, the hybrid strategies again outperform the others i.e., the hybrid strategies take advantage of the VLC link when the rate in the RF link goes down drastically, and utilize the RF link when the channel conditions are better again. While the VLC link provides stability and decreases the delay bounds, the RF link increases the range of average data arrival rate that can be supported. Finally, in Figure 5.11 we see that increasing the average power can potentially improve the system performance in terms of the buffering delay.

6

CONCLUSIONS AND FUTURE WORK

In this thesis, we have explored the cross-layer interaction between the physical and data-link layers in different wireless networks. Specifically, we have investigated the impacts of different parameters at the physical layer when statistical quality of service constraints are imposed at the data-link layer as limits on the buffer overflow and delay violation probabilities. Noting that such constraints are crucial for 5G networks, in this thesis we have mainly focused on three wireless technologies that are foreseen as disruptive technologies for 5G and beyond 5G. These technologies are massive multiple-input multiple-output (MIMO), power-domain non-orthogonal multiple access (NOMA), and visible light communications (VLC).

First, we have studied the throughput and energy efficiency in a general class of MIMO systems with arbitrary inputs when they are subject to statistical QoS constraints. In this part, we have adopted the effective capacity, which defines the maximum *constant* arrival rate that the service (channel) process can support while satisfying the required QoS needs, as the main performance metric. We have obtained the optimal power allocation policies across transmit antennas when there is a short-term average power constraint. Moreover, we have analyzed the system performance in the low signal-to-noise ratio and massive-antenna (massive MIMO) regimes. We have attained the first and second derivatives of the effective capacity when the signal-to-noise ratio approaches zero. Using these characterizations, we have revealed that the minimum energy-per-bit does not depend on the input distribution and the QoS constraints but the slope does. In the massive MIMO regime, we have identified that the gap between the effective capacity and the average transmission rate in the channel decreases with the increasing number of antennas.

Second, we have investigated the optimal power allocation policies that maximize the effective capacity region of a two-user power-domain NOMA system with arbitrarily distributed input signals. We have formulated the relationship between the minimum mean square error (MMSE) and the first derivative of the mutual information with respect to the power allocation policies. We have provided an algorithm that determines the optimal normalized power allocation policies. We have established the optimal decision region boundaries for successive interference cancellation at the receiver for given power allocation policies. Through numerical techniques, we have justified that the Gaussian input signaling has better performance and that the performance gap increases in higher signal-to-noise ratio regime.

Third, we have explored the performance of a VLC system with a fixed-rate transmission and an ON-OFF data source when statistical QoS constraints are applied as limits on the buffer overflow and delay violation probabilities. Regarding the physical and data-link layers, we have provided

a cross-layer study by regarding the maximum average arrival rate at the transmitter buffer and the non-asymptotic delay bounds as the performance measures. To this end, we have adopted a two-state (ON-OFF) Markov process to model the fixed-rate transmission scenario, where the channel is assumed in the ON (OFF) state when a reliable transmission is (not) guaranteed. As the main conclusion in this part of the thesis, we have shown that transmitting with fixed rates is preferred when stricter QoS constraints are required. This indeed has a potential impact in practice since transmitting with a fixed rate can significantly simplify the implementation complexity.

Fourth, and finally we have studied the performance of a multi-mechanism transmission scenario in which RF and VLC technolgies can be used for data transmission in the same indoor environment. Following the cross-layer approach to concern the statistical QoS constraints at the data-link layer, we have employed the maximum average arrival rate at the transmitter buffer and the non-asymptotic delay bounds as the main performance measures. As the access technology is the main physical parameter to be regarded in this part, we have proposed and analyzed three strategies in which RF and VLC links are utilized for data transmission. We have further formulated the performance levels achieved by each of the proposed strategies. We have demonstrated that RF technology can be beneficial when there are lower average power constraints and/or looser QoS requirements. Moreover, we have shown that utilizing the VLC technology for data transmission, either alone or in a hybrid transmission strategy, can potentially enhance the system performance in terms of delay performance. It lowers the buffering delay bounds, when compared to the RF technology. Particularly, when data arrival rates at the transmitter buffer is low, VLC links provide lower queueing delays than RF links do, but RF links support higher data arrival rates at the transmitter buffer.

6.1 FUTURE WORK

It is obvious that cross-layer analysis regarding the physical and data-link layers has been gaining an increasing attention in the recent years. In this thesis, we have tried to fill some gabs in the existing literature studies concerning cross-layer analysis for some of the new wireless technologies suggested for 5G and beyond 5G. Nevertheless, there are still other research opportunities in this direction, some of which we highlight in the following as future research ideas.

- We have noticed a gap in the existing studies concerning cross-layer analysis for multi-carrier, multi-user and multi-antenna systems. In this direction, the authors in [10] considered an orthogonal frequency division multiplexing (OFDM)-based multicarrier, multiuser, and multi-antenna (MIMO) system and derived an adaptive resource allocation method. This method can jointly adapt subcarrier allocation, power distribution, and bit distribution with respect to instantaneous

channel conditions to maximize the system energy efficiency, while no delay concerns have been regarded.

- Although OFDM is simple and effective against inter-symbol interference, it suffers from many problems such as high peak-to-average ratio, poor spectral efficiency due to the insertion of cyclic prefix, and high out-of-band radiations. To overcome such problems, some alternatives for OFDM have been proposed in the literature. Filter band multi-carrier (FBMC) [187, 188] and generalized OFDM (GOFDM) [189] are the most promising candidates in this regard. However, to the best of our knowledge the cross-layer aspects of these techniques have not been yet investigated.

- In the hybrid RF/VLC scenario we considered a selection process from the user-perspective only, which means that the selected link is the one that satisfies the user's requirements. However, access points (APs) can also have different selection criteria based on the system design and requirements, e.g., to serve more users or to support certain QoS levels. Such concerns are practically important in multi-user and multi-AP scenarios in which different users also have different QoS needs. Here, cross-layer-based load balancing techniques are required to optimize the system design. In this framework, machine learning methods can be employed.

Part II

APPENDIX

A

PROOF OF THEOREM 1

Note that the logarithm in (2.8) is a monotonic function of $\mathcal{I}(\mathbf{x}_t;\mathbf{y}_t)$. Hence, we can write the optimization problem as

$$\min_{\mathbf{K}_t} \mathbb{E}_{\widehat{\mathbf{H}}}\left\{e^{-\theta T\mathcal{I}(\mathbf{x}_t;\mathbf{y}_t)}\right\} \tag{A.1}$$

such that

$$\mathbf{tr}\{\mathbf{K}_t\} \leqslant 1 \quad \text{and} \quad \mathbf{K}_t \succeq 0.$$

Subsequently, we form the Lagrange function as

$$\mathcal{L}(\mathbf{K}_t,\lambda,\Phi) = \mathbb{E}_{\widehat{\mathbf{H}}}\big\{e^{-\theta T\mathcal{I}(\mathbf{x}_t;\mathbf{y}_t)} - \lambda(1-\mathbf{tr}\{\mathbf{K}_t\}) - \mathbf{tr}\{\Phi\mathbf{K}_t\}\big\},$$

where λ and $\Phi \succeq 0$ are the Lagrange multipliers to the problem constraints. Then, evaluating its gradient with respect to $\mathbf{K}_t$, we obtain

$$-\theta T e^{-\theta T\mathcal{I}(\mathbf{x}_t;\mathbf{y}_t)}\frac{\partial\mathcal{I}(\mathbf{x}_t;\mathbf{y}_t)}{\partial\mathbf{K}_t} + \lambda\mathbf{I} - \Phi = 0, \tag{A.2}$$

where $\lambda(1-\mathbf{tr}\{\mathbf{K}_t\}) = 0$ for $\lambda \geqslant 0$, and $\mathbf{tr}\{\Phi\mathbf{K}_t\} = 0$ for $\Phi \succeq 0$ and $\mathbf{K}_t \succeq 0$. Moreover, we know from [116, Eq. (25)] that

$$\frac{\partial\mathcal{I}(\mathbf{x}_t;\mathbf{y}_t)}{\partial\mathbf{K}_t}\mathbf{K}_t = P\widehat{\mathbf{H}}^\dagger\boldsymbol{\Sigma}_{\widetilde{\mathbf{w}}}^{-1}\widehat{\mathbf{H}}\mathbf{mmse}_t. \tag{A.3}$$

Since we consider the worst-case noise assumption, we have $\boldsymbol{\Sigma}_{\widetilde{\mathbf{w}}} = \sigma_{\widetilde{w}}^2\mathbf{I}_{N\times N}$ as the noise covariance matrix.

Now, plugging (A.3) into (A.2), we have

$$-\theta T e^{-\theta T\mathcal{I}(\mathbf{x}_t;\mathbf{y}_t)}\gamma\widehat{\mathbf{H}}_t^\dagger\widehat{\mathbf{H}}_t\mathbf{mmse}_t + \lambda\mathbf{K}_t - \Phi\mathbf{K}_t = 0. \tag{A.4}$$

where $\gamma = P/\sigma_{\widetilde{w}}^2$. Moreover, we can further express (A.4) by multiplying both sides with $\mathbf{K}_t^{\frac{1}{2}}$ as follows:

$$-\theta T e^{-\theta T\mathcal{I}(\mathbf{x}_t;\mathbf{y}_t)}\gamma\mathbf{K}_t^{\frac{1}{2}}\widehat{\mathbf{H}}_t^\dagger\widehat{\mathbf{H}}_t\mathbf{mmse}_t + \lambda\mathbf{K}_t^{\frac{3}{2}} - \mathbf{K}_t^{\frac{1}{2}}\Phi\mathbf{K}_t^{\frac{1}{2}}\mathbf{K}_t^{\frac{1}{2}} = 0.$$

Noting that $\mathbf{tr}\{\Phi\mathbf{K}_t\} = \mathbf{tr}\{\mathbf{K}_t^{\frac{1}{2}}\Phi\mathbf{K}_t^{\frac{1}{2}}\} = 0$, we know that $\mathbf{K}_t^{\frac{1}{2}}\Phi\mathbf{K}_t^{\frac{1}{2}}$ is forced to be a null matrix [111]. Consequently, the optimal input covariance matrix, $\mathbf{K}_t \succeq 0$, is the solution of the following expression:

$$\mathbf{K}_t = \frac{\theta T\gamma e^{-\theta T\mathcal{I}(\mathbf{x}_t;\mathbf{y}_t)}}{\lambda}\widehat{\mathbf{H}}_t^\dagger\widehat{\mathbf{H}}_t\mathbf{mmse}_t. \tag{A.5}$$

This concludes the proof.

B

PROOF OF THEOREM 2

With the input-output channel model given in (2.10), we have component-wise independent channels, i.e.,

$$\widetilde{y}_t(i) = \sqrt{\gamma d_t(i)}\widetilde{x}_t(i) + \widetilde{n}_t(i) \text{ for } i = 1, \cdots, \min\{M, N\},$$

where $\sqrt{d_t(i)}$ is the i^{th} non-zero diagonal of $\mathbf{D}_t$ and $d_t(i)$ is the i^{th} eigenvalue of $\widehat{\mathbf{H}}\widehat{\mathbf{H}}^\dagger$ and $\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}}$. Above, $\widetilde{x}_t(i)$, $\widetilde{y}_t(i)$ and $\widetilde{n}_t(i)$ are the i^{th} element of the input, output and noise vectors, respectively. We note that $\widetilde{x}_t(i) = 0$, $\widetilde{y}_t(i) = 0$, and $\widetilde{n}_t(i) = 0$ for $i > \min\{N, M\}$. Moreover, because we have $\mathcal{I}(\mathbf{x}_t; \mathbf{y}_t) = \mathcal{I}(\widetilde{\mathbf{x}}_t; \widetilde{\mathbf{y}}_t)$, the logarithm in (2.8) is a monotonic function of $\mathcal{I}(\widetilde{\mathbf{x}}_t; \widetilde{\mathbf{y}}_t)$ as well. Returning to the optimization problem in (A.1), we can see that when the minimum is obtained, $\mathcal{I}(\widetilde{\mathbf{x}}_t; \widetilde{\mathbf{y}}_t)$ is maximized at every time instant. So, the samples of $\widetilde{\mathbf{x}}$ should be independent of each other [9]. In particular, we should have $\widetilde{\mathbf{K}}_t = \mathbf{\Sigma}_t$, which is an $N \times M$ diagonal matrix with non-negative elements, $\{\sigma_t(i)\}|_{i=1}^{\min\{N,M\}}$. Consequently, the optimization problem becomes

$$\min_{\mathbf{\Sigma}_t} \mathbb{E}_{\widehat{\mathbf{H}}}\left\{e^{-\theta T \mathcal{I}(\widetilde{\mathbf{x}}_t; \widetilde{\mathbf{y}}_t)}\right\} \tag{B.1}$$

such that $\mathbf{tr}\{\mathbf{\Sigma}_t\} = \mathbf{tr}\{\widetilde{\mathbf{K}}_t\} = \mathbf{tr}\{\mathbf{V}_t^\dagger \mathbf{K}_t \mathbf{V}_t\} \leqslant 1$.

Herein, we benefit from the fact that the trace of a matrix is the sum of its eigenvalues and the fact that $\widetilde{\mathbf{K}}_t$ and $\mathbf{K}_t$ have the same eigenvalues because $\mathbf{V}_t$ is a unitary matrix. Subsequently, forming the Lagrange function as

$$\mathcal{L}(\Sigma_t, \lambda, \Phi) = \mathbb{E}_{\widehat{\mathbf{H}}}\{e^{-\theta T \mathcal{I}(\widetilde{\mathbf{x}}_t; \widetilde{\mathbf{y}}_t)} - \lambda(1 - \mathbf{tr}\{\mathbf{\Sigma}_t\}) - \mathbf{tr}\{\Phi \mathbf{\Sigma}_t\}\},$$

where λ and $\Phi \succeq 0$ are the Lagrange multipliers associated with the problem constraints, and taking its derivatives with respect to $\{\sigma_t(i)\}|_{i=1}^{\min\{M,N\}}$, we obtain

$$-\theta T e^{-\theta T \mathcal{I}(\widetilde{\mathbf{x}}_t, \widetilde{\mathbf{y}}_t)} \frac{\partial \mathcal{I}(\widetilde{x}_t(i); \widetilde{y}_t(i))}{\partial \sigma_t(i)} + \lambda - \phi(i, i) = 0, \tag{B.2}$$

where $\widetilde{x}_t(i)$ and $\widetilde{y}_t(i)$ are the i^{th} elements of $\widetilde{\mathbf{x}}_t$ and $\widetilde{\mathbf{y}}_t$, respectively, and $\phi(i, i)$ is the i^{th} diagonal element of Φ. Since $\mathbf{tr}\{\Phi\Sigma\} = 0$, we have $\phi(i, i) = 0$. Moreover, using [116, Eq. (25)], we show that

$$\frac{\partial \mathcal{I}(\widetilde{x}_t(i); \widetilde{y}_t(i))}{\partial \sigma_t(i)} = \frac{\gamma d_t(i)}{\sigma_t(i)} \mathbf{mmse}_t(i),$$

where

$$\mathbf{mmse}_t(i) = \mathbb{E}\left\{|\mathbb{E}\{\widetilde{x}_t(i)|\widetilde{y}_t(i)\} - \widetilde{x}_t(i)|^2\right\}.$$

Then, given $\sigma_t(i) \geqslant 0$, we have the optimal $\sigma_t(i)$ as the solution of the following:

$$\sigma_t(i) = \frac{\theta T \gamma d_t(i)}{\lambda} e^{-\theta T J(\tilde{x}_t, \tilde{y}_t)} \mathbf{mmse}_t(i). \tag{B.3}$$

If the solution in (B.3) is negative, we set $\sigma_t(i) = 0$. We further note that $e^{-\theta T J(\tilde{x}_t, \tilde{y}_t)}$ and $\mathbf{mmse}_t(i)$ are convex and monotonically decreasing functions of $\sigma_t(i)$. Therefore, the right-hand-side of (B.3) is also a convex function of $\sigma_t(i)$ and it is monotonically decreasing. Hence, it provides a unique and global solution.

C

PROOF OF THEOREM 3

The first derivative of the effective rate in (2.6), $R_E(\theta, P)$, with respect to the transmission power, P, when P approaches 0, is

$$\dot{R}_E(\theta, 0) = \lim_{P \to 0} \frac{\mathbb{E}_{\widehat{\mathbf{H}}} \left\{ \dot{\mathcal{I}}(P) e^{-\theta T \mathcal{I}(P)} \right\}}{N \mathbb{E}_{\widehat{\mathbf{H}}} \left\{ e^{-\theta T \mathcal{I}(P)} \right\}}, \tag{C.1}$$

where $\mathcal{I}(P)$ and $\dot{\mathcal{I}}(P)$ are the mutual information and its derivative, respectively, as a function of P. Noting the worst-case noise assumption, we can re-express (2.4) as follows:

$$\mathbf{y}_t = \sqrt{P} \widehat{\mathbf{H}}_t \mathbf{x}_t + \sqrt{P} \sigma_e \mathbf{n}_t + \mathbf{w}_t, \tag{C.2}$$

where $\mathbf{n}_t$ has zero-mean and unit-variance Gaussian random elements. Because $\widehat{\mathbf{H}}_t$ and $\widetilde{\mathbf{H}}_t$, and hence $\mathbf{n}_t$, are uncorrelated, we can see that the channel model in (C.2) is similar to the channel model described in [190, Eq. 7]. Therefore, the lower bound on the mutual information in the low signal-to-noise ratio regime, i.e., as $P \to 0$, is expressed as [190, Eq. 64]

$$\begin{aligned} \mathcal{I}(\mathbf{x}_t; \mathbf{y}_t) &= \frac{P}{\log_e 2} \mathbf{tr}\{\widehat{\mathbf{H}}_t \mathbf{K} \widehat{\mathbf{H}}_t^\dagger\} \\ &- \frac{P^2}{2 \log_e 2} \mathbf{tr}\{[\widehat{\mathbf{H}}_t \mathbf{K} \widehat{\mathbf{H}}_t^\dagger]^2 + 2\sigma_e^2 \widehat{\mathbf{H}}_t \mathbf{K} \widehat{\mathbf{H}}_t^\dagger\} + \mathcal{O}(P^2). \end{aligned} \tag{C.3}$$

Then, the first derivative of $\mathcal{I}(\mathbf{x}_t; \mathbf{y}_t)$ with respect to P in the low signal-to-noise ratio regime becomes

$$\dot{\mathcal{I}}(P) = \frac{\mathbf{tr}\{\widehat{\mathbf{H}}_t \mathbf{K} \widehat{\mathbf{H}}_t^\dagger\}}{\log_e 2} - \frac{P}{\log_e 2} \mathbf{tr}\{[\widehat{\mathbf{H}}_t \mathbf{K} \widehat{\mathbf{H}}_t^\dagger]^2 + 2\sigma_e^2 \widehat{\mathbf{H}}_t \mathbf{K} \widehat{\mathbf{H}}_t^\dagger\} + \mathcal{O}(P^2). \tag{C.4}$$

Then, we can re-write (C.1) as

$$\dot{R}_E(\theta, 0) = \frac{\mathbb{E}_{\widehat{\mathbf{H}}}\{\mathbf{tr}\{\widehat{\mathbf{H}} \mathbf{K} \widehat{\mathbf{H}}^\dagger\}\}}{N \log_e 2}. \tag{C.5}$$

We can easily observe that $\mathcal{I}(P) = 0$ when $P = 0$, and hence $e^{-\theta T \mathcal{I}(\gamma)} = 1$ in (C.1). Moreover, since the input covariance matrix, $\mathbf{K}$, is a positive semi-definite Hermitian matrix, we can express $\mathbf{K}$ as [191]

$$\mathbf{K} = \mathbf{U} \Sigma \mathbf{U}^\dagger = \sum_{i=1}^{M} \sigma_i \mathbf{u}_i \mathbf{u}_i^\dagger, \tag{C.6}$$

where $\mathbf{U}$ is the unitary matrix and Σ is the diagonal matrix. The unitary matrix is formed by the set of the eigenvectors of $\mathbf{K}$, i.e., $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_M]$, and the diagonal matrix is composed of the eigenvalues of $\mathbf{K}$ corresponding

to its eigenvectors, i.e., $\Sigma = \text{diag}\{\sigma_1, \cdots, \sigma_M\}$. Moreover, the eigenvectors form an orthonormal space, i.e., $\mathbf{u}_i^\dagger \mathbf{u}_j = 1$ for $i = j$ and $\mathbf{u}_i^\dagger \mathbf{u}_i = 0$ for $i \neq j$, and eigenvalues are greater than or equal to zero, i.e., $\sigma_i \geqslant 0$. Here, we assume that the system uses all the available energy for transmission, i.e., $\mathbf{tr}\{\mathbf{K}\} = 1$, and hence, we have $\sum_{i=1}^{M} \sigma_i = 1$. Now, we have the following:

$$\dot{R}_E(\theta, 0) = \frac{1}{N \log_e 2} \mathbb{E}_{\widehat{\mathbf{H}}}\{\mathbf{tr}\{\widehat{\mathbf{H}}\mathbf{K}\widehat{\mathbf{H}}^\dagger\}\} \tag{C.7}$$

$$= \frac{1}{N \log_e 2} \sum_{i=1}^{M} \sigma_i \mathbb{E}_{\widehat{\mathbf{H}}}\{\mathbf{tr}\{\widehat{\mathbf{H}}\mathbf{u}_i\mathbf{u}_i^\dagger\mathbf{H}^\dagger\}\} \tag{C.8}$$

$$= \frac{1}{N \log_e 2} \sum_{i=1}^{M} \sigma_i \mathbb{E}_{\widehat{\mathbf{H}}}\{\mathbf{u}_i^\dagger \widehat{\mathbf{H}}^\dagger \widehat{\mathbf{H}}\mathbf{u}_i\} \tag{C.9}$$

$$\leqslant \frac{1}{N \log_e 2} \mathbb{E}_{\widehat{\mathbf{H}}}\{\lambda_{\max}(\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}})\} = \dot{C}_E(\theta, 0), \tag{C.10}$$

where $\lambda_{\max}(\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}})$ is the maximum eigenvalue of $\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}}$. Above, (C.9) follows from the fact that

$$\mathbf{tr}\{\widehat{\mathbf{H}}\mathbf{u}_i\mathbf{u}_i^\dagger\widehat{\mathbf{H}}^\dagger\} = \mathbf{tr}\{\mathbf{u}_i^\dagger\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}}\mathbf{u}_i\} = \mathbf{u}_i^\dagger\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}}\mathbf{u}_i,$$

where $\mathbf{u}_i^\dagger\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}}\mathbf{u}_i$ is a scalar value. The upper bound in (C.10) can be achieved by choosing the normalized input covariance matrix as $\mathbf{K} = \mathbf{u}_{\max}\mathbf{u}_{\max}^\dagger$ and $\mathbf{u}_{\max}$ is the unit eigenvector of $\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}}$ that corresponds to the maximum eigenvalue of $\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}}$. This completes the first part of the proof.

The second derivative of the effective rate in (2.6), $R_E(\theta, P)$, with respect to the transmission power, P, when P approaches 0, is

$$\ddot{R}_E(\theta, 0) = \lim_{P\to 0} \frac{\mathbb{E}_{\widehat{\mathbf{H}}}\{\ddot{\jmath}(\gamma)e^{-\theta T \jmath(P)} - \theta T[\dot{\jmath}(P)]^2 e^{-\theta T \jmath(\gamma)}\}}{N\mathbb{E}_{\widehat{\mathbf{H}}}\{e^{-\theta T \jmath(\gamma)}\}} + \frac{\theta T \mathbb{E}^2_{\widehat{\mathbf{H}}}\left\{\dot{\jmath}(\gamma)e^{-\theta T \jmath(\gamma)}\right\}}{N\mathbb{E}^2_{\widehat{\mathbf{H}}}\left\{e^{-\theta T \jmath(\gamma)}\right\}}, \tag{C.11}$$

where $\ddot{\jmath}(P)$ is the second derivative of the lower bound on the mutual information with respect to P. From (C.3), we have

$$\ddot{\jmath}(0) = -\frac{1}{\log_e 2}\mathbf{tr}\{[\widehat{\mathbf{H}}_t\mathbf{K}\widehat{\mathbf{H}}_t^\dagger]^2\} - \frac{2}{\log_e 2}\mathbf{tr}\{\sigma_e^2\widehat{\mathbf{H}}_t\mathbf{K}\widehat{\mathbf{H}}_t^\dagger\}.$$

Then, we can re-write (C.11) as

$$\ddot{R}_E(\theta, 0) = \frac{\theta T}{N \log_e^2 2}\left[\mathbb{E}^2_{\widehat{\mathbf{H}}}\left\{\mathbf{tr}\{\widehat{\mathbf{H}}\mathbf{K}\widehat{\mathbf{H}}^\dagger\}\right\} - \mathbb{E}_{\widehat{\mathbf{H}}}\left\{\mathbf{tr}^2\{\widehat{\mathbf{H}}\mathbf{K}\widehat{\mathbf{H}}^\dagger\}\right\}\right] - \frac{1}{N \log_e 2}\mathbb{E}_{\widehat{\mathbf{H}}}\left\{\mathbf{tr}\{[\widehat{\mathbf{H}}\mathbf{K}\widehat{\mathbf{H}}^\dagger]^2 + 2\sigma_e^2\widehat{\mathbf{H}}\mathbf{K}\widehat{\mathbf{H}}^\dagger\}\right\}. \tag{C.12}$$

Now, let l be the multiplicity of $\lambda_{\max}(\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}})$. Hence, we can re-express $\mathbf{K}$ as follows: $\mathbf{K} = \sum_{i=1}^{l} \sigma_i\mathbf{u}_i\mathbf{u}_i^\dagger$, where $\sigma_i \in [0, 1]$ and $\sum_{i=1}^{l} \sigma_i = 1$. Above, $\{\mathbf{u}_i\}_{i=1}^{l}$ are the corresponding column vectors. Hence, we can show that

$$\mathbb{E}^2_{\widehat{\mathbf{H}}}\{\mathbf{tr}\{\widehat{\mathbf{H}}\mathbf{K}\widehat{\mathbf{H}}^\dagger\}\} = \mathbb{E}^2_{\widehat{\mathbf{H}}}\{\lambda_{\max}(\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}})\}$$

and

$$\mathbb{E}_{\widehat{\mathbf{H}}}\{\mathbf{tr}^2(\widehat{\mathbf{H}}\mathbf{K}\widehat{\mathbf{H}}^\dagger)\} = \mathbb{E}_{\widehat{\mathbf{H}}}\{\lambda^2_{\max}(\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}})\}.$$

Moreover, we have

$$\mathbb{E}_{\widehat{\mathbf{H}}}\{\mathbf{tr}\{[\widehat{\mathbf{H}}\mathbf{K}\widehat{\mathbf{H}}^\dagger]^2\}\} = \mathbb{E}_{\widehat{\mathbf{H}}}\{\mathbf{tr}\{\widehat{\mathbf{H}}\mathbf{K}\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}}\mathbf{K}\widehat{\mathbf{H}}^\dagger\}\} \tag{C.13}$$

$$= \mathbb{E}_{\widehat{\mathbf{H}}}\left\{\mathbf{tr}\left\{\widehat{\mathbf{H}}\sum_{i=1}^{l}\sigma_i\mathbf{u}_i\mathbf{u}_i^\dagger\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}}\sum_{j=1}^{l}\sigma_j\mathbf{u}_j\mathbf{u}_j^\dagger\widehat{\mathbf{H}}^\dagger\right\}\right\} \tag{C.14}$$

$$= \mathbb{E}_{\widehat{\mathbf{H}}}\left\{\mathbf{tr}\left\{\sum_{i,j}^{l}\sigma_i\sigma_j\mathbf{u}_i^\dagger\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}}\mathbf{u}_j\mathbf{u}_j^\dagger\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}}\mathbf{u}_i\right\}\right\} \tag{C.15}$$

$$= \mathbb{E}_{\widehat{\mathbf{H}}}\left\{\sum_{i,j}^{l}\sigma_i\sigma_j\left|\mathbf{u}_i^\dagger\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}}\mathbf{u}_j\right|^2\right\} \tag{C.16}$$

$$= \mathbb{E}_{\widehat{\mathbf{H}}}\left\{\sum_{i=1}^{l}\sigma_i^2\left|\mathbf{u}_i^\dagger\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}}\mathbf{u}_i\right|^2\right\} \tag{C.17}$$

$$= \mathbb{E}_{\widehat{\mathbf{H}}}\left\{\lambda^2_{\max}(\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}})\sum_{i=1}^{l}\sigma_i^2\right\} \tag{C.18}$$

$$\geqslant \frac{1}{l}\mathbb{E}_{\widehat{\mathbf{H}}}\left\{\lambda^2_{\max}(\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}})\right\}. \tag{C.19}$$

Above, (C.15) comes from the fact that $\mathbf{tr}\{\mathbf{AB}\} = \mathbf{tr}\{\mathbf{BA}\}$, where $\mathbf{A}$ and $\mathbf{B}$ are matrices. Moreover, since $\mathbf{u}_i^\dagger\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}}\mathbf{u}_j$ and $\mathbf{u}_j^\dagger\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}}\mathbf{u}_i$ are the complex conjugates of each other, we have the result in (C.16). Noting that $\mathbf{u}_i$ and $\mathbf{u}_j$ are orthonormal to each other, i.e., $\mathbf{u}_i^\dagger\mathbf{u}_j = 0$ given $i \neq j$ and $\mathbf{u}_i^\dagger\mathbf{u}_j = 1$ given $i = j$, we have (C.17). Moreover, we know that

$$\lambda^2_{\max}(\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}}) = \mathbf{u}_i^\dagger\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}}\mathbf{u}_i.$$

Subsequently, we have (C.18). Finally, $\sum_i^l \sigma_i^2$ is minimized when $\sigma_i = \frac{1}{l}$. Therefore, we have the lower bound in (C.19). As a result, the second derivative of the effective rate, $\ddot{R}(\theta, P)$, when P diminishes to zero, is upper-bounded as follows:

$$\begin{aligned}\ddot{R}_E(\theta,0) \leqslant\ & \frac{\theta T}{N\log_e^2 2}\left[\mathbb{E}^2_{\widehat{\mathbf{H}}}\{\lambda_{\max}(\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}})\} - \mathbb{E}_{\widehat{\mathbf{H}}}\{\lambda^2_{\max}(\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}})\}\right] \\ & - \frac{\mathbb{E}_{\widehat{\mathbf{H}}}\{\lambda^2_{\max}(\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}})\}}{lN\log_e 2} - \frac{2\sigma_e^2}{N\log_e 2}\mathbb{E}_{\widehat{\mathbf{H}}}\{\lambda_{\max}(\widehat{\mathbf{H}}^\dagger\widehat{\mathbf{H}})\} \\ =\ & \ddot{C}_E(\theta,0),\end{aligned}$$

which completes the second part of the proof.

D

PROOF OF THEOREM 4

Given an input covariance matrix, $\mathbf{K}$, the instantaneous mutual information between the channel input and output, defined in (2.5), can be expressed as follows:

$$\begin{aligned} r &= \mathbb{E}_{\mathbf{x},\mathbf{y}}\left\{ \log_2 \frac{f_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})}{f_{\mathbf{y}}(\mathbf{y})} \right\} \\ &= \mathbb{E}_{\mathbf{x},\mathbf{y}}\left\{ \log_2 f_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) \right\} - \mathbb{E}_{\mathbf{y}}\left\{ \log_2 f_{\mathbf{y}}(\mathbf{y}) \right\} \\ &= -\frac{N}{\log_e 2} - \mathbb{E}_{\mathbf{y}}\left\{ \log_2 \mathbb{E}_{\mathbf{x}}\left\{ e^{-\frac{1}{\sigma_{\tilde{w}}^2}||\mathbf{y}-\sqrt{P}\widehat{\mathbf{H}}\mathbf{x}||^2} \right\} \right\}. \end{aligned} \tag{D.1}$$

Now, by inserting (D.1) into (2.6) and taking the limit when M goes to infinity, we can write the effective rate as follows:

$$\lim_{M\to\infty} R_E(\theta, P) = \lim_{M\to\infty} -\frac{1}{\theta NT}$$
$$\times \log_e \mathbb{E}_{\widehat{\mathbf{H}}}\left\{ e^{\frac{\theta TN}{\log_e 2}} e^{\theta T \mathbb{E}_{\mathbf{y}}\left\{ \log_2 \mathbb{E}_{\mathbf{x}}\left\{ e^{-\frac{1}{\sigma_{\tilde{w}}^2}||\mathbf{y}-\sqrt{P}\widehat{\mathbf{H}}\mathbf{x}||^2} \right\} \right\}} \right\} \tag{D.2}$$

$$= \lim_{M\to\infty} \left\{ -\frac{1}{\log_e 2} - \frac{1}{\theta NT} \log_e \right.$$
$$\left. \mathbb{E}_{\widehat{\mathbf{H}}}\left\{ e^{\theta T \mathbb{E}_{\mathbf{y}}\left\{ \log_2 \mathbb{E}_{\mathbf{x}}\left\{ e^{-\frac{1}{\sigma_{\tilde{w}}^2}||\mathbf{y}-\sqrt{P}\widehat{\mathbf{H}}\mathbf{x}||^2} \right\} \right\}} \right\} \right\} \tag{D.3}$$

$$= \lim_{M\to\infty} \left\{ -\frac{1}{\log_e 2} - \frac{1}{\theta NT} \log_e \right.$$
$$\left. \mathbb{E}_{\widehat{\mathbf{H}}}\left\{ e^{\theta TM \mathbb{E}_{\mathbf{y}}\left\{ \frac{1}{M} \log_2 \mathbb{E}_{\mathbf{x}}\left\{ e^{-\frac{1}{\sigma_{\tilde{w}}^2}||\mathbf{y}-\sqrt{P}\widehat{\mathbf{H}}\mathbf{x}||^2} \right\} \right\}} \right\} \right\} \tag{D.4}$$

$$= \lim_{M\to\infty} \left\{ -\frac{1}{\log_e 2} - \frac{1}{\theta NT} \log_e \right.$$
$$\left. \mathbb{E}_{\widehat{\mathbf{H}}}\left\{ e^{\theta TM \mathbb{E}_{\mathbf{y},\mathbf{H}}\left\{ \frac{1}{M} \log_2 \mathbb{E}_{\mathbf{x}}\left\{ e^{-\frac{1}{\sigma_{\tilde{w}}^2}||\mathbf{y}-\sqrt{P}\widehat{\mathbf{H}}\mathbf{x}||^2} \right\} \right\}} \right\} \right\} \tag{D.5}$$

$$= \lim_{M\to\infty} -\frac{1}{\log_e 2}$$
$$-\frac{M}{N}\mathbb{E}_{\mathbf{y},\widehat{\mathbf{H}}}\left\{ \frac{1}{M} \log_2 \mathbb{E}_{\mathbf{x}}\left\{ e^{-\frac{1}{\sigma_{\tilde{w}}^2}||\mathbf{y}-\sqrt{P}\widehat{\mathbf{H}}\mathbf{x}||^2} \right\} \right\} \tag{D.6}$$

$$= \lim_{M\to\infty} \frac{1}{N}\mathbb{E}_{\widehat{\mathbf{H}}}\left\{ -\frac{N}{\log_e 2} \right.$$
$$\left. -\mathbb{E}_{\mathbf{y}}\left\{ \log_2 \mathbb{E}_{\mathbf{x}}\left\{ e^{-\frac{1}{\sigma_{\tilde{w}}^2}||\mathbf{y}-\sqrt{P}\widehat{\mathbf{H}}\mathbf{x}||^2} \right\} \right\} \right\} \tag{D.7}$$

$$= \lim_{M\to\infty} \frac{1}{N}\mathbb{E}_{\widehat{\mathbf{H}}}\{r\}. \tag{D.8}$$

In (D.4), we benefit from the connection between the free energy and the mutual information and employ the *self-averaging* property, which provides us the following [192]:

$$\begin{aligned}&\lim_{M\to\infty} \mathbb{E}_{\mathbf{y}}\left\{\frac{1}{M}\log_2 \mathbb{E}_{\mathbf{x}}\left\{e^{-\frac{1}{\sigma_{\tilde{w}}^2}||\mathbf{y}-\sqrt{P}\widehat{\mathbf{H}}\mathbf{x}||^2}\right\}\right\}\\ &= \lim_{M\to\infty} \mathbb{E}_{\mathbf{y},\widehat{\mathbf{H}}}\left\{\frac{1}{M}\log_2 \mathbb{E}_{\mathbf{x}}\left\{e^{-\frac{1}{\sigma_{\tilde{w}}^2}||\mathbf{y}-\sqrt{P}\widehat{\mathbf{H}}\mathbf{x}||^2}\right\}\right\},\end{aligned} \tag{D.9}$$

which is a result of the assumption of the *self-averaging* property, in which the free energy converges in probability to its expectation over the distribution of the random variables $\widehat{\mathbf{H}}$ and $\mathbf{y}$ in the large-system limit [192]. Moreover, the expression inside the first bracket in (D.7) is same with the expression in (D.1), we have the result in (D.8). Then, we have

$$\lim_{M\to\infty} R_E(\theta, P) = \lim_{M\to\infty} \frac{1}{N}\mathbb{E}_{\widehat{\mathbf{H}}}\{r\}.$$

Similarly, when N goes to infinity or both M and N go to infinity, the solution is trivial. We can again use the reformulation performed in (D.4) and engage the property stated in (D.9). Since the aforementioned proof is valid for any input covariance matrix, we can complete the proof with (2.19).

E

PROOF OF THEOREM 5

Recall that $\alpha_1 = \frac{P_1}{\overline{P}}$ and $\alpha_2 = \frac{P_2}{\overline{P}}$, and

$$f(y) = \sum_x p(x) f(y|x), \tag{E.1}$$

where $x = (x_1, x_2)$. Since our analysis is performed in the complex plane, we can express $f(y|x)$ as

$$f(y|x) = \frac{1}{\pi} \exp\Big\{ -\left(y_r - \sqrt{P_1} c_{1r} - \sqrt{P_2} c_{2r}\right)^2 - \left(y_i - \sqrt{P_1} c_{1i} - \sqrt{P_2} c_{2i}\right)^2 \Big\}, \tag{E.2}$$

where $y = y_r + j y_i$, $h_1 x_1 = c_{1r} + j c_{1i}$ and $h_2 x_2 = c_{2r} + j c_{2i}$. The derivative of the pdf with respect to P_1 is given as

$$\frac{df(y|x)}{dP_1} = \frac{f(y|x)}{\sqrt{P_1}} \begin{pmatrix} c_{1r} & c_{1i} \end{pmatrix} \begin{pmatrix} y_r - \sqrt{P_1} c_{1r} - \sqrt{P_2} c_{2r} \\ y_i - \sqrt{P_1} c_{1i} - \sqrt{P_2} c_{2i} \end{pmatrix},$$

and

$$\frac{df(y|x)}{dy} = \dot{f}(y|x) = -2 f(y|x) \begin{pmatrix} y_r - \sqrt{P_1} c_{1r} - \sqrt{P_2} c_{2r} \\ y_i - \sqrt{P_1} c_{1i} - \sqrt{P_2} c_{2i} \end{pmatrix}. \tag{E.3}$$

Hence, we have

$$\frac{df(y|x)}{dP_1} = \frac{-1}{2\sqrt{P_1}} \begin{pmatrix} c_{1r} & c_{1i} \end{pmatrix} \dot{f}(y|x). \tag{E.4}$$

Now, we can express [1]

$$\frac{d\mathcal{I}(x_1; y)}{dP_1} = \frac{d\mathcal{I}(x; y)}{dP_1} - \underbrace{\frac{d\mathcal{I}(x_2; y_2)}{dP_1}}_{=0}. \tag{E.5}$$

Invoking the marginal pdf $f(y|x)$ in (E.2), we can rewrite the mutual information $\mathcal{I}(x; y)$ expressed in (3.2) as

$$\mathcal{I}(x; y) = -\log(\pi e) - \int f(y) \log(f(y)) dy.$$

Consequently, we can write (E.5) as

$$\frac{d\mathcal{I}(x_1; y)}{dP_1} = \frac{d\mathcal{I}(x; y)}{dP_1} = -\frac{d}{dP_1} \int f(y) \log(f(y)) dy,$$

1 This is based on the known relation $\mathcal{I}(x; y) = \mathcal{I}(x_j; y) + \mathcal{I}(x_m; y_m)$ for $j, m \in \{1, 2\}$ and $j \neq m$ [46].

$$= -\int [1 + \log(f(y))] \frac{df(y)}{dP_1} dy. \tag{E.6}$$

Substituting (E.1) and (E.4) in (E.6), we obtain

$$\frac{d\mathcal{I}(x_1; y)}{dP_1} = \frac{1}{2\sqrt{P_1}} \sum_x p(x) \begin{pmatrix} c_{1r} & c_{1i} \end{pmatrix} \int [1 + \log(f(y))] \dot{f}(y|x) dy. \tag{E.7}$$

Let $m = [1 + \log(f(y))]$ and $dn = \dot{f}(y|x)dy$, then the integration in (E.7) can be evaluated using integration by part such that $\int m dn = mn \int n dm$. By noting that $mn = 0$ as $y \to \infty$, we can write (E.7) as

$$\frac{d\mathcal{I}(x_1; y)}{dP_1} = \frac{-1}{2\sqrt{P_1}} \sum_x p(x) \begin{pmatrix} c_{1r} & c_{1i} \end{pmatrix} \int \frac{f(y|x)}{f(y)} \dot{f}(y|x) dy. \tag{E.8}$$

By plugging (E.3) in (E.8), we have

$$\begin{aligned}
\frac{d\mathcal{I}(x_1; y)}{dP_1} &= \frac{1}{\sqrt{P_1}} \sum_x p(x) \begin{pmatrix} a_{1r} & a_{1i} \end{pmatrix} \int \frac{f(y|x)}{f(y)} \sum_x p(x) f(y|x) \\
&\quad \times \begin{pmatrix} y_r - \sqrt{P_1} a_{1r} - \sqrt{P_2} a_{2r} \\ y_i - \sqrt{P_1} a_{1i} - \sqrt{P_2} a_{2i} \end{pmatrix} dy, \\
&= \frac{1}{\sqrt{P_1}} \sum_x p(x) \begin{pmatrix} a_{1r} & a_{1i} \end{pmatrix} \int \frac{f(y|x)}{f(y)} \sum_x p(x) f(y|x) \begin{pmatrix} y_r \\ y_i \end{pmatrix} dy \\
&\quad - \sum_x p(x) \begin{pmatrix} a_{1r} & a_{1i} \end{pmatrix} \int \frac{f(y|x)}{f(y)} \sum_x p(x) f(y|x) \begin{pmatrix} a_{1r} \\ a_{1i} \end{pmatrix} dy \\
&\quad - \frac{\sqrt{P_2}}{\sqrt{P_1}} \sum_x p(x) \begin{pmatrix} a_{1r} & a_{1i} \end{pmatrix} \int \frac{f(y|x)}{f(y)} \sum_x p(x) f(y|x) \begin{pmatrix} a_{2r} \\ a_{2i} \end{pmatrix} dy, \\
&= \frac{1}{\sqrt{P_1}} \sum_x p(x) \begin{pmatrix} a_{1r} & a_{1i} \end{pmatrix} \int f(y|x) \begin{pmatrix} y_r \\ y_i \end{pmatrix} dy \\
&\quad - \int f(y) \begin{pmatrix} \hat{a}_{1r} & \hat{a}_{1i} \end{pmatrix} \begin{pmatrix} \hat{a}_{1r} \\ \hat{a}_{1i} \end{pmatrix} dy - \frac{\sqrt{P_2}}{\sqrt{P_1}} \int f(y) \begin{pmatrix} \hat{a}_{1r} & \hat{a}_{1i} \end{pmatrix} \begin{pmatrix} \hat{a}_{2r} \\ \hat{a}_{2i} \end{pmatrix} dy, \\
&= \frac{1}{\sqrt{P_1}} \sum_x p(x) \begin{pmatrix} a_{1r} & a_{1i} \end{pmatrix} \begin{pmatrix} \sqrt{P_1} a_{1r} + \sqrt{P_2} a_{2r} \\ \sqrt{P_1} a_{1i} + \sqrt{P_2} a_{2i} \end{pmatrix} \\
&\quad - \mathbb{E}\left\{\hat{a}_{1r}^2 + \hat{a}_{1i}^2\right\} - \frac{\sqrt{P_2}}{\sqrt{P_1}} \mathbb{E}\left\{\hat{a}_{1r}\hat{a}_{2r} + \hat{a}_{1i}\hat{a}_{2i}\right\}, \\
&= \mathbb{E}\left\{a_{1r}^2 + a_{1i}^2\right\} + \frac{\sqrt{P_2}}{\sqrt{P_1}} \mathbb{E}\left\{a_{1r}a_{2r} + a_{1i}a_{2i}\right\} \\
&\quad - \mathbb{E}\left\{\hat{a}_{1r}^2 + \hat{a}_{1i}^2\right\} - \frac{\sqrt{P_2}}{\sqrt{P_1}} \mathbb{E}\left\{\hat{a}_{1r}\hat{a}_{2r} + \hat{a}_{1i}\hat{a}_{2i}\right\}, \\
&= |h_1|^2 \mathbb{E}\left\{|x_1|^2\right\} + \frac{\sqrt{P_2}}{\sqrt{P_1}} \mathrm{Re}\left(h_1 h_2^\star \mathbb{E}\left\{x_1 x_2^\star\right\}\right)
\end{aligned}$$

$$
\begin{aligned}
&- |h_1|^2 \mathbb{E}\left\{|\widehat{x}_1|^2\right\} - \frac{\sqrt{P_2}}{\sqrt{P_1}} \mathrm{Re}\left(h_1 h_2^\star \mathbb{E}\{\widehat{x}_1 \widehat{x}_2^\star\}\right), \\
&= |h_1|^2 \mathrm{MMSE}(x_1; y) + \frac{\sqrt{P_2}}{\sqrt{P_1}} \mathrm{Re}\left(h_1 h_2^\star \mathbb{E}\{x_1 x_2^\star - \widehat{x}_1 \widehat{x}_2^\star\}\right).
\end{aligned}
$$

Now, by applying the chain rule $\frac{d(\cdot)}{d\alpha_1} = \overline{P} \frac{d(\cdot)}{dP_1}$, we have

$$
\frac{d\mathcal{I}(x_1; y)}{d\alpha_1} = \overline{P} z_1 \mathrm{MMSE}(x_1; y) + \overline{P} \sqrt{\frac{\alpha_2}{\alpha_1}} \mathrm{Re}\left(h_1 h_2^\star \mathbb{E}\{x_1 x_2^\star - \widehat{x}_1 \widehat{x}_2^\star\}\right).
$$

Note that when the data of Transmitter 2 is subtracted from the received signal y, i.e, $\alpha_2 = 0$, the above equation is reduced to

$$
\frac{d\mathcal{I}(x_1; y_1)}{d\alpha_1} = \overline{P} z_1 \mathrm{MMSE}(x_1; y_1).
$$

In a similar way, we can show that

$$
\frac{d\mathcal{I}(x_2; y)}{dP_2} = \overline{P} z_2 \mathrm{MMSE}(x_2; y) + \overline{P} \sqrt{\frac{\alpha_1}{\alpha_2}} \mathrm{Re}\left(h_2 h_1^\star \mathbb{E}\{x_2 x_1^\star - \widehat{x}_2 \widehat{x}_1^\star\}\right).
$$

and

$$
\frac{d\mathcal{I}(x_2; y_2)}{d\alpha_2} = \overline{P} z_2 \mathrm{MMSE}(x_2; y_2).
$$

F

PROOF OF THEOREM 6

We initially start by expressing the effective capacity of each user in (3.13) and (3.14) in the integration form and with respect to $z_2^\star$ and $\theta_1 = \theta_2 = \theta$ as

$$C_1(\theta, z_2^\star) = \frac{-1}{\theta n} \log_e \left(\int_0^\infty \int_{z_2^\star}^\infty e^{-\theta n \mathcal{I}(x_1; y_1|z_1)} p_z(z) dz_2 dz_1 + \int_0^\infty \int_0^{z_2^\star} e^{-\theta n \mathcal{I}(x; y|z_1, g(z_1))} e^{\theta n \mathcal{I}(x_2; y_2|g(z_1))} p_z(z) dz_2 dz_1 \right),$$

and

$$C_2(\theta, z_2^\star) = \frac{-1}{\theta n} \log_e \left(\int_0^\infty \int_0^{z_2^\star} e^{-\theta n \mathcal{I}(x_2; y_2|g(z_1))} p_z(z) dz_2 dz_1 + \int_0^\infty \int_{z_2^\star}^\infty e^{-\theta n \mathcal{I}(x; y|z_1, g(z_1))} e^{\theta n \mathcal{I}(x_1; y_1|z_1)} p_z(z) dz_2 dz_1 \right),$$

where $n = TB$. Let $\mathcal{B}(\hat{z_2}) = \lambda_1 C_1(\theta, \hat{z_2}) + \lambda_2 C_2(\theta, \hat{z_2})$, where $\hat{z_2} = z_2^\star + e\xi$, $z_2^\star$ is the optimal decoding function that solve the optimization problem (3.8), e is a constant and ξ represents an arbitrary deviation. Consequently, the following condition should be satisfied[193]:

$$\frac{d}{de} \mathcal{B}(\hat{z_2}) \Big|_{e=0} = 0. \tag{F.1}$$

By noting that this condition holds for any ξ and that $\frac{d\hat{z_2}}{de} = \xi$, solving (F.1) results in the following:

$$e^{-\theta n \mathcal{I}(x; y|z_1, z_2^\star)} \left\{ \frac{-\lambda_1}{\psi_1} e^{\theta n \mathcal{I}(x_2; y_2|z_2^\star)} + \frac{\lambda_2}{\psi_2} e^{\theta n \mathcal{I}(x_1; y_1|z_1)} \right\} = \frac{\lambda_2}{\psi_2} e^{-\theta n \mathcal{I}(x_2; y_2|z_2^\star)} - \frac{\lambda_1}{\psi_1} e^{-\theta n \mathcal{I}(x_1; y_1|z_1)}. \tag{F.2}$$

Now, let us denote $\mathcal{I}_{12} = \mathcal{I}(x; y|z_1, z_2^\star)$, $\mathcal{I}_1 = \mathcal{I}(x_1; y_1|z_1)$ and $\mathcal{I}_2 = \mathcal{I}(x_2; y_2|z_2^\star)$. Consequently, we can express (F.2) as

$$e^{-\theta n \mathcal{I}_{12}} \left\{ -\psi_2 \lambda_1 e^{\theta n \mathcal{I}_2} + \psi_1 \lambda_2 e^{\theta n \mathcal{I}_1} \right\} = -\psi_2 \lambda_1 e^{-\theta n \mathcal{I}_1} + \psi_1 \lambda_2 e^{-\theta n \mathcal{I}_2}. \tag{F.3}$$

Let us further define $A = e^{-\theta n \mathcal{I}_2}$ and $D = e^{-\theta n \mathcal{I}_1}$. Then, (F.3) can be rewritten as

$$e^{-\theta n \mathcal{I}_{12}} \left\{ \frac{-\psi_2 \lambda_1}{A} + \frac{\psi_1 \lambda_2}{D} \right\} = -\psi_2 \lambda_1 D + \psi_1 \lambda_2 A,$$

which can be further simplified as

$$e^{-\theta n \mathcal{I}_{12}} = AD = e^{-\theta n\{\mathcal{I}_1 + \mathcal{I}_2\}}. \tag{F.4}$$

Note that (F.4) implies that $\mathcal{I}_{12} = \mathcal{I}_1 + \mathcal{I}_2$ which is equivalent to having

$$\mathcal{I}(x; y|z_1, z_2^\star) = \mathcal{I}(x_1; y_1|z_1) + \mathcal{I}(x_2; y_2|z_2^\star).$$

G

DERIVATION OF $\rho_R(\theta)$ IN (5.12)

Recall that λ is the data arrival rate at the buffer when the data source is in the ON state, and that $p_{ON} = \frac{\beta}{\alpha+\beta}$ is the steady-state probability of the ON state. Thus, the average data arrival rate is $\lambda p_{ON} = \frac{\beta}{\alpha+\beta}\lambda$. Moreover, we have $\Lambda_a(\theta_r^\star) = -\Lambda_r(-\theta_r^\star)$ in steady-state, i.e.,

$$\log_e\left\{\frac{1-\beta+(1-\alpha)e^{\theta_r^\star\lambda}}{2}+\frac{\sqrt{[1-\beta+(1-\alpha)e^{\theta_r^\star\lambda}]^2-4(1-\alpha-\beta)e^{\theta_r^\star\lambda}}}{2}\right\}$$
$$=-\log_e\mathbb{E}_h\left\{e^{-\theta_r^\star R_l}\right\}. \tag{G.1}$$

Solving the aforementioned equation for λ with any given $\theta > 0$, we obtain

$$\lambda = \frac{1}{\theta}\log_e\left\{\frac{1-(1-\beta)\mathbb{E}_h\left\{e^{-\theta R_l}\right\}}{(1-\alpha)\mathbb{E}_h\{e^{-\theta R_l}\}-(1-\alpha-\beta)\mathbb{E}_h^2\{e^{-\theta R_l}\}}\right\}. \tag{G.2}$$

As a result, we formulate the maximum average data arrival rate as $\rho_r(\theta) = \frac{\beta}{\alpha+\beta}\lambda$ for any $\theta > 0$. As for $\rho_v(\theta)$, we set $\Lambda_a(\theta_v^\star) = -\Lambda_v(-\theta_v^\star)$ and follow the same steps.

H

PROOF OF PROPOSITION 2

Based on the link selection process, the VLC link is selected only when $\rho_v(\theta) > \rho_r(\theta)$, where $\rho_v(\theta)$ is the maximum average arrival rate supported by the VLC link given in (5.13) and $\rho_r(\theta)$ is the maximum average arrival rate supported by the RF link given in (5.12). Since logarithm is a monotonic increasing function, this condition is satisfied when we have

$$\begin{aligned} &\frac{e^{2\theta V} - (1-\beta)e^{\theta V}}{(1-\alpha)e^{\theta V} - (1-\alpha-\beta)} \\ &> \frac{1-(1-\beta)\mathbb{E}_h\left\{e^{-\theta R_l}\right\}}{(1-\alpha)\mathbb{E}_h\{e^{-\theta R_l}\} - (1-\alpha-\beta)\mathbb{E}_h^2\{e^{-\theta R_l}\}}, \end{aligned} \tag{H.1}$$

Now, let $\chi = \mathbb{E}_h\{e^{-\theta R_l}\}$ and $\mathcal{O} = e^{\theta V}$. Then, (H.1) can be expressed by the following quadratic inequality:

$$\mathcal{O}^2 - (1-\beta+(1-\alpha)\xi)\mathcal{O} + \xi(1-\alpha-\beta) > 0. \tag{H.2}$$

where $\xi = \frac{1-(1-\beta)\chi}{(1-\alpha)\chi-(1-\alpha-\beta)\chi^2}$. Solving the above equation results in two solutions:

$$\mathcal{O}_1 = \frac{1-\beta+(1-\alpha)\xi - \sqrt{[1-\beta+(1-\alpha)\xi]^2 - 4(1-\alpha-\beta)\xi}}{2} \tag{H.3}$$

and

$$\mathcal{O}_2 = \frac{1-\beta+(1-\alpha)\xi + \sqrt{[1-\beta+(1-\alpha)\xi]^2 - 4(1-\alpha-\beta)\xi}}{2} \tag{H.4}$$

where $\mathcal{O}_2 > \mathcal{O}_1$. V has two ranges $0 < V < \log_e\{\mathcal{O}_1\}$ and $V \geqslant \log_e\{\mathcal{O}_2\}$. Setting $\mathcal{O} = 1$ in (H.2), we have $\beta(1-\xi) > 0$. Note that $\xi > 1$ because

$$\rho_r(\theta) = \frac{\beta}{(\alpha+\beta)\theta}\log_e\{\xi\} > 0.$$

Hence, we have $\beta(1-\xi) < 0$, which implies that $\mathcal{O}_1 < 1$ and $\log_e\{\mathcal{O}_1\} < 0$. Therefore, we have only $V \geqslant \log_e\{\mathcal{O}_2\}$ as the solution region, which completes the proof.

BIBLIOGRAPHY

[1] Sriram Vishwanath, Syed Ali Jafar, and Andrea Goldsmith. Adaptive resource allocation in composite fading environments. In *Proc. IEEE Global Telecommun. (GLOBECOM)*, volume 2, pages 1312–1316, 2001.

[2] Jia Tang and Xi Zhang. Quality-of-service driven power and rate adaptation over wireless links. *IEEE Trans. Wireless Commun.*, 6(8), 2007.

[3] David N. C. Tse and Stephen V Hanly. Multiaccess fading channels. I. polymatroid structure, optimal resource allocation and throughput capacities. *IEEE Trans. Inform. Theory*, 44(7):2796–2815, 1998.

[4] Lifang Li and Andrea J Goldsmith. Capacity and optimal resource allocation for fading broadcast channels. II. outage capacity. *IEEE Trans. Inf. Theory*, 47(3):1103–1127, 2001.

[5] DN Tse. Optimal power allocation over parallel gaussian broadcast channels. In *Proc. Int. Sym. Inf. Theory*, page 27, 1997.

[6] Zukang Shen, Jeffrey G Andrews, and Brian L Evans. Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints. *IEEE Trans. Wireless Commun.*, 4(6):2726–2737, 2005.

[7] Mehdi Mohseni, Rui Zhang, and John M Cioffi. Optimized transmission for fading multiple-access and broadcast channels with multiple antennas. *IEEE J. Sel. Areas Commun.*, 24(8):1627–1639, 2006.

[8] Angel Lozano, Antonia Maria Tulino, and Sergio Verdú. Optimum power allocation for parallel Gaussian channels with arbitrary input distributions. *IEEE Trans. Inf. Theory*, 52(7):3033–3051, 2006.

[9] Emre Telatar. Capacity of multi-antenna Gaussian channels. *European trans. on telecommun.*, 10(6):585–595, 1999.

[10] Ying Jun Zhang and Khaled Ben Letaief. An efficient resource-allocation scheme for spatial multiuser access in MIMO/OFDM systems. *IEEE Trans. Commun.*, 53(1):107–116, 2005.

[11] Andrea J Goldsmith and Soon-Ghee Chua. Variable-rate variable-power MQAM for fading channels. *IEEE Trans. Commun.*, 45(10):1218–1230, 1997.

[12] Andre Noll Barreto and Simeon Furrer. Adaptive bit loading for wireless OFDM systems. In *Proc. IEEE Int. Sym. Personal, Indoor and Mobile Radio Commun.*, volume 2, pages G–G, 2001.

[13] Thomas Keller and Lajos Hanzo. Adaptive modulation techniques for duplex OFDM transmission. *IEEE trans. Veh. Technol.*, 49(5):1893–1906, 2000.

[14] Salvatore D'Alessandro, Andrea M Tonello, and Lutz Lampe. Bit-loading algorithms for OFDM with adaptive cyclic prefix length in PLC channels. In *Proc. IEEE Sym. Power Line Commun. and Its Applications (ISPLC)*, pages 177–181, 2009.

[15] Enzo Baccarelli, Antonio Fasano, and Mauro Biagi. Novel efficient bit-loading algorithms for peak-energy-limited ADSL-type multicarrier systems. *IEEE trans. on signal processing*, 50(5):1237–1247, 2002.

[16] Kinda Khawam, Samer Lahoud, Marc Ibrahim, Mohamad Yassin, Steven Martin, Melhem El Helou, and Farah Moety. Radio access technology selection in heterogeneous networks. *Physical Communication*, 18:125–139, 2016.

[17] Marceau Coupechoux, Jean-Marc Kelif, and Philippe Godlewski. Network controlled joint radio resource management for heterogeneous networks. In *Proc. IEEE Veh. Technol. Conf. Spring*, pages 1771–1775, 2008.

[18] Qingyang Song and Abbas Jamalipour. Network selection in an integrated wireless LAN and UMTS environment using mathematical modeling and computing techniques. *IEEE wireless commun.*, 12(3):42–48, 2005.

[19] Jijun Luo, Rahul Mukerjee, Markus Dillinger, Eiman Mohyeldin, and Egon Schulz. Investigation of radio resource scheduling in WLANs coupled with 3G cellular network. *IEEE Commun. Mag.*, 41(6):108–115, 2003.

[20] K Premkumar and Anurag Kumar. Optimum association of mobile wireless devices with a WLAN-3G access network. In *Proc. IEEE Int. Conf. Commun. (ICC)*, volume 5, pages 2002–2008, 2006.

[21] Dushyantha A Basnayaka and Harald Haas. Design and analysis of a hybrid radio frequency and visible light communication system. *IEEE Trans. Commun.*, 65(10):4334–4347, 2017.

[22] Mohamed Kashef, Muhammad Ismail, Mohamed Abdallah, Khalid A Qaraqe, and Erchin Serpedin. Energy efficient resource allocation for mixed RF/VLC heterogeneous wireless networks. *IEEE J. Sel. Areas Commun.*, 34(4):883–893, 2016.

[23] Hossein Kazemi, Murat Uysal, and Farid Touati. Outage analysis of hybrid FSO/RF systems based on finite-state markov chain modeling. In *Proc. IEEE Int. Workshop Optical Wireless Commun. (IWOW)*, pages 11–15, 2014.

[24] Yunlu Wang and Harald Haas. Dynamic load balancing with handover in hybrid Li-Fi and Wi-Fi networks. *J. Lightwave Techn.*, 33(22):4671–4682, 2015.

[25] Cisco visual networking index: Global mobile data traffic forecast update, 2016–2021, white paper at cisco.com.

[26] Afif Osseiran et al. Scenarios for 5G mobile and wireless communications: The vision of the METIS project. *IEEE Commun. Mag.*, 52(5):26–35, 2014.

[27] Jeffrey G Andrews, Stefano Buzzi, Wan Choi, Stephen V Hanly, Angel Lozano, Anthony CK Soong, and Jianzhong Charlie Zhang. What will 5G be? *IEEE J. Sel. Areas Commun.*, 32(6):1065–1082, 2014.

[28] Federico Boccardi, Robert W Heath, Angel Lozano, Thomas L Marzetta, and Petar Popovski. Five disruptive technology directions for 5G. *IEEE Commun. Mag.*, 52(2):74–80, 2014.

[29] Gerard J Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell labs technical journal*, 1(2):41–59, 1996.

[30] Lizhong Zheng and David NC Tse. Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels. *IEEE Trans. Inform. Theory*, 49(5):1073–1096, 2003.

[31] Yan Chen, Shunqing Zhang, Shugong Xu, and Geoffrey Ye Li. Fundamental trade-offs on green wireless networks. *IEEE Commun. Mag.*, 49(6):30–37, 2011.

[32] Fabien Heliot, Muhammad Ali Imran, and Rahim Tafazolli. On the energy efficiency-spectral efficiency trade-off over the MIMO Rayleigh fading channel. *IEEE Trans. Commun.*, 60(5):1345–1356, 2012.

[33] Thomas L Marzetta. Noncooperative cellular wireless with unlimited numbers of base station antennas. *IEEE Trans. Wireless Commun.*, 9(11):3590–3600, 2010.

[34] Erik Larsson, Ove Edfors, Fredrik Tufvesson, and Thomas Marzetta. Massive MIMO for next generation wireless systems. *IEEE Commun. Mag.*, 52(2):186–195, 2014.

[35] Bertrand M Hochwald, Thomas L Marzetta, and Vahid Tarokh. Multiple-antenna channel hardening and its implications for rate feedback and scheduling. *IEEE Trans. Inf. Theory*, 50(9):1893–1909, 2004.

[36] Hien Quoc Ngo and Erik G Larsson. No downlink pilots are needed in TDD massive MIMO. *IEEE Trans. Wireless Commun.*, 16(5):2921–2935, 2017.

[37] Angel Lozano and Antonia Maria Tulino. Capacity of multiple-transmit multiple-receive antenna architectures. *IEEE Trans. Inf. Theory*, 48(12):3117–3128, 2002.

[38] Aris L Moustakas, Steven H Simon, and Anirvan M Sengupta. MIMO capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large N analysis. *IEEE Trans. Inf. Theory*, 49(10):2545–2561, 2003.

[39] Emil Bjornson, Jakob Hoydis, Marios Kountouris, and Merouane Debbah. Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits. *IEEE Trans. Inf. Theory*, 60(11):7112–7139, 2014.

[40] Hien Quoc Ngo, Erik G Larsson, and Thomas L Marzetta. Energy and spectral efficiency of very large multiuser MIMO systems. *IEEE Trans. Commun.*, 61(4):1436–1449, 2013.

[41] Juei-Chin Shen, Jun Zhang, and Khaled B Letaief. Downlink user capacity of massive MIMO under pilot contamination. *IEEE Trans. Wireless Commun.*, 14(6):3183–3193, 2015.

[42] Jakob Hoydis, Stephan Ten Brink, and Mérouane Debbah. Massive MIMO in the UL/DL of cellular networks: How many antennas do we need? *IEEE J. Sel. Areas Commun.*, 31(2):160–171, 2013.

[43] Chao-Kai Wen, Shi Jin, and Kai-Kit Wong. On the sum-rate of multiuser MIMO uplink channels with jointly-correlated Rician fading. *IEEE Trans. Commun.*, 59(10):2883–2895, 2011.

[44] Joao Vieira, Steffen Malkowsky, Karl Nieman, Zachary Miers, Nikhil Kundargi, Liang Liu, Ian Wong, Viktor Öwall, Ove Edfors, and Fredrik Tufvesson. A flexible 100-antenna testbed for massive MIMO. In *Globecom Workshops (GC Wkshps), 2014*, pages 287–293, 2014.

[45] Steffen Malkowsky, Joao Vieira, Liang Liu, Paul Harris, Karl Nieman, Nikhil Kundargi, Ian Wong, Fredrik Tufvesson, Viktor Owall, and Ove Edfors. The world's first real-time testbed for massive MIMO: Design, implementation, and validation. *IEEE Access*, 2017.

[46] David Tse and Pramod Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.

[47] Linglong Dai, Bichai Wang, Yifei Yuan, Shuangfeng Han, I Chih-Lin, and Zhaocheng Wang. Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends. *IEEE Commun. Mag.*, 53(9):74–81, 2015.

[48] Jianchao Chen, Liang Yang, and Mohamed-Slim Alouini. Performance analysis of cooperative NOMA schemes in spatially random Relaying networks. *IEEE Access*, 2018.

[49] Lu Lv, Jian Chen, Qiang Ni, Zhiguo Ding, and Hai Jiang. Cognitive non-orthogonal multiple access with cooperative Relaying: A new wireless frontier for 5G spectrum sharing. *IEEE Commun. Mag.*, 56(4):188–195, 2018.

[50] Pramod Viswanath, David N. C. Tse, and Venkat Anantharam. Asymptotically optimal water-filling in vector multiple-access channels. *IEEE Trans. Inform. Theory*, 47(1):241–267, 2001.

[51] Samah AM Ghanem. MAC Gaussian channels with arbitrary inputs: Optimal precoding and power allocation. In *Int. Conf. Wireless Commun. Signal Process. (WCSP)*, pages 1–6, 2012.

[52] Jagadeesh Harshan and Bikash Sundar Rajan. On two-user Gaussian multiple access channels with finite input constellations. *IEEE Trans. Inform. Theory*, 57(3):1299–1327, 2011.

[53] Thomas M Cover. Broadcast channels. *IEEE Trans. Inform. Theory*, 18(1):2–14, 1972.

[54] Patrick P Bergmans. Random coding theorem for broadcast channels with degraded components. *IEEE Trans. Inform. Theory*, 19(2):197–207, 1973.

[55] Nihar Jindal, Sriram Vishwanath, and Andrea Goldsmith. On the duality of Gaussian multiple-access and broadcast channels. *IEEE Trans. Inform. Theory*, 50(5):768–783, 2004.

[56] Federal Communications Commission et al. Millimeter wave propagation: spectrum management implications. *Bulletin*, 70, 1997.

[57] Theodore S Rappaport, Shu Sun, Rimma Mayzus, Hang Zhao, Yaniv Azar, Kangping Wang, George N Wong, Jocelyn K Schulz, Mathew Samimi, and Felix Gutierrez. Millimeter wave mobile communications for 5G cellular: It will work! *Access, IEEE*, 1:335–349, 2013.

[58] Hany Elgala, Raed Mesleh, and Harald Haas. Indoor optical wireless communication: potential and state-of-the-art. *IEEE Commun. Mag.*, 49(9):56–62, 2011.

[59] Mostafa Z Afgani, Harald Haas, Hany Elgala, and Dietmar Knipp. Visible light communication using OFDM. In *Proc. IEEE Int. Conf. TRIDENTCOM*, pages 6–pp, 2006.

[60] Paul Waide et al. Phase out of incandescent lamps: Implications for international supply and demand for regulatory compliant lamps. Technical report, OECD Publishing, 2010.

[61] Moussa Ayyash, Hany Elgala, Abdallah Khreishah, Volker Jungnickel, Thomas Little, Sihua Shao, Michael Rahaim, Dominic Schulz, Jonas Hilt, and Ronald Freund. Coexistence of WiFi and LiFi toward 5G: concepts, opportunities, and challenges. *IEEE Commun. Mag.*, 54(2):64–71, 2016.

[62] Michael B Rahaim, Anna Maria Vegni, and Thomas DC Little. A hybrid radio frequency and broadcast visible light communication system. In *Proc. IEEE Global Telecommun. (GLOBECOM) Workshops*, pages 792–796, 2011.

[63] Yunlu Wang, Dushyantha A Basnayaka, Xiping Wu, and Harald Haas. Optimization of load balancing in hybrid LiFi/RF networks. *IEEE Trans. Commun.*, 65(4):1708–1720, 2017.

[64] Xuan Li, Rong Zhang, and Lajos Hanzo. Cooperative load balancing in hybrid visible light communications and WiFi. *IEEE Trans. Commun.*, 63(4):1319–1329, 2015.

[65] Sanjay Shakkottai, Theodore S Rappaport, and Peter C Karlsson. Cross-layer design for wireless networks. *IEEE Commun. Mag.*, 41(10):74–80, 2003.

[66] Andrea Goldsmith. *Wireless communications*. Cambridge university, 2005.

[67] Sami Akin and Mustafa Cenk Gursoy. Effective capacity analysis of cognitive radio channels for quality of service provisioning. *IEEE Trans. Wireless Commun.*, 9(11):3354–3364, 2010.

[68] Mustafa Cenk Gursoy. MIMO wireless communications under statistical queueing constraints. *IEEE Trans. Inf. Theory*, 57(9):5897–5917, 2011.

[69] Mustafa Ozmen and M Cenk Gursoy. Wireless throughput and energy efficiency with random arrivals and statistical queuing constraints. *IEEE Trans. Inf. Theory*, 62(3):1375–1395, 2016.

[70] Sami Akin and Markus Fidler. Backlog and delay reasoning in HARQ system. In *Proc. IEEE Int. Teletraffic Congress (ITC)*, pages 185–193, 2015.

[71] Fan Jin, Xuan Li, Rong Zhang, Chen Dong, and Lajos Hanzo. Resource allocation under delay-guarantee constraints for visible-light communication. *IEEE Access*, 4:7301–7312, 2016.

[72] Fan Jin, Rong Zhang, and Lajos Hanzo. Resource allocation under delay-guarantee constraints for heterogeneous visible-light and RF femtocell. *IEEE Trans. Wireless Commun.*, 14(2):1020–1034, 2015.

[73] Marwan Hammouda, Sami Akin, and Jürgen Peissig. Effective capacity in cognitive radio broadcast channels. In *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, pages 1071–1077, 2014.

[74] Marwan Hammouda, Sami Akin, M Cenk Gursoy, and Jürgen Peissig. Effective capacity in MIMO channels with arbitrary inputs. *IEEE trans. Veh. Technol.*, 2017.

[75] Marwan Hammouda, Sami Akin, and Jürgen Peissig. Effective capacity in multiple access channels with arbitrary inputs. In *Proc. IEEE Int. Conf. Wireless and Mobile Computing, Netw. and Commun. (WiMob)*, pages 406–413, 2015.

[76] Marwan Hammouda, Sami Akin, Anna Maria Vegni, Harald Haas, and Jurgen Peissig. Link selection in hybrid RF/VLC systems under statistical queueing constraints. *IEEE Trans. Wireless Commun.*, 2018.

[77] Sami Akin, Marwan Hammouda, and Jürgen Peissig. QoS analysis of cognitive radios employing HARQ. In *Proc. IEEE Int. Conf. Commun. (ICC)*, pages 1–6, 2017.

[78] Marwan Hammouda and Jürgen Peissig. VLC systems with fixed-rate transmissions under statistical queueing constraints. *arXiv preprint arXiv:1804.11097*, 2018.

[79] Marwan Hammouda, Sami Akin, and Jürgen Peissig. Effective capacity in broadcast channels with arbitrary inputs. In *Proc. Int. Conf. on Wired/Wireless Internet Commun.*, pages 323–334. Springer, 2016.

[80] Cheng-Shang Chang. Stability, queue length, and delay of deterministic and stochastic queueing networks. *IEEE Trans. Autom. Control*, 39(5):913–931, 1994.

[81] Dapeng Wu and Rohit Negi. Effective capacity: a wireless link model for support of quality of service. *IEEE Trans. Wireless Commun.*, 2(4):630–643, 2003.

[82] Qingwen Liu, Shengli Zhou, and Georgios B Giannakis. Cross-layer scheduling with prescribed QoS guarantees in adaptive wireless networks. *IEEE J. Sel. Areas Commun.*, 23(5):1056–1066, 2005.

[83] Qingwen Liu, Shengli Zhou, and Georgios B Giannakis. Cross-layer modeling of adaptive wireless links for QoS support in multimedia networks. In *First International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks*, pages 68–75. IEEE, 2004.

[84] Jia Tang and Xi Zhang. Cross-layer-model based adaptive resource allocation for statistical QoS guarantees in mobile wireless networks. *IEEE Trans. Wireless Commun.*, 7(6), 2008.

[85] Cheng-Shang Chang and Tim Zajic. Effective bandwidths of departure processes from queues with time varying capacities. In *Proc. IEEE INFOCOM*, volume 3, pages 1001–1009, 1995.

[86] Cheng-Shang Chang. *Performance Guarantees in Communication Networks*. Springer, 2000.

[87] Andrea Goldsmith, Syed Ali Jafar, Nihar Jindal, and Sriram Vishwanath. Capacity limits of MIMO channels. *IEEE J. Select. Areas Commun.*, 21(5):684–702, 2003.

[88] Da-Shan Shiu, Gerard J Foschini, Michael J Gans, and Joseph M Kahn. Fading correlation and its effect on the capacity of multielement antenna systems. *IEEE Trans. Commun.*, 48(3):502–513, 2000.

[89] Syed Ali Jafar and Andrea Goldsmith. Transmitter optimization and optimality of beamforming for multiple antenna systems. *IEEE Trans. Wireless Commun.*, 3(4):1165–1175, 2004.

[90] Eduard A Jorswieck and Holger Boche. Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback. *IEEE Trans. Wireless Commun.*, 3(5):1543–1553, 2004.

[91] Mai Vu and Arogyaswami Paulraj. Capacity optimization for Rician correlated MIMO wireless channels. In *Proc. 39th Asilomar Conf. Signal Syst. Comput. (ASILOMAR)*, pages 133–138, 2005.

[92] Antonia Maria Tulino, Angel Lozano, and Sergio Verdú. Capacity-achieving input covariance for single-user multi-antenna channels. *IEEE Trans. Wireless Commun.*, 5(3):662–671, 2006.

[93] Wonjong Rhee and John M Cioffi. On the capacity of multiuser wireless channels with multiple antennas. *IEEE Trans. Inform. Theory*, 49(10):2580–2595, 2003.

[94] Michel Ivrlac, T Kurpjuhn, Christopher Brunner, and J Nossek. On channel capacity of correlated MIMO channels. *ITG Fokusprojekt: Mobilkommunikation Systeme mit intelligenten Antennen*, Mar. 2001.

[95] Sivarama Venkatesan, Steven H Simon, and Reinaldo A Valenzuela. Capacity of a Gaussian MIMO channel with nonzero mean. In *Proc. IEEE Veh. Technol. Conf. Fall (VTC-FALL)*, volume 3, pages 1767–1771, 2003.

[96] D Hosli and Amos Lapidoth. The capacity of a MIMO Ricean channel is monotonic in the singular values of the mean. *ITG FACHBERICHT*, pages 381–386, 2004.

[97] Ming Kang and Mohamed-Slim Alouini. Capacity of MIMO Rician channels. *IEEE Trans. Wireless Commun.*, 5(1):112–122, 2006.

[98] Sergio Verdú. Spectral efficiency in the wideband regime. *IEEE Trans. Inform. Theory*, 48(6):1319–1343, 2002.

[99] Angel Lozano, Antonia M Tulino, and Sergio Verdú. Multiple-antenna capacity in the low-power regime. *IEEE Trans. Inform. Theory*, 49(10):2527–2544, 2003.

[100] Pasquale Memmolo, Marco Lops, Antonia M Tulino, and Reinaldo A Valenzuela. Up-link multi-user MIMO capacity in low-power regime. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 2308–2312, JUn. 2010.

[101] Vasanthan Raghavan and Akbar M Sayeed. Achieving coherent capacity of correlated MIMO channels in the low-power regime with non-flashy signaling schemes. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 906–910, Adelaide, Australia, Sep. 2005.

[102] Oluwakayode Onireti, Fabien Heliot, and Muhammad Ali Imran. On the energy efficiency-spectral efficiency trade-off of distributed MIMO systems. *IEEE Trans. Commun.*, 61(9):3741–3753, 2013.

[103] Thomas L Marzetta, Giuseppe Caire, Merouane Debbah, I Chih-Lin, and Saif K Mohammed. Special issue on massive MIMO. *J. Commun. Networks*, 15(4):333–337, 2013.

[104] Sami Akin and M Cenk Gursoy. On the throughput and energy efficiency of cognitive MIMO transmissions. *IEEE Trans. Veh. Technol.*, 62(7):3245–3260, 2013.

[105] Eduard A Jorswieck, Rami Mochaourab, and Martin Mittelbach. Effective capacity maximization in multi-antenna channels with covariance feedback. *IEEE Trans. Wireless Commun.*, 9(10):2988–2993, 2010.

[106] Sami Akin. The interplay between data transmission power and transmission link utilization. *IEEE Commun. Lett.*, 19(11):1953–1956, 2015.

[107] Fernando Pérez-Cruz, Miguel RD Rodrigues, and Sergio Verdú. MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation. *IEEE Trans. Inform. Theory*, 56(3):1070–1084, 2010.

[108] M Meritxell Lamarca Orozco. Linear precoding for mutual information maximization in MIMO systems. In *Proc. 6th IEEE Int. Symp. Wireless Commun. Syst. (ISWCS)*, pages 26–30, 2010.

[109] Miquel Payaró and Daniel P Palomar. On optimal precoding in linear vector Gaussian channels with arbitrary input distribution. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 1085–1089, 2009.

[110] Chengshan Xiao, Yahong Rosa Zheng, and Zhi Ding. Globally optimal linear precoders for finite alphabet signals over complex vector Gaussian channels. *IEEE Trans. Signal Processing*, 59(7):3301–3314, 2011.

[111] Miguel RD Rodrigues, Fernando Pérez-Cruz, and Sergio Verduy. Multiple-input multiple-output Gaussian channels: Optimal covariance for non-Gaussian inputs. In *Proc. IEEE Inf. Theory Workshop (ITW)*, pages 445–449, 2008.

[112] Yongpeng Wu, Chao-Kai Wen, Chengshan Xiao, Xiqi Gao, and Robert Schober. Linear MIMO precoding in jointly-correlated fading multiple access channels with finite alphabet signaling. In *Proc. IEEE Int. Conf. Commun. (ICC)*, pages 5306–5311, 2014.

[113] Mingxi Wang, Weiliang Zeng, and Chengshan Xiao. Linear precoding for MIMO multiple access channels with finite discrete inputs. *IEEE Trans. Wireless Commun.*, 10(11):3934–3942, 2011.

[114] Yongpeng Wu, Chao-Kai Wen, Chengshan Xiao, Xiqi Gao, and Robert Schober. Linear precoding for the MIMO multiple access channel with finite alphabet inputs and statistical CSI. *IEEE Trans. Wireless Commun.*, 14(2):983–997, 2015.

[115] Dongning Guo, Shlomo Shamai, and Sergio Verdú. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Trans. Inform. Theory*, 51(4):1261–1282, 2005.

[116] Daniel Pérez Palomar and Sergio Verdú. Gradient of mutual information in linear vector Gaussian channels. *IEEE Trans. Inform. Theory*, 52(1):141–154, 2006.

[117] Emil Björnson, Jakob Hoydis, Marios Kountouris, and Mérouane Debbah. Hardware impairments in large-scale MISO systems: Energy efficiency, estimation, and capacity limits. In *Proc. IEEE Int. Conf. Digital Signal Processing (DSP)*, pages 1–6, 2013.

[118] Ulf Gustavsson, Cesar Sanchéz-Perez, Thomas Eriksson, Fredrik Athley, Giuseppe Durisi, Per Landin, Katharina Hausmair, Christian Fager, and Lars Svensson. On the impact of hardware impairments on massive MIMO. In *Globecom Workshops (GC Wkshps), 2014*, pages 294–300, 2014.

[119] ICT-317669 METIS Project. Scenarios, requirements and KPIs for 5G mobile and wireless system. *Del. D1.1*, May, 2013.

[120] Ezio Biglieri, Robert Calderbank, Anthony Constantinides, Andrea Goldsmith, Arogyaswami Paulraj, and H Vincent Poor. *MIMO wireless communications*. Cambridge university press, 2007.

[121] Chen-Nee Chuah, David N. C. Tse, Joseph M Kahn, and Reinaldo A Valenzuela. Capacity scaling in MIMO wireless systems under correlated fading. *IEEE Trans. on Inf. Theory*, 48(3):637–650, 2002.

[122] Claude Oestges, Bruno Clerckx, Maxime Guillaud, and Mérouane Debbah. Dual-polarized wireless communications: from propagation models to system performance evaluation. *IEEE Tran. on Wireless Commun.*, 7(10), 2008.

[123] Sergey L Loyka. Channel capacity of MIMO architecture using the exponential correlation matrix. *IEEE Communications letters*, 5(9):369–371, 2001.

[124] Taesang Yoo and Andrea Goldsmith. Capacity and power allocation for fading MIMO channels with channel estimation error. *IEEE Trans. Inf. Theory*, 52(5):2203–2214, 2006.

[125] Leila Musavian, Mohammad Reza Nakhai, Mischa Dohler, and A Hamid Aghvami. Effect of channel uncertainty on the mutual information of MIMO fading channels. *IEEE trans. Veh. Technol.*, 56(5):2798–2806, 2007.

[126] Thierry E Klein and Robert G Gallager. Power control for the additive white Gaussian noise channel under channel estimation errors. In *Proc. IEEE Int. Sym. Inf. Theory*, page 304, 2001.

[127] Muriel Medard. The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel. *IEEE Trans. Inf. theory*, 46(3):933–946, 2000.

[128] Babak Hassibi and Bertrand M Hochwald. How much training is needed in multiple-antenna wireless links? *IEEE Trans. Inf. Theory*, 49(4):951–963, 2003.

[129] S. Akin and M. Fidler. On the transmission rate strategies in cognitive radios. *IEEE Trans. Wireless Commun.*, 15(3):2335–2350, March 2016.

[130] Dimitri P Bertsekas, Robert G Gallager, and Pierre Humblet. *Data networks*, volume 2. Prentice-Hall International New Jersey, 1992.

[131] Cheng-Shang Chang. *Performance Guarantees in Communication Networks*. Springer-Verlag, 2000.

[132] Yuming Jiang and Yong Liu. *Stochastic network calculus*, volume 1. Springer, 2008.

[133] Markus Fidler and Amr Rizk. A guide to the stochastic network calculus. 17(1):92–105, 2015.

[134] Kam Lee. Performance bounds in communication networks with variable-rate links. In *ACM SIGCOMM Comput. Commun. Review*, volume 25, pages 126–136. ACM, 1995.

[135] Markus Fidler, Ralf Lübben, and Nico Becker. Capacity–delay–error boundaries: A composable model of sources and systems. *IEEE Trans. Wireless Commun.*, 14(3):1280–1294, 2015.

[136] Ugo Fano. Ionization yield of radiations. II. the fluctuations of the number of ions. *Physical Review*, 72(1):26, 1947.

[137] Xiaofeng Tao, Xiaodong Xu, and Qimei Cui. An overview of cooperative communications. *IEEE Commun. Mag.*, 50(6):65–71, 2012.

[138] Zhiguo Ding, Yuanwei Liu, Jinho Choi, Qi Sun, Maged Elkashlan, H Vincent Poor, et al. Application of non-orthogonal multiple access in LTE and 5G networks. *arXiv preprint arXiv:1511.08610*, 2015.

[139] Ezio Biglieri and László Györfi. *Multiple Access Channels: Theory and Practice*. IOS press, 2007.

[140] Gautam A Gupta and Stavros Toumpis. Power allocation over parallel Gaussian multiple access and broadcast channels. *IEEE Trans. Inform. Theory*, 52(7):3274–3282, 2006.

[141] Raymond Knopp and Pierre A Humblet. Information capacity and power control in single-cell multiuser communications. In *IEEE Int. Commun. Conf. (ICC)*, volume 1, pages 331–335, 1995.

[142] Sriram Vishwanath, S Jafar, and Andrea Goldsmith. Optimum power and rate allocation strategies for multiple access fading channels. In *IEEE Veh. Technol. Conf. Spring (VTC-SPRING)*, volume 4, pages 2888–2892, 2001.

[143] Khoa D Nguyen, Albert Guillen i Fabregas, and Lars K Rasmussen. Outage exponents of block-fading channels with power allocation. *IEEE Trans. Inform. Theory*, 56(5):2373–2381, 2010.

[144] Gozde Ozcan and M Cenk Gursoy. Optimal power control for fading channels with arbitrary input distributions and delay-sensitive traffic. *IEEE Trans. Commun.*, 2018.

[145] Gozde Ozcan, Mustafa Ozmen, and M Cenk Gursoy. QoS-driven energy-efficient power control with random arrivals and arbitrary input distributions. *IEEE Trans. Wireless Commun.*, 16(1):376–388, 2017.

[146] Robert G Gallager. *Information theory and reliable communication*, volume 2. Springer, 1968.

[147] Deli Qiao, Mustafa Cenk Gursoy, and Senem Velipasalar. Achievable throughput regions of fading broadcast and interference channels under QoS constraints. *IEEE Trans. Commun.*, 61(9):3730–3740, 2013.

[148] Deli Qiao, Mustafa Cenk Gursoy, and Senem Velipasalar. Transmission strategies in multiple-access fading channels with statistical QoS constraints. *IEEE Trans. Inform. Theory*, 58(3):1578–1593, 2012.

[149] Gozde Ozcan and M Cenk Gursoy. QoS-driven power control for fading channels with arbitrary input distributions. In *IEEE Int. Symp. Inform. Theory (ISIT)*, pages 1381–1385, 2014.

[150] Won Ho Jeong, Joo Seock Kim, Myoung-won Jung, and Kyung-Seok Kim. MIMO channel measurement and analysis for 4G mobile communication. pages 676–682. 2012.

[151] Marwan Hammouda, Jürgen Peissig, and Anna Maria Vegni. Design of a cognitive VLC network with illumination and handover requirements. In *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, pages 451–456, 2017.

[152] Marwan Hammouda, Anna Maria Vegni, Jürgen Peissig, and Mauro Biagi. Resource allocation in a multi-color DS-OCDMA VLC cellular architecture. *Optics Express*, 26(5):5940–5961, 2018.

[153] Jean Armstrong. OFDM for optical communications. *Journal of lightwave technology*, 27(3):189–204, 2009.

[154] Svilen Dimitrov, Sinan Sinanovic, and Harald Haas. Signal shaping and modulation for optical wireless communication. *J. of lightwave techn.*, 30(9):1319–1328, 2012.

[155] Jia Tang and Xi Zhang. Cross-layer modeling for quality of service guarantees over wireless links. *IEEE Trans. Wireless Commun.*, 6(12), 2007.

[156] Toshihiko Komine and Masao Nakagawa. Fundamental analysis for visible-light communication system using LED lights. *IEEE Trans. on Consumer Electronics*, 50(1):100–107, 2004.

[157] John R Barry, Joseph M Kahn, William J Krause, et al. Simulation of multipath impulse response for indoor wireless optical channels. *IEEE J. Sel. Areas Commun.*, 11(3):367–379, 1993.

[158] Anas Chaaban, Zouheir Rezki, and Mohamed-Slim Alouini. Fundamental limits of parallel optical wireless channels: Capacity results and outage formulation. *IEEE Trans. on Commun.*, 65(1):296–311, 2017.

[159] Amos Lapidoth, Stefan M Moser, and Michele A Wigger. On the capacity of free-space optical intensity channels. *IEEE Trans. Inf. Theory*, 55(10):4449–4461, 2009.

[160] Nick Letzepis and Albert Guillen I Fabregas. Outage probability of the Gaussian MIMO free-space optical channel with PPM. *IEEE Trans. on Commun.*, 57(12), 2009.

[161] Lawrence H Ozarow, Shlomo Shamai, and Aaron D Wyner. Information theoretic considerations for cellular mobile radio. *IEEE trans. Veh. Technol.*, 43(2):359–378, 1994.

[162] Deli Qiao, Mustafa Cenk Gursoy, and Senem Velipasalar. Energy efficiency of fixed-rate wireless transmissions under QoS constraints. In *Proc. IEEE Conf. Inter. Commun. (ICC)*, pages 1–5, 2009.

[163] Fang Wang, Zhaocheng Wang, Chen Qian, Linglong Dai, and Zhixing Yang. Efficient vertical handover scheme for heterogeneous VLC-RF systems. *J. Optical Commun. and Net.*, 7(12):1172–1180, 2015.

[164] Harry Heffes and David Lucantoni. A markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. *IEEE J. Sel. Areas Commun.*, 4(6):856–868, 1986.

[165] Abdelnaser Adas. Traffic models in broadband networks. *IEEE commun. Mag.*, 35(7):82–89, 1997.

[166] H Chowdhury and M Katz. Cooperative data download on the move in indoor hybrid (radio-optical) WLAN-VLC hotspot coverage. *Trans. on Emerging Telecommun. Technol.*, 25(6):666–677, 2014.

[167] Dushyantha A Basnayaka and Harald Haas. Hybrid RF and VLC systems: Improving user data rate performance of VLC systems. In *Proc. IEEE Veh. Technol. Conf. Spring (VTC-Spring)*, pages 1–5, 2015.

[168] Xu Bao, Xiaorong Zhu, Tiecheng Song, and Yanqiu Ou. Protocol design and capacity analysis in hybrid network of visible light communication and OFDMA systems. *IEEE Trans. Veh. Technol.*, 63(4):1770–1778, 2014.

[169] Sihua Shao, Abdallah Khreishah, Michael B Rahaim, Hany Elgala, Moussa Ayyash, Thomas DC Little, and Jie Wu. An indoor hybrid WiFi-VLC internet access system. In *Proc. IEEE Int. Conf. on Mobile Ad Hoc and Sensor Sys.*, pages 569–574, 2014.

[170] Anna Maria Vegni and Thomas DC Little. Handover in VLC systems with cooperating mobile devices. In *Proc. Int. Conf. Computing, Netw. Commun. (ICNC)*, pages 126–130, 2012.

[171] Shufei Liang, Hui Tian, Bo Fan, and Ronglin Bai. A novel vertical handover algorithm in a hybrid visible light communication and LTE system. In *Proc. IEEE Veh. Technol. Conf. Fall (VTC-FALL)*, pages 1–5, 2015.

[172] Sihua Shao and Abdallah Khreishah. Delay analysis of unsaturated heterogeneous omnidirectional-directional small cell wireless networks: The case of RF-VLC coexistence. *IEEE Trans. on Wireless Commun.*, 15(12):8406–8421, 2016.

[173] Irina Stefan, Harald Burchardt, and Harald Haas. Area spectral efficiency performance comparison between VLC and RF femtocell networks. In *Proc. IEEE Int. Conf. Commun. (ICC)*, pages 3825–3829, 2013.

[174] Nestor D Chatzidiamantis, George K Karagiannidis, Emmanouil E Kriezis, and Michail Matthaiou. Diversity combining in hybrid RF/FSO systems with PSK modulation. In *Proc. IEEE Int. Conf. Commun. (ICC)*, pages 1–6, 2011.

[175] Shlomo Shamai and Israel Bar-David. The capacity of average and peak-power-limited quadrature Gaussian channels. *IEEE Trans. Inf. Theory*, 41(4):1060–1071, 1995.

[176] Joseph M Kahn and John R Barry. Wireless infrared communications. *Proceedings of the IEEE*, 85(2):265–298, 1997.

[177] Volker Pohl, Volker Jungnickel, and Clemens Von Helmolt. A channel model for wireless infrared communication. In *PIMRC*, pages 297–303, 2000.

[178] V. Jungnickel et al. A European view on the next generation optical wireless communication standard. In *Proc. IEEE Conf. Standards for Commun. and Networking (CSCN)*, pages 106–111, Oct 2015.

[179] Svilen Dimitrov and Harald Haas. *Principles of LED Light Communications: Towards Networked Li-Fi*. Cambridge University Press, 2015.

[180] Haoran Yu, Man Hon Cheung, Longbo Huang, and Jianwei Huang. Power-delay tradeoff with predictive scheduling in integrated cellular and Wi-Fi networks. *IEEE J. Sel. Areas Commun.*, 34(4):735–742, 2016.

[181] Florin Ciucu, Almut Burchard, and Jörg Liebeherr. Scaling properties of statistical end-to-end bounds in the network calculus. *IEEE Trans. Inf. Theory*, 52(6):2300–2312, 2006.

[182] Weihua Wu, Fen Zhou, and Qinghai Yang. Dynamic network resource optimization in hybrid VLC and radio frequency networks. In *Proc. Int. Conf. Sel. Topics in Mobile and Wireless Netw. (MoWNeT)*, pages 1–7, 2017.

[183] Mohammad Dehghani Soltani, Xiping Wu, Majid Safari, and Harald Haas. On limited feedback resource allocation for visible light communication networks. In *Proc. Int. Works.Visible Light Commun. Sys.*, pages 27–32. ACM, 2015.

[184] Robert G Akl, Dinesh Tummala, and Xinrong Li. *Indoor propagation modeling at 2.4 GHz for IEEE 802.11 networks*. International Association of Science and Technology for Development, 2006.

[185] Theodore S Rappaport and Clare D McGillem. UHF fading in factories. *IEEE J. Sel. Areas Commun.*, 7(1):40–48, 1989.

[186] Aleksandar Neskovic, Natasa Neskovic, and George Paunovic. Modern approaches in modeling of mobile radio systems propagation environment. *IEEE Commun. Surveys and Tutorials*, 3(3):2–12, 2000.

[187] Behrouz Farhang-Boroujeny. OFDM versus filter bank multicarrier. *IEEE Signal Processing Mag.*, 28(3):92–112, 2011.

[188] Martin Fuhrwerk, Jurgen Peissig, and Malte Schellmann. Performance comparison of CP-OFDM and OQAM-OFDM systems based on LTE parameters. In *Proc. IEEE Int. Conf. Wireless and Mobile Computing, Netw. and Commun. (WiMob)*, pages 604–610, 2014.

[189] Gerhard Fettweis, Marco Krondorf, and Steffen Bittner. GFDM-generalized frequency division multiplexing. In *Proc. IEEE Veh. Technol. Conf. Spring*, pages 1–4, 2009.

[190] Vyacheslav V. Prelov and Sergio Verdú. Second-order asymptotics of mutual information. *IEEE Trans. Inf. Theory*, 50(8):1567–1580, 2004.

[191] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[192] Chao-Kai Wen, Pangan Ting, and Jiunn-Tsair Chen. Asymptotic analysis of MIMO wireless systems with spatial correlation at the receiver. *IEEE Trans. Commun.*, 54(2):349–363, 2006.

[193] George B Arfken. *Mathematical methods for physicists*. Academic press, 2013.

JOURNALS

M. Hammouda, A. M. Vegni, H. Haas, and J. Peissig, "Resource Allocation and Interference Management in OFDMA-based VLC Networks," In *Special Issue on Optical Wireless Commun., Physical Commun., Elsevier*, Apr. 2018.

M. Hammouda, S. Akın, A. M. Vegni, H. Haas, and J. Peissig, "Link Selection in Hybrid RF/VLC Systems under Statistical Queueing Constraints," In *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2738-2754, April 2018.

M. Hammouda, A. M. Vegni, J. Peissig, and M. Biagi, "Resource Allocation in a multi-color DS-OCDMA VLC Cellular Architecture," In *Optics express*, 26(5), pp.5940-5961.

M. Hammouda, S. Akın, M. C. Gursoy, and J. Peissig, "Effective Capacity in MIMO Channels with Arbitrary Inputs," In *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3252-3268, April 2018.

CONFERENCES

M. Hammouda, S. Akın, A. M. Vegni, H. Haas, and J. Peissig, "Hybrid RF/VLC Systems under QoS Constraints," In *Proc. IEEE ICT*, Saint-Malo, France, June 26-28 2018.

M. Hammouda, and J. Peissig, "VLC Systems with Fixed-Rate Transmissions Under Statistical Queueing Constraints," In *Proc. IEEE/IET CSNDSP*, Budapest, Hungary, July 18-20 2018.

S. Akın, **M. Hammouda**, and J. Peissig, "QoS analysis of cognitive radios employing HARQ," In *Proc. IEEE ICC*, Paris, France, 2017, pp. 1-6.

M. Hammouda, J. Peissig, and A. M. Vegni, "Design of a cognitive VLC network with illumination and handover requirements," In *Proc. IEEE ICC Workshops*, Paris, France, 2017, pp. 451-456.

M. Hammouda, S. Akın, and J. Peissig, "Effective Capacity in Broadcast Channels with Arbitrary Inputs," In *Proc. Int. Conf. Wired/Wireless Internet Commun.*, Thessaloniki Greece, 2016, pp. 323-334, Springer.

M. Hammouda, S. Akın, and J. Peissig, "Effective capacity in multiple access channels with arbitrary inputs," In *Proc. IEEE WiMob*, Abu Dhabi, UAE, 2015, pp. 406-413.

M. Hammouda, S. Akın, and J. Peissig, "Performance analysis of energy-detection-based massive SIMO," In *Proc. IEEE BlackSeaCom*, Constanta Romania, 2015, pp. 152-156.

M. Hammouda, S. Akın, and J. Peissig, "Effective capacity in cognitive radio broadcast channels," In *Proc. IEEE GLOBECOM*, Austin, USA, 2014, pp. 1071-1077.

M. Hammouda, and J. Wallace, "Noise uncertainty in cognitive radio sensing: Analytical modeling and detection performance," In *Proc. IEEE Int. ITG Workshop on Smart Antennas (WSA)*, TU-Dresden, Germany, 2012, pp. 287-293.

R. Mesleh, H. Elgala, **M. Hammouda**, I. Stefan, and H. Haas, "Optical spatial modulation with transmitter-receiver alignments," In *Proc. IEEE European Conf. Netw. and Opt. Commun. (NOC)*, 2011, pp. 1-4

CURRICULUM VITAE

Name	Marwan Hammouda
Day of birth	20.12.1985
Education	
since 11/2012	**Ph.D.** student Leibniz Universität Hannover Hannover, Germany Thesis Title: Resource Allocation for 5G Technologies under Statistical Queueing Constraints
09/2010 - 09/2012	**M.Sc.** in Electrical Engineering Jacobs University Bremen Bremen, Germany Thesis Title: Noise uncertainty in cognitive radio sensing: Analytical modeling and detection performance
09/2003 - 07/2007	**B.Sc.** in Electrical Engineering Islamic University of Gaza Gaza, Palestine Thesis Title: A Mimicking Human Arm Controlled by LabVIEW
Work Experience	
since 11/2012	**Research Assistant** Leibniz Universität Hannover Hannover, Germany

07/2011 - 01/2012	**Internship: Wireless Power Transfer for Lighting Systems** Philips Research Eindhoven, Netherlands
09/2009 - 07/2010	**Head of Engineering Professions Department** Namaa College for Science and Technology Jabalia, Palestine
10/2008 - 09/2009	**Project Coordinator** Islamic University of Gaza Gaza, Palestine
Teaching Experience	
Academic Year of 2014-2016 Summer Semesters	Exercises on "Modulation Methods"
Academic Year of 2015-2018 Winter Semesters	Exercises on "Digital Communication Systems"
Academic Year of 2014-2018 Summer Semesters	Lab Exercises on "Digital Communication Systems"
Since 05/2013	Co-advising several BA and MSc theses