Spring 5-2012

Wireless Communications under QoS Constraints: Energy Efficiency, Power and Rate Control, and Throughput

Deli Qiao
University of Nebraska-Lincoln, dqiao726@huskers.unl.edu

WIRELESS COMMUNICATIONS UNDER QOS CONSTRAINTS: ENERGY EFFICIENCY, POWER AND RATE CONTROL, AND THROUGHPUT

by

Deli Qiao

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfilment of Requirements

For the Degree of Doctor of Philosophy

Major: Engineering

Under the Supervision of

Professors Mustafa Cenk Gursoy and Senem Velipasalar

Lincoln, Nebraska

May, 2012

WIRELESS COMMUNICATIONS UNDER QOS CONSTRAINTS: ENERGY EFFICIENCY, POWER AND RATE CONTROL, AND THROUGHPUT

Deli Qiao, Ph. D.

University of Nebraska, 2012

Adviser: Mustafa Cenk Gursoy and Senem Velipasalar

This dissertation deals with various issues in wireless communications under statistical quality of service (QoS) constraints. Effective capacity, which provides the maximum constant arrival rate that a wireless channel can sustain while satisfying statistical QoS constraints, is adopted as the performance metric. Energy efficiency of point-to-point links is first studied by characterizing the spectral efficiency-bit energy tradeoff in the low-power and wideband regimes. Different transmission strategies (with variable or fixed rate) and power policies are studied. Then, the effective capacity region for fading multiple-access channels (MAC) is investigated for different transmission strategies: Superposition coding with successive decoding and time division multiple acess (TDMA). With fixed power, it is shown that varying the decoding order with respect to the channel states can significantly increase the achievable throughput region. In the two-user case, the optimal decoding strategy is determined for the scenario in which the users have the same QoS constraints. The optimal power allocation policies for any partition of the channel state space are identified. With the characterization of effective capacity regions, the energy efficiency of MAC is investigated by quantizing the minimum bit energy and wideband slope regions for different transmission strategies. In addition, the throughput for the two-hop wireless communication links with individual QoS constraints at the source and relay nodes is determined as a function of the QoS parameters and signal-to-noise ratios at the

source and relay, and the fading distributions of the links. The analysis is performed for both full-duplex and half-duplex relaying. Finally, the throughput with finite blocklength channel codes is analyzed for variable-rate and fixed-rate transmissions in single-user settings. The optimum error probability for variable-rate transmission and the optimum coding rate for fixed-rate transmission are shown to be unique.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisers, Professors Mustafa Cenk Gursoy and Senem Velipasalar for their guidance, encouragement and support throughout my Ph.D study. I am especially indebted to Professor Mustafa Cenk Gursoy, without the help of whom this dissertation would be impossible. I have benefited tremendously from his wisdom, insights and inspiring discussions about the research topics. In addition, his encouragement and endless patience when I was in front of challenges and difficulties have made the impossible tasks possible. I feel so fortunate to have a rewarding graduate study and have sharpened my comprehensive skills.

I thank Professors Lance C Pérez and Byrav Ramamurthy for serving as my thesis readers and defense committee members. Their valuable feedback and insights have strengthened this dissertation in every aspect.

I would like to thank my officemates and friends: Junwei Zhang, Qing Chen, Youlu Wang, Zhe Zhang, Li He, Bo Liang, Sami Akin, Mauricio Casares, and Alvaro Pinto, with whom I have experienced a great time in my graduate study. I would also like to thank my friends in Lincoln who have left unforgettable memories, especially Yunbo Wang, Xingling Dai, Yijia Zhao, and Dong Lin for their help during my first year in Lincoln.

Finally, I am grateful to my family for their understanding and support: my mother, my father, my brothers, my sisters, my aunts, and my uncles. No words could express my deepest gratitude for my grandmother, who has a far-reaching impact on my life. To them I dedicate this dissertation.

Contents

List of Figures

List of Algorithms

Chapter 1

Introduction

Fueled by the fourth generation (4G) wireless standards, smart phones and tablets, social networking tools and video-sharing sites, wireless transmission of multimedia content has significantly increased in volume and is expected to be the dominant traffic in data communications. Such wireless multimedia traffic requires certain quality-of-service (QoS) guarantees so that acceptable performance and quality levels can be met for the end-users. For instance, in voice over IP (VoIP), interactive-video (e.g., videoconferencing), and streaming-video applications in wireless systems, latency is a key QoS metric. In such cases, information has to be communicated with minimal delay. Hence, certain constraints on the queue length should be imposed in order to have the data not wait too long in the buffer at the transmitter. At the same time, satisfying these QoS considerations is challenging in wireless communication scenarios. Due to mobility, changing environment and multipath fading, the power of the received signal, and hence the instantaneous rates supported by the channel, fluctuate randomly [1]. In such a volatile environment, providing deterministic delay guarantees either is not possible or, when it is possible, requires the system to operate pessimistically and achieve low performance underutilizing the resources. Therefore,

wireless systems are better suited to support statistical QoS guarantees.

While providing powerful results, information-theoretic studies generally do not address delay and QoS constraints [2]. For instance, results on the channel capacity give insights on the performance levels achieved when the blocklength of codes becomes large [3]. The impact upon the queue length and queueing delay of transmission using codes with large blocklength can be significant. Situation is even further exacerbated in wireless channels in which the ergodic capacity has an operational meaning only if the codewords are long enough to span all fading states. Now, we also have dependence on fading, and in slow fading environments, large delays can be experienced in order to achieve the ergodic capacity. Due to these considerations, performance metrics such as capacity versus outage [4] and delay limited capacity [5] have been considered in the literature for slow fading scenarios. For a given outage probability constraint, outage capacity gives the maximum transmission rate that satisfies the outage constraint. Delay-limited capacity is defined as the outage capacity associated with zero outage probability, and is a performance level that can be attained regardless of the values of the fading states. Hence, delay limited capacity can be seen as a deterministic service guarantee. However, delay limited capacity can be low or even zero, for instance in Rayleigh fading channels even if both the receiver and transmitter have perfect channel side information.

More recently, delay constraints are more explicitly considered and their impact on communication over fading channels is analyzed in [6] and [7]. In these studies, the tradeoff between the average transmission power and average delay is identified. In [6], this tradeoff is analyzed by considering an optimization problem in which the weighted combination of the average power and average delay is minimized over transmission policies that determine the transmission rate by taking into account the arrival state, buffer occupancy, channel state jointly together.

In this thesis, we follow a different approach to solving problems arising in wireless communications under QoS constraints. we employ the notion of effective capacity [8], which can be seen as the maximum throughput that can be achieved by the given energy levels while providing statistical QoS guarantees. Effective capacity formulation uses the large deviations theory and incorporates the statistical QoS constraints by capturing the rate of decay of the buffer occupancy probability for large queue lengths (see e.g., [9], [10], [11], and [12]).

1.1 Effective Capacity

In [8] [13] [14] [15], Wu and Negi defined the effective capacity as the maximum constant arrival rate that a given service process can support in order to guarantee a statistical QoS requirement specified by the QoS exponent θ. The effective capacity is formulated as

$$C(\theta) = -\lim_{t\to\infty} \frac{1}{\theta t} \log_e \mathbb{E}\{e^{-\theta S[t]}\} \quad \text{bits/s}, \tag{1.1}$$

where the expectation is with respect to $S[t] = \sum_{i=1}^{t} s[i]$, which is the time-accumulated service process, and $\{s[i], i = 1, 2, \ldots\}$ denotes the discrete-time stationary and ergodic stochastic service process.

Operational meaning of the effective capacity is the following. If the constant arrival rate to the buffer is equal to the effective capacity $C(\theta)$, then the queue length process converges in distribution to a random variable Q that satisfies

$$\lim_{q\to\infty} \frac{\log \Pr(Q \geq q)}{q} = -\theta. \tag{1.2}$$

Above, Q can be seen as the stationary queue length, and θ as the asymptotic decay

rate of the tail distribution of the queue length Q. Hence, effective capacity specifies the maximum constant arrival rate that can be supported by the time-varying channel while the queue-overflow probability is guaranteed to behave as $\Pr(Q \geq q) \approx e^{-\theta q}$ for large overflow-threshold q. Therefore, the QoS exponent θ can be regarded as a parameter that specifies the asymptotic exponential decay-rate of the overflow probability and describes how strict the QoS constraints are. For instance, larger θ corresponds to more strict QoS constraints while smaller θ implies looser QoS guarantees. As noted in [16], when we have $\Pr(Q \geq q) \approx e^{-\theta q}$, then the delay violation probability can be approximated as $\Pr(D \geq d) \approx e^{-\theta \xi d}$ for large d, where D denotes the steady-state delay experienced in the buffer and ξ is determined by the arrival and service processes. In a more specific scenario in which the arrival rate is constant, Liu and Chamberland in [17] showed that $\Pr(D > d) \leq c\sqrt{\Pr(Q > q)}$ where c is some constant, $q = ad$, and a is the constant arrival rate.

Since the average arrival rate is equal to the average departure rate when the queue is in steady-state [18], effective capacity, which characterizes the maximum constant arrival rate, can also be seen as the maximum throughput in the presence of constraints on the buffer or delay violation probabilities. Note that requiring the tail probabilities of buffer/delay violations to decay exponentially is a stronger condition than stability or having the average buffer length or delay to be finite. Therefore, throughput in the presence of QoS limitations will in general be less than the throughput under stability constraints.

In the following, in order to simplify the analysis while considering general fading distributions, we assume that the fading coefficients stay constant over the frame duration T and vary independently for each frame and each user. In this scenario, $s[i] = TR[i]$, where $R[i]$ is the instantaneous service rate in the ith frame duration

$[iT;(i+1)T)$. Then, (1.1) can be written as

$$C(\theta) = -\frac{1}{\theta T}\log_e \mathbb{E}_{\mathbf{z}}\{e^{-\theta T R[i]}\} \quad \text{bits/s}, \tag{1.3}$$

where $R[i]$ is in general a function of the fading state $\mathbf{z}$, which will be discussed in details in later chapters. (1.3) is obtained using the fact that instantaneous rates $\{R[i]\}$ vary independently from one frame to another. It is interesting to note that as $\theta \to 0$ and hence QoS constraints relax, effective capacity approaches the ergodic capacity, i.e., $C(\theta) \to \mathbb{E}_{\mathbf{z}}\{R[i]\}$. On the other hand, as shown in [19], $C(\theta)$ converges to the delay limited capacity as θ grows without limit (i.e., $\theta \to \infty$) and QoS constraints become increasingly more strict. Therefore, effective capacity enables us to study the performance levels between the two extreme cases of delay limited capacity, which can be seen as a deterministic service guarantee or equivalently as a performance level attained under hard QoS limitations, and ergodic capacity, which is achieved in the absence of any QoS considerations.

Throughout the rest of the dissertation, we use the effective capacity normalized by bandwidth B, which is denoted by

$$\mathsf{C}(\theta) = \frac{C(\theta)}{B} \quad \text{bits/s/Hz}. \tag{1.4}$$

1.2 Spectral Efficiency-Bit Energy Tradeoff in the Low-SNR regime

In wireless systems, mobile wireless systems can only be equipped with limited energy resources, and hence energy efficient operation is a crucial requirement in most cases. Indeed, one of the features of fourth generation (4G) wireless systems is the ability

to support multimedia services at low transmission costs [20, Chap. 23, available online]. To measure and compare the energy efficiencies of different systems and transmission schemes, one can choose as a metric the energy required to reliably send one bit of information. Information-theoretic studies show that energy-per-bit requirement is generally minimized, and hence the energy efficiency is maximized, if the system operates at low signal-to-noise ratio (SNR) levels and hence in the low-power or wideband regimes. Recently, Verdú in [21] has determined the minimum bit energy required for reliable communication over a general class of channels, and studied the spectral efficiency–bit energy tradeoff in the wideband regime while also providing novel tools that are useful for analysis at low SNRs.

In this section, we focus on the energy efficiency aspect of wireless transmissions under the aforementioned statistical queueing constraints. Since energy efficient operation generally requires operation at low-SNR levels, our analysis in Chapters 2, 3 and 4 is carried out in the low-SNR regime. We define $\mathrm{SNR} = \frac{\bar{P}}{N_0 B}$, which can be seen more clearly in the following chapters. Therefore, low SNR means either low average power $\bar{P}$ or high bandwidth B. Then, it can be easily seen that $\frac{E_b}{N_0}_{\min}$ under QoS constraints can be obtained from [21]

$$\frac{E_b}{N_0}_{\min} = \lim_{\mathrm{SNR}\to 0} \frac{\mathrm{SNR}}{\mathsf{C}_E(\mathrm{SNR})} = \frac{1}{\dot{\mathsf{C}}_E(0)}. \tag{1.5}$$

At $\frac{E_b}{N_0}_{\min}$, the slope $\mathcal{S}_0$ of the spectral efficiency versus E_b/N_0 (in dB) curve is defined as [21]

$$\mathcal{S}_0 = \lim_{\frac{E_b}{N_0}\downarrow \frac{E_b}{N_0}_{\min}} \frac{\mathsf{C}_E(\frac{E_b}{N_0})}{10\log_{10}\frac{E_b}{N_0} - 10\log_{10}\frac{E_b}{N_0}_{\min}} 10\log_{10} 2. \tag{1.6}$$

Considering the expression for normalized effective capacity, the wideband slope can

be found from[1]

$$\mathcal{S}_0 = -\frac{2(\dot{\mathsf{C}}_E(0))^2}{\ddot{\mathsf{C}}_E(0)} \log_e 2 \tag{1.7}$$

where $\dot{\mathsf{C}}_E(0)$ and $\ddot{\mathsf{C}}_E(0)$ are the first and second derivatives, respectively, of the function $\mathsf{C}_E(\text{SNR})$ in bits/s/Hz at zero SNR [21]. $\frac{E_b}{N_0\,\min}$ and $\mathcal{S}_0$ provide a linear approximation of the spectral efficiency curve at low spectral efficiencies, i.e.,

$$\mathsf{C}_E\left(\frac{E_b}{N_0}\right) = \frac{\mathcal{S}_0}{10\log_{10} 2}\left(\frac{E_b}{N_0}\bigg|_{dB} - \frac{E_b}{N_{0\,\min}}\bigg|_{dB}\right) + \epsilon \tag{1.8}$$

where $\frac{E_b}{N_0}\Big|_{dB} = 10\log_{10}\frac{E_b}{N_0}$ and $\epsilon = o\left(\frac{E_b}{N_0} - \frac{E_b}{N_{0\,\min}}\right)$. Note that the expressions in (1.5) and (1.7) for fixed rate transmissions studied in Chapters 3 and 4 will change accordingly to accommodate the transmission schemes.

1.3 Review of Research on Multiple-access fading channels

In wireless networks, the design and analysis of efficient transmissions strategies have been of significant interest for many years. In particular, fading multiple access channels (MAC) have been extensively studied from an information-theoretic point of view [3][22][23][24][25][26][27]. For instance, Tse and Hanly [23] have characterized the capacity region of and determined the optimal resource allocation policies for multiple access fading channels. They have shown that the boundary surface points are in general achieved by superposition coding and successive decoding techniques, and obtaining each boundary point can be associated with an optimization problem in which a weighted sum rate is maximized. Vishwanath *et al.* [26] derived the explicit

[1] We note that the expressions in (1.5) and (1.7) differ from those in [21] by a constant factor due to the fact that we assume that the units of C_E is bits/s/Hz rather than nats/s/Hz.

optimal power and rate allocation schemes (similar to *waterfilling*) by considering that the users are successively decoded in the same order for all channel states. For the convex capacity region, the unique decoding order was shown to be the reverse order of the priority weight. While superposition coding and successive decoding strategies provide superior performance, time-division multiple access (TDMA) may in certain cases be preferred due to its simplicity. Note that the performance of TDMA approaches that of the optimal strategy as the signal-to-noise ratio (SNR) vanishes but, as shown by Caire *et al.* in [27], TDMA is strictly suboptimal when SNR is low but nonzero.

While establishing the fundamental performance limits, the aforementioned studies have not explicitly taken into account buffer constraints and random arrivals. In [28] and [29], Yeh and Cohen considered multiaccess fading channels with random packet arrivals to buffered transmitters, and characterized rate and power allocation strategies that maximize the stable throughput of the system. The maximum stable throughput region was shown in [30] to be the same as the MAC capacity region. In [31], the same authors investigated rate allocation policies that minimize the average packet delay in multiaccess fading channels again under the assumption of randomly arriving packets. More recently, Ehsan and Javidi in [32] studied delay optimal rate allocation strategies as well in two-user multiaccess channels but in the presence of asymmetric arrival processes, processing rates, and packet length distributions. Yang and Ulukus in [33] also considered an asymmetric setting and analyzed how to control the transmission probabilities in order to minimize the average delay in a two-user multiaccess scenario.

In Chapter 5, we also investigate the performance under buffer constraints but provide a perspective different from those of previous studies. In particular, we consider statistical quality of service (QoS) constraints in the form of limitations on the buffer

violation probabilities, and study the achievable rate region under such constraints in multiaccess fading channels. In [34], Liu *et al.* considered a two-user cooperative multiple access fading channel and analyzed the rate region achieved with frequency-division multiplexing when the users are operating under QoS constraints in the form of limitations on buffer overflow probabilities. In this study, cooperation among the users is shown to significantly improve the achievable rate region if the quality of the wireless link between the users is better than those of the links between the users and the destination. We note that since the transmitters are assumed to not know the channel conditions, power and rate adaptation policies are not studied in [34]. Additionally, since orthogonal transmission schemes are considered, superposition coding and successive decoding strategies are not addressed in detail.

1.4 Main Contributions

The analysis and application of effective capacity in various settings has attracted much interest recently (see e.g., [16][17][34][35][36][37][38][39][40][41] and references therein). Motivated by this observation, we attempt to make progress towards a better understanding of wireless communications under QoS constraints and draw valuable insights for the design of communication systems.

Chapter 2 deals with the energy efficiency in fading channels with variable rate transmissions [19]. Spectral efficiency–bit energy tradeoff is analyzed in the low-power and wideband regimes by employing the effective capacity formulation, rather than the Shannon capacity. Through this analysis, energy requirements under QoS constraints are identified. The analysis is conducted under two assumptions: perfect channel side information (CSI) available only at the receiver and perfect CSI available at both the receiver and transmitter. In particular, it is shown in the low-power

regime that the minimum bit energy required under QoS constraints is the same as that attained when there are no such limitations. However, this performance is achieved as the transmitted power vanishes. Through the wideband slope analysis, the increased energy requirements at low but nonzero power levels in the presence of QoS constraints are determined. A similar analysis is also conducted in the wideband regime. The minimum bit energy and wideband slope expressions are obtained. In this regime, the required bit energy levels are found to be strictly greater than those achieved when Shannon capacity is considered.

Chapter 3 solves the problem of energy efficiency with fixed rate transmissions [42]. When only the receiver has CSI, transmitter is assumed to send the information at a fixed rate. A two-state (ON-OFF) transmission model is adopted, where information is transmitted reliably at a fixed rate in the ON state while no reliable transmission occurs in the OFF state. We obtain the bit energy required at zero spectral efficiency and the wideband slope in both wideband and low-power regimes. Initially, the wideband regime with multipath sparsity is investigated, and the minimum bit energy and wideband slope expressions are found. It is shown that the minimum bit energy requirements increase as the QoS constraints become more stringent. Subsequently, the low-power regime, which is also equivalent to the wideband regime with rich multipath fading, is analyzed. In this case, bit energy requirements are quantified through the expressions of bit energy required at zero spectral efficiency and wideband slope. It is shown for a certain class of fading distributions that the bit energy required at zero spectral efficiency is indeed the minimum bit energy for reliable communications. Moreover, it is proven that this minimum bit energy is attained in all cases regardless of the strictness of the QoS limitations. The impact upon the energy efficiency of multipath sparsity and richness is quantified.

In Chapter 4, we consider the energy efficiency when neither the transmitter or the

receiver has the perfect CSI [43]. In this case, the channel coefficients are estimated at the receiver via minimum mean-square-error (MMSE) estimation with the aid of training symbols. The optimal fraction of power allocated to training is identified. It is shown that the bit energy increases without bound in the low-power regime as the average power vanishes. A similar conclusion is reached in the wideband regime if the number of noninteracting subchannels grow without bound with increasing bandwidth. On the other hand, it is proven that if the number of resolvable independent paths and hence the number of noninteracting subchannels remain bounded as the available bandwidth increases, the bit energy diminishes to its minimum value in the wideband regime. For this case, expressions for the minimum bit energy and wideband slope are derived. Overall, energy costs of channel uncertainty and queueing constraints are identified, and the impact of multipath richness and sparsity is determined.

Chapter 5 studies the effective capacity region of fading MACs in the presence of quality of service (QoS) constraints [44]. Perfect channel side information (CSI) is assumed to be available at both the transmitters and the receiver. With fixed power, the performance achieved by superposition coding with successive decoding techniques is investigated. It is shown that varying the decoding order with respect to the channel states can significantly increase the achievable throughput region. In the two-user case, the optimal decoding strategy is determined for the scenario in which the users have the same QoS constraints. The performance of orthogonal transmission strategies is also analyzed. It is shown that for certain QoS constraints, time-division multiple-access (TDMA) can achieve better performance than superposition coding if fixed successive decoding order is used at the receiver side. When power control policies are incorporated, we identify the optimal power allocation policies for any fixed decoding order over all channel states. For a given variable decoding order strategy,

the conditions that the optimal power control policies must satisfy are determined, and an algorithm that can be used to compute these optimal policies is provided.

In Chapter 6, the minimum bit energy levels and wideband slope regions of fading MAC have been characterized for different transmission and reception strategies, namely time-division multiple-access (TDMA), superposition coding with fixed decoding order, and superposition coding with variable decoding order [45]. In the low-power regime, it has been shown that the minimum received bit energies achieved by these different strategies are the same and independent of the QoS constraints. For the case of superposition coding, it has been found that varying the decoding order at the receiver with the fading realizations does not enlarge the wideband slope region. Also, the suboptimality of TDMA with respect to superposition schemes has been proved except for the special case in which the fading states are linearly dependent. In the wideband regime, the minimum bit energies achieved by different strategies are the same for each user but vary with the QoS constraints. One stark difference from the results in the low-power regime is that varying the decoding order at the receiver with the fading realizations expands the wideband slope region. Also, unlike in the low-power regime, it is shown that TDMA can interestingly outperform superposition coding with fixed decoding order when wideband slope regions are considered. Conversely, the condition under which TDMA slope region is inside that of superposition coding with fixed decoding order has been determined.

In Chapter 7, a two-hop wireless communication link in which a source sends data to a destination with the aid of an intermediate relay node is studied [46]. It is assumed that there is no direct link between the source and the destination, and the relay forwards the information to the destination by employing the decode-and-forward scheme. Both the source and intermediate relay nodes are assumed to operate under statistical quality of service (QoS) constraints imposed as limitations

on the buffer overflow probabilities. The maximum constant arrival rates that can be supported by this two-hop link in the presence of QoS constraints are characterized by determining the effective capacity of such links as a function of the QoS parameters and signal-to-noise ratios at the source and relay, and the fading distributions of the links. The analysis is performed for both full-duplex and half-duplex relaying. Through this study, the impact upon the throughput of having buffer constraints at the source and intermediate relay nodes is identified. The interactions between the buffer constraints in different nodes and how they affect the performance are studied. The optimal time-sharing parameter in half-duplex relaying is determined, and performance with half-duplex relaying is investigated.

Finally, in Chapter 8, we study the finite-blocklength channel codes for transmission [47]. A block fading model, in which fading stays constant in each coherence block and changes independently between blocks, is considered. It is assumed that channel coding is performed over multiple coherence blocks. An approximate lower bound on the transmission rate is obtained from Feintein's Lemma. This lower bound is considered as the service rate and is incorporated into the effective capacity formulation. The optimum error probability for variable-rate transmission and the optimum coding rate for fixed-rate transmission are shown to be unique. The tradeoffs and interactions between the throughput, the number of blocks over which channel coding is performed, error probabilities, channel coherence duration, and queueing constraints are identified.

Chapter 2

Energy Efficiency for Variable Rate Transmissions

We first study the variable-rate/variable-power and variable-rate/fixed-power transmission schemes with different assumptions on the availability of channel side information (CSI) at the transmitter and receiver [48]. We obtain the minimum bit energy and wideband slope expressions, and in the variable-power case, we analyze the impact of power control policies on energy efficiency.

2.1 System Model

We consider a point-to-point communication system in which there is one source and one destination. The general system model is depicted in Fig.2.1, and is similar to the one studied in [35]. In this model, it is assumed that the source generates data sequences which are divided into frames of duration T. These data frames are initially stored in the buffer before they are transmitted over the wireless channel. The discrete-time channel input-output relation in the i^{th} symbol duration is given

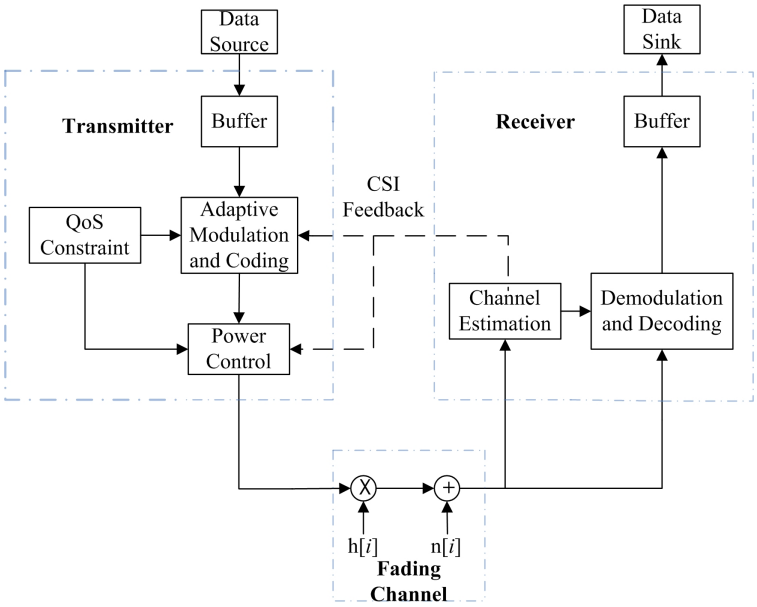

Figure 2.1: The system model

by

$$y[i] = h[i]x[i] + n[i] \quad i = 1, 2, \ldots . \tag{2.1}$$

where $x[i]$ and $y[i]$ denote the complex-valued channel input and output, respectively. The channel input is subject to an average power constraint $\mathbb{E}\{|x[i]|^2\} \leq \bar{P}$ for all i, and we assume that the bandwidth available in the system is B. Above, $n[i]$ is a zero-mean, circularly symmetric, complex Gaussian random variable with variance $\mathbb{E}\{|n[i]|^2\} = N_0$. The additive Gaussian noise samples $\{n[i]\}$ are assumed to form an independent and identically distributed (i.i.d.) sequence. Finally, $h[i]$ denotes the channel fading coefficient, and $\{h[i]\}$ is a stationary and ergodic discrete-time process. We assume that perfect channel state information (CSI) is available at the receiver while the transmitter has either *no* or *perfect* CSI. The availability of CSI at the transmitter is facilitated through CSI feedback from the receiver. Note that

if the transmitter knows the channel fading coefficients, it employs power and rate adaptation. Otherwise, the signals are sent with constant power.

Note that in the above system model, the average transmitted signal-to-noise ratio is $\text{SNR} = \bar{P}/(N_0 B)$. We denote the magnitude-square of the fading coefficient by $z[i] = |h[i]|^2$, and its distribution function by $p_z(z)$. When there is only receiver CSI, instantaneous transmitted power is $P[i] = \bar{P}$ and instantaneous received SNR is expressed as $\gamma[i] = \bar{P} z[i]/(N_0 B)$. Moreover, the maximum instantaneous service rate $R[i]$ is

$$R[i] = B \log_2 \Big(1 + \text{SNR} z[i]\Big) \quad \text{bits/s}. \tag{2.2}$$

We note that although the transmitter does not know $z[i]$, recently developed rateless codes such as LT [49] and Raptor [50] codes enable the transmitter to adapt its rate to the channel realization and achieve $R[i]$ without requiring CSI at the transmitter side [51], [52]. For systems that do not employ such codes, service rates are smaller than that in (2.2), and the results in this chapter serve as upper bounds on the performance.

When the transmitter also has CSI, the instantaneous service rate is

$$R[i] = B \log_2 \Big(1 + \mu_{\text{opt}}(\theta, z[i]) z[i]\Big) \quad \text{bits/s} \tag{2.3}$$

where $\mu_{\text{opt}}(\theta, z)$ is the power-adaptation policy that maximizes the effective capacity, which has been discussed in Section 1.1. This optimal power policy is determined in [35]:

$$\mu_{\text{opt}}(\theta, z) = \begin{cases} \frac{1}{\alpha^{\frac{1}{\beta+1}} z^{\frac{\beta}{\beta+1}}} - \frac{1}{z} & z \geq \alpha \\ 0 & z < \alpha \end{cases} \tag{2.4}$$

where θ is the QoS exponent defined in (1.2), $\beta = \frac{\theta TB}{\log_e 2}$ is the normalized QoS exponent and α is the channel threshold chosen to satisfy the average power constraint:

$$\text{SNR} = \mathbb{E}\{\mu_{\text{opt}}(\theta, z)\} = \mathbb{E}\left\{\left[\frac{1}{\alpha^{\frac{1}{\beta+1}} z^{\frac{\beta}{\beta+1}}} - \frac{1}{z}\right]\tau(\alpha)\right\} \tag{2.5}$$

where $\tau(\alpha) = 1\{z \geq \alpha\} = \begin{cases} 1 & \text{if } z \geq \alpha \\ 0 & \text{if } z < \alpha \end{cases}$ is the indicator function. Note that $\mu_{\text{opt}}(\theta, z)$ depends on the average power constraint only through the threshold α. Moreover, power allocation strategy $\mu_{\text{opt}}(\theta, z)$, while varying with the instantaneous values of the fading coefficients, depends on the queueing constraints statistically only through the QoS exponent θ, and hence is not a function of the instantaneous queue lengths.

We finally note that since the maximum service rates are equal to the instantaneous channel capacity values, we assume through information-theoretic arguments that when the transmitter transmits at the rate $R[i]$ given in (2.2) and (2.3), the information is reliably received at the receiver and no retransmissions are required.

It can be easily shown that effective capacity specializes to the Shannon capacity and delay-limited capacity in the asymptotic regimes. As θ approaches to 0, constraints on queue length and queueing delay relax, and effective capacity converges to the Shannon ergodic capacity:

$$\lim_{\theta \to 0} C_E(\text{SNR}, \theta) = \begin{cases} \mathbb{E}\{B \log_2(1 + \text{SNR} z)\} & \text{CSI at the RX} \\ \mathbb{E}\left\{B \log_2\left(1 + \mu_{\text{opt}}(0, z) z\right)\right\} & \text{CSI at the RX and TX} \end{cases} \tag{2.6}$$

where expectations are with respect to z. Note that in (2.6), $\mu_{\text{opt}}(0, z)$ is the water-filling power adaptation policy, which maximizes the Shannon capacity. On the other hand, as $\theta \to \infty$, QoS constraints become more and more strict and effective capacity

approaches the delay-limited capacity which as described before can be seen as a deterministic service guarantee:

$$\lim_{\theta\to\infty} C_E(\mathrm{SNR},\theta) = \begin{cases} B\log_2(1+\mathrm{SNR}z_{\min}) & \text{CSI at the RX} \\ B\log_2(1+\sigma) & \text{CSI at the RX and TX} \end{cases} \tag{2.7}$$

where $\sigma = \frac{\mathrm{SNR}}{\mathbb{E}\{1/z\}}$ and $z_{\min}$ is the minimum value of the random variable z, i.e., $z \geq z_{\min} \geq 0$ with probability 1. Note that in Rayleigh fading, $\sigma = 0$ and $z_{\min} = 0$, and hence the delay-limited capacities are zero in both cases and no deterministic guarantees can be provided.

We first have the following preliminary result.

Proposition 1 *The normalized effective capacity, $\mathsf{C}_E(SNR)$, given in (1.4) is a concave function of SNR with the transmission schemes described above.*

Proof: It can be easily seen that $e^{-\theta TR[i]}$, where $R[i] = B\log_2(1+\mathrm{SNR}z[i])$, is a log-convex function of SNR because $-R[i]$ is a convex function of SNR. Since log-convexity is preserved under sums, $g(x) = \int f(x,y)dy$ is log-convex in x if $f(x,y)$ is log-convex in x for each y [53]. From this fact, we immediately conclude that $\mathbb{E}\{e^{-\theta TR[i]}\}$ is also a log-convex function of SNR. Hence, $\log_e \mathbb{E}\{e^{-\theta TR[i]}\}$ is convex and $-\log_e \mathbb{E}\{e^{-\theta TR[i]}\}$ is concave in SNR.

When the transmitter also has CSI, we have $R[i] = B\log_2(1+\mu_{\mathrm{opt}}(\theta, z[i])z[i])$. In this case, the concavity of C_E in SNR can be easily proven using the facts that $\mathbb{E}\{e^{-\theta TR[i]}\}$ is an non-decreasing, concave function of the threshold value α (specified in (2.4)) and approaches zero as α diminishes to zero, and α is a non-increasing function of SNR. □

2.2 Energy Efficiency in the Low-Power Regime

As discussed in the previous section, the minimum bit energy is achieved as $\text{SNR} = \frac{\bar{P}}{N_0 B} \to 0$, and hence energy efficiency improves if one operates in the low-power regime in which $\bar{P}$ is small, or the high-bandwidth regime in which B is large. From the Shannon capacity perspective, similar performances are achieved in these two regimes, which therefore can be seen as equivalent. However, as we shall see in this chapter, considering the effective capacity leads to different results at low power and high bandwidth levels. In this section, we consider the low-power regime for fixed bandwidth, B, and study the spectral efficiency vs. bit energy tradeoff by finding the minimum bit energy and the wideband slope.

2.2.1 CSI at the Receiver Only

We initially consider the case in which only the receiver knows the channel conditions. Substituting (2.2) into (1.4), we obtain the spectral efficiency given θ as a function of SNR:

$$\mathsf{C}_E(\text{SNR}) = -\frac{1}{\theta TB}\log_e \mathbb{E}\{e^{-\theta TB\log_2(1+\text{SNR}z)}\} = -\frac{1}{\theta TB}\log_e \mathbb{E}\{(1+\text{SNR}z)^{-\beta}\} \quad (2.8)$$

where again $\beta = \frac{\theta TB}{\log_e 2}$. Note that since the analysis is performed for fixed θ throughout the chapter, we henceforth express the effective capacity only as a function of SNR to simplify the expressions. The following result provides the minimum bit energy and the wideband slope.

Theorem 1 *When only the receiver has perfect CSI, the minimum bit energy and*

wideband slope are

$$\frac{E_b}{N_0}_{min} = \frac{\log_e 2}{\mathbb{E}\{z\}} \text{ and } \mathcal{S}_0 = \frac{2}{(\beta+1)\frac{\mathbb{E}\{z^2\}}{\Big(\mathbb{E}\{z\}\Big)^2} - \beta}. \tag{2.9}$$

Proof: The first and second derivative of $\mathsf{C}_E(\text{SNR})$ with respect to SNR are given by

$$\dot{\mathsf{C}}_E(\text{SNR}) = \frac{1}{\log_e 2}\frac{\mathbb{E}\{(1+\text{SNR}z)^{-(\beta+1)}z\}}{\mathbb{E}\{(1+\text{SNR}z)^{-\beta}\}} \quad \text{and,} \tag{2.10}$$

$$\ddot{\mathsf{C}}_E(\text{SNR}) = \frac{\beta}{\log_e 2}\left(\frac{\mathbb{E}\{(1+\text{SNR}z)^{-(\beta+1)}z\}}{\mathbb{E}\{(1+\text{SNR}z)^{-\beta}\}}\right)^2 - \frac{\beta+1}{\log_e 2}\frac{\mathbb{E}\{(1+\text{SNR}z)^{-(\beta+2)}z^2\}}{\mathbb{E}\{(1+\text{SNR}z)^{-\beta}\}}, \tag{2.11}$$

respectively, which result in the following expressions when $\text{SNR} = 0$:

$$\dot{\mathsf{C}}_E(0) = \frac{\mathbb{E}\{z\}}{\log_e 2} \quad \text{and} \quad \ddot{\mathsf{C}}_E(0) = -\frac{1}{\log_e 2}\Big((\beta+1)\mathbb{E}\{z^2\} - \beta\Big(\mathbb{E}\{z\}\Big)^2\Big). \tag{2.12}$$

Substituting the expressions in (2.12) into (1.5) and (1.7) provides the desired result. □

From the above result, we immediately see that $\frac{E_b}{N_0}_{\min}$ does not depend on θ and the minimum *received* bit energy is $\frac{E_b^r}{N_0}_{\min} = \frac{E_b}{N_0}_{\min}\mathbb{E}\{z\} = \log_e 2 = -1.59$ dB. Note that if the Shannon capacity is used in the analysis, i.e., if $\theta = 0$ and hence $\beta = 0$, $\frac{E_b^r}{N_0}_{\min} = -1.59$ dB and $\mathcal{S}_0 = 2/(\mathbb{E}\{z^2\}/\mathbb{E}^2\{z\})$. Therefore, we conclude from Theorem 1 that as the average power $\bar{P}$ decreases, energy efficiency approaches the performance achieved by a system that does not have QoS limitations. However, we note that wideband slope is smaller if $\theta > 0$. Hence, the presence of QoS constraints decreases the spectral efficiency or equivalently increases the energy requirements for fixed spectral efficiency values at low but nonzero SNR levels.

Fig. 2.2 plots the spectral efficiency as a function of the bit energy for different values of θ in the Rayleigh fading channel with $\mathbb{E}\{|h|^2\} = \mathbb{E}\{z\} = 1$. Note that the

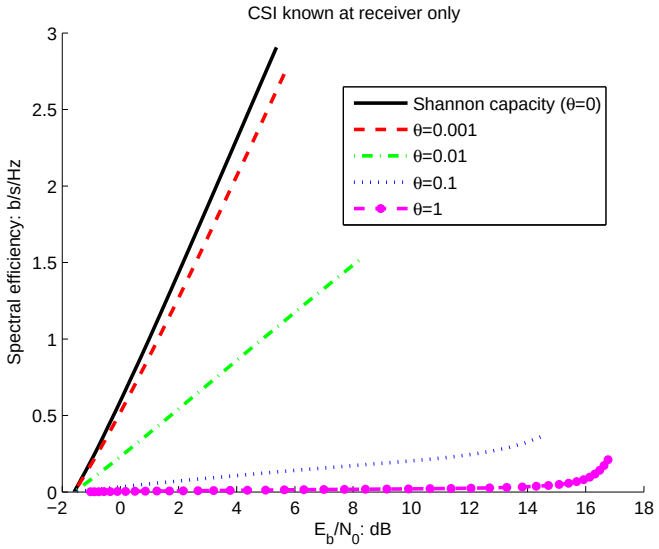

Figure 2.2: Spectral efficiency vs. E_b/N_0 in the Rayleigh fading channel with fixed B; CSI known at the receiver only.

curve for $\theta = 0$ corresponds to the Shannon capacity. Throughout the chapter, we set the frame duration to $T = 2$ms in the numerical results. For the fixed bandwidth case, we have assumed $B = 10^5$ Hz. In Fig. 2.2, we observe that all curves approach $\frac{E_b}{N_0}_{\min} = -1.59$ dB as predicted. On the other hand, we note that the wideband slope decreases as θ increases. Therefore, at low but nonzero spectral efficiencies, more energy is required as the QoS constraints become more stringent. Considering the linear approximation in (1.8), we can easily show for fixed spectral efficiency $\mathsf{C}\left(\frac{E_b}{N_0}\right)$ for which the linear approximation is accurate that the increase in the bit energy in dB, when the QoS exponent increases from θ_1 to θ_2, is $\frac{E_b}{N_0}\Big|_{dB,\theta_2} - \frac{E_b}{N_0}\Big|_{dB,\theta_1} = \left(\frac{1}{\mathcal{S}_{0,\theta_2}} - \frac{1}{\mathcal{S}_{0,\theta_1}}\right) \mathsf{C}\left(\frac{E_b}{N_0}\right) 10 \log_{10} 2$.

2.2.2 CSI at both the Transmitter and Receiver

We now consider the case in which both the transmitter and receiver have perfect CSI. Substituting (2.3) into (1.4), we have

$$\begin{aligned} \mathsf{C}_E(\text{SNR}) &= -\frac{1}{\theta TB}\log_e \mathbb{E}\Big\{ e^{-\theta TB\log_2\left(1+\mu_{\text{opt}}(\theta,z)z\right)} \Big\} \\ &= -\frac{1}{\theta TB}\log_e \left(F(\alpha) + \mathbb{E}\Big\{ \left(\frac{z}{\alpha}\right)^{-\frac{\beta}{\beta+1}} \tau(\alpha) \Big\} \right) \end{aligned} \tag{2.13}$$

where $F(\alpha) = \mathbb{E}\{1\{z < \alpha\}\}$. We note that the normalized effective capacity expression in (2.13) is obtained assuming that the optimal power-adaptation policy $\mu_{\text{opt}}(\theta, z)$ given in (2.4) is employed in the system. Maximizing the effective capacity, this optimal power allocation policy minimizes the bit energy requirements. For this case, following an approach similar to that in [54], we obtain the following result.

Theorem 2 *When both the transmitter and receiver have perfect CSI, the minimum bit energy with optimal power control and rate adaptation becomes*

$$\frac{E_b}{N_0}_{min} = \frac{\log_e 2}{z_{\max}} \tag{2.14}$$

where $z_{\max}$ is the essential supremum of the random variable z, i.e., $z \leq z_{\max}$ with probability 1.

Proof: We assume that $z_{\max}$ is the maximum value that the random variable z can take, i.e., $P(z \leq z_{\max}) = 1$. From (2.5), we can see that as SNR vanishes, α increases to $z_{\max}$, because otherwise while SNR approaches zero, the right most side of (2.5) does not. Then, we can suppose for small enough SNR that $\alpha = z_{\max} - \eta$ where $\eta \to 0$

as SNR $\to 0$. Replacing α by $z_{\max} - \eta$ in (2.5) and (2.13), we get

$$\frac{E_b}{N_{0\,\min}} = \lim_{\mathrm{SNR}\to 0} \frac{\mathrm{SNR}}{\mathsf{C}(\mathrm{SNR})}$$

$$= \lim_{\eta\to 0} \frac{\mathbb{E}\left\{\left[\frac{1}{(z_{\max}-\eta)^{\frac{1}{\beta+1}} z^{\frac{\beta}{\beta+1}}} - \frac{1}{z}\right]\tau(z_{\max}-\eta)\right\}}{-\frac{1}{\theta TB}\log_e\left(F(z_{\max}-\eta) + \mathbb{E}\left\{\left(\frac{z}{z_{\max}-\eta}\right)^{-\frac{\beta}{\beta+1}}\tau(z_{\max}-\eta)\right\}\right)} \tag{2.15}$$

$$= \lim_{\eta\to 0} \frac{\int_{z_{\max}-\eta}^{z_{\max}} \left(\frac{1}{(z_{\max}-\eta)^{\frac{1}{\beta+1}} z^{\frac{\beta}{\beta+1}}} - \frac{1}{z}\right) p_z(z)dz}{-\frac{1}{\theta TB}\log_e\left(\int_0^{z_{\max}-\eta} p_z(z)dz + \int_{z_{\max}-\eta}^{z_{\max}} \left(\frac{z}{z_{\max}-\eta}\right)^{-\frac{\beta}{\beta+1}} p_z(z)dz\right)} \tag{2.16}$$

$$= \lim_{\eta\to 0} \frac{\frac{1}{\beta+1}(z_{\max}-\eta)^{-\frac{\beta+2}{\beta+1}} \int_{z_{\max}-\eta}^{z_{\max}} \frac{p_z(z)}{z^{\frac{\beta}{\beta+1}}}dz}{-\frac{1}{\beta\log_e 2}\frac{-\frac{\beta}{\beta+1}(z_{\max}-\eta)^{-\frac{1}{\beta+1}} \int_{z_{\max}-\eta}^{z_{\max}} \frac{p_z(z)}{z^{\frac{\beta}{\beta+1}}}dz}{\int_0^{z_{\max}-\eta} p_z(z)dz + \int_{z_{\max}-\eta}^{z_{\max}} \left(\frac{z}{z_{\max}-\eta}\right)^{-\frac{\beta}{\beta+1}} p_z(z)dz}} \tag{2.17}$$

$$= \lim_{\eta\to 0} \frac{\left(\int_0^{z_{\max}-\eta} p_z(z)dz + \int_{z_{\max}-\eta}^{z_{\max}} \left(\frac{z}{z_{\max}-\eta}\right)^{-\frac{\beta}{\beta+1}} p_z(z)dz\right)\log_e 2}{z_{\max}-\eta} \tag{2.18}$$

$$= \frac{\log_e 2}{z_{\max}} \tag{2.19}$$

where p_z is the distribution of channel gain z. (2.16) is obtained by expressing the expectations in (2.15) as integrals. (2.17) follows by using the L'Hospital's Rule and applying Leibniz Integral Rule. (2.18) is obtained after straightforward algebraic simplifications and the result follows immediately.

Above, we have implicitly assumed that $z_{\max}$ is finite. For fading distributions with unbounded support, $z_{\max} = \infty$. In this case, the result can be shown by replacing in (2.16) $z_{\max}$ by ∞, and $z_{\max} - \eta$ by the threshold α, and letting $\alpha \to \infty$. After these steps, the final expression, which is akin to that in (2.19), becomes $\lim_{\alpha\to\infty} \frac{\log_e 2}{\alpha} = 0$, proving that (2.14) also holds for the case in which $z_{\max} = \infty$. □

Note that for distributions with unbounded support, we have $z_{\max} = \infty$ and hence $\frac{E_b}{N_0}_{\min} = 0 = -\infty$ dB. In this case, it is easy to see that the wideband slope is $S_0 = 0$.

Example 1 *Specifically, for the Rayleigh fading channel, as in [55], it can be shown that* $\lim_{SNR\to 0} \frac{C_E(SNR)}{SNR \log_e(\frac{1}{SNR}) \log_e 2} = 1$. *Then, spectral efficiency can be written as* $C_E(SNR) \approx SNR \log_e(\frac{1}{SNR}) \log_e 2$, *so*

$$\frac{E_b}{N_0}_{min} = \lim_{SNR\to 0} \frac{SNR}{C_E(SNR)} = \lim_{SNR\to 0} \frac{1}{\log_e(\frac{1}{SNR}) \log_e 2} = 0$$

which also verifies the above result.

Remark: We note that as in the case in which there is CSI at the receiver, the minimum bit energy achieved under QoS constraints is the same as that achieved by the Shannon capacity [54]. Hence, the energy efficiency again approaches the performance of an unconstrained system as power diminishes. Searching for an intuitive explanation of this observation, we note that arrival rates that can be supported vanishes with decreasing power levels. As a result, the impact of buffer occupancy constraints on the performance lessens. Note that in contrast, increasing the bandwidth increases the arrival rates supported by the system. Therefore, limitations on the buffer occupancy will have significant impact upon the energy efficiency in the wideband regime as will be discussed in Section 2.3.

Fig. 2.3 plots the spectral efficiency vs. bit energy for different values of θ in the Rayleigh fading channel with $\mathbb{E}\{z\} = 1$. In all cases, we observe that the bit energy goes to $-\infty$ as the spectral efficiency decreases. We also note that at small but nonzero spectral efficiencies, the required energy is higher as θ increases.

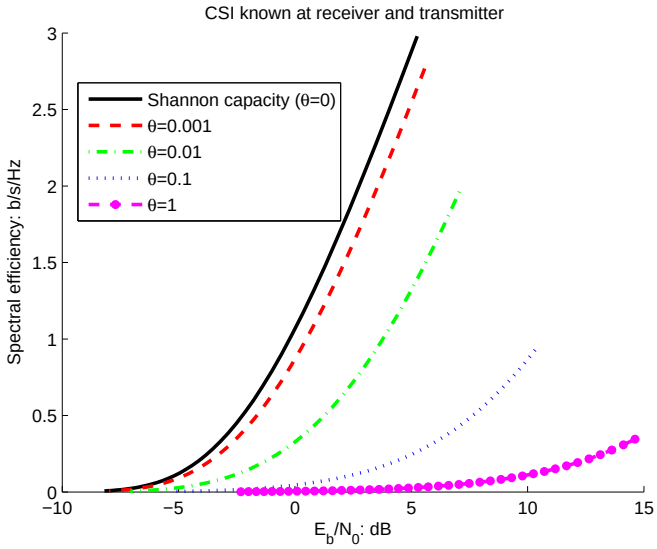

Figure 2.3: Spectral efficiency vs. E_b/N_0 in the Rayleigh fading channel with fixed B; CSI known at the transmitter and receiver.

2.3 Energy Efficiency in the Wideband Regime

In this section, we study the performance at high bandwidths while the average power $\bar{P}$ is kept fixed. We investigate the impact of θ on $\frac{E_b}{N_0}_{\min}$ and the wideband slope $\mathcal{S}_0$ in this wideband regime. Note that as the bandwidth increases, the average signal-to-noise ratio $\text{SNR} = \bar{P}/(N_0 B)$ and the spectral efficiency decreases.

2.3.1 CSI at the Receiver Only

We define $\zeta = \frac{1}{B}$ and express the spectral efficiency (2.8) as a function of ζ:

$$\mathsf{C}_E(\zeta) = -\frac{\zeta}{\theta T} \log_e \mathbb{E}\{e^{-\frac{\theta T}{\zeta} \log_2(1+\frac{\bar{P}\zeta}{N_0} z)}\}. \tag{2.20}$$

The bit energy is again defined as

$$\frac{E_b}{N_0} = \frac{\text{SNR}}{\mathsf{C}_E(\text{SNR})} = \frac{\frac{\bar{P}\zeta}{N_0}}{\mathsf{C}_E(\zeta)} = \frac{\frac{\bar{P}}{N_0}}{\mathsf{C}_E(\zeta)/\zeta}. \tag{2.21}$$

It can be readily verified that $\mathsf{C}_E(\zeta)/\zeta$ monotonically increases as $\zeta \to 0$ (or equivalently as $B \to \infty$) (see Appendix A). Therefore

$$\frac{E_b}{N_0}_{\min} = \lim_{\zeta \to 0} \frac{\bar{P}\zeta/N_0}{\mathsf{C}_E(\zeta)} = \frac{\bar{P}/N_0}{\dot{\mathsf{C}}_E(0)} \tag{2.22}$$

where $\dot{\mathsf{C}}_E(0)$ is the first derivative of the spectral efficiency with respect to ζ at $\zeta = 0$. The wideband slope $\mathcal{S}_0$ can be obtained from the formula (1.7) by using the first and second derivatives of the spectral efficiency $\mathsf{C}_E(\zeta)$ with respect to ζ.

Theorem 3 *When only the receiver has CSI, the minimum bit energy and wideband slope, respectively, in the wideband regime are given by*

$$\frac{E_b}{N_0}_{min} = -\frac{\frac{\theta T \bar{P}}{N_0}}{\log_e \mathbb{E}\{e^{-\frac{\theta T \bar{P}}{N_0 \log_e 2} z}\}}, \quad \textit{and} \tag{2.23}$$

$$\mathcal{S}_0 = 2\Big(\frac{N_0 \log_e 2}{\theta T \bar{P}}\Big)^2 \frac{\mathbb{E}\{e^{-\frac{\theta T \bar{P}}{N_0 \log_e 2} z}\}\Big(\log_e \mathbb{E}\{e^{-\frac{\theta T \bar{P}}{N_0 \log_e 2} z}\}\Big)^2}{\mathbb{E}\{e^{-\frac{\theta T \bar{P}}{N_0 \log_e 2} z} z^2\}}. \tag{2.24}$$

Proof: The first and second derivative of $\mathsf{C}_E(\zeta)$ are given by

$$\dot{\mathsf{C}}_E(\zeta) = -\frac{1}{\theta T} \log_e \mathbb{E}\{e^{-\frac{\theta T}{\zeta} \log_2(1+\frac{\bar{P}\zeta z}{N_0})}\} - \frac{\mathbb{E}\left\{e^{-\frac{\theta T}{\zeta} \log_2(1+\frac{\bar{P}\zeta z}{N_0})}\left[\frac{\log_2(1+\frac{\bar{P}\zeta z}{N_0})}{\zeta} - \frac{\frac{\bar{P} z}{N_0 \log_e 2}}{1+\frac{\bar{P}\zeta z}{N_0}}\right]\right\}}{\mathbb{E}\{e^{-\frac{\theta T}{\zeta} \log_2(1+\frac{\bar{P}\zeta z}{N_0})}\}}, \tag{2.25}$$

$$\ddot{\mathsf{C}}_E(\zeta) = \frac{\theta T}{\zeta}\left(\frac{\mathbb{E}\left\{e^{-\frac{\theta T}{\zeta}\log_2(1+\frac{\bar{P}\zeta z}{N_0})}\left[\frac{\log_2(1+\frac{\bar{P}\zeta z}{N_0})}{\zeta} - \frac{\frac{\bar{P}z}{N_0\log_e 2}}{1+\frac{\bar{P}\zeta z}{N_0}}\right]\right\}}{\mathbb{E}\{e^{-\frac{\theta T}{\zeta}\log_2(1+\frac{\bar{P}\zeta z}{N_0})}\}}\right)^2 - \frac{\mathbb{E}\left\{e^{-\frac{\theta T}{\zeta}\log_2(1+\frac{\bar{P}\zeta z}{N_0})}\left[\frac{\theta T}{\zeta}\left(\frac{\log_2(1+\frac{\bar{P}\zeta z}{N_0})}{\zeta} - \frac{\frac{\bar{P}z}{N_0\log_e 2}}{1+\frac{\bar{P}\zeta z}{N_0}}\right)^2 + \log_e 2\left(\frac{\frac{\bar{P}z}{N_0\log_e 2}}{1+\frac{\bar{P}\zeta z}{N_0}}\right)^2\right]\right\}}{\mathbb{E}\{e^{-\frac{\theta T}{\zeta}\log_2(1+\frac{\bar{P}\zeta z}{N_0})}\}}. \tag{2.26}$$

First, we define the function $f(\zeta) = \frac{\log_2(1+\frac{\bar{P}\zeta z}{N_0})}{\zeta^2} - \frac{\frac{\bar{P}z}{N_0\zeta\log_e 2}}{1+\frac{\bar{P}\zeta z}{N_0}}$. Then, we can show that

$$\begin{aligned}\lim_{\zeta\to 0} f(\zeta) &= \lim_{\zeta\to 0}\frac{\frac{\log_2(1+\frac{\bar{P}\zeta z}{N_0})}{\zeta} - \frac{\frac{\bar{P}z}{N_0\log_e 2}}{1+\frac{\bar{P}\zeta z}{N_0}}}{\zeta}\\ &= \lim_{\zeta\to 0}\left(-\frac{\log_2(1+\frac{\bar{P}\zeta z}{N_0})}{\zeta^2} + \frac{\frac{\bar{P}z}{N_0\log_e 2}}{1+\frac{\bar{P}\zeta z}{N_0}} + \left(\frac{\frac{\bar{P}z}{N_0\log_e 2}}{1+\frac{\bar{P}\zeta z}{N_0}}\right)^2\log_e 2\right)\\ &= -\lim_{\zeta\to 0} f(\zeta) + \frac{1}{\log_e 2}\left(\frac{\bar{P}z}{N_0}\right)^2\end{aligned}$$

which yields

$$\lim_{\zeta\to 0} f(\zeta) = \frac{1}{2\log_e 2}\left(\frac{\bar{P}z}{N_0}\right)^2 \tag{2.27}$$

Using (2.27), we can easily find from (2.25) that

$$\lim_{\zeta\to 0}\dot{\mathsf{C}}_E(\zeta) = -\frac{1}{\theta T}\log_e \mathbb{E}\left\{e^{-\frac{\theta T\bar{P}}{N_0\log_e 2}z}\right\} \tag{2.28}$$

from which (2.23) follows immediately. Moreover, from (2.26), we can derive

$$\lim_{\zeta\to 0}\ddot{\mathsf{C}}_E(\zeta) = -\frac{1}{\log_e 2}\left(\frac{\bar{P}}{N_0}\right)^2\frac{\mathbb{E}\{e^{-\frac{\theta T\bar{P}}{N_0\log_e 2}z}z^2\}}{\mathbb{E}\{e^{-\frac{\theta T\bar{P}}{N_0\log_e 2}z}\}}. \tag{2.29}$$

Evaluating (1.7) with (2.28) and (2.29) provides (2.24). □

It is interesting to note that unlike the low-power regime results, we now have

$$\frac{E_b}{N_{0\,\min}} = \frac{-\frac{\theta T\bar{P}}{N_0}}{\log_e \mathbb{E}\{e^{-\frac{\theta T\bar{P}}{N_0 \log_e 2}z}\}} \geq \frac{-\frac{\theta T\bar{P}}{N_0}}{\mathbb{E}\{\log_e e^{-\frac{\theta T\bar{P}}{N_0 \log_e 2}z}\}} = \frac{\log_e 2}{\mathbb{E}\{z\}}$$

where Jensen's inequality is used. Therefore, we will be operating above -1.59 dB unless there are no QoS constraints and hence $\theta = 0$. For the Rayleigh channel, we can specialize (2.23) and (2.24) to obtain

$$\frac{E_b}{N_{0\,\min}} = \frac{\frac{\theta T\bar{P}}{N_0}}{\log_e(1+\frac{\theta T\bar{P}}{N_0 \log_e 2})} \quad \text{and}$$

$$\mathcal{S}_0 = \left(\frac{N_0 \log_e 2}{\theta T\bar{P}} \log_e(1+\frac{\theta T\bar{P}}{N_0 \log_e 2}) + \log_e(1+\frac{\theta T\bar{P}}{N_0 \log_e 2})\right)^2. \tag{2.30}$$

It can be easily seen that in the Rayleigh channel, the minimum bit energy monotonically increases with increasing θ. Fig. 2.4 plots the spectral efficiency curves as a function of bit energy in the Rayleigh channel. In all the curves, we set $\bar{P}/N_0 = 10^4$. We immediately observe that more stringent QoS constraints and hence higher values of θ lead to higher minimum bit energy values and also higher energy requirements at other nonzero spectral efficiencies. The wideband slope values are found to be equal to $\mathcal{S}_0 = \{1.0288, 1.2817, 3.3401, 12.3484\}$ for $\theta = \{0.001, 0.01, 0.1, 1\}$, respectively. Note that the wideband slope increases with increasing θ, indicating that the increment in the bit energy required to increase the spectral efficiency by a fixed amount in the wideband regime is smaller when θ is larger. We also note that despite this observation, since the minimum bit energy is also higher for larger θ, the absolute bit energy requirements at a given spectral efficiency are higher when θ is increased.

We finally note that $\frac{E_b}{N_{0\,\min}}$ and $\mathcal{S}_0$ now depend on θ and $\frac{\bar{P}}{N_0}$. Fig. 2.5 plots $\frac{E_b}{N_{0\,\min}}$ as a function of these two parameters. Probing into the inherent relationships among these parameters can give us some interesting results, which are helpful in

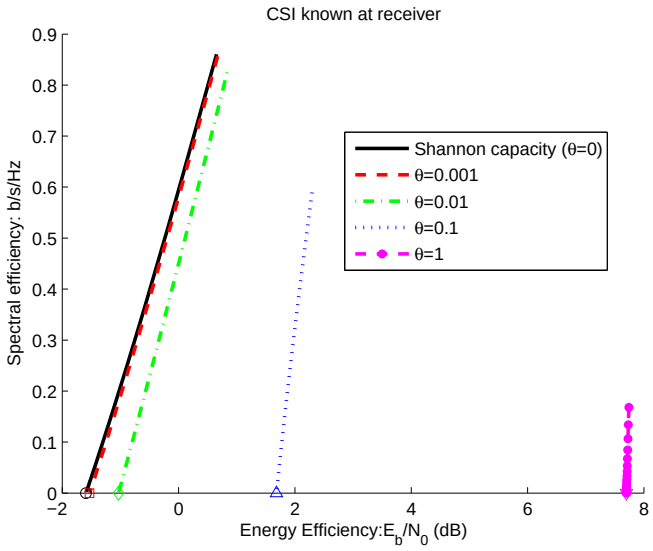

Figure 2.4: Spectral efficiency vs. E_b/N_0 in the Rayleigh fading channel with fixed $\bar{P}$; CSI known at the receiver only.

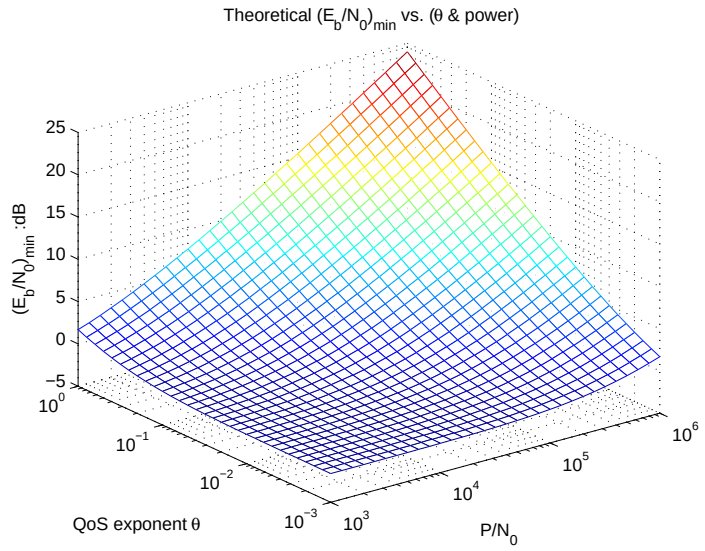

Figure 2.5: $\frac{E_b}{N_0}_{\min}$ vs. θ and $\bar{P}/N_0$ in the Rayleigh fading channel; CSI known at the receiver only.

designing wireless networks. For instance, for some $\bar{P}/N_0$ required to achieve some specific transmission rate, we can find the most stringent QoS guarantee possible while attaining a certain efficiency in the usage of energy, or if a QoS requirement θ is specified, we can find the minimum power $\bar{P}$ to achieve a specific bit energy.

2.3.2 CSI at both the Transmitter and Receiver

To analyze $\frac{E_b}{N_0}_{\min}$ in this case, we initially obtain the following result and identify the limiting value of the threshold α as the bandwidth increases to infinity.

Theorem 4 *In wideband regime, the threshold α in the optimal power adaptation scheme (2.4) satisfies*

$$\lim_{\zeta \to 0} \alpha(\zeta) = \alpha^* \tag{2.31}$$

where α^* is the solution to

$$\mathbb{E}\left\{\left[\log_e\left(\frac{z}{\alpha^*}\right)\frac{1}{z}\right]\tau(\alpha^*)\right\} = \frac{\theta T \bar{P}}{N_0 \log_e 2}. \tag{2.32}$$

Moreover, for $\theta > 0$, $\alpha^*_{opt} < \infty$.

Proof: Recall from (2.5) that the optimal power adaptation rule should satisfy the average power constraint:

$$\text{SNR} = \frac{\bar{P}\zeta}{N_0} = \mathbb{E}\left\{\left(\frac{1}{\alpha^{\frac{1}{\beta+1}} z^{\frac{\beta}{\beta+1}}} - \frac{1}{z}\right)\tau(\alpha)\right\} = \mathbb{E}\left\{\left[\left(\left(\frac{z}{\alpha}\right)^{\frac{1}{\beta+1}} - 1\right)\frac{1}{z}\right]\tau(\alpha)\right\} \tag{2.33}$$

where $\beta = \frac{\theta T B}{\log_e 2} = \frac{\theta T}{\zeta \log_e 2}$. For the case in which $\theta = 0$, if we let $\zeta \to 0$, we obtain from (2.33) that

$$0 = \mathbb{E}\left\{\left[\left(\frac{z}{\alpha^*_{\text{opt}}} - 1\right)\frac{1}{z}\right]\tau(\alpha^*_{\text{opt}})\right\} \tag{2.34}$$

where $\alpha^*_{\text{opt}} = \lim_{\zeta\to 0}\alpha(\zeta)$. Using the fact that $\log_e x \le x-1$ for $x \ge 1$, we have $\log_e\left(\frac{z}{\alpha^*_{\text{opt}}}\right) \le \frac{z}{\alpha^*_{\text{opt}}} - 1$ for $z \ge \alpha^*_{\text{opt}}$ which implies that

$$0 \le \mathbb{E}\left\{\left[\log_e\left(\frac{z}{\alpha^*}\right)\frac{1}{z}\right]\tau(\alpha^*)\right\} \le \mathbb{E}\left\{\left[\left(\frac{z}{\alpha^*_{\text{opt}}}-1\right)\frac{1}{z}\right]\tau(\alpha^*_{\text{opt}})\right\} = 0$$
$$\implies \mathbb{E}\left\{\left[\log_e\left(\frac{z}{\alpha^*}\right)\frac{1}{z}\right]\tau(\alpha^*)\right\} = 0$$

proving (2.32) for the case of $\theta = 0$.

In the following, we assume $\theta > 0$. We first define $g(\zeta) = \left(\frac{z}{\alpha}\right)^{\frac{1}{\beta+1}} = \left(\frac{z}{\alpha}\right)^{\frac{\zeta\log_e 2}{\zeta\log_e 2+\theta T}}$ and take the logarithm of both sides to obtain

$$\log_e g(\zeta) = \frac{\zeta\log_e 2}{\zeta\log_e 2+\theta T}\log_e\frac{z}{\alpha}. \tag{2.35}$$

Differentiation over both sides leads to

$$\frac{\dot{g}(\zeta)}{g(\zeta)} = \frac{\theta T\log_e 2}{(\zeta\log_e 2+\theta T)^2}\log_e\frac{z}{\alpha} - \frac{\zeta\log_e 2}{\zeta\log_e 2+\theta T}\frac{\dot{\alpha}}{\alpha} \tag{2.36}$$

where $\dot{g}$ and $\dot{\alpha}$ denote the first derivatives g and α, respectively, with respect to ζ. Noting that $g(0) = 1$, we can see from (2.36) that as $\zeta \to 0$, we have

$$\dot{g}(0) = \frac{\log_e 2}{\theta T}\log_e\frac{z}{\alpha^*_{\text{opt}}} \tag{2.37}$$

where $\alpha^*_{\text{opt}} = \lim_{\zeta\to 0}\alpha(\zeta)$. For small values of ζ, the function g admits the following Taylor series:

$$g(\zeta) = \left(\frac{z}{\alpha}\right)^{\frac{1}{\beta+1}} = g(0) + \dot{g}(0)\zeta + o(\zeta) = 1 + \dot{g}(0)\zeta + o(\zeta). \tag{2.38}$$

Therefore, we have

$$\left(\frac{z}{\alpha}\right)^{\frac{1}{\beta+1}} - 1 = \frac{\log_e 2}{\theta T} \log_e \left(\frac{z}{\alpha^*_{\text{opt}}}\right) \zeta + o(\zeta). \tag{2.39}$$

Then, from (2.33), we can write

$$\text{SNR} = \mathbb{E}\left\{\left[\left(\frac{\log_e 2}{\theta T} \log_e \left(\frac{z}{\alpha}\right) \zeta + o(\zeta)\right) \frac{1}{z}\right] \tau(\alpha)\right\}. \tag{2.40}$$

If we divide both sides of (2.40) by $\text{SNR} = \frac{\bar{P}\zeta}{N_0}$ and let $\zeta \to 0$, we obtain

$$\lim_{\zeta\to 0} \frac{\text{SNR}}{\text{SNR}} = \lim_{\zeta\to 0} \frac{\text{SNR}}{\frac{\bar{P}\zeta}{N_0}} = 1 = \frac{N_0 \log_e 2}{\theta T \bar{P}} \mathbb{E}\left\{\left[\log_e \left(\frac{z}{\alpha^*_{\text{opt}}}\right) \frac{1}{z}\right] \tau(\alpha^*_{\text{opt}})\right\} \tag{2.41}$$

from which we conclude that $\mathbb{E}\left\{\left[\log_e \left(\frac{z}{\alpha^*}\right) \frac{1}{z}\right] \tau(\alpha^*)\right\} = \frac{\theta T \bar{P}}{N_0 \log_e 2}$, proving (2.32) for $\theta > 0$.

Using the fact that $\log_e \left(\frac{z}{\alpha}\right) < \frac{z}{\alpha}$ for $z \geq 0$, we can write

$$0 \leq \mathbb{E}\left\{\left[\log_e \left(\frac{z}{\alpha}\right) \frac{1}{z}\right] \tau(\alpha)\right\} \leq \mathbb{E}\left\{\frac{1}{\alpha} \tau(\alpha)\right\} \leq \frac{1}{\alpha}. \tag{2.42}$$

Assume now that $\lim_{\zeta\to 0} \alpha(\zeta) = \alpha^*_{\text{opt}} = \infty$. Then, the rightmost side of (2.42) becomes zero in the limit as $\zeta \to 0$ which implies that $\mathbb{E}\left\{\left[\log_e \left(\frac{z}{\alpha^*_{\text{opt}}}\right) \frac{1}{z}\right] \tau(\alpha^*_{\text{opt}})\right\} = 0$. From (2.32), this is clearly not possible for $\theta > 0$. Hence, we have proved that $\alpha^*_{\text{opt}} < \infty$ when $\theta > 0$. □

Remark: As noted before, wideband and low-power regimes are equivalent when $\theta = 0$. Hence, as in the proof of Theorem 2, we can easily see in the wideband regime that the threshold α approaches the maximum fading value $z_{\max}$ as $\zeta \to 0$ when $\theta = 0$. Hence, for fading distributions with unbounded support, $\alpha \to \infty$ with vanishing ζ. The threshold being very large means that the transmitter waits sufficiently long until

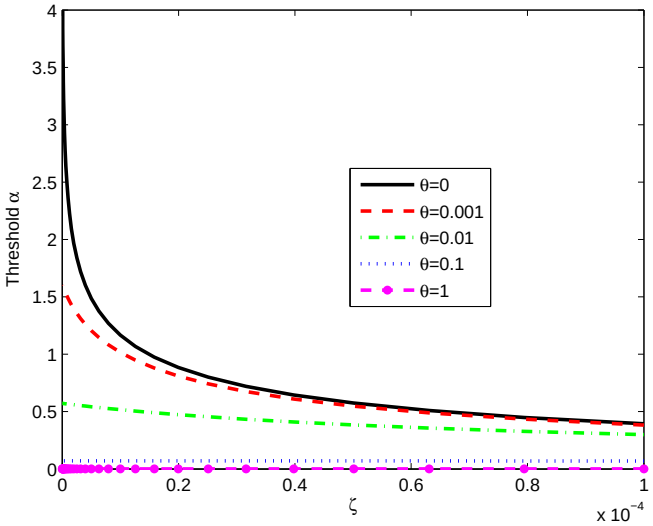

Figure 2.6: Threshold of channel gain α vs. ζ in the Rayleigh fading channel; CSI known at the transmitter and receiver.

the fading assumes very large values and becomes favorable. That is how arbitrarily small bit energy values can be attained. However, in the presence of QoS constraints, arbitrarily long waiting times will not be permitted. As a result, α approaches a finite value (i.e., $\alpha^*_{\text{opt}} < \infty$) as $\zeta \to 0$ when $\theta > 0$. Moreover, from (2.32), we can immediately note that as θ increases, α^*_{opt} has to decrease. This fact can also be observed in Fig. 2.6 in which α vs. ζ is plotted in the Rayleigh fading channel. Consequently, arbitrarily small bit energy values will no longer be possible when $\theta > 0$ as will be shown in Theorem 5.

The spectral efficiency with optimal power adaptation is now given by

$$\mathsf{C}_E(\zeta) = -\frac{\zeta}{\theta T}\log_e\left(F(\alpha) + \mathbb{E}\left\{\left(\frac{z}{\alpha}\right)^{-\frac{\theta T}{\theta T+\zeta\log_e 2}}\tau(\alpha)\right\}\right) \tag{2.43}$$

where again $F(\alpha) = \mathbb{E}\{1\{z < \alpha\}\}$ and $\tau(\alpha) = 1\{\tau \geq \alpha\}$.

Theorem 5 *When both the receiver and transmitter have CSI, the minimum bit en-*

ergy and wideband slope in the wideband regime are given by

$$\frac{E_b}{N_0}_{\min} = -\frac{\frac{\theta T \bar{P}}{N_0}}{\log_e \xi} \text{ and } \mathcal{S}_0 = \frac{\xi(\log_e \xi)^2 \log_e 2}{\theta T(\frac{\bar{P}\alpha^*_{opt}}{N_0} + \dot{\alpha}(0)\mathbb{E}\{\frac{1}{z}\tau(\alpha^*_{opt})\})} \tag{2.44}$$

where $\xi = F(\alpha^*) + \mathbb{E}\{\frac{\alpha^*}{z}\tau(\alpha^*)\}$, *and* $\dot{\alpha}(0)$ *is the derivative of* α *with respect to* ζ, *evaluated at* $\zeta = 0$.

Proof: Substituting (2.43) into (2.22) leads to

$$\begin{aligned}\frac{E_b}{N_0}_{\min} &= \lim_{\zeta \to 0} \frac{\bar{P}\zeta/N_0}{-\frac{\zeta}{\theta T}\log_e\left(F(\alpha) + \mathbb{E}\left\{\left(\frac{z}{\alpha}\right)^{-\frac{\theta T}{\theta T + \zeta \log_e 2}}\tau(\alpha)\right\}\right)} \\ &= -\frac{\theta T \bar{P}}{N_0 \log_e\left(F(\alpha^*) + \mathbb{E}\left\{\frac{\alpha^*}{z}\tau(\alpha^*)\right\}\right)}.\end{aligned} \tag{2.45}$$

After denoting $\xi = F(\alpha^*) + \mathbb{E}\{\frac{\alpha^*}{z}\tau(\alpha^*)\}$, we obtain the expression for minimum bit energy in (2.44).

Meanwhile, $\mathsf{C}_E(\zeta)$ has the following Taylor series expansion up to second order:

$$\mathsf{C}_E(\zeta) = \dot{\mathsf{C}}_E(0)\zeta + \frac{1}{2}\ddot{\mathsf{C}}_E(0)\zeta^2 + o(\zeta^2). \tag{2.46}$$

Therefore, the second derivative of C_E with respect to ζ at $\zeta = 0$ can be computed from

$$\ddot{\mathsf{C}}_E(0) = 2\lim_{\zeta \to 0} \frac{\mathsf{C}_E(\zeta) - \dot{\mathsf{C}}_E(0)\zeta}{\zeta^2}. \tag{2.47}$$

From the derivation of (2.45) and (2.22), we know that

$$\dot{\mathsf{C}}_E(0) = -\frac{1}{\theta T}\log_e\left(F(\alpha^*) + \mathbb{E}\left\{\frac{\alpha^*}{z}\tau(\alpha^*)\right\}\right). \tag{2.48}$$

Then,

$$\ddot{\mathsf{C}}_E(0) = 2\lim_{\zeta\to 0}\left(\frac{-\frac{\zeta}{\theta T}\log_e\left(F(\alpha)+\mathbb{E}\left\{\left(\frac{z}{\alpha}\right)^{-\frac{\theta T}{\theta T+\zeta\log_e 2}}\tau(\alpha)\right\}\right)}{\zeta^2}\right.$$
$$\left.+\frac{\frac{\zeta}{\theta T}\log_e\left(F(\alpha^*)+\mathbb{E}\left\{\frac{\alpha^*_{\text{opt}}}{z}\tau(\alpha^*_{\text{opt}})\right\}\right)}{\zeta^2}\right)$$
$$= -\frac{2}{\theta T}\lim_{\zeta\to 0}\frac{\log_e\frac{F(\alpha)+\mathbb{E}\left\{\left(\frac{\alpha}{z}\right)^{\frac{\theta T}{\theta T+\zeta\log_e 2}}\tau(\alpha)\right\}}{F(\alpha^*_{\text{opt}})+\mathbb{E}\left\{\frac{\alpha^*_{\text{opt}}}{z}\tau(\alpha^*_{\text{opt}})\right\}}}{\zeta} \tag{2.49}$$
$$= -\frac{2}{\theta T}\lim_{\zeta\to 0}\frac{\mathbb{E}\left\{\left(\frac{\alpha}{z}\right)^{\frac{\theta T}{\theta T+\zeta\log_e 2}}\left(-\frac{\theta T\log_e 2}{(\theta T+\zeta\log_e 2)^2}\log_e\left(\frac{\alpha}{z}\right)+\frac{\theta T\dot{\alpha}}{(\theta T+\zeta\log_e 2)\alpha}\right)\tau(\alpha)\right\}}{F(\alpha^*_{\text{opt}})+\mathbb{E}\left\{\frac{\alpha^*_{\text{opt}}}{z}\tau(\alpha^*_{\text{opt}})\right\}} \tag{2.50}$$
$$= -\frac{2\log_e 2}{(\theta T)^2}\frac{\mathbb{E}\left\{\frac{\alpha^*_{\text{opt}}}{z}\log_e\left(\frac{z}{\alpha^*_{\text{opt}}}\right)\tau(\alpha^*_{\text{opt}})\right\}+\frac{\theta T\dot{\alpha}(0)}{\log_e 2}\mathbb{E}\{\frac{1}{z}\tau(\alpha^*_{\text{opt}})\}}{F(\alpha^*_{\text{opt}})+\mathbb{E}\left\{\frac{\alpha^*_{\text{opt}}}{z}\tau(\alpha^*_{\text{opt}})\right\}}, \tag{2.51}$$

where $\dot{\alpha}$ is the derivative of α with respect to ζ. Above, (2.50) is obtained by using L'Hospital's Rule. Evaluating (1.7) with (2.48) and (2.51), and combining with the result in (2.32), we obtain the expression for $\mathcal{S}_0$ in (2.44). □

It is interesting to note that the minimum bit energy is strictly greater than zero for $\theta > 0$. Hence, we see a stark difference between the wideband regime and low-power regime in which the minimum bit energy is zero for fading distributions with unbounded support. Fig. 2.7 plots the spectral efficiency curves in the Rayleigh fading channel and is in perfect agreement with the theoretical results. Obviously, the plots are drastically different from those in the low-power regime (Fig. 2.3) where all curves approach $-\infty$ as the spectral efficiency decreases. In Fig. 2.7, the minimum bit energy is finite for the cases in which $\theta > 0$. The wideband slope

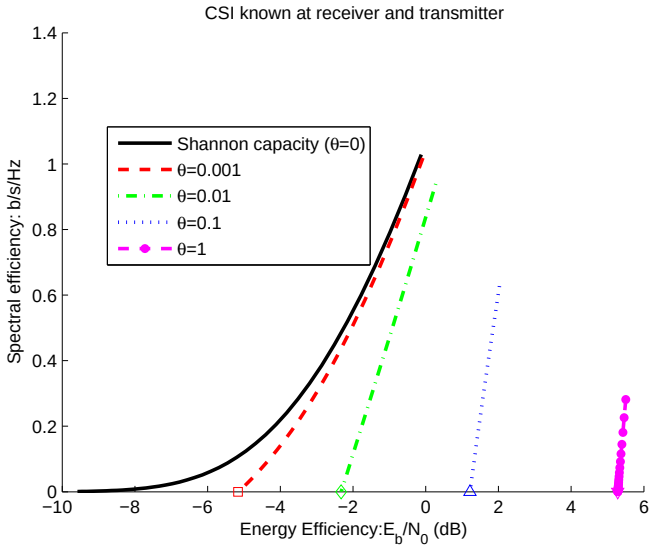

Figure 2.7: Spectral efficiency vs. E_b/N_0 in the Rayleigh fading channel with fixed $\frac{\bar{P}}{N_0} = 10^4$; CSI known at the transmitter and receiver.

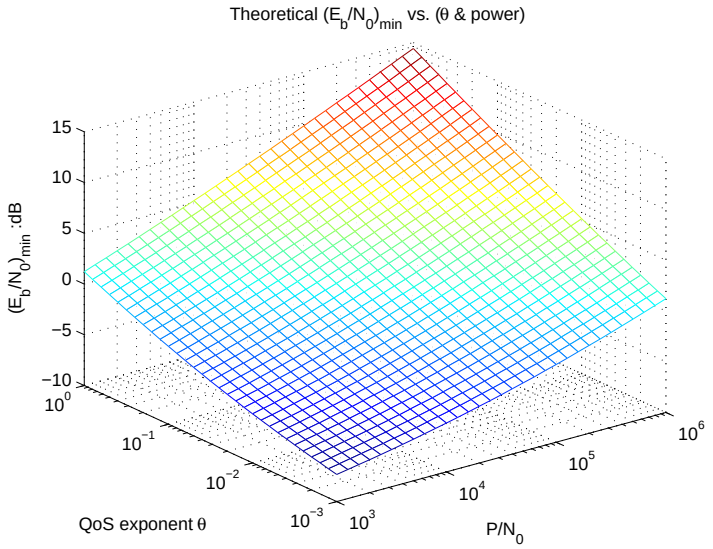

Figure 2.8: $\frac{E_b}{N_0\,\min}$ vs. θ and $\bar{P}/N_0$ in the Rayleigh fading channel; CSI known at the transmitter and receiver.

values are computed to be equal to $\mathcal{S}_0 = \{0.3081, 1.0455, 2.5758, 4.1869\}$. Fig. 2.8 plots the $\frac{E_b}{N_0}_{\min}$ as a function of θ and $\bar{P}/N_0$. Generally speaking, due to power and rate adaptation, $\frac{E_b}{N_0}_{\min}$ in this case is smaller compared to that in the case in which only the receiver has CSI. This can be observed in Fig. 2.9 where the minimum bit energies are compared. From Fig. 2.9, we note that the presence of CSI at the transmitter is especially beneficial for very small and also large values of θ. While the bit energy in the CSIR case approaches -1.59 dB with vanishing θ, it decreases to $-\infty$ dB when the transmitter also knows the channel. On the other hand, when $\theta \approx 10^{-3}$, we interestingly observe that there is not much to be gained in terms of the minimum bit energy by having CSI at the transmitter. More specifically, power adaptation in this case does not result in significant improvements in the asymptotic value of the (unnormalized) effective capacity C_E achieved as $B \to \infty$. We note from (2.23) and (2.44) that the minimum bit energy expressions have a common expression in the numerator while the expressions in the denominator are proportional to the asymptotic value of C_E. When $\bar{P}/N_0 = 10^6$, $T = 2$ms and $\theta = 10^{-3}$, we can easily compute for the Rayleigh channel that $-\log_e \mathbb{E}\{e^{-\frac{\theta T \bar{P}}{N_0 \log_e 2} z}\} = 1.357$. In the case of CSIT, we have $\alpha^*_{\text{opt}} = 0.0716$ and $-\log_e \xi = 1.507$, verifying our conclusion above. For $\theta > 10^{-3}$, we again start having improvements with the presence of CSIT.

Throughout the chapter, numerical results are provided for the Rayleigh fading channel. However, note that the theoretical results hold for general stationary and ergodic fading processes. Hence, other fading distributions can easily be accommodated as well. In Fig. 2.10, we plot the spectral efficiency vs. bit energy curves for the Nakagami-m fading channel with $m = 2$.

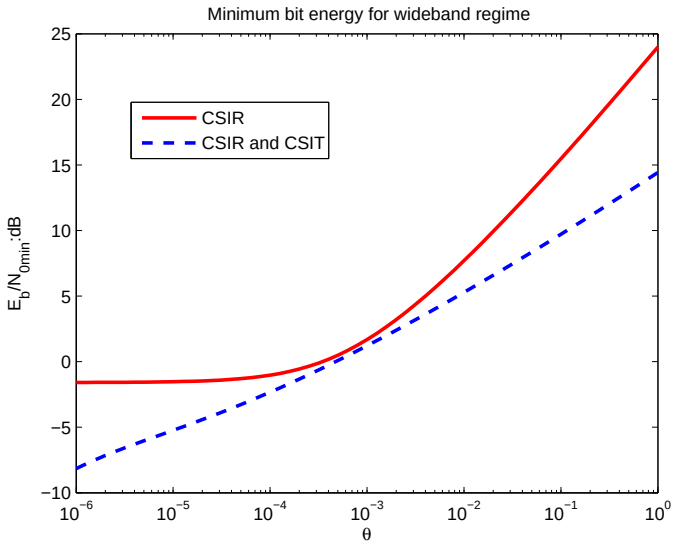

Figure 2.9: $\frac{E_b}{N_0}_{\min}$ vs. θ in the Rayleigh fading channel. $\frac{\bar{P}}{N_0} = 10^6$.

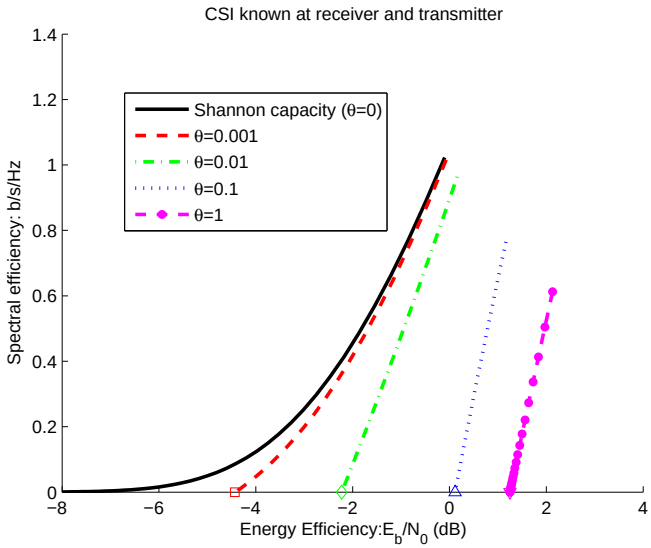

Figure 2.10: Spectral efficiency vs. E_b/N_0 in the Nakagami-m fading channel with $m = 2$; $\frac{\bar{P}}{N_0} = 10^4$;CSI known at the transmitter and receiver.

2.4 Conclusion

In this chapter, we have analyzed the energy efficiency in fading channel under QoS constraints by considering the effective capacity as a measure of the maximum throughput under certain statistical QoS constraints, and analyzing the bit energy levels. Our analysis has provided a characterization of the energy-bandwidth-delay tradeoff. In particular, we have investigated the spectral efficiency vs. bit energy tradeoff in the low-power and wideband regimes under QoS constraints. We have elaborated the analysis under two scenarios: perfect CSI available at the receiver and perfect CSI available at both the receiver and transmitter. We have obtained expressions for the minimum bit energy and wideband slope. Through this analysis, we have quantified the increased energy requirements in the presence of delay-QoS constraints. While the bit energy levels in the low-power regime can approach those that can be attained in the absence of QoS constraints, we have shown that strictly higher bit energy values are needed in the wideband regime. We have provided numerical results by considering the Rayleigh and Nakagami fading channels and verified the theoretical conclusions.

Chapter 3

Energy Efficiency for Fixed Rate Transmissions

In this chapter, we assume that the transmitter does not have channel knowledge and it sends the information at a fixed rate and fixed power. For this fixed rate scenario, we adopt a two state (ON-OFF) transmission model, where information is transmitted reliably at a fixed rate in the ON state while no transmission occurs in the OFF state. We investigate the wideband regime in sparse multipath fading, in which the number of subchannels remains bounded as bandwidth increases, and also in rich multipath fading, in which the number of non-interacting subchannels increases without bound with increasing bandwidth. The minimum bit energy and wideband slope expressions are found for the wideband regime with multipath sparsity. The expressions for bit energy required at zero spectral efficiency and wideband slope are quantified for the low-power regime, which is also equivalent to the wideband regime with rich multipath fading. It is shown for a certain class of fading distributions that the bit energy required at zero spectral efficiency is indeed the minimum bit energy for reliable communications.

3.1 System Model

We consider a point-to-point wireless link in which there is one source and one destination. The system model is depicted in Figure 3.1. It is assumed that the source generates data sequences which are divided into frames of duration T. These data frames are initially stored in the buffer before they are transmitted over the wireless channel. The discrete-time channel input-output relation in the i^{th} symbol duration is given by

$$y[i] = h[i]x[i] + n[i] \quad i = 1, 2, \ldots. \tag{3.1}$$

where $x[i]$ and $y[i]$ denote the complex-valued channel input and output, respectively. We assume that the bandwidth available in the system is B and the channel input is subject to the following average energy constraint: $\mathbb{E}\{|x[i]|^2\} \leq \bar{P}/B$ for all i. Since the bandwidth is B, symbol rate is assumed to be B complex symbols per second, indicating that the average power of the system is constrained by $\bar{P}$. Above in (3.1), $n[i]$ is a zero-mean, circularly symmetric, complex Gaussian random variable with variance $\mathbb{E}\{|n[i]|^2\} = N_0$. The additive Gaussian noise samples $\{n[i]\}$ are assumed to form an independent and identically distributed (i.i.d.) sequence. Finally, $h[i]$ denotes the channel fading coefficient, and $\{h[i]\}$ is a stationary and ergodic discrete-time process. We denote the magnitude-square of the fading coefficients by $z[i] = |h[i]|^2$.

In this chapter, we consider the scenario in which the receiver has perfect channel side information and hence perfectly knows the instantaneous values of $\{h[i]\}$ while the transmitter has no such knowledge. In this case, the instantaneous channel capacity with channel gain $z[i]$=$|h[i]|^2$ is

$$C[i] = B \log_2(1 + \text{SNR} z[i]) \text{ bits/s} \tag{3.2}$$

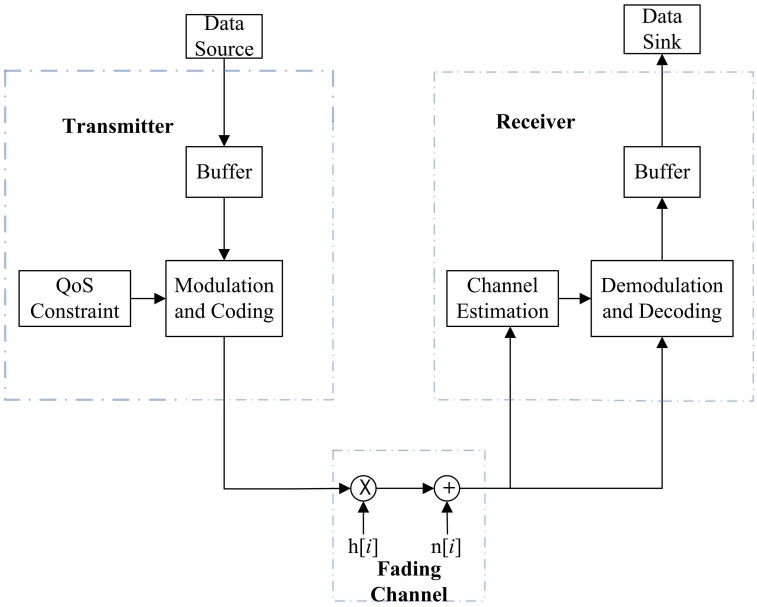

Figure 3.1: The general system model.

where SNR $= \bar{P}/(N_0 B)$ is the average transmitted signal-to-noise ratio. Since the transmitter is unaware of the channel conditions, information is transmitted at a fixed rate of r bits/s. When $r < C$, the channel is considered to be in the ON state and reliable communication is achieved at this rate. From information-theoretic arguments, this is possible if strong codes with large blocklength is employed in the system. Since there are TB symbols in each block, we assume TB is large enough to establish reliable communication. If, on the other hand, $r \geq C$, outage occurs. In this case, channel is in the OFF state and reliable communication at the rate of r bits/s cannot be attained. Hence, effective data rate is zero and information has to be resent. We assume that a simple ARQ mechanism is incorporated in the communication protocol to acknowledge the reception of data and to ensure that the erroneous data is retransmitted [38].

Fig. 3.2 depicts the two-state transmission model together with the transition probabilities. In this chapter, we assume that the channel fading coefficients stay

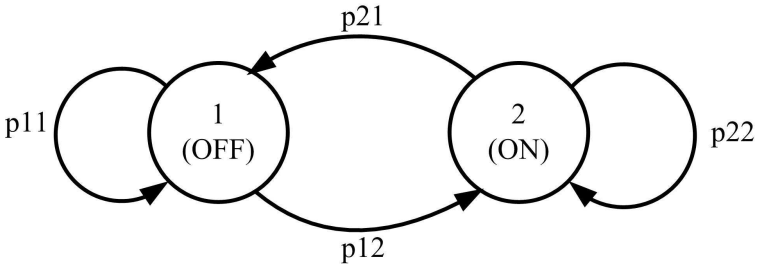

Figure 3.2: ON-OFF state transition model.

constant over the frame duration T. Hence, the state transitions occur at every T seconds. Now, the probability of staying in the ON state, p_{22}, is defined as follows[1]:

$$\begin{aligned} p_{22} &= P\{r < C[i+TB] \,\big|\, r < C[i]\} \\ &= P\{z[i+TB] > \alpha \,\big|\, z[i] > \alpha\} \end{aligned} \tag{3.3}$$

where

$$\alpha = \frac{2^{\frac{r}{B}} - 1}{\text{SNR}}. \tag{3.4}$$

Note that p_{22} depends on the joint distribution of $(z[i+TB], z[i])$. For the Rayleigh fading channel, the joint density function of the fading amplitudes can be obtained in closed-form [56]. In this chapter, with the goal of simplifying the analysis and providing results for arbitrary fading distributions, we assume that fading realizations are independent for each frame[2]. Hence, we basically consider a block-fading channel model. Note that in block-fading channels, the duration T over which the fading coefficients stay constant can be varied to model fast or slow fading scenarios.

Under the block fading assumption, we now have $p_{22} = P\{z[i+TB] > \alpha\} =$

[1]The formulation in (3.3) assumes as before that the symbol rate is B symbols/s and hence we have TB symbols in a duration of T seconds.

[2]This assumption also enables us to compare the results of this chapter with those in Chapter 2 in which variable-rate/variable-power and variable-rate/fixed-power transmission schemes are studied for block fading channels.

$$\frac{\Lambda(\theta)}{\theta} = \frac{1}{\theta}\log_e\left(\frac{1}{2}\left(p_{11} + p_{22}e^{\theta Tr} + \sqrt{(p_{11} + p_{22}e^{\theta Tr})^2 + 4(p_{11} + p_{22} - 1)e^{\theta Tr}}\right)\right). \quad (3.6)$$

$P\{z > \alpha\}$. Similarly, the other transition probabilities become

$$\begin{aligned} p_{11} = p_{21} = P\{z \leq \alpha\} &= \int_0^{\alpha} p_z(z)dz \quad \text{and} \\ p_{22} = p_{12} = P\{z > \alpha\} &= \int_{\alpha}^{\infty} p_z(z)dz \end{aligned} \quad (3.5)$$

where p_z is the probability density function of z. Throughout the chapter, we assume that both $p_z(z)$ and the cumulative distribution function $P\{z \leq \alpha\}$ are differentiable. We finally note that rT bits are successfully transmitted and received in the ON state, while the effective transmission rate in the OFF state is zero.

3.2 Preliminary

For this chapter, with (1.1) in mind, we know that $s[i] = rT$ or 0 depending on the channel state being ON or OFF, respectively. In [10] and [12, Section 7.2, Example 7.2.7], it is shown that for such an ON-OFF model, the log-moment generating function normalized by θ, i.e., $\frac{\Lambda(\theta)}{\theta}$, is given by (3.6) at the top of the page.

Using the formulation in (3.6) and noting that $p_{11} + p_{22} = 1$ in our model, we express the effective capacity normalized by the frame duration T and bandwidth B, or equivalently spectral efficiency in bits/s/Hz, for a given statistical QoS constraint

θ, as [12]

$$
\begin{aligned}
&\mathsf{R}_E(\mathrm{SNR},\theta) \\
&= \frac{1}{TB}\max_{r\geq 0}\Big\{-\frac{\Lambda(-\theta)}{\theta}\Big\} && (3.7)\\
&= \max_{r\geq 0}\Big\{-\frac{1}{\theta TB}\log_e\Big(p_{11}+p_{22}e^{-\theta Tr}\Big)\Big\} && (3.8)\\
&= \max_{r\geq 0}\Big\{-\frac{1}{\theta TB}\log_e\Big(1-P\{z>\alpha\}(1-e^{-\theta Tr})\Big)\Big\} && (3.9)\\
&= -\frac{1}{\theta TB}\log_e\Big(1-P\{z>\alpha_{\mathrm{opt}}\}\Big(1-e^{-\theta Tr_{\mathrm{opt}}}\Big)\Big)\ \text{bits/s/Hz} && (3.10)
\end{aligned}
$$

where r_{opt} is the maximum fixed transmission rate that solves (3.9) and $\alpha_{\mathrm{opt}} = (2^{\frac{r_{\mathrm{opt}}}{B}}-1)/\mathrm{SNR}$. Note that both α_{opt} and r_{opt} are functions of SNR and θ.

The normalized effective capacity, R_E, provides the maximum throughput under statistical QoS constraints in the fixed-rate transmission model. It can be easily shown that

$$\lim_{\theta\to 0}\mathsf{R}_E(\mathrm{SNR},\theta) = \max_{r\geq 0}\ \frac{r}{B}\,P\{z>\alpha\}. \tag{3.11}$$

Hence, the QoS requirements relax, the maximum constant arrival rate approaches the average transmission rate. On the other hand, for $\theta > 0$, $\mathsf{R}_E < \frac{1}{B}\max_{r\geq 0} rP\{z>\alpha\}$ in order to avoid violations of QoS constraints.

In this chapter, we focus on the energy efficiency of wireless transmissions under the aforementioned statistical QoS limitations. Since energy efficient operation generally requires operation at low-SNR levels, our analysis throughout the chapter is carried out in the low-SNR regime. In this regime, the tradeoff between the normalized effective capacity (i.e, spectral efficiency) R_E and bit energy $\frac{E_b}{N_0} = \frac{\mathrm{SNR}}{\mathsf{R}_E(\mathrm{SNR})}$ is a key tradeoff in understanding the energy efficiency, and is characterized by the bit

energy at zero spectral efficiency and wideband slope provided, respectively, by

$$\begin{aligned} \left.\frac{E_b}{N_0}\right|_{\mathsf{R}_E=0} &= \lim_{\text{SNR}\to 0} \frac{\text{SNR}}{\mathsf{R}_E(\text{SNR})} = \frac{1}{\dot{\mathsf{R}}_E(0)}, \quad \text{and} \\ \mathcal{S}_0 &= -\frac{2(\dot{\mathsf{R}}_E(0))^2}{\ddot{\mathsf{R}}_E(0)} \log_e 2 \end{aligned} \tag{3.12}$$

where $\dot{\mathsf{R}}_E(0)$ and $\ddot{\mathsf{R}}_E(0)$ are the first and second derivatives with respect to SNR, respectively, of the function $\mathsf{R}_E(\text{SNR})$ at zero SNR [21]. $\left.\frac{E_b}{N_0}\right|_{\mathsf{R}_E=0}$ and $\mathcal{S}_0$ provide a linear approximation of the spectral efficiency curve at low spectral efficiencies, i.e.,

$$\mathsf{R}_E\left(\frac{E_b}{N_0}\right) = \frac{\mathcal{S}_0}{10\log_{10} 2}\left(\left.\frac{E_b}{N_0}\right|_{dB} - \left.\frac{E_b}{N_0}\right|_{\mathsf{R}_E=0,dB}\right) + \epsilon \tag{3.13}$$

where $\epsilon = o\left(\frac{E_b}{N_0} - \left.\frac{E_b}{N_0}\right|_{\mathsf{R}_E=0}\right)$.

Above, $\left.\frac{E_b}{N_0}\right|_{dB} = 10\log_{10}\frac{E_b}{N_0}$. When the spectral efficiency R_E is a non-decreasing concave function of SNR, the bit energy $\frac{E_b}{N_0}$ diminishes with decreasing spectral efficiency. Hence, in this case, the bit energy required at zero spectral efficiency is indeed the minimum one, i.e., $\left.\frac{E_b}{N_0}\right|_{\mathsf{R}_E=0} = \frac{E_b}{N_0\min}$.

3.3 Energy Efficiency in the Wideband Regime

In this section, we consider the wideband regime in which the bandwidth is large. We assume that the average power $\bar{P}$ is kept constant. Note that as the bandwidth B increases, $\text{SNR} = \frac{\bar{P}}{N_0 B}$ approaches zero and we operate in the low-SNR regime.

Following the approach generally employed in information-theoretic analyses, we assume that the wideband channel is decomposed into N parallel subchannels. We further assume that each subchannel has a bandwidth that is equal to the coherence

bandwidth, B_c. Therefore, independent flat-fading is experienced in each subchannel, and we have $B = NB_c$. Similar to (3.1), the input-output relation in the k^{th} subchannel can be written as

$$y_k[i] = h_k[i]x_k[i] + n_k[i] \quad i = 1, 2, \ldots \text{ and } \ k = 1, 2, \ldots, N. \tag{3.14}$$

The fading coefficients $\{h_k\}_{k=1}^n$ in different subchannels are assumed to be independent. The signal-to-noise ratio in the k^{th} subchannel is $\text{SNR}_k = \frac{\bar{P}_k}{N_0 B_c}$ where $\bar{P}_k$ denotes the power allocated to the k^{th} subchannel and we have $\sum_{k=1}^N \bar{P}_k = \bar{P}$. Over each subchannel, the same transmission strategy as described in Section 3.1 is employed. Therefore, the transmitter, not knowing the fading coefficients of the subchannels, sends the data over each subchannel at the fixed rate of r. If $r < B_c \log(1+\text{SNR}_k z_k[i])$ where $z_k = |h_k|^2$, then transmission over the k^{th} subchannel is successful. Otherwise, retransmission is required. Hence, we have an ON-OFF state model for each subchannel. On the other hand, for the transmission over N subchannels, we have a state-transition model with $N+1$ states because we have overall the following $N+1$ possible total transmission rates: $\{0, rT, 2rT, \ldots, NrT\}$. For instance, if all N subchannels are in the OFF state simultaneously, the total rate is zero. If j out of N subchannels are in the ON state, then the rate is jrT.

Now, assume that the states are enumerated in the increasing of order of the total transmission rates supported by them. Hence, in state $j \in \{1, \ldots, N+1\}$, the transmission rate is $(j-1)rT$. The transition probability from state $i \in \{1, \ldots, N+1\}$ to state $j \in \{1, \ldots, N+1\}$ is given by (3.15) on the next page where $\mathcal{I}_{j-1}$ denotes a subset of the index set $\{1, \ldots, N\}$ with $j-1$ elements. The summation in (3.15) is over all such subsets. Moreover, in (3.15), $\mathcal{I}_{j-1}^c$ denotes the complement of the set $\mathcal{I}_{j-1}$, and $\alpha_k = \frac{2^{\frac{r}{B_c}}-1}{\text{SNR}_k}$. Note in the above formulation that the transition probabilities,

$$p_{ij} = p_j = P\{(j-1) \text{ subchannels out of } N \text{ subchannels are in the ON state}\}$$
$$= \sum_{\mathcal{I}_{j-1} \subset \{1,\ldots,N\}} \left(\prod_{k \in \mathcal{I}_{j-1}} P\{z_k > \alpha_k\} \prod_{k \in \mathcal{I}_{j-1}^c} (1 - P\{z_k > \alpha_k\}) \right) \tag{3.15}$$

---

p_{ij}, do not depend on the initial state i due to the block-fading assumption. If, in addition to being independent, the fading coefficients and hence $\{z_k\}_{k=1}^N$ in different subchannels are identically distributed, then p_{ij} in (3.15) simplifies and becomes a binomial probability:

$$p_{ij} = p_j = \begin{pmatrix} N \\ j-1 \end{pmatrix} (P\{z > \alpha\})^{j-1} (1 - P\{z > \alpha\})^{N-j+1}. \tag{3.16}$$

Note that if the fading coefficients are i.i.d., the total power should be uniformly distributed over the subchannels. Hence, in this case, we have $\bar{P}_k = \frac{\bar{P}}{N}$ and therefore $\text{SNR}_k = \frac{\bar{P}_k}{N_0 B_c} = \frac{\bar{P}/N}{N_0 B/N} = \frac{\bar{P}}{N_0 B} = \text{SNR}$ which is equal to the original SNR definition used in (3.2). Now, we have the same $\alpha = \frac{2^{\frac{r}{B_c}}-1}{\text{SNR}}$ for each subchannel.

The effective capacity of this wideband channel model is given by the following result.

Theorem 6 *For the wideband channel with N parallel noninteracting subchannels each with bandwidth B_c and independent flat fading, the normalized effective capacity in bits/s/Hz is*

$$R_E(SNR, \theta) = \max_{\substack{r \geq 0 \\ \bar{P}_k \geq 0 \text{ s.t. } \sum \bar{P}_k \leq \bar{P}}} \left\{ -\frac{1}{\theta T B} \log_e \left(\sum_{j=1}^{N+1} p_j \, e^{-\theta(j-1)rT} \right) \right\} \tag{3.17}$$

where p_j is given in (3.15). If $\{z_k\}_{k=1}^N$ are identically distributed, then the normalized

effective capacity expression simplifies to

$$\mathsf{R}_E(SNR,\theta) = \max_{r\geq 0}\left\{-\frac{1}{\theta T B_c}\log_e\left(1-P\{z>\alpha\}(1-e^{-\theta T r})\right)\right\}. \tag{3.18}$$

where $\alpha = \frac{2^{\frac{r}{B_c}}-1}{SNR}$ *and* $SNR = \frac{\bar{P}}{N_0 B}$.

Proof: See Appendix B.

Theorem 6 shows that the effective capacity of a wideband channel with N subchannels each with i.i.d. flat fading has an expression similar to that in (3.9), which provides the effective capacity of a single channel experiencing flat fading. The only difference between (3.9) and (3.18) is that B is replaced in (3.18) by B_c, which is the bandwidth of each subchannel.

In this section, we consider the wideband regime in which the overall bandwidth of the system, B, is large. In particular, we analyze the performance in the scenario of sparse multipath fading. Motivated by the recent measurement studies in the ultrawideband regime, the authors in [57] and [58] considered sparse multipath fading channels and analyzed the performance under channel uncertainty, employing the Shannon capacity formulation as the performance metric. In particular, [57] and [58] noted that the number of independent resolvable paths in sparse multipath channels increase at most sublinearly with the bandwidth, which in turn causes the coherence bandwidth B_c to increase with increasing bandwidth. To characterize the performance of sparse fading channels in the wideband regime, we assume in this section that $B_c \to \infty$ as $B \to \infty$. We further assume that the the number of subchannels N remains bounded and hence the degrees of freedom are limited. For instance, this case arises if the number of resolvable paths are bounded even at infinite bandwidth. Such a scenario is considered in [59] where the capacity and mutual information are characterized under channel uncertainty in the wideband regime with bounded

number of paths.

The case of rich multipath fading in which the B_c remains fixed and N grows without bound and the scenario in which both B_c and N increase to infinity are treated in Section 3.4 because each subchannel in these cases operates in the low-power regime as N increases.

We first introduce the notation $\zeta = \frac{1}{B_c}$. Note that as $B_c \to \infty$, we have $\zeta \to 0$. Moreover, with this notation, the normalized effective capacity in (3.18) given for i.i.d. fading can, after maximization, be expressed as[3]

$$\mathsf{R}_E(\mathrm{SNR}) = -\frac{\zeta}{\theta T} \log_e \Big(1 - P\{z > \alpha_{\mathrm{opt}}\}\Big(1 - e^{-\theta T r_{\mathrm{opt}}}\Big)\Big). \tag{3.19}$$

Note that α_{opt} and r_{opt} are also in general dependent on B_c and hence ζ. The following result provides the expressions for the minimum bit energy, which is achieved at zero spectral efficiency (i.e., as $B \to \infty$ and $B_c \to \infty$), and the wideband slope, and characterizes the spectral efficiency-bit energy tradeoff in the wideband regime when multipath fading is sparse, the number of subchannels is bounded, and the fading coefficients are i.i.d. in different subchannels.

Theorem 7 *In sparse multipath fading wideband channels with bounded number of subchannels each with i.i.d. fading coefficients, the minimum bit energy and wideband slope are given by*

$$\frac{E_b}{N_0}_{\min} = \frac{-\delta \log_e 2}{\log_e \xi} \quad \textit{and} \tag{3.20}$$

$$\mathcal{S}_0 = \frac{2\xi \log_e^2 \xi}{(\delta \alpha^*_{opt})^2 P\{z > \alpha^*_{opt}\} e^{-\delta \alpha^*_{opt}}}, \tag{3.21}$$

[3]Since the results in the chapter are generally obtained for fixed but arbitrary θ, the normalized effective capacity is often expressed in the chapter as $\mathsf{R}_E(\mathrm{SNR})$ instead of $\mathsf{R}_E(\mathrm{SNR}, \theta)$ to avoid cumbersome expressions.

respectively, where $\delta = \frac{\theta T \bar{P}}{N N_0 \log_e 2}$ *and* $\xi = 1 - P\{z > \alpha^*_{opt}\}(1 - e^{-\delta \alpha^*_{opt}})$. α^*_{opt} *is defined as* $\alpha^*_{opt} = \lim_{\zeta \to 0} \alpha_{opt}$ *and* α^*_{opt} *satisfies*

$$\delta \alpha^*_{opt} = \log_e \left(1 + \delta \frac{P\{z > \alpha^*_{opt}\}}{p_z(\alpha^*_{opt})} \right). \tag{3.22}$$

Proof: See Appendix C.

Remark: Theorem 7, through the minimum bit energy and wideband slope expressions, quantifies the bit energy requirements in the wideband regime when the system is operating subject to statistical QoS constraints specified by θ. Note that both $\frac{E_b}{N_0 \min}$ and $\mathcal{S}_0$ depend on the QoS exponent θ through δ. As will be observed in the numerical results, $\frac{E_b}{N_0 \min}$ and the bit energy requirements at nonzero spectral efficiency values generally increase with increasing θ. Moreover, when compared with the results in Section 3.4, it will be seen that sparse multipath fading and having a bounded number of subchannels incur energy penalty in the presence of QoS constraints while performances do not depend on the multipath sparsity when there are no such constraints and hence $\theta = 0$.

Having analytically characterized the spectral efficiency–bit energy tradeoff in the wideband regime in Theorem 7, we now provide numerical results to illustrate the theoretical findings. Fig. 3.3 plots the spectral efficiency curves as a function of the bit energy in the Rayleigh channel. In all the curves, we have $\bar{P}/(NN_0) = 10^4$. Moreover, we set $T = 2$ ms in the numerical results throughout the chapter. As predicted by the result of Theorem 7, $\frac{E_b}{N_0}\Big|_{\mathsf{R}_E=0} = \frac{E_b}{N_0 \min}$ in all cases in Fig. 3.3. It can be found that $\alpha^*_{\text{opt}} = \{1, 0.9858, 0.8786, 0.4704, 0.1177\}$ from which we obtain $\frac{E_b}{N_0 \min} = \{2.75, 2.79, 3.114, 5.061, 10.087\}$dB for $\theta = \{0, 0.001, 0.01, 0.1, 1\}$, respectively. For the same set of θ values in the same sequence, we compute the wideband slope values as $\mathcal{S}_0 = \{0.7358, 0.7463, 0.8345, 1.4073, 3.1509\}$. We immediately observe that

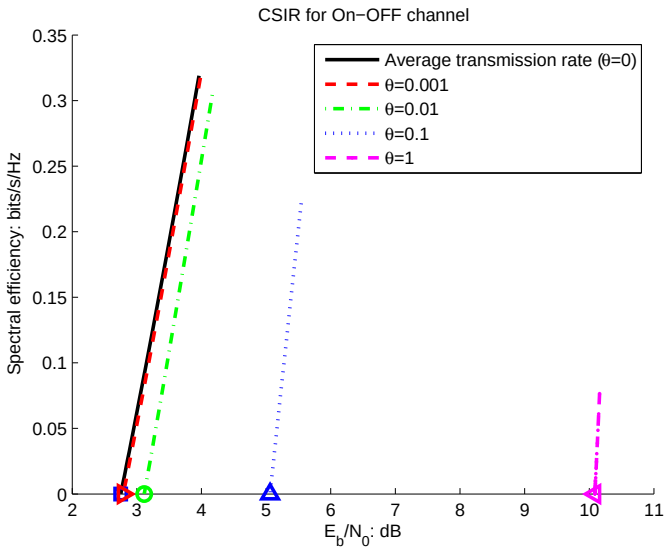

Figure 3.3: Spectral efficiency vs. E_b/N_0 in the Rayleigh channel.

more stringent QoS constraints and hence higher values of θ lead to higher minimum bit energy values and also higher energy requirements at other nonzero spectral efficiencies. Fig. 3.4 provides the spectral efficiency curves for Nakagami-m fading channels for different values of m. In this figure, we set $\theta = 0.01$. For $m = 0.6, 1, 2, 5$, we find that $\alpha^*_{\text{opt}} = \{1.0567, 0.8786, 0.7476, 0.6974\}$, $\frac{E_b}{N_0}_{\min} = \{3.618, 3.114, 2.407, 1.477\}$, and $\mathcal{S}_0 = \{0.6382, 0.8345, 1.1220, 1.4583\}$, respectively. Note that as m increases and hence the channel conditions improve, the minimum bit energy decreases and the wideband slope increases, improving the energy efficiency both at zero spectral efficiency and at nonzero but small spectral efficiency values. As $m \to \infty$, the performance approaches that of the unfaded additive Gaussian noise channel (AWGN) for which we have $\frac{E_b}{N_0}_{\min} = -1.59$ dB and $\mathcal{S}_0 = 2$ [21].

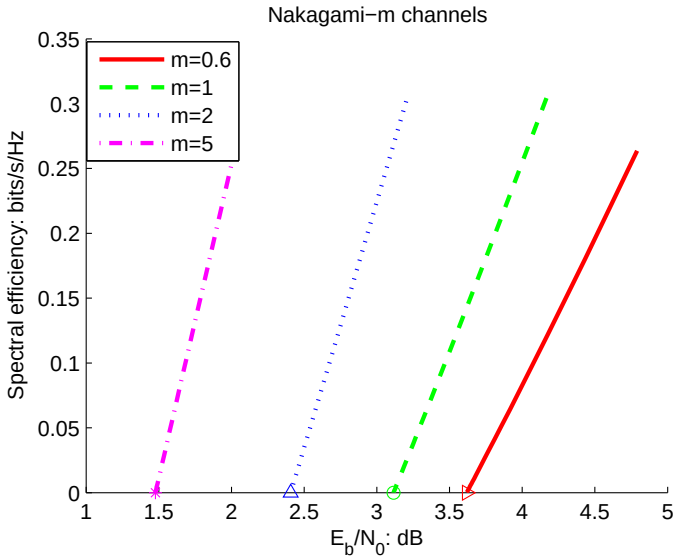

Figure 3.4: Spectral efficiency vs. E_b/N_0 in Nakagami-m channels; $\theta = 0.01$, $m = 0.6, 1, 2, 5$.

3.4 Energy Efficiency in the Low-Power Regime

In this section, we investigate the spectral efficiency–bit energy tradeoff in a single flat-fading channel as the average power $\bar{P}$ diminishes. We assume that the bandwidth allocated to the channel is fixed. Note that $\text{SNR} = \bar{P}/(N_0 B)$ vanishes with decreasing $\bar{P}$, and we again operate in the low-SNR regime similarly as in Section 3.3. Note further from (3.10) that the effective capacity of a flat-fading channel is given by

$$\mathsf{R}_E(\text{SNR}) = -\frac{1}{\theta T B} \log_e \Big(1 - P\{z > \alpha_{\text{opt}}\}\big(1 - e^{-\theta T r_{\text{opt}}}\big)\Big). \tag{3.23}$$

On the other hand, we remark that the results derived here also apply to the wide-band regime under the assumption that the number of non-interacting subchannels increases without bound with increasing bandwidth. Note that in such a case, each subchannel operates in the low-power regime.

The following result provides the expressions for the bit energy at zero spectral

efficiency and the wideband slope.

Theorem 8 *In the low-power regime, the bit energy at zero spectral efficiency and wideband slope are given by*

$$\left.\frac{E_b}{N_0}\right|_{R_E=0} = \frac{\log_e 2}{\alpha^*_{opt} P\{z > \alpha^*_{opt}\}} \quad \textit{and} \tag{3.24}$$

$$\mathcal{S}_0 = \frac{2P\{z > \alpha^*_{opt}\}}{1 + \beta(1 - P\{z > \alpha^*_{opt}\})}, \tag{3.25}$$

respectively, where $\beta = \frac{\theta TB}{\log_e 2}$ *is the normalized QoS constraint. In the above formulation,* α^*_{opt} *is defined as* $\alpha^*_{opt} = \lim_{SNR\to 0} \alpha_{opt}$*, and* α^*_{opt} *satisfies*

$$\alpha^*_{opt} p_z(\alpha^*_{opt}) = P\{z > \alpha^*_{opt}\}. \tag{3.26}$$

Proof: See Appendix D.

Corollary 1 *The same bit energy and wideband slope expressions as in (3.24) and (3.25) are achieved in the wideband regime as* $B \to \infty$ *if the fading coefficients in different subchannels are i.i.d. and also if the number of subchannels* N *increases linearly with increasing bandwidth (as in rich multipath fading channels), keeping the coherence bandwidth fixed.*

Under the assumptions stated in Corollary 1, the effective capacity is given by (3.18). Moreover, as $B \to \infty$, we have B_c fixed while $N \to \infty$. Hence, $\text{SNR} = \frac{\bar{P}/N}{N_0 B_c} \to 0$. This setting is exactly the same as the low-power regime considered in Theorem 8. Therefore, the results of Theorem 8 apply immediately.

Next, we show that equation (3.26) that needs to be satisfied by α^*_{opt} has a unique solution for a certain class of fading distributions.

Theorem 9 *Assume that the probability density function of z, denoted by $p_z(\cdot)$, is differentiable, and both $p_z(\cdot)$ and its derivative $\dot{p}_z(\cdot)$ at the origin do not contain impulses or higher-order singularities and are finite. Assume further that the support of $p_z(\cdot)$ is $[0, \infty)$. Under these assumptions, if $2p_z(x)+x\dot{p}_z(x) = 0$ is solved at a single point $x_0 > 0$ among all $x \in (0, \infty)$, then the equation $\alpha^*_{opt} p_z(\alpha^*_{opt}) = P\{z > \alpha^*_{opt}\}$ has a unique solution.*

Proof: We first define $f(x) = xp_z(x) - P\{z > x\}$ for $x \geq 0$.Under the conditions stated in Theorem 9, we can easily see that $f(0) = -1$ and $f(\infty) = 0$. Moreover, $f(x) \geq -1$ for all $x \geq 0$ because $p_z(x) \geq 0$ and $P\{z > x\} \leq 1$. It can also be seen that $\int_0^\infty f(x)dx = \int_0^\infty xp_z(x)dx - \int_0^\infty P\{z > x\}dx = E\{z\} - E\{z\} = 0$. Therefore, there exists $x > 0$ such that $f(x) > 0$.

Differentiating $f(x)$ with respect to x gives $\dot{f}(x) = 2p_z(x) + x\dot{p}_z(x)$. Note that $\dot{f}(0) = 2p_z(0) \geq 0$. Since $f(x) \geq -1$ and $f(0) = -1$, f is necessarily an increasing function initially. Hence, $\dot{f}(x) > 0$ for all $x \in (0, x_0)$ where x_0 is the point at which $\dot{f}(x_0) = 0$. Since x_0 is the only positive point for which the derivative is zero, and $f(x) > 0$ for some x as discussed above and $f(x)$ has to approach zero as $x \to \infty$, we conclude that $f(x)$ is a decreasing function for all $x > x_0$, and hence $\dot{f}(x) < 0$ for all $x > x_0$. Otherwise, if $\dot{f}(x) > 0$ for some x, $\dot{f}(x)$ never becomes zero again, and $f(x)$ increases indefinitely. Furthermore, we can see that $f(x) > 0$ for all $x \geq x_0$ because if $f(x) < 0$ for some $x \geq x_0$, $f(x)$ should start increasing to zero as $x \to \infty$. However, this is not possible because $\dot{f}(x) < 0$ for all $x > x_0$.

Therefore, we have concluded that $f(0) = -1$ and $f(x)$ is an increasing function in the range $x \in (0, x_0)$. Moreover, $f(x_0) > 0$ and $f(x)$ decreases to zero without being negative as $x \to \infty$. From this, we conclude that $f(x)$ intersects the horizontal axis only once at an x value in between 0 and x_0. Therefore, $f(x) = 0$ has a unique

solution. □

Remark: The conditions of Theorem 9 are satisfied by a general class of distributions, including the Gamma distribution,

$$p_z(z) = \frac{z^{\alpha-1}e^{-\frac{z}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)},$$

where $z, \alpha, \beta > 0$, and Lognormal distribution,

$$p_z(z) = \frac{1}{\sigma z\sqrt{2\pi}}e^{-\frac{(\log_e z - m)^2}{2\sigma^2}},$$

where $z > 0$, $-\infty < m < \infty$, and $\sigma > 0$. Note that in Nakagami-m and Rayleigh fading channels, the distribution of $z = |h|^2$ can be seen as special cases of the Gamma distribution. In Fig. 3.5 and 3.6, where the function $f(\cdot)$ is plotted for Gamma and Lognormal distributions, we indeed observe that these distributions satisfy the conditions of Theorem 9 and the function $f(\cdot)$ is equal to zero at a unique point.

Remark: Theorem 8 shows that the $\frac{E_b}{N_0}\Big|_{\mathsf{R}_E=0}$ for any $\theta \geq 0$ depends only on α^*_{opt}. From Theorem 9, we know under certain conditions that α^*_{opt} is unique and hence is the same for all $\theta \geq 0$. We immediately conclude from these results that $\frac{E_b}{N_0}\Big|_{\mathsf{R}_E=0}$ also has the same value for all $\theta \geq 0$ and therefore does not depend on θ for the class of distributions and channels given in the above Remark.

Moreover, using the results of Theorem 9 above and Theorem 7 in Section 3.3, we can further show that $\frac{E_b}{N_0}\Big|_{\mathsf{R}_E=0}$ is the minimum bit energy. Note that this implies that the same minimum bit energy can be attained regardless of how strict the QoS constraint is. On the other hand, we note that the wideband slope $\mathcal{S}_0$ in general varies with θ.

Corollary 2 *In the low-power regime, when $\theta = 0$, the minimum bit energy is*

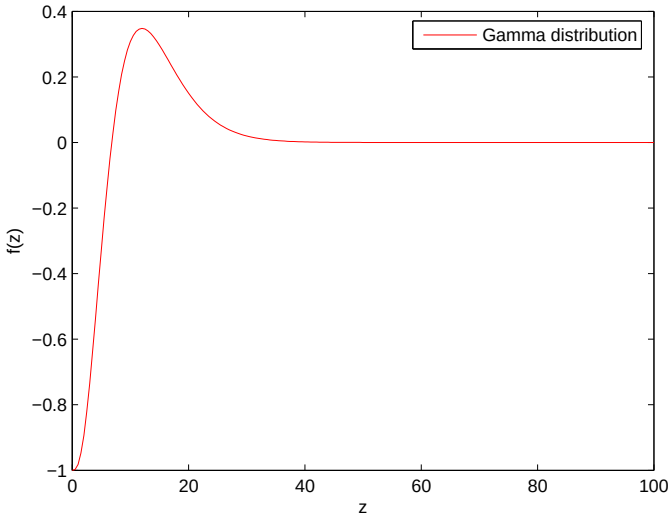

Figure 3.5: The plot of the function $f(z)$ for Gamma distribution $p_z(z) = \frac{z^{\alpha-1}e^{-\frac{z}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)}$ with $\alpha = \beta = 3$.

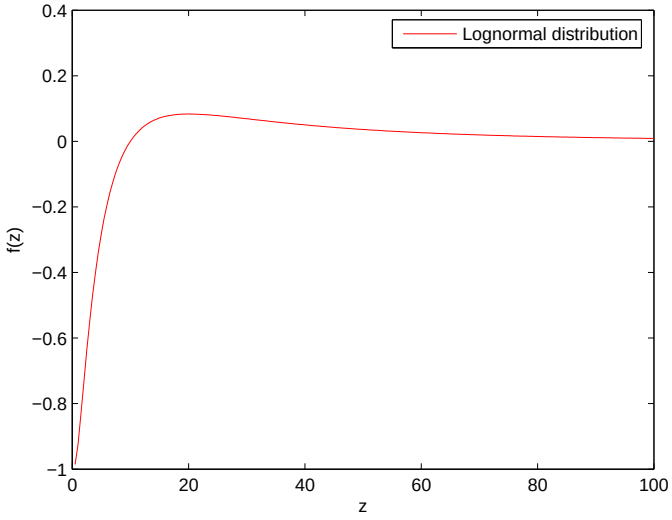

Figure 3.6: The plot of the function $f(z)$ for Lognormal distribution $p_z(z) = \frac{1}{\sigma x\sqrt{2\pi}}e^{-\frac{(\log_e x-m)^2}{2\sigma^2}}$ with $\sigma = 1, m = 2$.

achieved as $\bar{P} \to 0$*, i.e.,* $\frac{E_b}{N_0}\Big|_{\mathsf{R}_E=0} = \frac{E_b}{N_0}_{\min}$*. Moreover, if the probability density function of* z *satisfies the conditions stated in Theorem 9, then the minimum bit energy is achieved as* $\bar{P} \to 0$*, i.e.* $\frac{E_b}{N_0}\Big|_{\mathsf{R}_E=0} = \frac{E_b}{N_0}_{\min}$*, for all* $\theta \geq 0$*.*

Proof: Recall from (3.11) that in the limit as $\theta \to 0$,

$$\mathsf{R}_E(\mathrm{SNR}, 0) = \lim_{\theta \to 0} \mathsf{R}_E(\mathrm{SNR}, \theta) = \max_{r \geq 0} \ \frac{r}{B} P\left\{ z > \frac{2^{\frac{r}{B}} - 1}{\mathrm{SNR}} \right\}. \tag{3.27}$$

Since the optimization is performed over all $r \geq 0$, it can be easily seen that the above maximization problem can be recast as follows:

$$\mathsf{R}_E(\mathrm{SNR}, 0) = \max_{x \geq 0} \ x\, P\left\{ z > \frac{2^{x} - 1}{\mathrm{SNR}} \right\}. \tag{3.28}$$

From (3.28), we note that $\mathsf{R}_E(\mathrm{SNR}, 0)$ depends on B only through $\mathrm{SNR} = \frac{\bar{P}}{N_0 B}$. Therefore, increasing B has the same effect as decreasing $\bar{P}$. Hence, low-power and wideband regimes are equivalent when $\theta = 0$. Consequently, the result of Theorem 7, which shows that the minimum bit energy is achieved as $B \to \infty$, implies that the minimum bit energy is also achieved as $\bar{P} \to 0$.

Note that $\mathsf{R}_E(\mathrm{SNR}, \theta) \leq \mathsf{R}_E(\mathrm{SNR}, 0)$ for $\theta > 0$. Therefore, the bit energy required when $\theta > 0$ is larger than that required when $\theta = 0$. On the other hand, as we have proven in Theorem 9, α^*_{opt} is unique and the bit energy required as $\bar{P} \to 0$ is the same for all $\theta \geq 0$ when p_z satisfies certain conditions. Since the minimum bit energy in the case of $\theta = 0$ is achieved as $\bar{P} \to 0$, and the same bit energy is attained for all $\theta > 0$, we immediately conclude that $\frac{E_b}{N_0}\Big|_{\mathsf{R}_E=0} = \frac{E_b}{N_0}_{\min}$ for all $\theta \geq 0$ □

Next, we provide numerical results which confirm the theoretical conclusions and illustrate the impact of QoS constraints on the energy efficiency. We set $B = 10^5$ Hz in the computations. Fig. 3.7 plots the spectral efficiency as a function of the bit energy

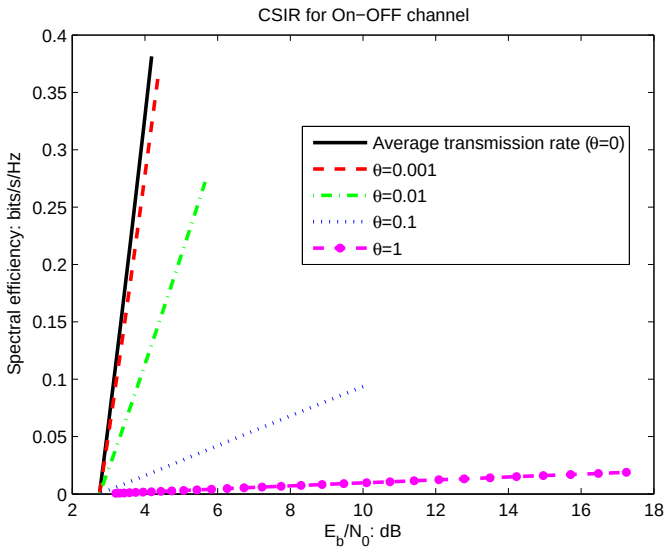

Figure 3.7: Spectral efficiency vs. E_b/N_0 in the Rayleigh channel (equivalently Nakagami-m channel with $m = 1$).

for different values of θ in the Rayleigh fading channel (or equivalently Nakagami-m fading channel with $m = 1$) for which $\mathbb{E}\{|h|^2\} = \mathbb{E}\{z\} = 1$. In all cases in Fig. 3.7, we readily note that $\frac{E_b}{N_0}\Big|_{\mathsf{R}_E=0} = \frac{E_b}{N_0}_{\min}$. Moreover, as predicted, the minimum bit energy is the same and is equal to the one achieved when there are no QoS constraints (i.e., when $\theta = 0$). From the equation $\alpha^*_{\text{opt}} p_z(\alpha^*_{\text{opt}}) = P\{z > \alpha^*_{\text{opt}}\}$, we can find that $\alpha^*_{\text{opt}} = 1$ in the Rayleigh channel for which $p_z(\alpha^*_{\text{opt}}) = P\{z > \alpha^*_{\text{opt}}\} = e^{-\alpha^*_{\text{opt}}}$. Hence, the minimum bit energy is $\frac{E_b}{N_0}_{\min} = 2.75$ dB. On the other hand, the wideband slopes are $\mathcal{S}_0 = \{0.7358, 0.6223, 0.2605, 0.0382, 0.0040\}$ for $\theta = \{0, 0.001, 0.01, 0.1, 1\}$, respectively. Hence, $\mathcal{S}_0$ decreases with increasing θ and consequently more bit energy is required at a fixed nonzero spectral efficiency. Assuming that the minimum bit energies are the same and considering the linear approximation in (3.13), we can easily show for fixed spectral efficiency R_E $\left(\frac{E_b}{N_0}\right)$ for which the linear approximation is accurate that the increase in the bit energy in dB, when the QoS exponent increases

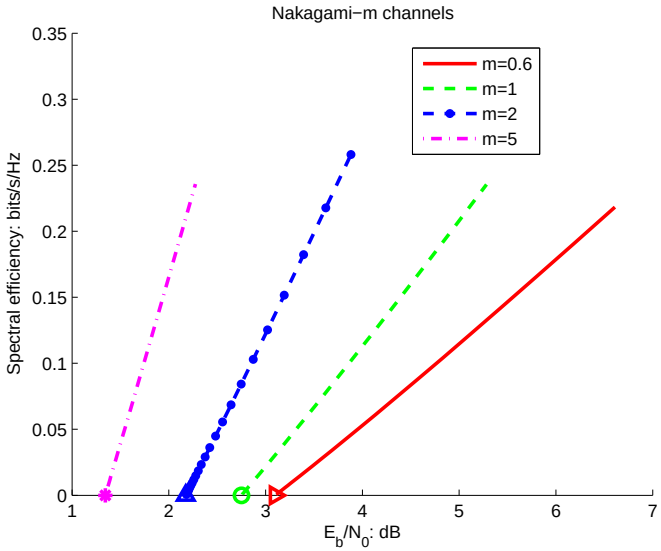

Figure 3.8: Spectral efficiency vs. E_b/N_0 in Nakagami-m channels; $\theta = 0.01$, $m = 0.6, 1, 2, 5$.

from θ_1 to θ_2, is

$$\left.\frac{E_b}{N_0}\right|_{dB,\theta_2} - \left.\frac{E_b}{N_0}\right|_{dB,\theta_1} = \left(\frac{1}{\mathcal{S}_{0,\theta_2}} - \frac{1}{\mathcal{S}_{0,\theta_1}}\right) \mathsf{R}_E\left(\frac{E_b}{N_0}\right) 10\log_{10} 2. \tag{3.29}$$

As observed in Fig. 3.7 (and also as will be seen in Fig. 3.8), spectral efficiency curves are almost linear in the low-power regime, validating the accuracy of the linear approximation in (3.13) obtained through $\left.\frac{E_b}{N_0}\right|_{\mathsf{R}_E=0}$ and $\mathcal{S}_0$.

Fig. 3.8 plots the spectral efficiency curves as a function of the bit energy for Nakagami-m channels for different values of m. θ is set to be 0.01. For $m = \{0.6, 1, 2, 5\}$, we compute that $\alpha^*_{\text{opt}} = \{1.2764, 1, 0.809, 0.7279\}$,

$$\frac{E_b}{N_{0\,\min}} = \{3.099, 2.751, 2.176, 1.343\}$$

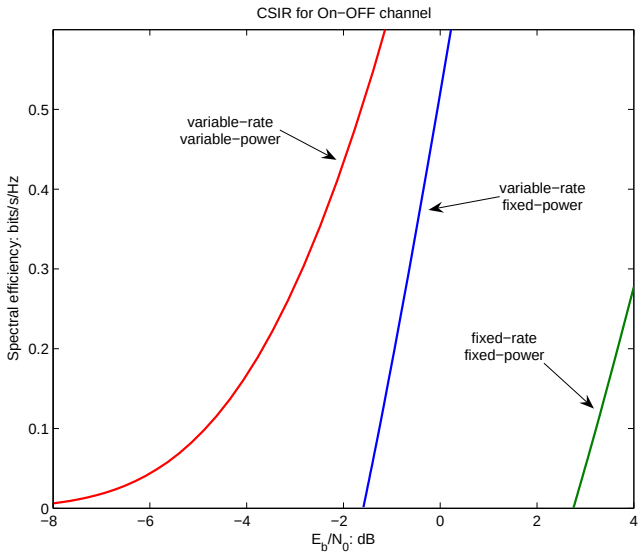

Figure 3.9: Spectral efficiency vs. E_b/N_0 in the Rayleigh channel; $\theta = 0.001$.

, and

$$\mathcal{S}_0 = \{0.1707, 0.2605, 0.4349, 0.7479\}$$

, respectively. We observe that as m increases and hence the channel quality improves, lower bit energies are required. Finally, in Fig. 3.9, we plot the spectral efficiency vs. E_b/N_0 for different transmission strategies. The variable-rate/variable-power and variable-rate/fixed-power strategies are studied in Chapter 2. We immediately see that substantially more energy is required for fixed-rate/fixed-power transmission schemes considered in this chapter.

Remark: From the result of Corollary 1, we note that the analytical and numerical results in this section apply to wideband channels with rich multipath fading. Comparison of Fig. 3.7 with Fig. 3.3, where sparse multipath fading scenario is considered, leads to several insightful observations. Note that in both figures, the performance is the same when $\theta = 0$. Hence, in the absence of QoS constraints, mul-

tipath sparsity or richness has no effect. This also confirms the claim in the proof of Corollary 2 that low-power and wideband regimes are equivalent when θ. However, we see a stark difference when $\theta > 0$. We observe that multipath sparsity and having the number of subchannels bounded in the wideband regime increases the bit energy requirements significantly especially when θ is large. Moreover, while the minimum bit energy is the same for all θ in Fig. 3.7, the minimum bit energy increases with increasing θ in Fig. 3.3.

In Section 3.3, the number of subchannels are assumed to be bounded. In this section, we have considered the rich multipath fading channels in which the number of subchannels increases linearly with bandwidth. A scenario in between these two cases is the one in which the number of subchannels N increases but only sublinearly with increasing bandwidth. As N increases, each subchannel is allocated less power and operate in the low-power regime. At the same time, since N increases sublinearly with B, the coherence bandwidth $B_c = B/N$ also increases. Therefore, the minimum bit energy and wideband slope expressions for this scenario can be obtained by letting B in the results of Theorem 8 go to infinity. Note that under the conditions of Theorem 9, α^*_{opt} is unique and hence does not depend on the bandwidth.

Corollary 3 *In the wideband regime, if the number of subchannels N increases sublinearly with B and if fading coefficients in different subchannels are i.i.d. and the probability density function p_z satisfies the conditions in Theorem 9, then the minimum bit energy and wideband slope are given by*

$$\frac{E_b}{N_{0\,\min}} = \frac{\log_e 2}{\alpha^*_{opt} P\{z > \alpha^*_{opt}\}} \quad \textit{and} \tag{3.30}$$

$$\mathcal{S}_0 = \begin{cases} 2P\{z > \alpha^*_{opt}\} & \theta = 0 \\ 0 & \theta > 0 \end{cases}. \tag{3.31}$$

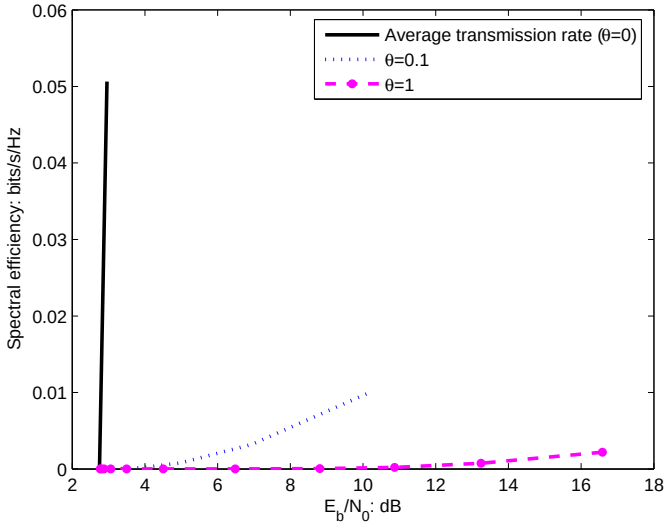

Figure 3.10: Spectral efficiency vs. E_b/N_0 in the Rayleigh channel. The number of subchannels N increases sublinearly with bandwidth.

In this result, we see that although the same minimum bit energy is attained for all $\theta \geq 0$, approaching this minimum energy level is extremely slow and demanding when $\theta > 0$ due to zero wideband slope. This result is illustrated numerically in Fig. 3.10.

3.5 Conclusion

In this chapter, we have considered the effective capacity as a measure of the maximum throughput under statistical QoS constraints, and analyzed the energy efficiency of fixed-rate transmission schemes over fading channels. In particular, we have investigated the spectral efficiency–bit energy tradeoff in the low-power and wideband regimes. We have obtained expressions for the bit energy at zero spectral efficiency and the wideband slope, which provide a linear approximation to the spectral efficiency curve at low SNRs. In the initial analysis of the wideband regime with bounded

number of resolvable paths and hence bounded number of subchannels, we have determined that the bit energy required at zero spectral efficiency (or equivalently at infinite bandwidth) is the minimum bit energy. In this case, we have noted that the minimum bit energy and wideband slope in general depend on the QoS exponent θ. As the QoS constraints become more stringent and hence θ is increased, we have observed in the numerical results that the required minimum bit energy increases. Subsequently, we have considered the low-power regime, which can also be equivalently regarded as the wideband regime with rich multipath fading. We have obtained expressions for the bit energy required at zero spectral efficiency, and wideband slope. For a certain class of fading distributions, we have shown that the bit energy at zero spectral efficiency is indeed the minimum bit energy and is achieved regardless of how strict the QoS constraints are. However, we have also noted that the wideband slope decreases as θ increases, increasing the energy requirements at nonzero spectral efficiency values. Overall, we have quantified the increased energy requirements in the presence of QoS constraints in both wideband and low-power regimes, and identified the impact upon the energy efficiency of multipath sparsity and richness in the wideband regime.

Chapter 4

Energy Efficiency for Training Based Transmissions

In this chapter, we consider the scenario in which neither the transmitter nor the receiver has CSI prior to transmission and the channel coefficients are estimated at the receiver via minimum mean-square-error (MMSE) estimation with the aid of training symbols. For this scenario, we identify the optimal fraction of power allocated to training. We show that the bit energy increases without bound in the low-power regime as the average power vanishes. A similar conclusion is reached in the wideband regime if the number of noninteracting subchannels grows without bound with increasing bandwidth. On the other hand, it is proven that if the number of resolvable independent paths and hence the number of noninteracting subchannels remain bounded as the available bandwidth increases, the bit energy diminishes to its minimum value in the wideband regime.

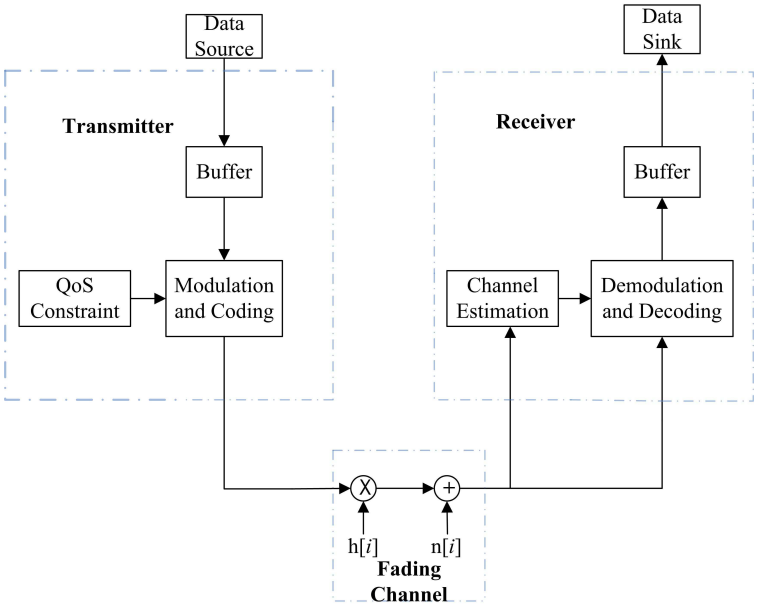

Figure 4.1: The general system model.

4.1 Channel Model

We consider a point-to-point wireless link. Figure 4.1 illustrates the functional diagram of the system. It is assumed that the source generates data sequences which are divided into frames of duration T. These data frames are initially stored in the buffer before they are transmitted over the wireless channel. The discrete-time channel input-output relation in the i^{th} symbol duration is given by

$$y[i] = h[i]x[i] + n[i] \quad i = 1, 2, \ldots. \tag{4.1}$$

where $x[i]$ and $y[i]$ denote the complex-valued channel input and output, respectively. We assume that the bandwidth available in the system is B and the channel input is subject to the following average energy constraint: $\mathbb{E}\{|x[i]|^2\} \leq \bar{P}/B$ for all i. Since the bandwidth is B, symbol rate is assumed to be B complex symbols per

second, indicating that the average power of the system is limited by $\bar{P}$. Above in (4.1), $n[i]$ is a zero-mean, circularly symmetric, complex Gaussian random variable with variance $\mathbb{E}\{|n[i]|^2\} = N_0$, i.e., $n[i] \sim \mathcal{CN}(0, N_0)$. The additive Gaussian noise samples $\{n[i]\}$ are assumed to form an independent and identically distributed (i.i.d.) sequence. Finally, $h[i]$, which denotes the channel fading coefficient, is assumed to be a zero-mean Gaussian random variable with variance $E\{|h|^2\} = \gamma$. Therefore, the wireless channel is modeled as a Rayleigh fading channel. We consider a block-fading channel model. Hence, we assume that the fading coefficients stay constant during the frame duration of T seconds and change independently from one frame to another. Finally, we assume that neither the transmitter nor the receiver has channel side information prior to transmission. While the transmitter remains unaware of the actual realizations of the fading coefficients throughout the transmission, the receiver attempts to learn them through training.

4.2 Training and Data Transmission

4.2.1 Training Phase

The system operates in two phases: training phase and data transmission phase. In the training phase, known pilot symbols are transmitted to enable the receiver to estimate the channel conditions, albeit imperfectly. We assume that minimum mean-square-error (MMSE) estimation is employed at the receiver to estimate the channel coefficient $h[i]$. Since the MMSE estimate depends only on the training energy and not on the training duration [60] and the fading coefficients are assumed to stay constant during the frame duration of T seconds, it can be easily seen that transmission of a single pilot at every T seconds is optimal. Note that in every frame duration of T

seconds, we have TB symbols and the overall available energy is $\bar{P}T$. We now assume that each frame consists of a pilot symbol and $TB-1$ data symbols. The energies of the pilot and data symbols are

$$\mathcal{E}_p = \rho \bar{P} T, \quad \text{and} \quad \mathcal{E}_s = \frac{(1-\rho)\bar{P}T}{TB-1}, \tag{4.2}$$

respectively, where ρ is the fraction of total energy allocated to training. Note that the data symbol energy $\mathcal{E}_s$ is obtained by uniformly allocating the remaining energy among the data symbols.

In the training phase, the transmitter sends the pilot symbol $x_p = \sqrt{\mathcal{E}_p} = \sqrt{\rho \bar{P} T}$ and the receiver obtains[1]

$$y[1] = h\sqrt{\mathcal{E}_p} + n[1]. \tag{4.3}$$

Based on the received signal in this phase, the receiver obtains the MMSE estimate $h_{est} = \mathbb{E}\{h \big| y[1]\}$ which can be easily seen to be a circularly symmetric, complex, Gaussian random variable with mean zero and variance $\frac{\gamma^2 \mathcal{E}_p}{\gamma \mathcal{E}_p + N_0}$, i.e., $h_{est} \sim \mathcal{CN}\left(0, \frac{\gamma^2 \mathcal{E}_p}{\gamma \mathcal{E}_p + N_0}\right)$[61]. Now, the channel fading coefficient h can be expressed as $h = h_{est} + h_{err}$ where h_{err} is the estimate error and $h_{err} \sim \mathcal{CN}(0, \frac{\gamma N_0}{\gamma \mathcal{E}_p + N_0})$.

4.2.2 Data Transmission Phase and Capacity Lower Bound

Data transmission follows the training phase. Since the receiver is now equipped with the channel estimate, the channel input-output relation in one frame in the data

[1]Since the analysis in this section focuses on a single frame in which the fading stays constant, we drop the time index in $h[i]$ and express the fading coefficient as h. In (4.3), $y[1]$ and $n[1]$ denote the received symbol and noise sample, respectively, in the training phase. Note that the first symbol duration in each frame is allocated for the training phase in which a single pilot symbol is sent.

transmission phase can be expressed as

$$y[i] = h_{est}x[i] + h_{err}x[i] + n[i] \quad i = 2, 3, \ldots, TB. \tag{4.4}$$

Since finding the capacity of the channel in (4.4) is a difficult task [61], a capacity lower bound is generally obtained by treating $h_{err}x[i] + n[i]$ as Gaussian distributed noise with variance $\mathbb{E}\{|h_{err}x[i] + n[i]|^2\} = \sigma^2_{h_{err}}\mathcal{E}_s + N_0$ where $\sigma^2_{h_{err}} = \mathbb{E}\{|h_{err}|^2\} = \frac{\gamma N_0}{\gamma\mathcal{E}_p + N_0}$ is the variance of the estimate error. Under these assumptions, a lower bound on the instantaneous capacity is given by [60], [61]

$$\begin{aligned} C_L &= \frac{TB-1}{T}\log_2\left(1 + \frac{\mathcal{E}_s}{\sigma^2_{h_{err}}\mathcal{E}_s + N_0}|h_{est}|^2\right) \\ &= \frac{TB-1}{T}\log_2\left(1 + \text{SNR}_{\text{eff}}|w|^2\right) \text{ bits/s} \end{aligned} \tag{4.5}$$

where the effective SNR is

$$\text{SNR}_{\text{eff}} = \frac{\mathcal{E}_s\sigma^2_{h_{est}}}{\sigma^2_{h_{err}}\mathcal{E}_s + N_0}, \tag{4.6}$$

and $\sigma^2_{h_{est}} = \mathbb{E}\{|h_{est}|^2\} = \frac{\gamma^2\mathcal{E}_p}{\gamma\mathcal{E}_p + N_0}$ is the variance of the estimate h_{est}. Note that the expression in (4.5) is obtained by defining $h_{est} = \sigma_{h_{est}}w$ where w is a standard complex Gaussian random variable with zero mean and unit variance, i.e., $w \sim \mathcal{CN}(0, 1)$. Henceforth, we base our analysis on C_L to understand the impact of the imperfect channel estimate.

4.2.3 Fixed-Rate Transmission and ON-OFF Model

Since the transmitter is unaware of the channel conditions, it is assumed that information is transmitted at a fixed rate of r bits/s. When $r < C_L$, the channel is considered to be in the ON state and reliable communication is achieved at this rate.

Note that under the block-fading assumption, the channel stays in the ON state for T seconds and the number of bits transmitted in this duration is rT. If, on the other hand, $r \geq C_L$, we assume that outage occurs. In this case, channel is in the OFF state during the frame duration and reliable communication at the rate of r bits/s cannot be attained. Hence, effective data rate is zero and information has to be resent. The probability of the channel being in the OFF state is

$$p_{\text{off}} = \Pr\{r \geq C_L\} = 1 - e^{-\alpha} \tag{4.7}$$

where

$$\alpha = \frac{2^{\frac{rT}{TB-1}} - 1}{\text{SNR}_{\text{eff}}}. \tag{4.8}$$

Rightmost expression in (4.7) follows from the fact that $|w|^2$ is an exponential random variable with mean 1. Noting that $|w|^2$ gives the normalized estimated channel strength, we see that the channel is in the OFF state if this channel strength is less than the threshold α. Similarly, the probability of being in the ON state is

$$p_{\text{on}} = \Pr\{r < C_L\} = \Pr\{|w|^2 > \alpha\} = e^{-\alpha}. \tag{4.9}$$

We finally remark that since the fading coefficients (and consequently h_{est}, w, and C_L) change independently from one frame to another under the block-fading assumption, the channel, in any given frame, is either in the ON or OFF state independently of its previous state.

4.3 Preliminary

With the transmission scheme described above, similar to Chapter 3, for the ON-OFF channel model described in Section 4.2.3, the effective capacity normalized by the frame duration T and bandwidth B, or equivalently spectral efficiency in bits/s/Hz, for a given QoS delay constraint specified by θ is given by[2],

$$\begin{aligned}
\mathsf{R}_E(\mathrm{SNR},\theta) &= \max_{\substack{r\geq 0\\ 0\leq\rho\leq 1}} -\frac{1}{\theta TB}\log_e\left(p_{\mathrm{off}}+p_{\mathrm{on}}e^{-\theta Tr}\right)\\
&= \max_{\substack{r\geq 0\\ 0\leq\rho\leq 1}} -\frac{1}{\theta TB}\log_e\left(1-e^{-\alpha}(1-e^{-\theta Tr})\right)\\
&= -\frac{1}{\theta TB}\log_e\left(1-e^{-\alpha_{\mathrm{opt}}}(1-e^{-\theta Tr_{\mathrm{opt}}})\right). \qquad (4.10)
\end{aligned}$$

where r_{opt} and α_{opt} are the optimal values of r and α, and p_{on} and p_{off}, as described in Section 4.2.3, are the probabilities of channel being in the ON and OFF states, respectively. Note that the optimal values r_{opt} and α_{opt} are functions of SNR in general. Note further that R_E is obtained by optimizing both the fixed transmission rate r and the fraction of power allocated to training, ρ. The dependence of the normalized effective capacity on ρ is through the threshold α which depends on $\mathrm{SNR}_{\mathrm{eff}}$.

[2]The formulation in (4.10) applies to the case in which the channel's currently being in the ON or OFF state is independent of its state in the previous frame. This arises due to block fading assumption. In a correlated fading scenario in which the current channel state has dependence on the previous one, we have a two-state (ON-OFF) Markov chain. For such a Markov model, using the result in [12, Section 7.2, Example 7.2.7], we can show that the effective capacity can be expressed as

$$\begin{aligned}
\max_{\substack{r\geq 0\\ 0\leq\rho\leq 1}} -\frac{1}{\theta TB}\log_e\Big(\frac{1}{2}\Big(&p_{\mathrm{off}}+p_{\mathrm{on}}e^{-\theta Tr}\\
&+\sqrt{(p_{\mathrm{off}}+p_{\mathrm{on}}e^{-\theta Tr})^2-4(p_{\mathrm{off}}+p_{\mathrm{on}}-1)e^{-\theta Tr}}\Big)\Big).
\end{aligned}$$

We can immediately see that the above expression specializes to (4.10) by noting that $p_{\mathrm{off}}+p_{\mathrm{on}}=1$ in the block fading scenario.

Also, it can easily be seen that

$$\begin{aligned}\mathsf{R}_E(\mathrm{SNR},0) &= \lim_{\theta\to 0}\mathsf{R}_E(\mathrm{SNR},\theta)\\ &= \max_{\substack{r\geq 0\\ 0\leq\rho\leq 1}} \frac{r}{B}\,\Pr\left\{|w|^2 > \frac{2^{\frac{rT}{TB-1}}-1}{\mathrm{SNR}_{\mathrm{eff}}}\right\}\\ &= \max_{\substack{r\geq 0\\ 0\leq\rho\leq 1}} \frac{r}{B}\,e^{-\frac{2^{\frac{rT}{TB-1}}-1}{\mathrm{SNR}_{\mathrm{eff}}}}. \end{aligned} \tag{4.11}$$

Hence, as the QoS requirements relax, the maximum constant arrival rate approaches the average transmission rate. On the other hand, for $\theta > 0$, $\mathsf{R}_E < \frac{1}{B}\max_{\substack{r\geq 0\\ 0\leq\rho\leq 1}} re^{-\alpha}$ in order to avoid violations of buffer constraints. Now, combine the discussion in Section 3.2, we can carry out the energy efficiency analysis for the transmission scheme described in this chapter.

4.3.1 Optimal Training Power

Before performing the energy efficiency analysis, we first obtain the following result on the optimal value of ρ, the fraction of the total energy allocated to training in the presence of QoS constraints.

Proposition 2 *At a given SNR level, the optimal fraction of power ρ_{opt} that solves the maximization problem above (4.10) does not depend on the QoS exponent θ and the transmission rate r, and is given by*

$$\rho_{opt} = \sqrt{\eta(\eta+1)} - \eta \tag{4.12}$$

where

$$\eta = \frac{\gamma TB\,SNR + TB - 1}{\gamma TB(TB-2)SNR} \quad \textit{and} \quad SNR = \frac{\bar{P}}{N_0 B}. \tag{4.13}$$

Proof: See Appendix E.

4.4 Energy Efficiency in the Low-Power Regime

In this section, we analyze the spectral-efficiency vs. bit energy tradeoff in the low power regime in which the average power of the system, $\bar{P}$, is small.

With the optimal value of ρ given in Proposition 2, we can now express the normalized effective capacity as

$$\mathsf{R}_E(\mathrm{SNR},\theta) = \max_{r\geq 0} -\frac{1}{\theta TB}\log_e\left(1 - e^{-\frac{2^{\frac{rT}{TB-1}}-1}{\mathrm{SNR}_{\mathrm{eff,opt}}}(1-e^{-\theta Tr})}\right) \tag{4.14}$$

$$= -\frac{1}{\theta TB}\log_e\left(1 - e^{-\frac{2^{\frac{r_{\mathrm{opt}}T}{TB-1}}-1}{\mathrm{SNR}_{\mathrm{eff,opt}}}(1-e^{-\theta Tr_{\mathrm{opt}}})}\right) \tag{4.15}$$

where r_{opt} is the optimal value of r that solves (4.14), and

$$\mathrm{SNR}_{\mathrm{eff,opt}} = \frac{\phi(\mathrm{SNR})\mathrm{SNR}^2}{\psi(\mathrm{SNR})\mathrm{SNR} + TB - 1}, \tag{4.16}$$

and

$$\begin{aligned}\phi(\mathrm{SNR}) &= \rho_{\mathrm{opt}}(1-\rho_{\mathrm{opt}})\gamma^2T^2B^2,\\ \psi(\mathrm{SNR}) &= (1+(TB-2)\rho_{\mathrm{opt}})\gamma TB.\end{aligned} \tag{4.17}$$

With these notations, we obtain the following result that shows us that operation at very low power levels is extremely energy inefficient and should be avoided.

Theorem 10 *In the presence of channel uncertainty, the bit energy for all* $\theta \geq 0$

increases without bound as the average power $\bar{P}$ and hence SNR vanishes, i.e.,

$$\left.\frac{E_b}{N_0}\right|_{R_E=0} = \lim_{SNR\to 0} \frac{E_b}{N_0} = \lim_{SNR\to 0} \frac{SNR}{R_E(SNR)} = \frac{1}{\dot{R}_E(0)} = \infty. \tag{4.18}$$

Proof: See Appendix F.

Remark: Theorem 10 shows that $\left.\frac{E_b}{N_0}\right|_{\mathsf{R}_E=0} = \infty$ for any $\theta \geq 0$. Note that this is a cautionary result. As will be evident in the numerical results, energy efficiency still improves if one operates at low power levels. However, if the power is reduced below a certain threshold, bit energy requirements start increasing and the required bit energy level grows without bound as power vanishes. One reason for this behavior is that although channel estimation at very low power levels does not provide reliable estimates, the receiver regards this estimate as perfect. Hence, in the low-power regime, we have both diminishing power and deteriorating channel estimate, which affect the performance adversely. The result of Theorem 10 also indicates that the minimum bit energy, which can be identified numerically, is achieved at a non-zero power level. In the numerical results, we will observe that both the minimum required bit energy and the other bit energy values required at a given level of spectral efficiency increase as the QoS constraints become more stringent.

Fig. 4.2 plots the spectral efficiency vs. bit energy for $\theta = \{1, 0.1, 0.01, 0.001\}$ when $B = 10^5$ Hz in Rayleigh channel with $\mathbb{E}\{|h|^2\} = \gamma = 1$. We notice that as spectral efficiency R_E decreases, the bit energy $\frac{E_b}{N_0}$ initially decreases. However, as predicted by the result of Theorem 10, the bit energy achieves its minimum value at a certain nonzero spectral efficiency below which $\frac{E_b}{N_0}$ starts increasing without bound. Hence, operation below the spectral efficiency or SNR level at which $\frac{E_b}{N_0\,\min}$ is attained should be avoided. We also note in Fig. 4.2 that the bit energy requirements in general and the minimum bit energy in particular increases with increasing θ value,

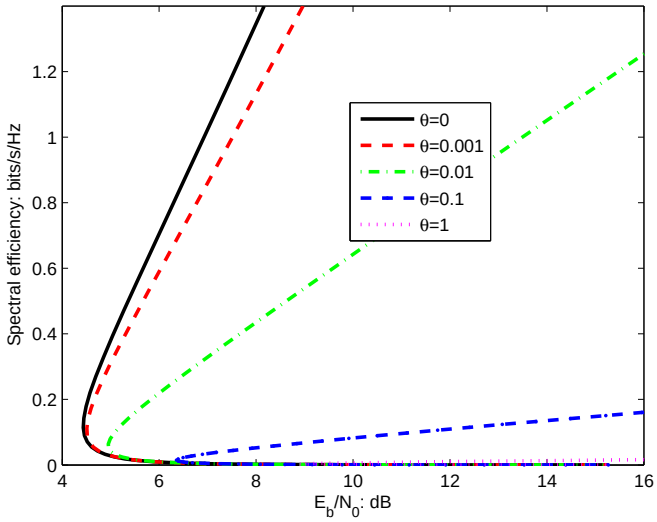

Figure 4.2: Spectral efficiency vs. E_b/N_0 in the Rayleigh channel with $\mathbb{E}\{|h|^2\} = 1$. $B = 10^5$.

indicating the increased energy costs as the QoS limitations become more stringent. In Fig. 4.3, we plot $\frac{E_b}{N_0}$ as a function of SNR for different bandwidth levels assuming $\theta = 0.01$. We again observe that the minimum bit energy is attained at a nonzero SNR value below which $\frac{E_b}{N_0}$ requirements start increasing. Furthermore, we see that as the bandwidth increases, the minimum bit energy tends to decrease and is achieved at a lower SNR level. Finally, we plot in Fig. 4.4 the minimum bit energy as a function of the bandwidth, B. We note that increasing B generally decreases $\frac{E_b}{N_0}_{\min}$ value. However, there is diminishing returns as B gets larger. Analysis in the wideband regime in the following section will provide more insight into the impact of large bandwidth.

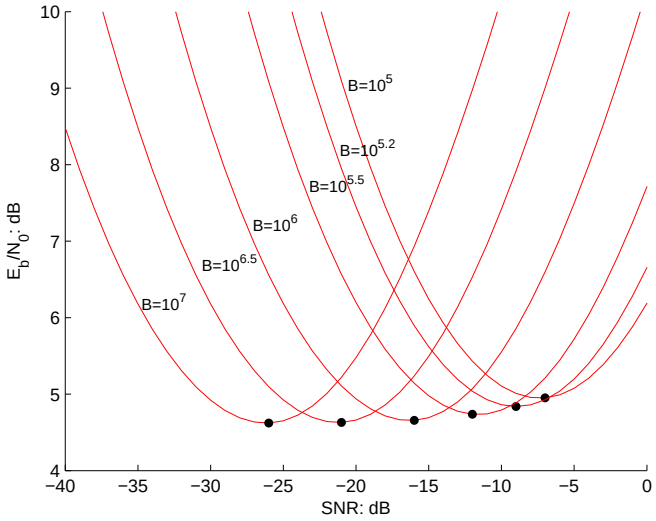

Figure 4.3: E_b/N_0 vs. SNR in the Rayleigh channel with $\mathbb{E}\{|h|^2\} = 1$. θ=0.01.

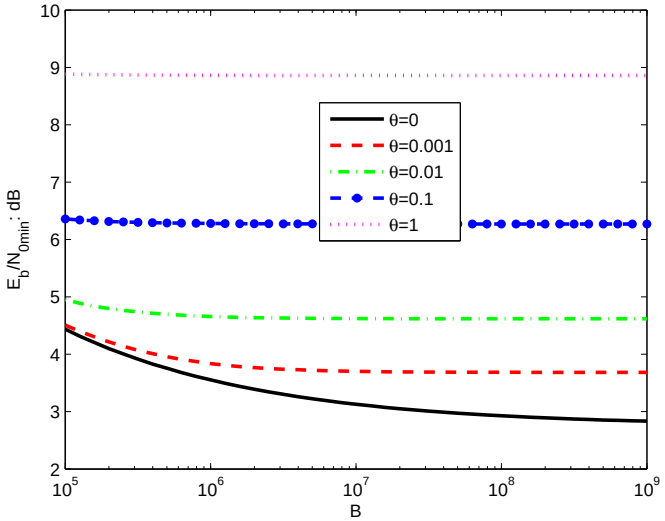

Figure 4.4: $\frac{E_b}{N_0}_{\min}$ vs. B in the Rayleigh channel with $\mathbb{E}\{|h|^2\} = 1$.

4.5 Energy Efficiency in the Wideband Regime

In this section, we consider the wideband regime in which the bandwidth is large. We assume that the average power $\bar{P}$ is kept constant. Note that as the bandwidth B increases, $\text{SNR} = \frac{\bar{P}}{N_0 B}$ approaches zero and we operate in the low-SNR regime.

4.5.1 Decomposing the Wideband Channel

In Section 4.1, we have described a flat fading channel model. However, flat fading assumption will not hold in the wideband regime as the bandwidth B increases without bound. On the other hand, if we decompose the wideband channel into N parallel subchannels, and suppose that each subchannel has a bandwidth that is equal to the coherence bandwidth, B_c, then we can assume that independent flat-fading is experienced in each subchannel. Note that we have $B = NB_c$. Similar to (4.1), the input-output relation in the k^{th} subchannel can be written as

$$y_k[i] = h_k[i]x_k[i] + n_k[i] \quad i = 1, 2, \ldots \quad \text{and} \quad k = 1, 2, \ldots, N. \tag{4.19}$$

The fading coefficients $\{h_k\}_{k=1}^N$ in different subchannels are assumed to be independent zero-mean Gaussian distributed with variances $\mathbb{E}\{|h_k|^2\} = \gamma_k$. The signal-to-noise ratio in the k^{th} subchannel is $\text{SNR}_k = \frac{\bar{P}_k}{N_0 B_c}$ where $\bar{P}_k$ denotes the power allocated to the k^{th} subchannel and we have $\sum_{k=1}^N \bar{P}_k = \bar{P}$ [3]. Over each subchannel, the same transmission strategy as described in Section 4.2.3 is employed. Therefore, the transmitter, not knowing the fading coefficients of the subchannels, sends the data over each subchannel at the fixed rate of r. Now, we can find that $C_{L,k}$ for each subchannel

[3] While not equipped with the knowledge of the instantaneous values of the fading coefficients, the transmitter is assumed to know the statistics of the fading coefficients, and possibly allocate different power levels to different subchannels with this knowledge.

is given by $\frac{TB_c-1}{T}\log_2\left(1+\text{SNR}_{\text{eff},k}|w|^2\right)$ bits/s, in which

$$\text{SNR}_{\text{eff},k}=\frac{\mathcal{E}_{s,k}\sigma^2_{h_{k,est}}}{\sigma^2_{h_{k,err}}\mathcal{E}_{s,k}+N_0} \tag{4.20}$$

where $\mathcal{E}_{s,k}=\frac{(1-\rho_k)T\bar{P}_k}{TB_c-1}$, $\mathcal{E}_{p,k}=\rho_k T\bar{P}_k$, $\sigma^2_{h_{k,err}}=\frac{\gamma_k N_0}{\gamma_k\mathcal{E}_{p,k}+N_0}$ and $\sigma^2_{h_{k,est}}=\frac{\gamma_k^2\mathcal{E}_{p,k}}{\gamma_k\mathcal{E}_{p,k}+N_0}$. Similarly as before, if $r<C_{L,k}$, then transmission over the k^{th} subchannel is successful. Otherwise, retransmission is required. Hence, we have ON and OFF states for each subchannel. On the other hand, for the transmission over N subchannels, we have a state model with $N+1$ states because we have overall the following $N+1$ possible total transmission rates: $\{0, rT, 2rT, \ldots, NrT\}$. For instance, if all N subchannels are in the OFF state simultaneously, the total rate is zero. If j out of N subchannels are in the ON state, then the rate is jrT. We note that such a decomposition strategy is also employed in [42] where the receiver is assumed to have perfect channel information. Although similar, this strategy is also discussed here for the sake of completeness.

Now, assume that the states are enumerated in the increasing order of the total transmission rates supported by them. Hence, in state $j\in\{1,\ldots,N+1\}$, the transmission rate is $(j-1)rT$. The probability of being in state $j\in\{1,\ldots,N+1\}$ is given by (4.22) in the next page, where $\mathcal{I}_{j-1}$ denotes a subset of the index set $\{1,\ldots,N\}$ with $j-1$ elements. The summation in (4.22) is over all such subsets. Also, in (4.22), $\mathcal{I}^c_{j-1}$ denotes the complement of the set $\mathcal{I}_{j-1}$, and $\alpha_k=\frac{2^{\frac{rT}{TB_c-1}}-1}{\text{SNR}_{\text{eff},k}}$. Note in the above formulation that, similarly as in Section 4.2.3, the probability of currently being in state j, i.e., q_j, does not depend on the state in the previous frame again due to the block-fading assumption. Moreover, the product form inside the summation in (4.21) is due to having noninteracting subchannels. If fading in different subchannels are correlated, q_j can be written as (4.23) in the next page, which, in general, depends on the joint distribution of $\{|w_1|^2,\ldots,|w_N|^2\}$.

$$
\begin{aligned}
q_j &= \Pr\{(j-1) \text{ subchannels out of } N \text{ subchannels are in the ON state}\} \\
&= \sum_{\mathcal{I}_{j-1}\subset\{1,\ldots,N\}} \left(\prod_{k\in\mathcal{I}_{j-1}} \Pr\{|w|^2 > \alpha_k\} \prod_{k\in\mathcal{I}_{j-1}^c} (1 - \Pr\{|w|^2 > \alpha_k\}) \right) \qquad (4.21) \\
&= \sum_{\mathcal{I}_{j-1}\subset\{1,\ldots,N\}} \left(\prod_{k\in\mathcal{I}_{j-1}} e^{-\alpha_k} \prod_{k\in\mathcal{I}_{j-1}^c} (1 - e^{-\alpha_k}) \right) \qquad (4.22)
\end{aligned}
$$

$$
q_j = \sum_{\mathcal{I}_{j-1}\subset\{1,\ldots,N\}} \left(\Pr\left\{ \left(\bigcap_{k\in\mathcal{I}_{j-1}} \{|w_k|^2 > \alpha_k\} \right) \bigcap \left(\bigcap_{k\in\mathcal{I}_{j-1}^c} \{|w_k|^2 \le \alpha_k\} \right) \right\} \right) \qquad (4.23)
$$

If, in addition to being independent, the fading coefficients h_k in different subchannels are identically distributed (i.e., the variances $\{\gamma_k\}_{k=1}^N$ are the same) and also if the total power is uniformly distributed over the subchannels and the fraction of energy, ρ_k, allocated to training in each subchannel is the same, then q_j in (4.22) simplifies and becomes a binomial probability:

$$
\begin{aligned}
q_j &= \binom{N}{j-1} \left(\Pr\{|w|^2 > \alpha\}\right)^{j-1} \left(1 - \Pr\{|w|^2 > \alpha\}\right)^{N-j+1} \\
&= \binom{N}{j-1} \left(e^{-\alpha}\right)^{j-1} \left(1 - e^{-\alpha}\right)^{N-j+1}. \qquad (4.24)
\end{aligned}
$$

Note that with equal power allocation, we have $\bar{P}_k = \frac{\bar{P}}{N}$ and therefore $\text{SNR}_k = \frac{\bar{P}_k}{N_0 B_c} = \frac{\bar{P}/N}{N_0 B/N} = \frac{\bar{P}}{N_0 B} = \text{SNR}$ which is equal to the original SNR used in (4.13). Since $\{\text{SNR}_{\text{eff},k}\}_{k=1}^N$ are also equal due to having equal ρ_k's, we have the same $\alpha = \frac{2^{\frac{rT}{TB_c-1}}-1}{\text{SNR}_{\text{eff}}}$ for each subchannel.

The effective capacity of this wideband channel model with N subchannels is given by the following result.

Corollary 4 *For the wideband channel with N parallel noninteracting subchannels each with bandwidth B_c and independent flat fading, the normalized effective capacity in bits/s/Hz is given by*

$$R_E(SNR,\theta) = \max_{\substack{r\geq 0 \\ \bar{P}_k\geq 0 \ s.t. \sum \bar{P}_k \leq \bar{P} \\ 0\leq \rho_k \leq 1 \ \forall k}} \left\{ -\frac{1}{\theta TB} \log_e \left(\sum_{j=1}^{N+1} q_j \, e^{-\theta(j-1)rT} \right) \right\} \tag{4.25}$$

where q_j is given in (4.22). If $\{h_k\}_{k=1}^N$ are identically distributed Gaussian random variables with zero mean and variance γ and the data and training energies are uniformly allocated over the subchannels, then the normalized effective capacity expression simplifies to

$$R_E(SNR,\theta) = \max_{\substack{r\geq 0 \\ 0\leq \rho \leq 1}} \left\{ -\frac{1}{\theta TB_c} \log_e \left(1 - e^{-\alpha}(1 - e^{-\theta Tr}) \right) \right\}. \tag{4.26}$$

where $\alpha = \frac{2^{\frac{rT}{TB_c-1}}-1}{SNR_{eff}}$ and $SNR_{eff} = \frac{\rho(1-\rho)\gamma^2T^2B_c^2SNR^2}{\rho\gamma TB_c(TB_c-2)SNR+\gamma TB_cSNR+TB_c-1}$, in which $SNR = \frac{\bar{P}}{N_0B} = \frac{\bar{P}}{NN_0B_c}$.

Proof: See Appendix B.

Remark: Although we concentrate on noninteracting subchannels, the effective capacity result in (4.25) is general and holds for the case in which the fading in different subchannels are correlated and q_j is given as in (4.23).

Remark: Corollary 4 shows that if the fading coefficients in different subchannels are i.i.d. and the data and training energies are uniformly allocated over the subchannels, then the effective capacity of a wideband channel has an expression similar to that in (4.10), which provides the effective capacity of a single channel experiencing flat fading. The only difference between (4.10) and (4.26) is that B is replaced in (4.26) by B_c, which is the bandwidth of each subchannel.

4.5.2 Rich and Sparse Multipath Fading Scenarios

After the characterization in Corollary 4, we henceforth limit our analysis to the case in which the effective capacity is given by (4.26) because optimization over the power allocation schemes and obtaining closed-form expressions are in general difficult tasks in the wideband regime in which the number of subchannels is potentially high. Under these assumptions, we investigate two scenarios:

1. *Rich multipath fading*: In this case, we assume that the number of independent resolvable paths increases linearly with the bandwidth. This in turn implies that as the bandwidth B increases, the number of noninteracting subchannels N increases while B_c stays fixed.

2. *Sparse multipath fading*: In this case, we assume that the number of independent resolvable paths increases *at most sublinearly* with the bandwidth. This assumption implies the coherence bandwidth $B_c = \frac{B}{N}$ increases with increasing bandwidth B [57], [58]. We can identify two subcases:

    a) If the number of resolvable paths remains bounded in the wideband regime (as considered for instance in [59]), then N remains bounded while B_c increases linearly with B.

    b) If the number of resolvable paths increases but only sublinearly with B, then both N and B_c grow without bound with B.

We first consider scenario (1) where rich multipath fading is assumed. In this case, as B increases, the signal-to-noise ratio $\text{SNR} = \frac{\bar{P}}{N_0 B} = \frac{\bar{P}}{N N_0 B_c}$ approaches zero while B_c stays fixed. From these facts and the similarity of the formulations in (4.10) and (4.26), we immediately conclude that the wideband regime analysis of the rich multipath case is the same as the low-power regime analysis conducted in Section 4.4. Therefore, as

$B \to \infty$ in the rich multipath fading scenario, we have $\frac{E_b}{N_0}\Big|_{\mathsf{R}_E=0} = \lim_{\mathrm{SNR}\to 0} \frac{E_b}{N_0} = \infty$ for all $\theta \geq 0$. Therefore, the minimum bit energy is attained at a high but finite bandwidth level that can be identified through numerical analysis. If the bandwidth is further increased, a penalty in energy efficiency starts to be experienced due to increased uncertainty. Note that we have high diversity in rich multipath fading as the number of noninteracting subchannels increase linearly with bandwidth. On the other hand, since independent fading coefficients are only imperfectly known and moreover the receiver's ability to estimate the subchannels diminishes with decreasing SNR, we have high uncertainty as well. Hence, uncertainty becomes the more dominant factor and extreme energy-inefficiency is experienced in the limit as $B \to \infty$.

Next, we analyze the performance in the scenario of sparse multipath fading. We note that the authors in [57] and [58], motivated by the recent measurement studies in the ultrawideband regime, considered sparse multipath fading channels and analyzed the performance under channel uncertainty, employing the Shannon capacity formulation as the performance metric. We in this chapter consider channel uncertainty and queueing constraints jointly and use the effective capacity to identify the performance. We first consider scenario (2a) where the the number of subchannels N remains bounded and the degrees of freedom are limited. The following result provides the expressions for the bit energy at zero spectral efficiency and the wideband slope, and characterize the spectral efficiency-bit energy tradeoff in the wideband regime when N is fixed and B_c grows linearly with B. It is shown that the bit energy required at zero spectral efficiency is indeed the minimum bit energy.

Theorem 11 *For sparse multipath fading channel with bounded number of independent resolvable paths, the minimum bit energy and wideband slope in the wideband*

regime are given by

$$\frac{E_b}{N_0}_{min} = \frac{E_b}{N_0}\bigg|_{R_E=0} = \lim_{SNR\to 0} \frac{E_b}{N_0} = \frac{-\delta}{\log_e \xi} \quad \textit{and} \tag{4.27}$$

$$\mathcal{S}_0 = \frac{\xi \log_e^2 \xi \log_e 2}{\theta T \alpha_{opt}^* (1-\xi)\left(\frac{1}{T}\left(\sqrt{1+\frac{\gamma \bar{P} T}{NN_0}}-1\right)+\frac{\varphi \alpha_{opt}^*}{2}\right)}, \tag{4.28}$$

respectively, where $\delta = \frac{\theta T \bar{P}}{NN_0}$, $\xi = 1 - e^{-\alpha_{opt}^*}(1 - e^{-\frac{\theta T \varphi \alpha_{opt}^*}{\log_e 2}})$, *and*

$$\varphi = \frac{\gamma \bar{P}}{NN_0}\left(\sqrt{1+\frac{NN_0}{\gamma \bar{P} T}} - \sqrt{\frac{NN_0}{\gamma \bar{P} T}}\right)^2$$

. α_{opt}^* *is defined as* $\alpha_{opt}^* = \lim_{\zeta\to 0} \alpha_{opt}$ *and* α_{opt}^* *satisfies*

$$\alpha_{opt}^* = \frac{\log_e 2}{\theta T \varphi} \log_e \left(1 + \frac{\theta T \varphi}{\log_e 2}\right). \tag{4.29}$$

Above, we define $\zeta = \frac{1}{B_c}$.

Proof: See Appendix G.

Remark: We note that the minimum bit energy in the sparse multipath case with bounded degrees of freedom is achieved as $B \to \infty$ and hence as SNR $\to 0$. This is in stark contrast to the results in the low-power regime and rich multipath cases in which the bit energy requirements grow without bound as SNR vanishes. This is due to the fact that in sparse fading with bounded number of independent resolvable paths, uncertainty does not grow without bound because the number of subchannels N is kept fixed as $B \to \infty$.

Remark: Theorem 11, through the minimum bit energy and wideband slope expressions, quantifies the bit energy requirements in the wideband regime when the system is operating subject to both statistical QoS constraints specified by θ and

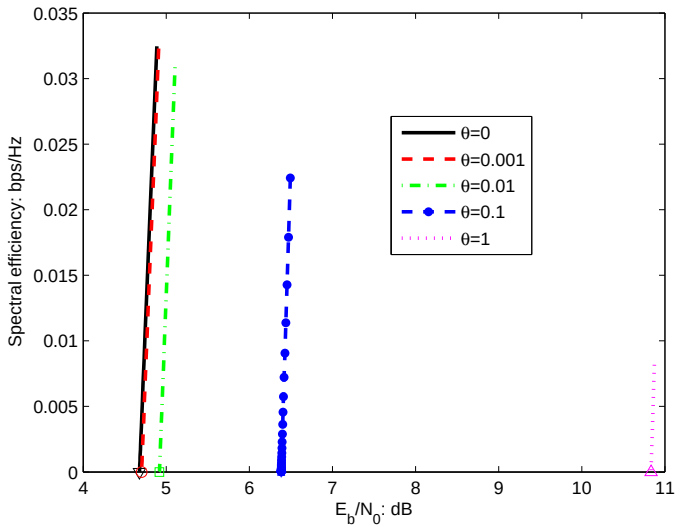

Figure 4.5: Spectral efficiency vs. E_b/N_0 in the Rayleigh channel with $E\{|h|^2\} = \gamma = 1$. $\bar{P}/NN_0 = 10^4$.

channel uncertainty. Note that both $\frac{E_b}{N_0}_{\min}$ and $\mathcal{S}_0$ depend on θ through δ and ξ. As will be observed in the numerical results, $\frac{E_b}{N_0}_{\min}$ and the bit energy requirements at nonzero spectral efficiency values generally increase with increasing θ. Moreover, when compared with the results in Section 4.4, it will be seen that sparse multipath fading and having a bounded number of subchannels incur energy penalty whether there are QoS constraints or not ($\theta = 0$), which is in stark contrast with previous results when there is perfect CSI at the receiver [42].

After having obtained analytical expressions for the minimum bit energy and wideband slope, we now provide numerical results. Fig. 4.5 plots the spectral efficiency–bit energy curve in the Rayleigh channel for different θ values. In the figure, we assume that $\bar{P}/(NN_0) = 10^4$. As predicted, the minimum bit energies are obtained as SNR and hence the spectral efficiency approach zero. $\frac{E_b}{N_0}_{\min}$ are computed to be equal to $\{4.6776, 4.7029, 4.9177, 6.3828, 10.8333\}$ dB for $\theta = \{0, 0.001, 0.01, 0.1, 1\}$, respectively. Moreover, the wideband slopes are $\mathcal{S}_0 = \{0.4720, 0.4749, 0.4978, 0.6151, 0.6061\}$

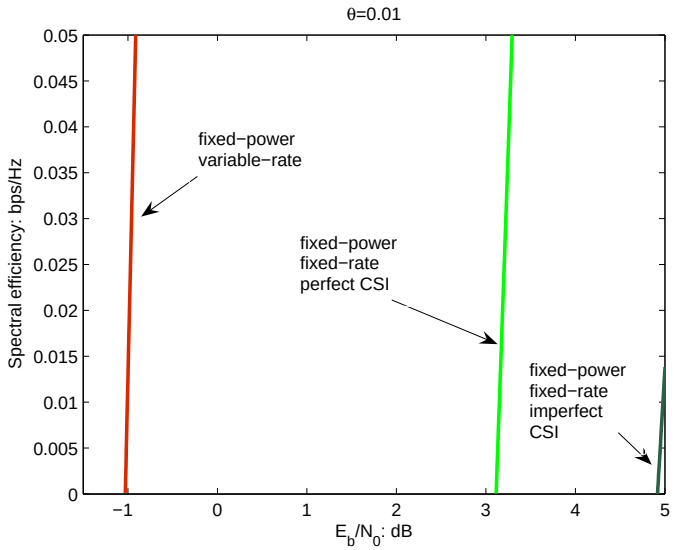

Figure 4.6: Comparison of spectral efficiency; $\bar{P}/NN_0 = 10^4$, $\theta = 0.01$, and $E\{|h|^2\} = \gamma = 1$.

for the same set of θ values. As can also be seen in the result of Theorem 11, the minimum bit energy and wideband slope in general depend on θ. In Fig. 4.5, we note that the bit energy requirements (including the minimum bit energy) increase with increasing θ, illustrating the energy costs of stringent queueing constraints. Finally, in this chapter, we have considered fixed-rate/fixed-power transmissions over imperfectly-known channels. In Fig. 4.6, we compare the performance of this system with those in which the channel is perfectly-known and fixed- or variable-rate transmission is employed. The latter models have been studied in Chapters 2 and 3. This figure demonstrates the energy costs of not knowing the channel and sending the information at fixed-rate.

We finally consider the sparse multipath fading scenario (2b) in which the number of subchannels N increases but only sublinearly with increasing bandwidth. Note that in this case, the bit energy required as $B \to \infty$ can be obtained by letting N in the result of Theorem 11, where N is assumed to be fixed, go to infinity.

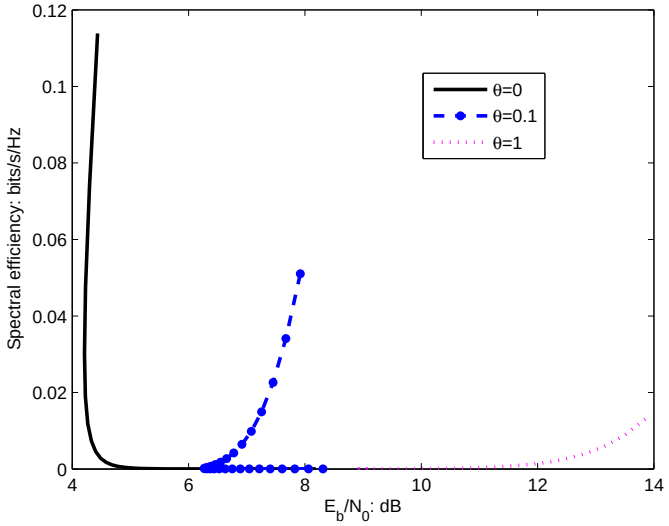

Figure 4.7: Spectral efficiency vs. E_b/N_0 in the Rayleigh channel with $E\{|h|^2\} = \gamma = 1$. $\bar{P}/NN_0 = 10^4$.

Corollary 5 *In the wideband regime, if the number of subchannels N increases sub-linearly with B, then the bit energy required in the limit as $B \to \infty$ is*

$$\left.\frac{E_b}{N_0}\right|_{R_E=0} = \infty \tag{4.30}$$

Remark: As N increases, each subchannel is allocated less power and operate in the low-power regime. Therefore, it is not surprising that we obtain the same bit energy result as in the low-power regime. Additionally, since the number of sub-channels N increases without bound, uncertainty in the wideband channel increases as well. Hence, similarly as in rich multipath fading, extreme energy-inefficiency is experienced as $B \to \infty$.

Fig. 4.7 confirms the theoretical results. In this figure, we observe that the bit energy requirements initially decrease with decreasing spectral efficiency. However, below a certain spectral efficiency level, $\frac{E_b}{N_0}$ starts growing without bound for all $\theta \geq 0$.

4.6 Conclusion

In this chapter, we have analyzed the energy efficiency of fixed-rate wireless transmissions for the communication scenario in which queueing constraints are present and the channel coefficients are estimated imperfectly by the receiver with the aid of training symbols. We have considered the effective capacity as a measure of the maximum throughput under statistical QoS constraints. We have identified the optimal fraction of power allocated to training and shown that this optimal fraction does not depend on the QoS exponent θ and the transmission rate. In particular, we have investigated the spectral efficiency–bit energy tradeoff in the low-power and wideband regimes. We have quantified the increased energy requirements in the presence of QoS constraints in the low-power and wideband regimes, and identified the impact upon the energy efficiency of channel uncertainty and multipath sparsity and richness. The key conclusions of this chapter on energy efficiency are the following:

1. Having very low power per degree of freedom has a detrimental impact on energy efficiency. Indeed, the bit energy requirements grow without bound as the power per degree of freedom vanishes by either letting the power in a narrowband channel become small or increasing the bandwidth and having the power per subchannel in a wideband scenario diminish. This is tightly linked to the fact that the system's ability to reliably estimate the channel conditions decreases as power gets small.

2. Although operating at low power levels or at wide bandwidths improves the energy efficiency, care should be exercised under channel uncertainty. In the low-power regime, the minimum bit energy is achieved at a certain small but non-zero power level. Unless sparse multipath fading with bounded number of independent resolvable paths is experienced, the minimum bit energy in the

wideband regime is attained at a large but finite bandwidth value. These critical power and bandwidth levels depend in general on the QoS constraints and can be obtained through numerical analysis.

If the power decreases or bandwidth increases beyond these minimum-bit-energy-achieving levels, energy efficiency starts degrading. These results have significant practical implications on wireless systems.

3. In the presence of QoS constraints and channel uncertainty, diversity in the frequency domain acts as a double-edged sword. Increasing the bandwidth and the number of noninteracting subchannels initially improves the energy efficiency by decreasing the required bit energy. This initial increase in the diversity is also beneficial in satisfying the QoS constraints. However, if the number of noninteracting subchannels increases without bound, the bit energy values eventually start growing without bound as well. Hence, beyond a certain threshold, the benefits of the presence of large number of subchannels are outweighed by the increased channel uncertainty due to the imperfect-knowledge of the conditions in these channels.

   Note that such a behavior is not exhibited if the number of subchannels remains bounded.

4. In general, required bit energy values increase as the QoS constraints become more stringent. The analysis in this chapter enable us to quantify these increases in the energy requirements.

Chapter 5

Power and Rate Control for Multiple-Access Fading Channels

In this chapter, we consider the scenario in which both the transmitters and the receiver have perfect channel side information (CSI). First, assuming that no power control is employed in the transmission, we characterize the rate regions for both superposition transmission strategies and TDMA. Unlike the results obtained in [22] and [26], varying the decoding order with respect to the channel states is shown to significantly increase the achievable rate region (i.e., *throughput region*) under QoS constraints. Also, it is demonstrated that time sharing strategies among the vertex of the rate regions can no longer achieve the boundary surface of the throughput region. Additionally, we show that if we take the sum-rate throughput, or the sum effective capacity, as the performance metric, TDMA can in certain cases even achieve better performance than superposition coding when a fixed decoding order is employed at the receiver. Next, we incorporate power control policies into the model. For this case, we first obtain closed-form expressions for the optimal power control policies under the assumption that the decoding order is fixed at the receiver side. When

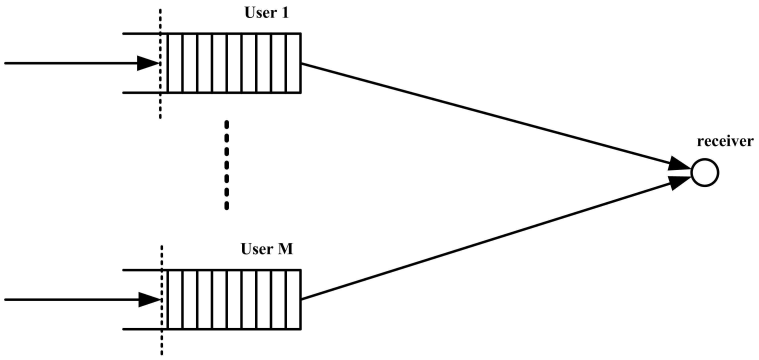

Figure 5.1: The system model.

the decoding order is variable, we identify the conditions the optimal power control policies should satisfy. We also describe an algorithm to determine such policies.

5.1 System Model and MAC Capacity Region

As shown in Figure 5.1, we consider an uplink scenario where M users with individual power and buffer constraints (i.e., QoS constraints) communicate with a single receiver. It is assumed that the transmitters generate data sequences which are divided into frames of duration T. These data frames are initially stored in the buffers before they are transmitted over the wireless channel. The discrete-time signal at the receiver in the i^{th} symbol duration is given by

$$Y[i] = \sum_{j=1}^{M} h_j[i]X_j[i] + n[i], \quad i = 1, 2, \ldots \tag{5.1}$$

where M is the number of users, $X_j[i]$ and $h_j[i]$ denote the complex-valued channel input and the fading coefficient of the jth user, respectively. We assume that $\{h_j[i]\}$'s are jointly stationary and ergodic discrete-time processes, and we denote the magnitude-square of the fading coefficients by $z_j[i] = |h_j[i]|^2$. Above, $n[i]$ is a

zero-mean, circularly symmetric, complex Gaussian random variable with variance $\mathbb{E}\{|n[i]|^2\} = N_0$. The additive Gaussian noise samples $\{n[i]\}$ are assumed to form an independent and identically distributed (i.i.d.) sequence. Finally, $Y[i]$ denotes the received signal.

The channel input of user j is subject to an average energy constraint $\mathbb{E}\{|x_j[i]|^2\} \leq \bar{P}_j/B$ for all j, where B is the bandwidth available in the system. Assuming that the symbol rate is B complex symbols per second, we can see that this formulation indicates that user j is subject to an average power constraint of $\bar{P}_j$. With these definitions, the average transmitted signal to noise ratio of user j is $\text{SNR}_j = \frac{\bar{P}_j}{N_0 B}$. Now, if we denote $P_j[\mathbf{z}]$ as the instantaneous transmit power as a function of the fading states $\mathbf{z} = (z_1, \cdots, z_M)$, the instantaneous transmitted SNR level becomes $\mu_j[\mathbf{z}] = \frac{P_j[\mathbf{z}]}{N_0 B}$. Then, the average power constraint is equivalent to the average SNR constraint $\mathbb{E}\{\mu_j[\mathbf{z}]\} \leq \text{SNR}_j$ for user j.

5.1.1 Fixed Power and Variable Rate

First, we consider the case in which the transmitters operate at fixed power and hence do not employ any power adaptation policies. The capacity region of this channel is given by [22], [23]

$$\mathcal{R}_{\text{MAC}} = \left\{ (R_{avg,1}, \ldots, R_{avg,M}) : \right.$$
$$\mathbf{R}_{avg}(S) \leq B\mathbb{E}_{\mathbf{z}} \left\{ \log_2 \left(1 + \sum_{j \in S} \text{SNR}_j z_j \right) \right\} ,$$
$$\left. \forall S \subset \{1, \ldots, M\} \right\} \tag{5.2}$$

where $\mathbf{R}_{avg}(S) = \sum_{i \in S} R_{avg,i}$. As well-known, there are $M!$ vertices of the polyhedron defined in (5.2). The vertex $\mathbf{R}_{avg,\pi} = \left(R_{avg,\pi(1)}, \cdots, R_{avg,\pi(M)}\right)$ corresponds to a permutation π, or the successive decoding order at the receiver, i.e., users are decoded in the order given by $\pi(1), \cdots, \pi(M)$. This vertex is specified by the average rates

$$R_{avg,\pi(k)} = B\mathbb{E}_{\mathbf{z}} \left\{ \log_2 \left(1 + \frac{\mathrm{SNR}_{\pi(k)} z_{\pi(k)}}{1 + \sum_{i=k+1}^{M} \mathrm{SNR}_{\pi(i)} z_{\pi(i)}} \right) \right\} \tag{5.3}$$

in bits/s for $k = 1, \cdots, M$. With this characterization, we see that for the given decoding order π, the maximum instantaneous service rate for user $\pi(k)$ is

$$R_{\pi(k)} = B \log_2 \left(1 + \frac{\mathrm{SNR}_{\pi(k)} z_{\pi(k)}}{1 + \sum_{i=k+1}^{M} \mathrm{SNR}_{\pi(i)} z_{\pi(i)}} \right) \text{ bits/s.} \tag{5.4}$$

Time sharing among these $M!$ permutations of decoding orders yields any point on the boundary surface of $\mathcal{R}_{\mathrm{MAC}}$ [1]. As also discussed in [26], it can be easily verified that varying the decoding order according to the channel states does not provide any improvement on the capacity region.

5.1.2 Variable Power and Variable Rate

Now, we suppose that dynamic power and rate allocation is performed according to time-variations in the channels. For a given set of power allocation policies $\mathcal{U} = \{\mu_1, \cdots, \mu_M\}$, where $\mu_j \geq 0$ is the power control policy of the jth user, the achievable rate region is described by [23]

$$\mathcal{R}(\mathcal{U}) = \left\{ \mathbf{R}_{avg} : \mathbf{R}_{avg}(S) \leq \mathbb{E}_{\mathbf{z}} \left\{ B \log_2 \left(1 + \sum_{j \in S} \mu_j(\mathbf{z}) z_j \right) \right\}, \right. \\ \left. \forall S \subset \{1, \cdots, M\} \right\}. \tag{5.5}$$

For a given decoding order at the receiver, the individual average and instantaneous rates of the users can be obtained similar to (5.3) and (5.4), respectively, with SNR replaced by μ. The capacity region is given by

$$\mathcal{R}_{\text{MAC}} = \bigcup_{\mathcal{U}\in\mathcal{F}} \mathcal{R}(\mathcal{U}) \tag{5.6}$$

where $\mathcal{F}$ is the set of all feasible power control policies that satisfy the average power constraint

$$\mathcal{F} \equiv \{\mathcal{U} : \mathbb{E}_{\mathbf{z}}\{\mu_j(\mathbf{z})\} \leq \text{SNR}_j, \mu_j \geq 0, \forall j\}. \tag{5.7}$$

5.1.3 TDMA

For simplicity, we assume that the time division strategy is fixed prior to transmission. Let δ_j denote the fraction of time allocated to user j. Note that we have $\sum_{j=1}^{M} \delta_j = 1$. In each frame, each user occupies the entire bandwidth to transmit the signal in the corresponding fraction of time. Then, the instantaneous service rate for user j in each frame is given by

$$R_j(\text{SNR}_j) = B\log_2\left(1 + \frac{\text{SNR}_j}{\delta_j} z_j\right) \text{ bits/s} \tag{5.8}$$

Above, note that user j is assumed to transmit with the higher average power of $\bar{P}_j/\delta_j$ in the allocated δ_j fraction of the time.

5.2 Throughput Region

At this point, it is also important to note that the transmission strategies (such as superposition coding schemes, time-division multiple-access methods, and power con-

trol policies) and reception strategies (such as the successive decoding order) will henceforth be designed and analyzed as functions of the fading states and the QoS exponent θ. Hence, our transmission and reception policies take into account the statistical queueing constraints through the QoS exponents but not the actual queue lengths and states. We note that the authors in [62], [63], and [64] have recently studied queue-length based policies in the context of wireless scheduling in broadcast scenarios. In these works, only one user at a time is served by the transmitter in a downlink model. Shakkottai in [62] investigated the effective capacity achieved by a greedy scheduling rule that picks the user with the highest channel rate and a max-queue rule that picks the user with the largest product of the queue length and the channel rate. Even though an i.i.d. channel model (akin to our block-fading assumption) is considered in this work, it is described that the main difficulty in the analysis of queue-length based policies arises from the fact that these policies statistically couple the rates allocated to various users across time. Therefore, due to correlation over time, the effective capacity formula in (1.1) cannot be simplified to that in (1.3). In such cases, effective capacity cannot be computed directly and certain technical difficulties are encountered. In particular, the techniques of sample path large deviations and calculus of variations are needed to determine the performance. In [62], these approaches are applied to relatively simple scenarios with two users, each of which experiences a two-state (ON-OFF) channel. More recently, using the sample-path large deviation principle, Venkataramanan and Lin in [63] studied wireless scheduling algorithms that maximize asymptotic decay rate of the queue-overflow probability in a more general downlink scenario with N users and M possible channel states.

In this chapter, we consider more complex channel models with continuous fading and more sophisticated transmission strategies such as superposition coding (rather than orthogonal transmissions) and power control techniques in a multiple-access

scenario. By addressing only the statistical queueing constraints, we formulate a tractable problem for more practically appealing system models. At the same time, we note that queue-length based policies have the potential to attain a higher effective capacity than those achieved by greedy policies that take into account only the channel states (see e.g., [62, Fig. 3]). Hence, from this perspective, our results, which incorporate the channel states and the QoS exponents but not the actual queue states in the transmission and reception, can be regarded as baselines with which the performances of queue-length based policies can be compared.

Suppose that $\Theta = (\theta_1, \cdots, \theta_M)$ is a vector composed of the QoS constraints of M users. Let $\mathsf{C}(\Theta) = (\mathsf{C}_1(\theta_1), \cdots, \mathsf{C}_M(\theta_M))$ denote the vector of the normalized effective capacities. We first have the following characterization.

Definition 1 *The* effective throughput region *is described as*

$$\mathcal{C}_{MAC}(\Theta, \boldsymbol{SNR}) = \bigcup_{\substack{\mathbf{R} \\ s.t.\ \mathbb{E}\{\mathbf{R}\} \in \mathcal{R}_{MAC}}} \left\{ \mathsf{C}(\Theta) \geq \mathbf{0} : \mathsf{C}_j(\theta_j) \leq -\frac{1}{\theta_j TB} \log_e \mathbb{E}_{\mathbf{z}} \left\{ e^{-\theta T R_j} \right\} \right\} \tag{5.9}$$

where $\mathbf{R} = \{R_1, R_2, \cdots, R_M\}$ *represents the vector composed of the instantaneous transmission (or equivalently service) rates of* M *users. Note that the union is over the distributions of the vector* $\mathbf{R}$ *such that the expected value* $\mathbb{E}\{\mathbf{R}\}$ *lies in the MAC capacity region.*

Remark: The *throughput region* given in Definition 1 represents the set of all vectors of constant arrival rates $\mathsf{C}(\theta)$ that can be supported in the fading multiple access channel in the presence QoS constraints specified by $\Theta = (\theta_1, \cdots, \theta_M)$. Since reliable communication is considered, the arrival rates are supported by instantaneous service rates whose expected values are in the MAC capacity region. For instance, in

the absence of power control, the maximum instantaneous service rates for a given decoding order are given by (5.4).

Using the convexity of the MAC capacity region $\mathcal{R}_{\mathrm{MAC}}$, we obtain the following preliminary result on the effective throughput region defined in (5.9).

Theorem 12 *The* throughput region $\mathcal{C}_{MAC}(\Theta, \boldsymbol{SNR})$ *is convex.*

Proof: Let the vectors $\mathsf{C}(\Theta)$ and $\mathsf{C}'(\Theta)$ belong to $\mathcal{C}_{\mathrm{MAC}}(\Theta, \mathbf{SNR})$. Then, there exist some rate vectors $\mathbf{R}$ and $\mathbf{R}'$ for $\mathsf{C}(\Theta)$ and $\mathsf{C}'(\Theta)$, respectively, such that $\mathbb{E}\{\mathbf{R}\}$ and $\mathbb{E}\{\mathbf{R}'\}$ are in the MAC capacity region. By a time sharing strategy, for any $\alpha \in (0,1)$, we know from the convexity of the MAC capacity region that $\mathbb{E}\{\alpha\mathbf{R} + (1-\alpha)\mathbf{R}'\} \in \mathcal{R}_{\mathrm{MAC}}$. Now, we can write

$$
\begin{aligned}
&\alpha\mathsf{C}(\Theta) + (1-\alpha)\mathsf{C}'(\Theta) \\
&\leq -\frac{1}{\Theta TB}\log_e\left(\mathbb{E}\left\{e^{-\Theta T\mathbf{R}}\right\}\right)^{\alpha}\left(\mathbb{E}\left\{e^{-\Theta T\mathbf{R}'}\right\}\right)^{1-\alpha} && (5.10)\\
&= -\frac{1}{\Theta TB}\log_e\left(\mathbb{E}\left\{\left(e^{-\Theta T\alpha\mathbf{R}}\right)^{\frac{1}{\alpha}}\right\}\right)^{\alpha} \\
&\qquad\qquad \times\left(\mathbb{E}\left\{\left(e^{-\Theta T(1-\alpha)\mathbf{R}'}\right)^{\frac{1}{1-\alpha}}\right\}\right)^{1-\alpha} && (5.11)\\
&\leq -\frac{1}{\Theta TB}\log_e\mathbb{E}\left\{e^{-\Theta T(\alpha\mathbf{R}+(1-\alpha)\mathbf{R}')}\right\}. && (5.12)
\end{aligned}
$$

Above, in (5.10) through (5.12), all operations, including the logarithm and exponential functions and expectations, are component-wise operations. For instance, the expression in (5.10) denotes a vector whose components are

$$
\left\{\frac{1}{\theta_j TB}\log_e\left(\mathbb{E}\left\{e^{-\theta_j TR_j}\right\}\right)^{\alpha}\left(\mathbb{E}\left\{e^{-\theta TR'_j}\right\}\right)^{1-\alpha}\right\}_{j=1}^{M}
$$

. Similarly, the inequalities in (5.10) and (5.12) are component-wise inequalities. The inequality in (5.10) follows from the definition in (5.9). Moreover, (5.12) follows from

Hölder's inequality and leads to the conclusion that $\alpha\mathsf{C} + (1-\alpha)\mathsf{C}'$ still lies in the *throughput region*, proving the convexity result. □

We are interested in the boundary of the region. Now that $\mathcal{C}_{\text{MAC}}(\Theta, \mathbf{SNR})$ is convex, we can characterize the boundary surface by considering the following optimization problem [23]:

$$\max \lambda \cdot \mathsf{C}(\Theta) \quad \text{subject to: } \mathsf{C}(\Theta) \in \mathcal{C}_{\text{MAC}}(\Theta, \mathbf{SNR}). \tag{5.13}$$

for all priority vectors $\lambda = (\lambda_1, \cdots, \lambda_M)$ in $\Re_+^M$ with $\sum_{j=1}^M \lambda_j = 1$.

5.3 Transmissions without Power Control

In this section, we assume that the signals are transmitted at a constant power level in each frame and hence power adaptation with respect to the fading states is not performed. Under this assumption, we initially consider the scenario in which the receiver decodes the users in a fixed order. Subsequently, we analyze the case of variable decoding order.

5.3.1 Fixed Decoding Order

We first assume that the receiver decodes the users in a fixed order in each frame. Hence, the decoding order does not change with respect to the realizations of the fading coefficients. If a single decoding order is used in the frame, it is obvious that only the vertices of the boundary region can be achieved. We consider a slightly more general case in which time sharing technique is employed in each frame among different decoding orders. Note that the time sharing strategy is also independent of the channel states and hence is fixed in different blocks. We denote the fraction of

time allocated to decoding order π_m as τ_m. Naturally, the fractions of time satisfy $\tau_m \geq 0$ and $\sum_{m=1}^{M!} \tau_m = 1$. Varying the values of τ_m enables us to characterize the throughput region. Under these assumptions, the effective capacity for each user on the boundary surface is

$$\mathsf{C}_j(\theta_j) = -\frac{1}{\theta_j TB} \log_e \mathbb{E}_{\mathbf{z}} \left\{ e^{-\theta_j T \sum_{m=1}^{M!} \tau_m R_{\pi_m^{-1}(j)}} \right\} \tag{5.14}$$

where $R_{\pi_m^{-1}(j)}$ represents the maximal instantaneous service rate of user j at a given decoding order π_m, which is given by

$$R_{\pi_m^{-1}(j)} = B \log_2 \left(1 + \frac{\mathrm{SNR}_j z_j}{1 + \sum_{\pi_m^{-1}(i) > \pi_m^{-1}(j)} \mathrm{SNR}_i z_i} \right) \tag{5.15}$$

where π_m^{-1} is the inverse trace function of π_m.

Remark: Note that $R_{\pi_m^{-1}(j)}$ is the maximum instantaneous service rate achieved with superposition coding and a particular decoding order. Hence, the corresponding effective capacities characterize the throughput achieved with this strategy in the presence of QoS constraints. Note also that $R_{\pi_m^{-1}(j)}$, which represents the information-theoretic limit for instantaneous rates, can be approached if codes with large blocklengths are employed. Therefore, in order to have operational significance in the results, we assume throughout the chapter that the number of symbols TB in a frame duration of T seconds is sufficiently large. If TB is relatively small, rates attained with finite blocklength channel codes in the presence of possible decoding errors should be considered as addressed in [65] and [47].

Remark: Throughout the rest of the chapter, we generally specify the effective capacity values on the boundary surface for simplicity and brevity. Effective capacity regions can immediately be specified using these boundary points. For instance, the

effective capacity (or equivalently throughput) region for superposition coding and fixed decoding order is

$$\bigcup_{\{\tau_m\}} \Bigg\{ \mathsf{C}(\Theta) \geq \mathbf{0} : \mathsf{C}_j(\theta_j) \leq -\frac{1}{\theta_j TB} \log_e \mathbb{E}_{\mathbf{z}} \left\{ e^{-\theta_j T \sum_{m=1}^{M!} \tau_m R_{\pi_m^{-1}(j)}} \right\} \Bigg\} \tag{5.16}$$

where the union is over different time allocation strategies.

Next, for comparison, we consider the TDMA case in which we also have similar time allocation strategies but only one user transmits in its specific fraction of time. We first have the following definition.

Definition 2 *The* throughput region *for TDMA can be seen as the achievable vectors of arrival rates with each component bounded by the effective capacity obtained when the instantaneous service rate is given by (5.8). More specifically, the maximum effective capacity for user j is*

$$\mathsf{C}_j^{TD}(\theta_j) = -\frac{1}{\theta_j TB} \log_e \mathbb{E} \left\{ e^{-\delta_j \theta_j TB \log_2 \left(1 + \frac{SNR_j}{\delta_j} z_j\right)} \right\} \tag{5.17}$$

where δ_j is the fraction of time allocated to user j, and $0 \leq \delta_j \leq 1$. We again assume that $\delta_j TB$ is sufficiently large so that the expression in (5.8) is a realistic representation of the service rate.

An immediate result can be obtained as follows:

Theorem 13 *The* throughput region *for TDMA is convex.*

Proof: Note that the points on the boundary surface is given in (5.17). Consider the function $f(\delta) = -\delta\theta TB \log_2 \left(1 + \frac{\mathrm{SNR}}{\delta} z\right)$. It can be easily verified that $f(\delta)$ is

$$\frac{\partial \mathcal{J}}{\partial \delta_j} = \lambda_j \frac{\mathbb{E}\left\{ e^{-\delta_j \theta_j TB \log_2\left(1+\frac{\mathrm{SNR}_j}{\delta_j} z_j\right)} \left(\log_2(1+\frac{\mathrm{SNR}_j}{\delta_j} z_j) - \frac{\frac{\mathrm{SNR}_j}{\delta_j} z_j}{1+\frac{\mathrm{SNR}_j}{\delta_j} z_j} \log_2 e \right) \right\}}{\mathbb{E}\left\{ e^{-\delta_j \theta_j TB \log_2\left(1+\frac{\mathrm{SNR}_j}{\delta_j} z_j\right)} \right\}} - \kappa = 0 \tag{5.19}$$

a convex function in δ. Then, $e^{f(\delta)}$ is a log-convex function. Since weighted non-negative sum preserves the log-convexity [53, Section 3.5], we know that $\mathbb{E}_z\{e^{f(\delta)}\}$ is log-convex. Then $-\frac{1}{\theta TB} \log_e \mathbb{E}\{e^{-\delta\theta TB \log_2\left(1+\frac{\mathrm{SNR}}{\delta} z\right)}\}$ is a concave function in δ. Hence, we immediately see that the *throughput region* for TDMA is convex. □

The optimal time allocation policy that maximizes the weighted sum can be obtained through the optimization problem

$$\max_{\{\delta_j\}} \sum_{j=1}^{M} -\frac{\lambda_j}{\theta_j TB} \log_e \mathbb{E}\left\{ e^{-\delta_j \theta_j TB \log_2\left(1+\frac{\mathrm{SNR}_j}{\delta_j} z_j\right)} \right\}, \tag{5.18}$$
$$\text{s.t. } \sum_{j=1}^{M} \delta_j = 1, \delta_j \geq 0.$$

The objective function in the above problem is concave, and we can use the Lagrangian maximization approach. Taking the derivative of the Lagrangian function with respect to δ_j, we obtain, for each user, the optimality condition given in (5.19) at the top of this page, where κ is the Lagrange multiplier whose value is chosen to satisfy the constraint $\sum_{j=1}^{M} \delta_j = 1$. If the optimal value of δ_j turns out to be negative, then the optimal value of δ_j should be 0. When $\lambda_1 = \lambda_2 = \cdots = \lambda_M$, the obtained values of $\{\delta_j\}$ are the ones that achieve the maximal sum-rate throughput, i.e., the sum of the effective capacities of the users. Although obtaining closed-form solutions is unlikely, the maximization problem in (5.18) can be easily solved numerically using

convex optimization tools. Numerical results are provided in Section 5.3.3.

5.3.2 Variable Decoding Order

We now study the case in which the receiver varies the decoding order with respect to the fading states $\mathbf{z} = (z_1, \ldots, z_M)$. In its most general form, we assume that the receiver, for each fading state $\mathbf{z}$, employs a time sharing of the decoding orders in which the fraction of time allocated to decoding order π_m is $\tau_m(\mathbf{z})$ for $m = 1, \ldots M!$. Hence, for each fading state $\mathbf{z}$, the receiver now has the freedom to use possibly a different decoding order or a different time sharing of multiple decoding orders. For a given time sharing policy $\{\tau_m(\mathbf{z})\}_{m=1}^{M!}$, the effective capacity of user j is

$$\mathsf{C}_j(\theta_j) = -\frac{1}{\theta_j TB} \log_e \mathbb{E}_{\mathbf{z}} \left\{ e^{-\theta_j T \sum_{m=1}^{M!} \tau_m(\mathbf{z}) R_{\pi_m^{-1}(j)}} \right\} \tag{5.20}$$

where $R_{\pi_m^{-1}(j)}$ is given by (5.15). In this scheme, the instantaneous transmission rates for the users are selected from any point on the dominant face of the MAC instantaneous capacity region.

A more restrictive but simpler scheme for the receiver is to eliminate the time-sharing and employ a particular single decoding order for each fading state $\mathbf{z}$. In this case, the instantaneous transmission rates are chosen from the vertices of the MAC instantaneous capacity region. More specifically, we assume that the vector space $\mathfrak{R}_+^M$ of the possible values for $\mathbf{z}$ is partitioned into $M!$ disjoint regions $\{\mathcal{Z}_m\}_{m=1}^{M!}$ with respect to decoding orders $\{\pi_m\}_{m=1}^{M!}$. Hence, each region corresponds to a unique decoding order. For instance, when $\mathbf{z} \in \mathcal{Z}_1$, the receiver decodes the information in the order π_1. Therefore, this scheme corresponds to the special case of the general time-sharing approach with $\tau_m(\mathbf{z}) = 1$ when $\mathbf{z} \in \mathcal{Z}_m$ and zero otherwise for all $m = 1, \ldots M!$.

Now, for a given partition $\{\mathcal{Z}_m\}_{m=1}^{M!}$, the maximum effective capacity that can be achieved by the jth user is

$$\mathsf{C}_j(\theta_j) = -\frac{1}{\theta_j TB}\log_e \mathbb{E}_{\mathbf{z}}\left\{e^{-\theta_j TR_j}\right\} \tag{5.21}$$

$$= -\frac{1}{\theta_j TB}\log_e\left(\sum_{m=1}^{M!}\int_{\mathbf{z}\in\mathcal{Z}_m} e^{-\theta_j TR_{\pi_m^{-1}(j)}} p_{\mathbf{z}}(\mathbf{z})d\mathbf{z}\right) \tag{5.22}$$

where $p_{\mathbf{z}}$ is the distribution function of $\mathbf{z}$ and $R_{\pi_m^{-1}(j)}$ is given in (5.15). Akin to the optimization in (5.13), the optimal partition $\{\mathcal{Z}_m\}_{m=1}^{M!}$ that maximizes the weighted sum of the effective capacities can be identified by solving the following optimization problem:

$$\max_{\{\mathcal{Z}_m\}} \lambda\cdot\mathsf{C}(\Theta) = \max_{\{\mathcal{Z}_m\}}\sum_{j=1}^{M}\lambda_j\mathsf{C}_j(\theta_j) \tag{5.23}$$

$$= \max_{\{\mathcal{Z}_m\}}\sum_{j=1}^{M} -\frac{\lambda_j}{\theta_j TB} \times\log_e\left(\sum_{m=1}^{M!}\int_{\mathbf{z}\in\mathcal{Z}_m} e^{-\theta_j TR_{\pi_m^{-1}(j)}} p_{\mathbf{z}}(\mathbf{z})d\mathbf{z}\right). \tag{5.24}$$

Note that the optimal partition depends on the weight vector $\lambda = (\lambda_1,\ldots,\lambda_M)$. By solving a sequence of optimization problems for different values of λ, we can trace the boundary of the effective throughput region.

Considering the expression for effective capacity and the optimization problem in (5.24), we note that finding closed-form analytical expressions for the optimal partitions of the channel state space seems intractable for a general scenario. With this in mind, we consider a simplified case in which all users have the same QoS constraint described by θ. This case arises, for instance, if users do not have priorities over others in terms of buffer limitations or delay constraints.

5.3.2.1 Two-user MAC

First, we consider the two-user MAC case and suppose that the two users have the same QoS exponent θ. Similar to the discussion in [41], finding an optimal decoding order function can be reduced to finding a function $z_2 = g(z_1)$ in the state space such that users are decoded in the order (1,2) if $z_2 < g(z_1)$ and users are decoded in the order (2,1) if $z_2 > g(z_1)$. Hence, the function g partitions the space of the possible values of $\mathbf{z} = (z_1, z_2)$. With this, the optimization problem in (5.23) becomes

$$\max_{g} \lambda_1 \mathsf{C}_1(\theta, g(z_1)) + (1-\lambda_1)\mathsf{C}_2(\theta, g(z_1)) \tag{5.25}$$

where $\mathsf{C}_1(\theta, g(z_1))$ and $\mathsf{C}_2(\theta, g(z_1))$ are expressed as

$$\begin{aligned}
&\mathsf{C}_1(\theta, g(z_1)) \\
&= \frac{-1}{\theta TB} \log_e \left(\int_0^{\infty}\int_{g(z_1)}^{\infty} e^{-\theta TB \log_2\left(1+\mathrm{SNR}_1 z_1\right)} p_{\mathbf{z}}(z_1, z_2) dz_2 dz_1 \right. \\
&\quad \left. + \int_0^{\infty}\int_0^{g(z_1)} e^{-\theta TB \log_2\left(1+\frac{\mathrm{SNR}_1 z_1}{1+\mathrm{SNR}_2 z_2}\right)} p_{\mathbf{z}}(z_1, z_2) dz_2 dz_1 \right),
\end{aligned} \tag{5.26}$$

$$\begin{aligned}
&\mathsf{C}_2(\theta, g(z_1)) \\
&= \frac{-1}{\theta TB} \log_e \left(\int_0^{\infty}\int_0^{g(z_1)} e^{-\theta TB \log_2\left(1+\mathrm{SNR}_2 z_2\right)} p_{\mathbf{z}}(z_1, z_2) dz_2 dz_1 \right. \\
&\quad \left. + \int_0^{\infty}\int_{g(z_1)}^{\infty} e^{-\theta TB \log_2\left(1+\frac{\mathrm{SNR}_2 z_2}{1+\mathrm{SNR}_1 z_1}\right)} p_{\mathbf{z}}(z_1, z_2) dz_2 dz_1 \right).
\end{aligned} \tag{5.27}$$

Note that the maximization in (5.25) is over the choice of the function $g(z_1)$. Implicitly, $g(z_1)$ should always be larger than zero as implied in (5.26) and (5.27). In cases in which this condition is not satisfied, we need to find a function $z_1 = f(z_2)$ instead, as will be specified below.

Theorem 14 *The optimal decoding order as a function of the fading state* $\mathbf{z} = (z_1, z_2)$ *for a specific common QoS constraint* θ *in the two-user case is characterized by the following functions:*

$$g(z_1) = \frac{(1 + SNR_1 z_1)K^{\frac{1}{\beta}} - 1}{SNR_2}, \quad \text{if } K \in [1, \infty) \text{ and} \tag{5.28}$$

$$f(z_2) = \frac{(1 + SNR_2 z_2)K^{-\frac{1}{\beta}} - 1}{SNR_1}, \quad \text{if } K \in [0, 1) \tag{5.29}$$

where $\beta = \frac{\theta TB}{\log_e 2}$ *and* $K \in [0, \infty)$ *is a constant that depends on the weight* λ_1 *in (5.25) and the values of the double integrals in (5.26) and (5.27). Note that the function used to partition the state space is either* g *or* f *depending on the value of* K*.*

Proof: Suppose that the optimal decoding order is specified by the function $z_2 = g(z_1)$. We define

$$\mathcal{J}(\hat{g}(z_1)) = \lambda_1 \mathsf{C}_1(\theta, \hat{g}(z_1)) + (1 - \lambda_1)\mathsf{C}_2(\theta, \hat{g}(z_1)) \tag{5.30}$$

where $\hat{g}(z_1) = g(z_1) + s\eta(z_1)$. $g(z_1)$ is the optimal function, s is any constant, and $\eta(z_1)$ represents arbitrary perturbation. A necessary condition that needs to be satisfied is [66]

$$\frac{d}{ds}\left(\mathcal{J}(\hat{g}(z_1))\right)\Big|_{s=0} = 0. \tag{5.31}$$

We define the following:

$$\phi_1 = \int_0^{\infty}\int_{g(z_1)}^{\infty} e^{-\theta TB\log_2\left(1+\mathrm{SNR}_1 z_1\right)} p_{\mathbf{z}}(z_1,z_2)dz_2dz_1 + \int_0^{\infty}\int_0^{g(z_1)} e^{-\theta TB\log_2\left(1+\frac{\mathrm{SNR}_1 z_1}{1+\mathrm{SNR}_2 z_2}\right)} p_{\mathbf{z}}(z_1,z_2)dz_2dz_1, \tag{5.32}$$

$$\phi_2 = \int_0^{\infty}\int_0^{g(z_1)} e^{-\theta TB\log_2\left(1+\mathrm{SNR}_2 z_2\right)} p_{\mathbf{z}}(z_1,z_2)dz_2dz_1 + \int_0^{\infty}\int_{g(z_1)}^{\infty} e^{-\theta TB\log_2\left(1+\frac{\mathrm{SNR}_2 z_2}{1+\mathrm{SNR}_1 z_1}\right)} p_{\mathbf{z}}(z_1,z_2)dz_2dz_1. \tag{5.33}$$

By noting that $\frac{d\hat{g}(z_1)}{ds} = \eta(z_1)$, and from (5.31)–(5.33), we can derive

$$\int_0^{\infty}\left(-\frac{\lambda_1}{\theta TB\phi_1}\left(\left(1+\frac{\mathrm{SNR}_1 z_1}{1+\mathrm{SNR}_2 g(z_1)}\right)^{-\beta} - (1+\mathrm{SNR}_1 z_1)^{-\beta}\right) - \frac{1-\lambda_1}{\theta TB\phi_2}\left((1+\mathrm{SNR}_2 g(z_1))^{-\beta} - \left(1+\frac{\mathrm{SNR}_2 g(z_1)}{1+\mathrm{SNR}_1 z_1}\right)^{-\beta}\right)\right) \cdot p_{\mathbf{z}}(z_1, g(z_1))\eta(z_1)dz_1 = 0 \tag{5.34}$$

Since the above equation holds for any $\eta(z_1)$, it follows that

$$-\frac{\lambda_1}{\theta TB\phi_1}\left(\left(1+\frac{\mathrm{SNR}_1 z_1}{1+\mathrm{SNR}_2 g(z_1)}\right)^{-\beta} - (1+\mathrm{SNR}_1 z_1)^{-\beta}\right) - \frac{1-\lambda_1}{\theta TB\phi_2}\left((1+\mathrm{SNR}_2 g(z_1))^{-\beta} - \left(1+\frac{\mathrm{SNR}_2 g(z_1)}{1+\mathrm{SNR}_1 z_1}\right)^{-\beta}\right) = 0 \tag{5.35}$$

which after rearranging and defining K as follows yields

$$\frac{\left(1+\frac{\mathrm{SNR}_1 z_1}{1+\mathrm{SNR}_2 g(z_1)}\right)^{-\beta} - (1+\mathrm{SNR}_1 z_1)^{-\beta}}{\left(1+\frac{\mathrm{SNR}_2 g(z_1)}{1+\mathrm{SNR}_1 z_1}\right)^{-\beta} - (1+\mathrm{SNR}_2 g(z_1))^{-\beta}} = \frac{(1-\lambda_1)\phi_1}{\lambda_1\phi_2} = K. \tag{5.36}$$

Obviously, $K \geq 0$. Notice that after a simple computation, (5.36) becomes

$$\left(\frac{1+\mathrm{SNR}_1 z_1}{1+\mathrm{SNR}_2 g(z_1)}\right)^{-\beta} = K \tag{5.37}$$

which leads to (5.28) after rearranging. Note here that if $K < 1$, $g(z_1) < 0$ for $z_1 < \frac{K^{-\frac{1}{\beta}}-1}{\mathrm{SNR}_1}$. Then, the expressions in (5.26) and (5.27) are not well-defined. In this case, we denote the optimal function as $z_1 = f(z_2)$ instead. Following a similar approach as shown in (5.26) through (5.37) yields (5.29). □

Remark: Above, we have assumed that the users are decoded in the order $(1, 2)$ when $z_2 < g(z_1)$ (or $z_1 > f(z_2)$ if $K < 1$) and decoded in the order $(2, 1)$ when $z_2 > g(z_1)$ (or $z_1 < f(z_2)$ if $K < 1$). It is interesting note that if we switch the decoding orders in the regions (i.e., if users are decoded in the order $(1, 2)$ when $z_2 > g(z_1)$), exactly the same partition functions as in (5.28) and (5.29) are obtained due to the symmetric nature of the problem. Hence, the structure of the optimal functions that partition the space of channel states (z_1, z_2) into two non-overlapping regions does not depend on which decoding order is used in which region.

Remark: Although the partition does not depend on the choice of the decoding orders in different regions, the performance definitely does. Our numerical computations show that the order selected originally at the beginning of our discussion (i.e., using the decoding order (1,2) when $z_2 < g(z_1)$ or $z_1 > f(z_2)$) provides a larger throughput region than otherwise. This observation leads to an interesting conclusion. Note that partition functions $g(z_1)$ in (5.28) and $f(z_2)$ in (5.29) are linear functions of z_1 and z_2, respectively. When $K \geq 1$ and

$$z_2 < g(z_1) = \frac{(1+\mathrm{SNR}_1 z_1)K^{\frac{1}{\beta}} - 1}{\mathrm{SNR}_2}, \tag{5.38}$$

user 1 is decoded first and user 2 is decoded last. Hence, for instance, when z_1 is much larger than z_2 and user 1 is enjoying much better channel conditions, user 1 is decoded first in the presence of interference caused by user 2's received signal. User 2, who has less favorable conditions, is decoded subsequently without experiencing any interference. Note that such an operation is the opposite of an opportunistic behavior and leads to a more fair treatment of users. This is rather insightful since the users are assumed to operate under similar QoS limitations (i.e., they have the same QoS exponent θ). Note that if the decoding orders are switched, users having favorable channel conditions will be decoded last and hence experience no interference. In such a case, there is a bias towards users with better channel conditions, which leads to inefficient performance when both users operate under similar buffer constraints.

Our observations above have led us to propose the following suboptimal decoding order strategy for a scenario with more than 2 users.

5.3.2.2 Suboptimal Decoding Order

In this section, we consider an arbitrary number of users. When all users have the same QoS constraint specified by θ, we propose a suboptimal decoding order given by

$$\frac{\lambda_{\pi(1)}}{z_{\pi(1)}} \leq \frac{\lambda_{\pi(2)}}{z_{\pi(2)}} \cdots \leq \frac{\lambda_{\pi(M)}}{z_{\pi(M)}}, \tag{5.39}$$

due to the observation that the user with the largest weight λ should be decoded last, and the fact that the higher the value of z, the less power is needed to achieve a specific effective capacity. Considering a two-user example, we, with this choice of

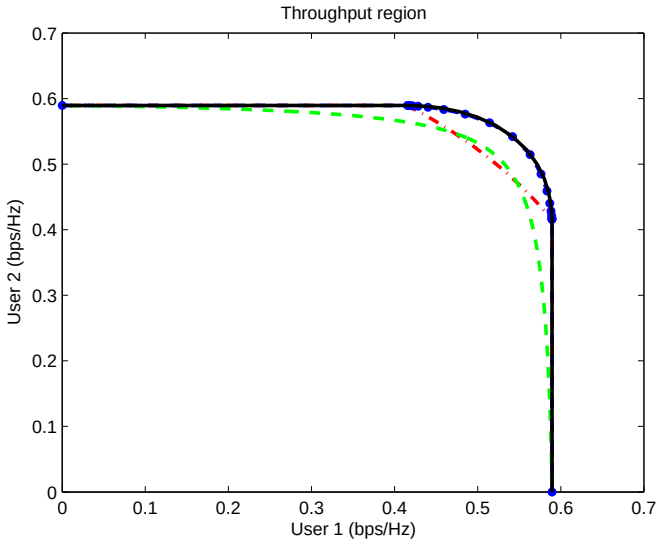

Figure 5.2: The throughput region of two-user MAC case. $\mathrm{SNR}_1 = \mathrm{SNR}_2 = 0$ dB. $\theta_1 = \theta_2 = 0.01$. The solid, dotted, dot-dashed, and dashed lines represent the regions achieved with optimal variable decoding order, suboptimal variable decoding order, fixed decoding with time sharing, and the TDMA respectively.

the decoding order, can express the points on the boundary surface as

$$\mathsf{C}_1(\theta) = -\frac{1}{\theta TB}\log_e\left(\int_0^\infty\int_{\frac{\lambda_2 z_1}{\lambda_1}}^\infty e^{-\theta TB\log_2(1+\mathrm{SNR}_1 z_1)}p_{\mathbf{z}}(z_1,z_2)dz_2dz_1 + \int_0^\infty\int_0^{\frac{\lambda_2 z_1}{\lambda_1}} e^{-\theta TB\log_2\left(1+\frac{\mathrm{SNR}_1 z_1}{1+\mathrm{SNR}_2 z_2}\right)}p_{\mathbf{z}}(z_1,z_2)dz_2dz_1\right) \tag{5.40}$$

$$\mathsf{C}_2(\theta) = -\frac{1}{\theta TB}\log_e\left(\int_0^\infty\int_0^{\frac{\lambda_2 z_1}{\lambda_1}} e^{-\theta TB\log_2(1+\mathrm{SNR}_2 z_2)}p_{\mathbf{z}}(z_1,z_2)dz_2dz_1 + \int_0^\infty\int_{\frac{\lambda_2 z_1}{\lambda_1}}^\infty e^{-\theta TB\log_2\left(1+\frac{\mathrm{SNR}_2 z_2}{1+\mathrm{SNR}_1 z_1}\right)}p_{\mathbf{z}}(z_1,z_2)dz_2dz_1\right). \tag{5.41}$$

5.3.3 Numerical Results

We have performed numerical analysis for independent Rayleigh fading channels with $\mathbb{E}\{\mathbf{z}\} = \mathbf{1}$. In Fig. 5.2, the throughput region of a two-user MAC is plotted for superposition strategies with different decoding ordering methods at the receiver, and also for TDMA. In the figure, the solid and dotted curves provide the throughput regions achieved by employing optimal and suboptimal variable decoding orders, respectively, at the receiver. Note that in the optimal strategy described by the results of Theorem 14, the receiver chooses the decoding order according to the channel states such that the weighted sum of effective capacities, i.e., summation of log-moment generating functions, is maximized. We see that the suboptimal strategy described in Section 5.3.2.2 can achieve almost the same rate region as the optimal strategy, indicating the efficiency of this approach. In the same figure, dot-dashed curve provides the throughput region achieved by employing a fixed decoding order for all channel states. Here, we observe that the strategy of using a fixed decoding order at the receiver is strictly suboptimal even when the users are operating under similar buffer constraints, and varying the decoding order with the respect to the channel gains can significantly increase the achievable region. Finally, the throughput region of TDMA is given by the dashed curve. We immediately note that TDMA can achieve some points outside of the throughput region attained with fixed decoding order at the receiver side. These numerical results show that markedly different strategies may need to be employed when systems are operating under buffer constraints. In the absence of such constraints, the performance is captured by the ergodic capacity region which cannot be improved by varying the decoding order with respect to the channel states [26]. Hence, using a fixed decoding order at the receiver is an optimal strategy when there are no QoS constraints. Moreover, TDMA is always suboptimal with respect

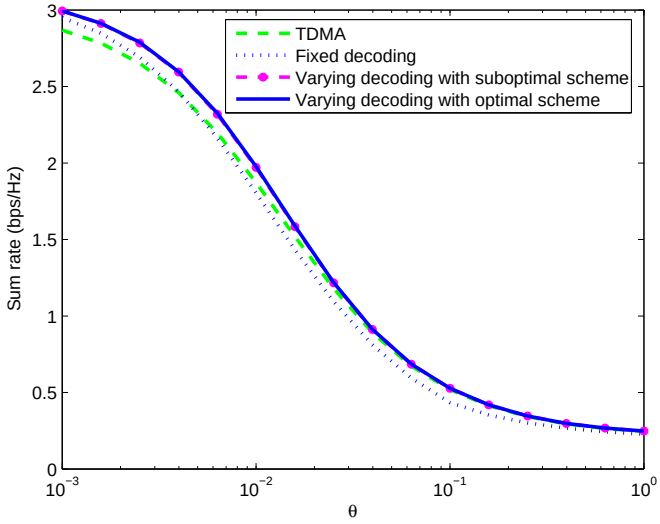

Figure 5.3: The sum-rate throughput as a function of θ. $\text{SNR}_1 = 10$ dB; $\text{SNR}_2 = 0$ dB.

to the superposition schemes regardless of the decoding-order strategy [27].

In Fig. 5.3, sum-rate throughput, i.e. the sum of the effective capacities, is plotted as a function of the QoS exponent θ. Here, we note that as θ increases, the curves of different strategies converge. In particular, TDMA performance approaches that of the superposition coding with variable decoding. Hence, orthogonal transmission strategies start being efficient in terms of attaining the sum rate under stringent buffer constraints. Note that the sum-rate throughput generally decreases with increasing θ, and we conclude from the figure that this diminished throughput can be captured by having each user concentrate its power in a certain fraction of time in the TDMA scheme. We also see that for approximately $\theta > 0.006$, TDMA starts outperforming superposition transmission when a fixed decoding order is employed at the receiver. Such an observation is also noted in the discussion of Fig. 5.2. In contrast, we observe that as θ approaches 0 and hence the QoS constraints relax, TDMA is the strategy with the worst performance. Note that when the performance metric is the

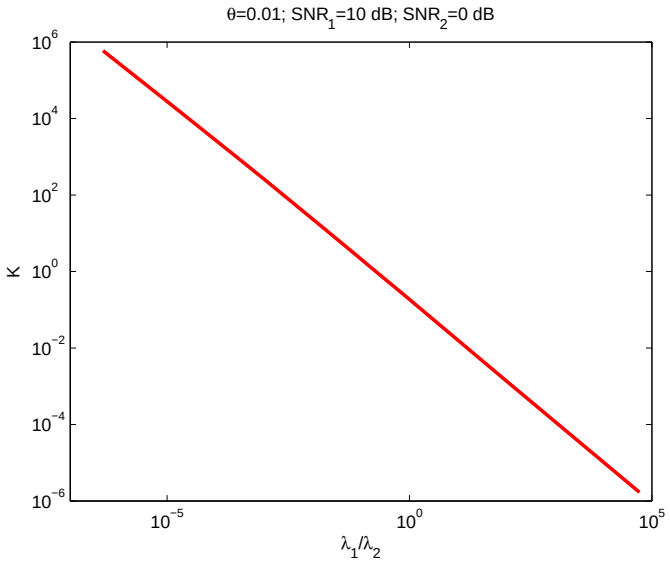

Figure 5.4: K vs. $\frac{\lambda_1}{\lambda_2}$. $\text{SNR}_1 = 10dB$. $\text{SNR}_2 = 0$ dB. $\theta_1 = \theta_2 = 0.01$.

ergodic capacity and hence no queueing constraints are considered, this suboptimality of TDMA with respect to superposition strategies is well-known (see e.g., [27]).

We are also interested in the values of parameter K that appears in the functions in Theorem 14 . In Fig. 5.4, we plot K as a function of $\frac{\lambda_1}{\lambda_2} = \frac{\lambda_1}{1-\lambda_1}$. It is interesting to note that $\log_e K$ seems to be linear with respect to $\log_e \left(\frac{\lambda_1}{1-\lambda_1}\right)$.

5.4 Transmissions with Power Control

In this section, we analyze the case in which the transmitter employs power control policies in the transmission. Similarly as before, we initially investigate the scenario in which the decoding order is fixed for all channel states. Subsequently, we study variable decoding order schemes. Note that varying the decoding order with respect to the channel states, according to the analysis in Section 5.3, has the potential to significantly affect the achievable rates.

5.4.1 Power Control Policy for Fixed Decoding Order

Here, we characterize the optimal power allocation policies when the decoding order is fixed for all channel states. Due to the convexity of $\mathcal{C}_{\text{MAC}}$, there exist Lagrange multipliers $\kappa = (\kappa_1, \ldots, \kappa_M) \in \mathfrak{R}_+^M$ such that $\mathsf{C}^*(\Theta)$ on the boundary surface can be obtained by solving the optimization problem

$$\max_{\mu} \lambda \cdot \mathsf{C}(\Theta, \mu) - \kappa \cdot \mathbb{E}\{\mu\} \tag{5.42}$$

where $\mu = (\mu_1, \ldots, \mu_M)$ represents the collection of the power control policies of all users, $\lambda = (\lambda_1, \ldots, \lambda_M)$ is the weight vector, and $\mathsf{C}(\Theta, \mu) = (\mathsf{C}_1(\theta_1, \mu), \ldots, \mathsf{C}_M(\theta_M, \mu))$ is the vector of maximum effective capacities of the users for given decoding order and power allocation policies. Note that $\mu_j = \frac{P_j}{N_0 B}$ (defined in Section 5.1 as the instantaneous transmitted SNR level) describes the power control policy of the jth user. For a given permutation π and set of power allocations μ, $\mathsf{C}_j(\theta_j, \mu)$ is given by

$$\mathsf{C}_j(\theta_j, \mu) = \\ -\frac{1}{\theta_j TB} \log_e \mathbb{E}\left\{ e^{-\theta_j TB \log_2\left(1+\frac{\mu_j z_j}{1+\sum_{\pi^{-1}(i)>\pi^{-1}(j)} \mu_i z_i}\right)} \right\}. \tag{5.43}$$

Now, the optimization problem (5.42) can be rewritten as

$$\max_{\mu} \sum_{j=1}^{M} -\lambda_j \frac{1}{\theta_j TB} \log_e \mathbb{E}\left\{ e^{-\theta_j TB \log_2\left(1+\frac{\mu_j z_j}{1+\sum_{\pi^{-1}(i)>\pi^{-1}(j)} \mu_i z_i}\right)} \right\} \\ -\sum_{j=1}^{M} \kappa_j \mathbb{E}\{\mu_j\}. \tag{5.44}$$

The following result identifies the optimal power adaptation policies that solve the above optimization problem.

Theorem 15 *Assume that the receiver, for all channel states, decodes the users in a fixed order specified by the permutation π. Then, the optimal power allocation allocation policies that solve the optimization problem in (5.44) are given by*

$$\mu_j = \left(\frac{\left(1 + \sum_{\pi^{-1}(i)>\pi^{-1}(j)} \mu_i z_i\right)^{\frac{\beta_j}{\beta_j+1}}}{\alpha_j^{\frac{1}{\beta_j+1}} z_j^{\frac{\beta_j}{\beta_j+1}}} - \frac{1 + \sum_{\pi^{-1}(i)>\pi^{-1}(j)} \mu_i z_i}{z_j} \right)^+ \text{ for } j = 1, 2, \ldots, M \tag{5.45}$$

where $\beta_j = \frac{\theta_j TB}{\log_e 2}$ is the normalized QoS exponent, $(x)^+ = \max\{x, 0\}$, and $(\alpha_1, \cdots, \alpha_M)$ are constants that are introduced to satisfy the average power constraints.

Proof: Note that with a fixed decoding order, the user $\pi(M)$ sees no interference from the other users, and hence the derivative of (5.44) with respect to $\mu_{\pi(M)}$ will only be related to the effective capacity formulation of user $\pi(M)$. Therefore, we can solve an equivalent problem by maximizing $\mathsf{C}_{\pi(M)}$ instead. After we derive $\mu_{\pi(M)}$, the derivative of (5.44) with respect to $\mu_{\pi(M-1)}$ will only be related to the effective capacity formulation of user $\pi(M-1)$. By repeated application of this procedure, for given λ, (5.44) can be further decomposed into the following M sequential optimization problems

$$\max_{\mu_j} -\lambda_j \frac{1}{\theta_j TB} \log_e \mathbb{E}\left\{ e^{-\theta_j TB \log_2\left(1+\frac{\mu_j z_j}{1+\sum_{\pi^{-1}(i)>\pi^{-1}(j)} \mu_i z_i}\right)} \right\} - \kappa_j \mathbb{E}\{\mu_j\} \quad j \in \{1, \cdots, M\} \tag{5.46}$$

in the inverse order of π. Similarly as in [35], due to the monotonicity of the logarithm, solving the above M optimizations is the same as solving

$$\min_{\mu_j} \mathbb{E}\left\{ e^{-\theta_j TB \log_2\left(1+\frac{\mu_j z_j}{1+\sum_{\pi^{-1}(i)>\pi^{-1}(j)} \mu_i z_i}\right)} \right\} + \kappa_j \mathbb{E}\{\mu_j\} \tag{5.47}$$

for $j \in \{1, \cdots, M\}$. Differentiating the above Lagrangian with respect to μ_j and setting the derivative to zero yield the intended result in (5.45). □

Remark: Exploiting the result in (5.45), we can find that instead of adapting the power according to only its channel state as in [35] where a single-user scenario is studied, the user adapts the power with respect to its channel state normalized by the observed interference and the noise.

Remark: To give an explicit idea of the power control policy, we consider a two-user example in which the decoding order is $(2, 1)$. For this case, we can easily find that

$$\mu_1 = \begin{cases} \frac{1}{\alpha_1^{\frac{1}{\beta_1+1}} z_1^{\frac{\beta_1}{\beta_1+1}}} - \frac{1}{z_1} & z_1 > \alpha_1, \\ 0 & \text{otherwise} \end{cases}, \tag{5.48}$$

and

$$\mu_2 = \begin{cases} \frac{1}{\alpha_2^{\frac{1}{\beta_2+1}} z_2^{\frac{\beta_2}{\beta_2+1}}} - \frac{1}{z_2} & z_1 \leq \alpha_1 \text{ and } z_2 > \alpha_2, \\ \frac{\left(\frac{z_1}{\alpha_1}\right)^{\frac{\beta_2}{(\beta_1+1)(\beta_2+1)}}}{\alpha_2^{\frac{1}{\beta_2+1}} z_2^{\frac{\beta_2}{\beta_2+1}}} - \frac{\left(\frac{z_1}{\alpha_1}\right)^{\frac{1}{\beta_1+1}}}{z_2} & z_1 > \alpha_1 \text{ and } \frac{z_2}{\alpha_2} > \left(\frac{z_1}{\alpha_1}\right)^{\frac{1}{\beta_1+1}} \\ 0 & \text{otherwise} \end{cases}, \tag{5.49}$$

where α_1 and α_2 are chosen to satisfy the average power constraints of the two users.

5.4.2 Power Control Policy for Variable Decoding Order

In this section, we study the optimal power allocation policy when the receiver varies the decoding order with respect to the channel fading states. We mainly concentrate on the two-user scenario. The key idea we introduce here is to consider the power allocation policy of each user j for each region $\mathcal{Z}_m$ (in which decoding is performed according to permutation π_m) while requiring the average power constraint to be satisfied by the joint power over all regions $\{\mathcal{Z}_m\}$.

For the two-user case, due to the convexity of the throughput region, there exist Lagrange multipliers $\kappa = (\kappa_1, \kappa_2) \in \Re_+^2$ such that $\mathsf{C}^*(\Theta)$ on the boundary surface can be obtained by solving the optimization problem

$$\max_{\mu} \lambda_1 \mathsf{C}_1(\mu, \mathcal{Z}) + \lambda_2 \mathsf{C}_2(\mu, \mathcal{Z}) - \kappa_1 \mathbb{E}\{\mu_1\} - \kappa_2 \mathbb{E}\{\mu_2\} \tag{5.50}$$

where $\mu = (\mu_1, \mu_2)$ are the power control policies, (λ_1, λ_2) are the weights in the weighted sum, and $\mathcal{Z} = (\mathcal{Z}_1, \mathcal{Z}_2)$ denotes a particular partition of the space of the positive values of $\mathbf{z} = (z_1, z_2)$ [1]. Hence, power control policies that solve (5.50) are the optimal ones for a given partition. In the following, since we assume $\mathcal{Z}$ is given, the notation $\mathsf{C}_j(\mu, \mathcal{Z})$ is replaced by $\mathsf{C}_j(\mu)$ for brevity.

Recalling the discussion in Section 5.3.2, we can express the effective capacities of the two users as in (5.26) and (5.27) by only replacing SNR_j with $\mu_j(\mathbf{z})$ in these expressions. The Lagrangian (which is the objective function in (5.50)) can now be expressed as in (5.51) given on the next page. In (5.51), the expressions in regions $\mathcal{Z}_1$ and $\mathcal{Z}_2$ are written separately due to the reason that possibly different power

[1] Similarly as discussed in Section 5.3.2, different decoding orders are employed in $\mathcal{Z}_1$ and $\mathcal{Z}_2$.

$$\begin{aligned}\mathcal{J} = &- \frac{\lambda_1}{\beta_1 \log_e 2} \log_e \left(\int_{\mathbf{z}\in\mathcal{Z}_1} \left(1 + \frac{\mu_1 z_1}{1+\mu_2 z_2}\right)^{-\beta_1} p_{\mathbf{z}}(z_1, z_2) dz_1 dz_2 \right. \\ &\left. + \int_{\mathbf{z}\in\mathcal{Z}_2} (1+\mu_1 z_1)^{-\beta_1} p_{\mathbf{z}}(z_1, z_2) dz_1 dz_2 \right) \\ &- \frac{\lambda_2}{\beta_2 \log_e 2} \log_e \left(\int_{\mathbf{z}\in\mathcal{Z}_2} \left(1 + \frac{\mu_2 z_2}{1+\mu_1 z_1}\right)^{-\beta_2} p_{\mathbf{z}}(z_1, z_2) dz_1 dz_2 \right. \\ &\left. + \int_{\mathbf{z}\in\mathcal{Z}_1} (1+\mu_2 z_2)^{-\beta_2} p_{\mathbf{z}}(z_1, z_2) dz_1 dz_2 \right) \\ &- \kappa_1(\mathbb{E}_{\mathbf{z}\in\mathcal{Z}_1}\{\mu_1\} + \mathbb{E}_{\mathbf{z}\in\mathcal{Z}_2}\{\mu_1\}) - \kappa_2(\mathbb{E}_{\mathbf{z}\in\mathcal{Z}_1}\{\mu_2\} + \mathbb{E}_{\mathbf{z}\in\mathcal{Z}_2}\{\mu_2\}). \end{aligned} \tag{5.51}$$

---

allocation strategies are employed in different regions. We define

$$\begin{aligned}\phi_1 = &\int_{\mathbf{z}\in\mathcal{Z}_1} \left(1 + \frac{\mu_1 z_1}{1+\mu_2 z_2}\right)^{-\beta_1} p_{\mathbf{z}}(z_1, z_2) dz_1 dz_2 \\ &+ \int_{\mathbf{z}\in\mathcal{Z}_2} (1+\mu_1 z_1)^{-\beta_1} p_{\mathbf{z}}(z_1, z_2) dz_1 dz_2, \end{aligned} \tag{5.52}$$

and

$$\begin{aligned}\phi_2 = &\int_{\mathbf{z}\in\mathcal{Z}_2} \left(1 + \frac{\mu_2 z_2}{1+\mu_1 z_1}\right)^{-\beta_2} p_{\mathbf{z}}(z_1, z_2) dz_1 dz_2 \\ &+ \int_{\mathbf{z}\in\mathcal{Z}_1} (1+\mu_2 z_2)^{-\beta_2} p_{\mathbf{z}}(z_1, z_2) dz_1 dz_2. \end{aligned} \tag{5.53}$$

Note that the values of these functions are obtained for given power control policies $\mu = (\mu_1, \mu_2)$ and given partition $\mathcal{Z} = (\mathcal{Z}_1, \mathcal{Z}_2)$.

Now, we consider the power control policy of each user in each decoding order region $\mathcal{Z}_i$, $i = 1, 2$. By differentiating the Lagrangian, we can find the following

optimality conditions:

$$1)\ \frac{\lambda_1}{\phi_1 \log_e 2}(1+\mu_1 z_1)^{-\beta_1-1} z_1 - \frac{\lambda_2}{\phi_2 \log_e 2}\left(1+\frac{\mu_2 z_2}{1+\mu_1 z_1}\right)^{-\beta_2-1} \frac{\mu_2 z_2 z_1}{(1+\mu_1 z_1)^2} - \kappa_1 = 0 \quad \forall \mathbf{z} \in \mathcal{Z}_1 \tag{5.54}$$

$$2)\ \frac{\lambda_2}{\phi_2 \log_e 2}\left(1+\frac{\mu_2 z_2}{1+\mu_1 z_1}\right)^{-\beta_2-1} \frac{z_2}{1+\mu_1 z_1} - \kappa_2 = 0 \quad \forall \mathbf{z} \in \mathcal{Z}_1 \tag{5.55}$$

$$3)\ \frac{\lambda_1}{\phi_1 \log_e 2}\left(1+\frac{\mu_1 z_1}{1+\mu_2 z_2}\right)^{-\beta_1-1} \frac{z_1}{1+\mu_2 z_2} - \kappa_1 = 0 \quad \forall \mathbf{z} \in \mathcal{Z}_2 \tag{5.56}$$

$$4)\ -\frac{\lambda_1}{\phi_1 \log_e 2}\left(1+\frac{\mu_1 z_1}{1+\mu_2 z_2}\right)^{-\beta_1-1} \frac{\mu_1 z_1 z_2}{(1+\mu_2 z_2)^2} + \frac{\lambda_2}{\phi_2 \log_e 2}(1+\mu_2 z_2)^{-\beta_2-1} z_2 - \kappa_2 = 0 \quad \forall \mathbf{z} \in \mathcal{Z}_2 \tag{5.57}$$

where (5.54) and (5.55) are obtained by differentiating the Lagrangian with respect to μ_1 and μ_2, respectively, over $\mathbf{z} \in \mathcal{Z}_1$. Similarly, (5.56) and (5.57) are obtained by differentiating with respect to μ_1 and μ_2, respectively, over $\mathbf{z} \in \mathcal{Z}_2$. Due to the convexity, whenever $\mu_i, i = 1, 2$ is negative valued, we set $\mu_i = 0, i = 1, 2$. Although obtaining closed form expressions from the optimality conditions seems to be unlikely, we can gather several insights on the power control policies by analyzing the equations (5.54)–(5.57).

Let us first define $\alpha_1 = \frac{\kappa_1 \phi_1 \log_e 2}{\lambda_1}$, $\alpha_2 = \frac{\kappa_2 \phi_2 \log_e 2}{\lambda_2}$, $\alpha_{12} = \frac{\kappa_2 \phi_1 \log_e 2}{\lambda_1}$, and $\alpha_{21} = \frac{\kappa_1 \phi_2 \log_e 2}{\lambda_2}$, where κ_1, κ_2 are the Lagrange multipliers whose values are chosen to satisfy the average power constraint (5.7) with equality, and ϕ_1 and ϕ_2 are defined in (5.52) and (5.53). Now, consider (5.54) and (5.55). The channel state lies in $\mathcal{Z}_1$. Through

a simple computation using (5.55), we can derive

$$\mu_2 = \frac{(1+\mu_1 z_1)^{\frac{\beta_2}{\beta_2+1}}}{\alpha_2^{\frac{1}{\beta_2+1}} z_2^{\frac{\beta_2}{\beta_2+1}}} - \frac{1+\mu_1 z_1}{z_2} \tag{5.58}$$

which tells us that $\mu_2 = 0$ if

$$\frac{z_2}{1+\mu_1 z_1} < \alpha_2. \tag{5.59}$$

If $\mu_2 = 0$, we have from (5.54) that

$$\frac{\lambda_1}{\phi_1 \log_e 2}(1+\mu_1 z_1)^{-\beta_1 - 1} z_1 - \kappa_1 = 0 \tag{5.60}$$

which gives us that

$$\mu_1 = \frac{1}{\alpha_1^{\frac{1}{\beta_1+1}} z_1^{\frac{\beta_1}{\beta_1+1}}} - \frac{1}{z_1} \tag{5.61}$$

which implies that $\mu_1 = 0$ if

$$z_1 < \alpha_1. \tag{5.62}$$

Now, if we substitute (5.58) into (5.54), we obtain the following additional condition for having $\mu_1 = 0$: the equation

$$\begin{aligned} &\frac{z_1}{\alpha_1}(1+\mu_1 z_1)^{-(\beta_1+1)} \\ &- \frac{z_1 \alpha_2}{z_2 \alpha_{12}} \left(\left(\frac{z_2}{\alpha_2(1+\mu_1 z_1)} \right)^{\frac{1}{\beta_2+1}} - 1 \right) - 1 = 0 \end{aligned} \tag{5.63}$$

has a solution that returns a negative or zero value for μ_1. The above discussion enables us to characterize the regions in which one user transmits while the other one is silent. We also have a closed-form formula in (5.61) for the optimal power

adaptation policy when only one user transmits. Indeed, this is the optimal power control policy derived in [35] for a single-user system. When both users transmit, the power control policies (μ_1, μ_2) are given directly by the non-negative solution of (5.54) and (5.55).

Note that the conditions and characterizations provided in (5.58)–(5.63) pertain to the case in which the channel state is in region $\mathcal{Z}_1$. Following a similar analysis of (5.56) and (5.57), we can obtain similar results for the cases in which the channel state is in $\mathcal{Z}_2$.

For a given partition $\{\mathcal{Z}_1, \mathcal{Z}_2\}$, the optimal power control policy can be determined numerically using the optimality conditions in (5.54) – (5.57). Additionally, the equations and inequalities in (5.58) through (5.63) can be used to guide the numerical algorithms as they specify under which conditions at most one user transmits, and provide the optimal power control policy in such cases. However, there is one difficulty. (5.58) – (5.63) depend on $\alpha_1, \alpha_2, \alpha_{12}$, and α_{21} which in turn depend on ϕ_1, ϕ_2, κ_1, and κ_2 which are in general functions of the power control policies. In such a situation, the following iterative procedure can be employed in search of the solution. We can first choose certain values for ϕ_1, ϕ_2, κ_1, and κ_2, and then determine the optimal power allocation policies for these selected values. Subsequently, we can check whether the obtained policy satisfies the average power constraint with equality. This enables us to determine if the selected κ_1 and κ_2 values are accurate. We can also compute ϕ_1 and ϕ_2 using the obtained policy and see if they agree with the initial values of ϕ_1 and ϕ_2. If there is no sufficient match or if the power constraint is not satisfied with equality, then we update the values of ϕ_1, ϕ_2, κ_1, and κ_2, and reiterate the search of the optimal policy.

With this insight, we propose the following algorithms that can be used to determine the optimal power allocated to each channel state:

input : λ_1, λ_2, the partition $\mathcal{Z}$, $\mathbf{z}$
output: Optimal μ^*
1 Initialize ϕ_1, ϕ_2;
2 Initialize κ_1 and κ_2;
3 Determine $\alpha_1 = \frac{\kappa_1 \phi_1 \log_e 2}{\lambda_1}$, $\alpha_2 = \frac{\kappa_2 \phi_2 \log_e 2}{\lambda_2}$, $\alpha_{12} = \frac{\kappa_2 \phi_1 \log_e 2}{\lambda_1}$, $\alpha_{21} = \frac{\kappa_1 \phi_2 \log_e 2}{\lambda_2}$;
4 Determine μ_1, μ_2 by Algorithm 2;
5 Check if the obtained power control policies μ_1 and μ_2 satisfy the power constraint with equality;
6 **if** *not satisfied with equality* **then**
7 | update the values of κ_1 and κ_2 and return to Step 2;
8 **else**
9 | move to Step 10;
10 Evaluate ϕ_1 and ϕ_2 with the obtained power control policies;
11 Check if the new values of ϕ_1 and ϕ_2 agree (up to a certain margin) with those used in Step 2;
12 **if** *do not agree* **then**
13 | update the values of ϕ_1 and ϕ_2 and return to Step 1;
14 **else**
15 | declare the obtained power allocation policies μ_1 and μ_2 as the optimal ones.
16

Algorithm 1: Power Control Algorithm

Note that we above have not specified how the values of $\kappa_1, \kappa_2, \phi_1$, and ϕ_2 are updated for each iteration in order to keep the algorithm 1 generic. In our numerical computations, we have updated κ_1 and κ_2 using the bisection search algorithm. The values of ϕ_1 and ϕ_2 are updated in Step 13 of the algorithm by assigning them the values evaluated in Step 11. Hence, the most recent values are carried over to the new iteration.

In Fig. 5.5, we plot the optimal power allocation policies μ_1 and μ_2 as functions of channel fading states z_1 and z_2. We assume that $\theta_1 = \theta_2 = 0.01$, $\text{SNR}_1 = \text{SNR}_2 = 0$ dB, and $\lambda_1 = \lambda_2 = 0.5$. We consider the partition specified by the suboptimal decoding order given in (5.39). Hence, since we have $\lambda_1 = \lambda_2 = 0.5$, decoding orders (1,2) and (2,1) are used when $z_2 < z_1$ and $z_2 > z_1$, respectively. Under these assumptions,

input : λ_1, λ_2, the partition $\mathcal{Z}$, $\mathbf{z}$, α_1, α_2, α_{12} and α_{21}
output: μ_1, μ_2

if $\mathbf{z} \in \mathcal{Z}_1$ **then**
  if $z_2 > \alpha_2$ **then**
    $\mu_2 = \frac{1}{\alpha_2^{\frac{1}{\beta_2+1}} z_2^{\frac{\beta_2}{\beta_2+1}}} - \frac{1}{z_2}$;
    if $\frac{z_1}{\alpha_1}\left(1+\mu_1 z_1\right)^{-(\beta_1+1)} - \frac{z_1\alpha_2}{z_2\alpha_{21}}\left(\left(\frac{z_2}{\alpha_2(1+\mu_1 z_1)}\right)^{\frac{1}{\beta_2+1}} - 1\right) - 1 = 0$ *returns nonpositive* μ_1 **then**
      $\mu_1 = 0$;
    else if $\frac{z_2}{\alpha_2} < \left(\frac{z_1}{\alpha_1}\right)^{\frac{1}{\beta_1+1}}$ **then**
      $\mu_2 = 0$, $\mu_1 = \left[\frac{1}{\alpha_1^{\frac{1}{\beta_1+1}} z_1^{\frac{\beta_1}{\beta_1+1}}} - \frac{1}{z_1}\right]^+$;
    else
      Compute μ_1, μ_2 from (5.54) and (5.55);
  else
    $\mu_2 = 0$, $\mu_1 = \left[\frac{1}{\alpha_1^{\frac{1}{\beta_1+1}} z_1^{\frac{\beta_1}{\beta_1+1}}} - \frac{1}{z_1}\right]^+$;
else if $\mathbf{z} \in \mathcal{Z}_2$ **then**
  if $z_1 > \alpha_1$ **then**
    $\mu_1 = \frac{1}{\alpha_1^{\frac{1}{\beta_1+1}} z_1^{\frac{\beta_1}{\beta_1+1}}} - \frac{1}{z_1}$;
    if $\frac{z_2}{\alpha_2}\left(1+\mu_2 z_2\right)^{-(\beta_2+1)} - \frac{z_2\alpha_1}{z_1\alpha_{21}}\left(\left(\frac{z_1}{\alpha_1(1+\mu_2 z_2)}\right)^{\frac{1}{\beta_1+1}} - 1\right) - 1 = 0$ *returns nonpositive* μ_2 **then**
      $\mu_2 = 0$;
    else if $\frac{z_1}{\alpha_1} < \left(\frac{z_2}{\alpha_2}\right)^{\frac{1}{\beta_2+1}}$ **then**
      $\mu_1 = 0$, $\mu_2 = \left[\frac{1}{\alpha_2^{\frac{1}{\beta_2+1}} z_2^{\frac{\beta_2}{\beta_2+1}}} - \frac{1}{z_2}\right]^+$;
    else
      Compute μ_1, μ_2 from (5.56) and (5.57);
  else
    $\mu_1 = 0$, $\mu_2 = \left[\frac{1}{\alpha_2^{\frac{1}{\beta_2+1}} z_2^{\frac{\beta_2}{\beta_2+1}}} - \frac{1}{z_2}\right]^+$;

Algorithm 2: Evaluating Power over All Channel State

we computed the optimal values as $\kappa_1^* = 0.0470$, $\kappa_2^* = 0.0462$, $\phi_1^* = 0.5550$, and $\phi_2^* = 0.5538$. In the figure, we observe that each user, not surprisingly, allocates most of its power to the regions in which it is decoded last and hence does not experience interference. However, due to the introduction of QoS constraints, we also note that each user also allocates certain power to the cases in which it is decoded first. This is performed in order to continue transmission and avoid buffer overflows.

So far, we have assumed that the partition $\mathcal{Z}$ is given. The optimal partition $\mathcal{Z}$ that maximizes the weighted sum-rate can be derived through the following optimization similarly as in [67]:

$$\mathsf{C}^* = \sup_{\mathcal{Z}} \lambda_1 \mathsf{C}_1(\mu, \mathcal{Z}) + \lambda_2 \mathsf{C}_2(\mu, \mathcal{Z}) \tag{5.64}$$

where C^* is the optimal weighted sum value for given pair of (λ_1, λ_2), and $\mu = (\mu_1, \mu_2)$ are the optimal power control policies for given $\mathcal{Z}$.

5.5 Conclusion

In this chapter, we have studied the achievable throughput regions in multiple access fading channels when users operate under QoS constraints. We have assumed that both the transmitters and the receiver have perfect CSI. We have employed the effective capacity as a measure of the throughput under buffer constraints. We have defined the effective capacity region and shown its convexity. We have considered different transmission and reception scenarios e.g., superposition coding, different strategies for the decoding order, and TDMA. When transmission with superposition coding is performed, we have shown that varying the decoding order at the receiver with respect to the fading states can significantly increase the achievable rate region

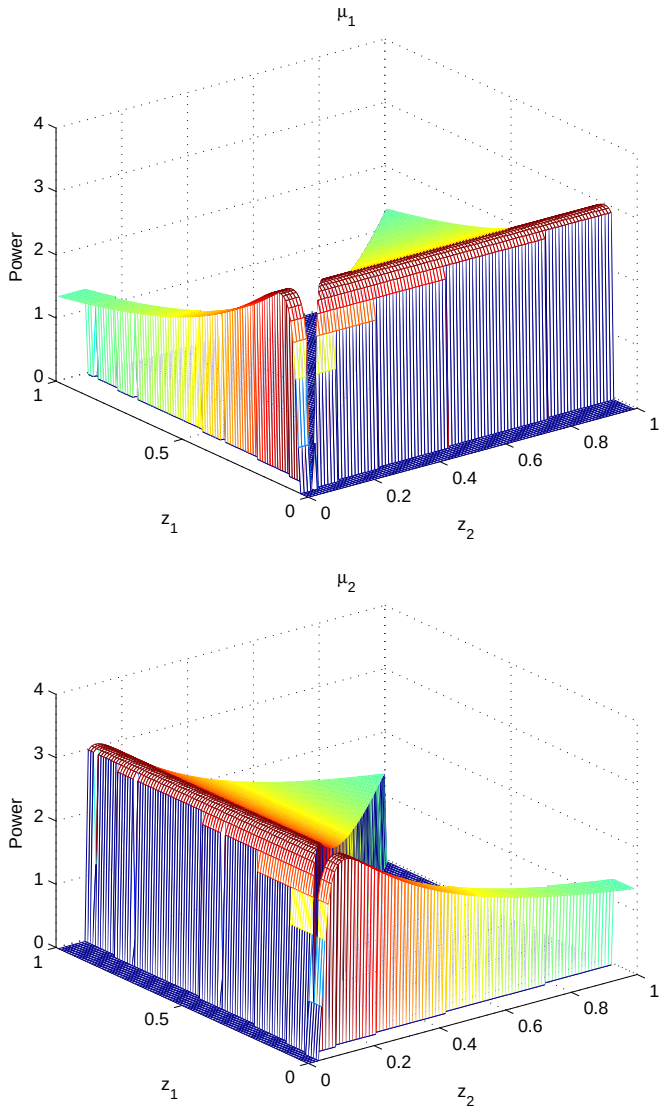

Figure 5.5: The optimal power control policies μ_1 and μ_1 of users 1 and 2, respectively, as a function of (z_1, z_2). $\lambda_1 = 0.5$, $\lambda_2 = 0.5$.

compared to that achieved with fixed decoding order schemes. For the case of two users with the same QoS constraints, we have derived the optimal strategy for varying the decoding order. We have noted that when the two users operate under similar QoS limitations and one user enjoys much more favorable channel conditions, the efficient strategy is to first decode the user with the better channel and subsequently decode the other user so that the user with worse channel conditions does not experience interference. Motivated by this observation, we have proposed for general multiple user scenarios a simpler suboptimal decoding rule which can almost perfectly match the optimal throughput region. We have also studied the performance of orthogonal transmission strategies by considering TDMA. In the numerical results, we have demonstrated that TDMA can perform better than superposition coding with fixed decoding order for certain QoS constraints. More specifically, we have noted that TDMA can support arrival rate pairs that are strictly outside the region achieved when fixed decoding order is employed at the receiver. We have also observed that the performance of TDMA approaches that of the optimal strategy of superposition coding with variable decoding order as θ increases (i.e., as the QoS constraints become more stringent).

In the second part of the chapter, we have incorporated power adaptation strategies into the model. For a given fixed decoding order at the receiver, we have identified the optimal power control policies. We have seen that the optimal schemes adapt the power by treating the observed interference as additional noise. Since the observed interference depends on the power control policies of the other users that will be decoded later, a coupling is introduced between the optimal policies. For cases in which a variable decoding order strategy is adopted by the transmitter, we have obtained the conditions that the optimal strategies should satisfy and described an algorithm to achieve these optimal schemes.

Chapter 6

Energy Efficiency in MAC

In this chapter, we employ the tools provided in [21] and [27] to investigate the bit energy and wideband slope regions of fading MAC under QoS constraints in the low-power and wideband regimes. With the results obtained in Chapter 5, the main contributions of this chapter are summarized in the following:

1. We show that different transmission and reception strategies do not affect the minimum bit energy levels required by each user. Additionally, we prove that while the minimum bit energies are independent of the QoS constraints in the low power regime, they vary with the QoS constraints in the wideband regime.

2. We determine that superposition coding with variable decoding order does not improve the performance in terms of slope region with respect to fixed decoding order in the low power regime, while it can achieve larger slope region in the wideband regime.

3. When wideband slope regions are considered, we show that TDMA is always suboptimal in the low power regime except the special case in which fading states are linearly dependent. On the other hand, TDMA in certain cases is

demonstrated to perform better than superposition coding with fixed decoding order in the wideband regime. We also identify the condition for TDMA to be suboptimal in this regime.

6.1 Effective Capacity Region of the MAC Channel

As shown in Chapter 5, the effective capacity regions for the different transmission strategies are given as follows.

The effective capacity regions of the multi-access channel for different scheduling policies have been characterized. The effective capacity region achieved by TDMA is

$$\bigcup_{\{\delta_j\}} \left\{ \mathsf{C}(\Theta) \geq \mathbf{0} : \mathsf{C}_j(\theta_j) \leq -\frac{1}{\theta_j TB} \log_e \mathbb{E}\left\{ e^{-\delta_j \theta_j TB \log_2\left(1+\frac{\mathrm{SNR}_j}{\delta_j} z_j\right)} \right\} \right\} \tag{6.1}$$

where δ_j is the fraction of time allocated to user j.

The effective capacity region achieved by superposition coding with fixed decoding order is given by

$$\bigcup_{\{\tau_m\}} \left\{ \mathsf{C}(\Theta) \geq \mathbf{0} : \mathsf{C}_j(\theta_j) \leq -\frac{1}{\theta_j TB} \log_e \mathbb{E}_{\mathbf{z}}\left\{ e^{-\theta_j T \sum_{m=1}^{M!} \tau_m R_{\pi_m^{-1}(j)}} \right\} \right\} \tag{6.2}$$

where τ_m is the fraction of time allocated to a specific decoding order π_m, $R_{\pi_m^{-1}(j)}$ represents the maximal instantaneous service rate of user j at a given decoding order π_m, which is given by

$$R_{\pi_m^{-1}(j)} = B \log_2 \left(1 + \frac{\mathrm{SNR}_j z_j}{1 + \sum_{\pi_m^{-1}(i) > \pi_m^{-1}(j)} \mathrm{SNR}_i z_i} \right) \tag{6.3}$$

where π_m^{-1} is the inverse trace function of π_m.

Decoding orders can be varied for each channel fading state $\mathbf{z}$. Suppose the vector space $\mathfrak{R}_+^M$ of the possible values for $\mathbf{z}$ is partitioned into $M!$ disjoint regions $\{\mathcal{Z}_m\}_{m=1}^{M!}$ with respect to decoding orders $\{\pi_m\}_{m=1}^{M!}$. Then, the maximum effective capacity that can be achieved by the jth user is

$$\begin{aligned}\mathsf{C}_j(\theta_j) &= -\frac{1}{\theta_j TB}\log_e \mathbb{E}_{\mathbf{z}}\left\{e^{-\theta_j TR_j}\right\} \\ &= -\frac{1}{\theta_j TB}\log_e\left(\sum_{m=1}^{M!}\int_{\mathbf{z}\in\mathcal{Z}_m} e^{-\theta_j TR_{\pi_m^{-1}(j)}} p_{\mathbf{z}}(\mathbf{z})d\mathbf{z}\right) \qquad (6.4)\end{aligned}$$

for $j = 1,\ldots,M$, where $p_{\mathbf{z}}$ is the distribution function of $\mathbf{z}$ and $R_{\pi_m^{-1}(j)}$ is given in (6.3).

6.2 Energy Efficiency in the Low-Power Regime

As described above, in order to transmit energy efficiently and achieve bit energy levels close to the minimum level, one needs to operate in the low-SNR regime in which either the power is low or bandwidth is large. In this section, we consider the low-power regime. We concentrate on the two-user multiaccess channel. Below, we first note the maximum effective capacities attained through different transmission strategies described in Section 5.3. Subsequently, we identify the corresponding minimum bit energies and the wideband slopes.

Now, for the two-user TDMA, if we fix the fraction of time allocated to user 1 as $\delta \in [0,1]$, the maximum effective capacities of the two-users in the TDMA region

given by (6.1) become

$$\mathsf{C}_1(\mathrm{SNR}_1) = -\frac{1}{\theta_1 TB}\log_e \mathbb{E}_{\mathbf{z}}\left\{e^{-\delta\theta_1 TB\log_2\left(1+\frac{\mathrm{SNR}_1 z_1}{\delta}\right)}\right\} \tag{6.5}$$

and

$$\mathsf{C}_2(\mathrm{SNR}_2) = -\frac{1}{\theta_2 TB}\log_e \mathbb{E}_{\mathbf{z}}\left\{e^{-(1-\delta)\theta_2 TB\log_2\left(1+\frac{\mathrm{SNR}_2 z_2}{1-\delta}\right)}\right\}, \tag{6.6}$$

respectively,

Next, consider superposition coding with fixed decoding order. We fix the ratio $\frac{\mathrm{SNR}_1}{\mathrm{SNR}_2} = \lambda$. Additionally, we let τ denote the fraction of time in which the decoding order $(2,1)$ is employed. Note that if the decoding order is $(2,1)$, the receiver first decodes the second user's signal in the presence of interference from first user's signal, and subsequently decodes the first user's signal with no interference. Note that the symmetric case occurs when the decoding order is $(1,2)$ in the remaining $(1-\tau)$ fraction of the time. When this strategy is used, the maximum effective capacities in the region described in (6.2) can now be expressed as

$$\mathsf{C}_1(\mathrm{SNR}_1) = -\frac{1}{\theta_1 TB}\log_e \mathbb{E}_{\mathbf{z}}\left\{e^{-\theta_1 TB\left(\tau\log_2(1+\mathrm{SNR}_1 z_1)+(1-\tau)\log_2\left(1+\frac{\mathrm{SNR}_1 z_1}{1+\mathrm{SNR}_1 z_2/\lambda}\right)\right)}\right\}, \tag{6.7}$$

$$\mathsf{C}_2(\mathrm{SNR}_2) = -\frac{1}{\theta_2 TB}\log_e \mathbb{E}_{\mathbf{z}}\left\{e^{-\theta_2 TB\left(\tau\log_2\left(1+\frac{\mathrm{SNR}_2 z_2}{1+\lambda\mathrm{SNR}_2 z_1}\right)+(1-\tau)\log_2(1+\mathrm{SNR}_2 z_2)\right)}\right\}. \tag{6.8}$$

Finally, we turn our attention to superposition coding with variable decoding order. In this case, the decoding order depends on the fading coefficients (z_1, z_2).

We define $z_2 = g(\mathrm{SNR}_1) = g(\lambda \mathrm{SNR}_2)$ as the partition function in the $z_1 - z_2$ space[1]. Depending on which decoding order is employed in each region, we have different effective capacity expressions. If users are decoded in the order (1,2) when $z_2 < g(\mathrm{SNR}_1)$ and are decoded in the order (2,1) when $z_2 > g(\mathrm{SNR}_1)$, the effective capacities are given by

$$\mathsf{C}_1(\mathrm{SNR}_1) = -\frac{1}{\theta_1 TB}\log_e\left(\int_0^{\infty}\int_{g(\mathrm{SNR}_1)}^{\infty} e^{-\theta_1 TB\log_2\left(1+\mathrm{SNR}_1 z_1\right)} p_{\mathbf{z}}(z_1,z_2)dz_2dz_1 + \int_0^{\infty}\int_0^{g(\mathrm{SNR}_1)} e^{-\theta_1 TB\log_2\left(1+\frac{\mathrm{SNR}_1 z_1}{1+\mathrm{SNR}_1 z_2/\lambda}\right)} p_{\mathbf{z}}(z_1,z_2)dz_2dz_1\right), \quad (6.9)$$

$$\mathsf{C}_2(\mathrm{SNR}_2) = -\frac{1}{\theta_2 TB}\log_e\left(\int_0^{\infty}\int_0^{g(\lambda\mathrm{SNR}_2)} e^{-\theta_2 TB\log_2\left(1+\mathrm{SNR}_2 z_2\right)} p_{\mathbf{z}}(z_1,z_2)dz_2dz_1 + \int_0^{\infty}\int_{g(\lambda\mathrm{SNR}_2)}^{\infty} e^{-\theta_2 TB\log_2\left(1+\frac{\mathrm{SNR}_2 z_2}{1+\lambda\mathrm{SNR}_2 z_1}\right)} p_{\mathbf{z}}(z_1,z_2)dz_2dz_1\right). \quad (6.10)$$

Similar effective capacity expressions can be derived if users are decoded in the order (2,1) if $z_2 < g(\mathrm{SNR}_1)$ and decoded in the order (1,2) if $z_2 > g(\mathrm{SNR}_1)$.

Assumption 1 *Throughout the chapter, we consider the partition functions $g(SNR_1)$ that satisfy the following properties:*

1. *$g(0)$ is finite.*
2. *The first and second derivatives of g with respect to SNR_1, $\dot{g}(SNR_1)$ and $\ddot{g}(SNR_1)$, exist. Moreover, $\dot{g}(0)$ and $\ddot{g}(0)$ are finite.*

Denote $\frac{E_{b,i}}{N_0} = \frac{\mathrm{SNR}_i}{\mathsf{C}_i}$ as the bit energy of user $i = 1, 2$. The received bit energy is

$$\frac{E_{b,i}^r}{N_0} = \frac{E_{b,i}}{N_0}\mathbb{E}\{z_i\}. \quad (6.11)$$

[1]The partition function can in general be a function of z_1 as well, i.e., $g(\mathrm{SNR}_1) = g(z_1, \mathrm{SNR}_1)$.

As the following result shows, the minimum received bit energies for the different strategies are the same.

Theorem 16 *For all* $\lambda = \frac{SNR_1}{SNR_2}$ *and all* $g(z_1, SNR_1)$ *satisfying the properties in Assumption 1, the minimum received bit energy for the multiaccess fading channel attained through TDMA, superposition coding with fixed decoding order, or superposition decoding with varying decoding order, is the same and is given by*

$$\frac{E_{b,1}^r}{N_0}_{min} = \frac{E_{b,2}^r}{N_0}_{min} = \log_e 2 = -1.59 \quad dB. \tag{6.12}$$

Proof: Consider the TDMA strategy. Taking the first derivative of the functions in (6.5) and (6.6) and letting $\mathrm{SNR}_1 = 0$, $\mathrm{SNR}_2 = 0$, we obtain

$$\dot{\mathsf{C}}_1(0) = \frac{\mathbb{E}\{z_1\}}{\log_e 2}, \tag{6.13}$$

$$\dot{\mathsf{C}}_2(0) = \frac{\mathbb{E}\{z_2\}}{\log_e 2}. \tag{6.14}$$

Substituting (6.13) and (6.14) into (1.5), we have

$$\frac{E_{b,1}}{N_0}_{\min} = \frac{\log_e 2}{\mathbb{E}\{z_1\}}, \tag{6.15}$$

$$\frac{E_{b,2}}{N_0}_{\min} = \frac{\log_e 2}{\mathbb{E}\{z_2\}} \tag{6.16}$$

which imply (6.12) according to (6.11).

For the superposition coding with fixed decoding, evaluating the first derivative

of (6.7) and (6.8) at $\mathrm{SNR}_1 = 0$ and $\mathrm{SNR}_2 = 0$, we immediately obtain

$$\dot{\mathsf{C}}_1(0) = \frac{\mathbb{E}\{z_1\}}{\log_e 2} \tag{6.17}$$

$$\dot{\mathsf{C}}_2(0) = \frac{\mathbb{E}\{z_2\}}{\log_e 2} \tag{6.18}$$

which again imply (6.12) taking into consideration (1.5) and (6.11).

Next, we prove the result for the variable decoding case. First, we consider (6.9) and (6.10) with the associated decoding order assignment. The first derivative of (6.9) can be expressed as

$$\begin{aligned}
\dot{\mathsf{C}}_1(\mathrm{SNR}_1) &= -\frac{\dot{\phi}_1}{\beta_1 \phi_1 \log_e 2} \\
&= -\frac{1}{\beta_1 \phi_1 \log_e 2}\Bigg(- \int_0^\infty (1+\mathrm{SNR}_1 z_1)^{-\beta_1} p(z_1, g(\mathrm{SNR}_1)) \dot{g}(\mathrm{SNR}) dz_1 \\
&\quad - \beta_1 \int_0^\infty \int_{g(\mathrm{SNR}_1)}^\infty (1+\mathrm{SNR}_1 z_1)^{-\beta_1 - 1} z_1 p(z_1, z_2) dz_2 z_1 \\
&\quad + \int_0^\infty \left(1 + \frac{\mathrm{SNR}_1 z_1}{1+\mathrm{SNR}_1 g(\mathrm{SNR}_1)/\lambda}\right)^{-\beta_1} p(z_1, g(\mathrm{SNR}_1)) \dot{g}(\mathrm{SNR}_1) dz_1 \\
&\quad - \beta_1 \int_0^\infty \int_0^{g(\mathrm{SNR}_1)} \left(1 + \frac{\mathrm{SNR}_1 z_1}{1+\mathrm{SNR}_1 z_2/\lambda}\right)^{-\beta_1 - 1} \frac{z_1}{(1+\mathrm{SNR}_1 z_2/\lambda)^2} p(z_1, z_2) dz_2 dz_1 \Bigg)
\end{aligned} \tag{6.19}$$

where $\dot{\phi}_1$ is the first derivative of ϕ_1, which is defined as

$$\begin{aligned}
\phi_1 = &\int_0^\infty \int_{g(z_1, \mathrm{SNR}_1)}^\infty e^{-\theta_1 TB \log_2\left(1+\mathrm{SNR}_1 z_1\right)} p_{\mathbf{z}}(z_1, z_2) dz_2 dz_1 \\
&+ \int_0^\infty \int_0^{g(z_1, \mathrm{SNR}_1)} e^{-\theta_1 TB \log_2\left(1+\frac{\mathrm{SNR}_1 z_1}{1+\mathrm{SNR}_1 z_2/\lambda}\right)} p_{\mathbf{z}}(z_1, z_2) dz_2 dz_1.
\end{aligned} \tag{6.20}$$

Under the assumptions that $g(0)$ and $\dot{g}(0)$ are finite, we can easily see from (6.19)

that letting $\text{SNR}_1 = 0$ leads to

$$\dot{\mathsf{C}}_1(0) = \frac{\mathbb{E}\{z_1\}}{\log_e 2}. \tag{6.21}$$

Similarly, taking the first derivative of (6.10) and letting $\text{SNR}_2 = 0$, we obtain

$$\dot{\mathsf{C}}_2(0) = \frac{\mathbb{E}\{z_2\}}{\log_e 2}. \tag{6.22}$$

Applying the definitions (1.5) and (6.11), we prove (6.12) for this decoding order assignment. For the reverse decoding order assignment (i.e., users are decoded in the order (2,1) if $z_2 < g(\text{SNR}_1)$ and decoded in the order (1,2) if $z_2 > g(\text{SNR}_1)$), following similar steps, we again obtain the result in (6.12). □

Remark: The result of Theorem 16 shows that different transmission strategies (e.g., TDMA or superposition coding) and different reception schemes (e.g., fixed or variable decoding orders) lead to the same fundamental limit on the minimum bit energy. Similarly as in [27], TDMA is optimally efficient in the asymptotic regime in which the signal-to-noise ratio vanishes. More interestingly, we note that this result is obtained in the presence of QoS constraints. Additionally, the minimum bit energy is clearly independent of the QoS limitations parametrized by the QoS exponents θ_1 and θ_2. Hence, the energy efficiency is not adversely affected by the buffer constraints in this asymptotic regime in which $\text{SNR} \to 0$.

Having shown that the minimum bit energies achieved by different transmission and reception strategies are the same for each user, we note that the wideband slope regions have become more interesting since they quantify the performance in the non-asymptotic regime in which SNRs are small but nonzero. With the analysis approach introduced in [27], we have the following results.

Theorem 17 *The multiaccess slope region achieved by TDMA is given by*

$$\mathcal{S} = \left\{(\mathcal{S}_1, \mathcal{S}_2) : 0 \le \mathcal{S}_1 \le \mathcal{S}_1^{up}, \quad 0 \le \mathcal{S}_2 \le \mathcal{S}_2^{up}, \quad \frac{\kappa_{11}\kappa_{12}}{\kappa_{11} - \mathcal{S}_1} + \frac{\kappa_{21}\kappa_{22}}{\kappa_{21} - \mathcal{S}_2} \le 1 + \kappa_{12} + \kappa_{22}\right\} \tag{6.23}$$

where

$$\mathcal{S}_1^{up} = \frac{2(\mathbb{E}\{z_1\})^2}{\beta_1\left(\mathbb{E}\{z_1^2\} - (\mathbb{E}\{z_1\})^2\right) + \mathbb{E}\{z_1^2\}},$$

$$\mathcal{S}_2^{up} = \frac{2(\mathbb{E}\{z_2\})^2}{\beta_2\left(\mathbb{E}\{z_2^2\} - (\mathbb{E}\{z_2\})^2\right) + \mathbb{E}\{z_2^2\}},$$

$$\kappa_{11} = \frac{2(\mathbb{E}\{z_1\})^2}{\beta_1\left(\mathbb{E}\{z_1^2\} - (\mathbb{E}\{z_1\})^2\right)},$$

$$\kappa_{12} = \frac{\mathbb{E}\{z_1^2\}}{\beta_1\left(\mathbb{E}\{z_1^2\} - (\mathbb{E}\{z_1\})^2\right)},$$

$$\kappa_{21} = \frac{2(\mathbb{E}\{z_2\})^2}{\beta_2\left(\mathbb{E}\{z_2^2\} - (\mathbb{E}\{z_2\})^2\right)},$$

$$\kappa_{22} = \frac{\mathbb{E}\{z_2^2\}}{\beta_2\left(\mathbb{E}\{z_2^2\} - (\mathbb{E}\{z_2\})^2\right)},$$

$\beta_1 = \theta_1 TB \log_2 e$ *and* $\beta_2 = \theta_2 TB \log_2 e$.

Proof: Taking the second derivatives of the functions in (6.5) and (6.6) and letting $\mathrm{SNR}_1 = 0$, $\mathrm{SNR}_2 = 0$, we obtain

$$\ddot{\mathsf{C}}_1(0) = \frac{1}{\log_e 2}\left(\beta_1\left((\mathbb{E}\{z_1\})^2 - \mathbb{E}\{z_1^2\}\right) - \frac{1}{\delta}\mathbb{E}\{z_1^2\}\right) \tag{6.24}$$

and

$$\ddot{\mathsf{C}}_2(0) = \frac{1}{\log_e 2}\left(\beta_2\left((\mathbb{E}\{z_2\})^2 - \mathbb{E}\{z_2^2\}\right) - \frac{1}{1-\delta}\mathbb{E}\{z_2^2\}\right). \tag{6.25}$$

Combining (6.13), (6.14), (6.24), and (6.25) with (1.7), we now get

$$\mathcal{S}_1 = \frac{2(\mathbb{E}\{z_1\})^2}{\beta_1\left(\mathbb{E}\{z_1^2\} - (\mathbb{E}\{z_1\})^2\right) + \frac{1}{\delta}\mathbb{E}\{z_1^2\}} \tag{6.26}$$

$$\mathcal{S}_2 = \frac{2(\mathbb{E}\{z_2\})^2}{\beta_2\left(\mathbb{E}\{z_2^2\} - (\mathbb{E}\{z_2\})^2\right) + \frac{1}{1-\delta}\mathbb{E}\{z_2^2\}} \tag{6.27}$$

which, after eliminating δ, provide us the third condition in (6.23). □

The following results provide the wideband slope expressions when superposition transmission is employed.

Theorem 18 *For any $\lambda = \frac{SNR_1}{SNR_2}$, the multiaccess slope region achieved by the superposition coding with fixed decoding order is*

$$\mathcal{S} = \Bigg\{(\mathcal{S}_1, \mathcal{S}_2) : 0 \le \mathcal{S}_1 \le \mathcal{S}_1^{up}, \quad 0 \le \mathcal{S}_2 \le \mathcal{S}_2^{up}, \\ \frac{\lambda(\mathbb{E}\{z_1\})^2}{\mathbb{E}\{z_1 z_2\}}\left(\frac{1}{\mathcal{S}_1} - \frac{1}{\mathcal{S}_1^{up}}\right) + \frac{(\mathbb{E}\{z_2\})^2}{\lambda\mathbb{E}\{z_1 z_2\}}\left(\frac{1}{\mathcal{S}_2} - \frac{1}{\mathcal{S}_2^{up}}\right) = 1\Bigg\}, \tag{6.28}$$

where $\mathcal{S}_1^{up}$ and $\mathcal{S}_2^{up}$ are the same as defined in Theorem 17.

Proof: The second derivatives of the functions (6.7) and (6.8) at zero signal-to-noise ratio are

$$\ddot{\mathsf{C}}_1(0) = \frac{1}{\log_e 2}\left(\beta_1(\mathbb{E}\{z_1\})^2 - (\beta_1+1)\mathbb{E}\{z_1^2\} - \frac{2(1-\tau)}{\lambda}\mathbb{E}\{z_1 z_2\}\right)$$

$$\ddot{\mathsf{C}}_2(0) = \frac{1}{\log_e 2}\left(\beta_2(\mathbb{E}\{z_2\})^2 - (\beta_2+1)\mathbb{E}\{z_2^2\} - 2\lambda\tau\mathbb{E}\{z_1 z_2\}\right). \tag{6.29}$$

Then, the wideband slopes are given by

$$\mathcal{S}_1 = \frac{2(\mathbb{E}\{z_1\})^2}{\beta_1\left(\mathbb{E}\{z_1^2\} - (\mathbb{E}\{z_1\})^2\right) + \mathbb{E}\{z_1^2\} + \frac{2(1-\tau)}{\lambda}\mathbb{E}\{z_1 z_2\}} \tag{6.30}$$

$$\mathcal{S}_2 = \frac{2(\mathbb{E}\{z_2\})^2}{\beta_2\left(\mathbb{E}\{z_2^2\} - (\mathbb{E}\{z_2\})^2\right) + \mathbb{E}\{z_2^2\} + 2\lambda\tau\mathbb{E}\{z_1 z_2\}}. \tag{6.31}$$

After solving for τ in (6.30) and (6.31) and subtracting the resulting equations, we obtain the third condition in (6.28). □

Theorem 19 *For any $\lambda = \frac{SNR_1}{SNR_2}$, and any $g(SNR_1)$ satisfying the properties in Assumption 1, the multiaccess slope region achieved by superposition coding with variable decoding order is*

$$\mathcal{S} = \Bigg\{(\mathcal{S}_1, \mathcal{S}_2) : 0 \le \mathcal{S}_1 \le \mathcal{S}_1^{up}, \quad 0 \le \mathcal{S}_2 \le \mathcal{S}_2^{up}, \\ \frac{\lambda(\mathbb{E}\{z_1\})^2}{\mathbb{E}\{z_1 z_2\}}\left(\frac{1}{\mathcal{S}_1} - \frac{1}{\mathcal{S}_1^{up}}\right) + \frac{(\mathbb{E}\{z_2\})^2}{\lambda\mathbb{E}\{z_1 z_2\}}\left(\frac{1}{\mathcal{S}_2} - \frac{1}{\mathcal{S}_2^{up}}\right) = 1\Bigg\}, \tag{6.32}$$

where $\mathcal{S}_1^{up}$ and $\mathcal{S}_2^{up}$ are the same as defined in Theorem 17.

Proof: See Appendix H.

Remark: Comparing (H.4) with (H.6) or (H.5) with (H.7), we see that different decoding orders do not change the wideband slope values for given user only if $g(0) = z_1$, i.e., the $z_1 - z_2$ space is equally divided. One more interesting remark is that if we compare the third conditions in (6.28) and (6.32), we notice that fixed decoding order achieves the same performance as variable decoding order.

Remark: It is interesting to note in the above results that, unlike the minimum bit energy levels, the wideband slopes depend on the QoS exponents θ_1 and θ_2 through β_1 and β_2. Indeed, as can be seen from the expressions of the upper bounds $\mathcal{S}_1^{up}$ and $\mathcal{S}_2^{up}$, the wideband slopes tend to diminish as QoS constraints become more stringent

and θ_1 and θ_2 increase. Therefore, a penalty in energy efficiency is experienced in the presence of buffer limitations.

In the following result, we establish the suboptimality of TDMA.

Theorem 20 *The wideband slope region of TDMA is inside the one attained with superposition coding.*

Proof: We only need to consider the third conditions of (6.23) and (6.28). Substituting (6.30) and (6.31) into the left-hand side (LHS) of the third constraint in (6.23), we obtain

$$\kappa_{12}+\kappa_{22}+\frac{\mathbb{E}\{z_1^2\}}{\mathbb{E}\{z_1^2\}+\frac{2(1-\tau)}{\lambda}\mathbb{E}\{z_1z_2\}}+\frac{\mathbb{E}\{z_2^2\}}{\mathbb{E}\{z_2^2\}+2\lambda\tau\mathbb{E}\{z_1z_2\}}. \tag{6.33}$$

Comparing the sum of the last two terms with 1 (or more precisely subtracting 1 from the sum), we can write

$$\begin{aligned}&\frac{\mathbb{E}\{z_1^2\}}{\mathbb{E}\{z_1^2\}+\frac{2(1-\tau)}{\lambda}\mathbb{E}\{z_1z_2\}}+\frac{\mathbb{E}\{z_2^2\}}{\mathbb{E}\{z_2^2\}+2\lambda\tau\mathbb{E}\{z_1z_2\}}-1\\&=\frac{\mathbb{E}\{z_1^2\}\mathbb{E}\{z_2^2\}-4\tau\left(\mathbb{E}\{z_1z_2\}\right)^2+4\left(\mathbb{E}\{z_1z_2\}\right)^2\tau^2}{\left(\mathbb{E}\{z_1^2\}+\frac{2(1-\tau)}{\lambda}\mathbb{E}\{z_1z_2\}\right)\left(\mathbb{E}\{z_2^2\}+2\lambda\tau\mathbb{E}\{z_1z_2\}\right)}.\end{aligned} \tag{6.34}$$

We are interested in the numerator which is a quadratic function of the parameter τ. We note that the discriminant of this quadratic function satisfies

$$\begin{aligned}\Delta&=16\left(\mathbb{E}\{z_1z_2\}\right)^4-16\left(\mathbb{E}\{z_1z_2\}\right)^2\mathbb{E}\{z_1^2\}\mathbb{E}\{z_2^2\}\\&=16\left(\mathbb{E}\{z_1z_2\}\right)^2\left(\left(\mathbb{E}\{z_1z_2\}\right)^2-\mathbb{E}\{z_1^2\}\mathbb{E}\{z_2^2\}\right)\le 0\end{aligned} \tag{6.35}$$

where the Cauchy-Schwarz inequality $\left(\mathbb{E}\{z_1z_2\}\right)^2\le\mathbb{E}\{z_1^2\}\mathbb{E}\{z_2^2\}$ is used. Thus, the numerator of (6.34) is always nonnegative, i.e., the slope region achieved by TDMA

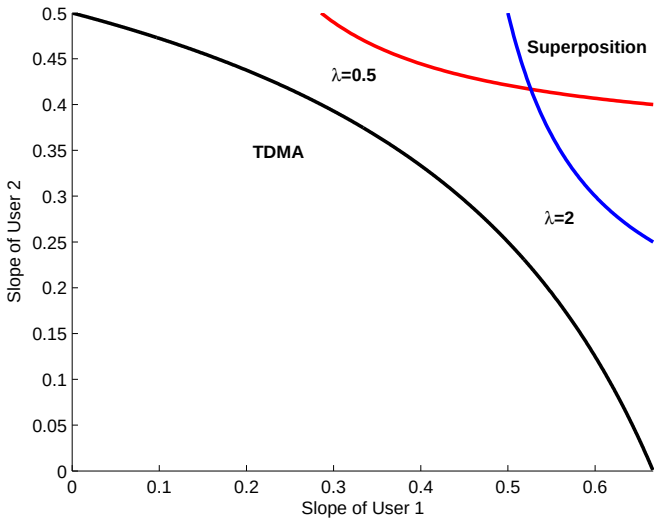

Figure 6.1: The slope regions for independent Rayleigh fading channels.

is inside the one achieved by superposition coding. The equality holds only if z_1 and z_2 are linearly dependent. □

In Fig. 6.1, we plot the slope regions in independent Rayleigh fading channels with variances $\mathbb{E}\{z_1\} = \mathbb{E}\{z_2\} = 1$. We assume $\beta_1 = 1$ and $\beta_2 = 2$. From the figure, we immediately observe the suboptimality of TDMA compared with superposition coding.

6.3 Energy Efficiency in the Wideband Regime

In this section, we consider the wideband regime in which the overall bandwidth of the system B is large. Let $\zeta = \frac{1}{B}$. Similar as in Chapters 2, we know that the minimum bit energy achieved in sparse multipath fading channels[2] as $B \to \infty$ (or

[2]As discussed in Chapters 2 and 3, wideband and low-power regimes are equivalent if rich multipath fading is experienced. Hence, in such a case, the same minimum bit energy and wideband slope expressions are obtained in both regimes.

equivalently $\zeta \to 0$) can be expressed as

$$\frac{E_{b,i}}{N_{0\ \min}} = \lim_{\zeta \to 0} \frac{\bar{P}_i \zeta / N_0}{\mathsf{C}_i(\zeta)} = \frac{\bar{P}_i / N_0}{\dot{\mathsf{C}}_i(0)},\ i = 1, 2. \tag{6.36}$$

To make the analysis more clear, below we first express the capacity expressions in (6.5)–(6.10) as functions of ζ. (6.5) and (6.6) can be rewritten as

$$\mathsf{C}_1(\zeta) = -\frac{\zeta}{\theta_1 T} \log_e \mathbb{E}_{\mathbf{z}} \left\{ e^{-\frac{\delta \theta_1 T}{\zeta} \log_2 \left(1 + \frac{\bar{P}_1 z_1 \zeta}{\delta N_0}\right)} \right\}, \tag{6.37}$$

and

$$\mathsf{C}_2(\zeta) = -\frac{\zeta}{\theta_2 T} \log_e \mathbb{E}_{\mathbf{z}} \left\{ e^{-\frac{(1-\delta) \theta_2 T}{\zeta} \log_2 \left(1 + \frac{\bar{P}_2 z_2 \zeta}{(1-\delta) N_0}\right)} \right\}, \tag{6.38}$$

respectively.

For superposition coding with fixed decoding order, and fixed $\lambda = \frac{\mathrm{SNR}_1}{\mathrm{SNR}_2} = \frac{\bar{P}_1 \zeta / N_0}{\bar{P}_2 \zeta / N_0} = \frac{\bar{P}_1}{\bar{P}_2}$, (6.7) and (6.8) now become

$$\mathsf{C}_1(\zeta) = -\frac{\zeta}{\theta_1 T} \log_e \mathbb{E}_{\mathbf{z}} \left\{ e^{-\frac{\theta_1 T}{\zeta} \left(\tau \log_2 (1 + \frac{\bar{P}_1 z_1 \zeta}{N_0}) + (1-\tau) \log_2 \left(1 + \frac{\frac{\bar{P}_1 z_1 \zeta}{N_0}}{1 + \frac{\bar{P}_2 z_2 \zeta}{N_0}} \right) \right)} \right\}, \tag{6.39}$$

$$\mathsf{C}_2(\zeta) = -\frac{\zeta}{\theta_2 T} \log_e \mathbb{E}_{\mathbf{z}} \left\{ e^{-\frac{\theta_2 T}{\zeta} \left(\tau \log_2 \left(1 + \frac{\frac{\bar{P}_2 z_2 \zeta}{N_0}}{1 + \frac{\bar{P}_1 z_1 \zeta}{N_0}} \right) + (1-\tau) \log_2 (1 + \frac{\bar{P}_2 z_2 \zeta}{N_0}) \right)} \right\}. \tag{6.40}$$

Note that we can write $g(\mathrm{SNR}_1)$ as $g\left(\frac{\bar{P}_1 \zeta}{N_0}\right)$, so similarly we can write (6.9) and

(6.10) as functions of ζ

$$\mathsf{C}_1(\zeta) = -\frac{\zeta}{\theta_1 T}\log_e\left(\int_0^\infty\int_{g(\frac{\bar{P}_1\zeta}{N_0})}^\infty e^{-\frac{\theta_1 T}{\zeta}\log_2\left(1+\frac{\bar{P}_1 z_1\zeta}{N_0}\right)} p_{\mathbf{z}}(z_1,z_2)dz_2dz_1 + \int_0^\infty\int_0^{g(\frac{\bar{P}_1\zeta}{N_0})} e^{-\frac{\theta_1 T}{\zeta}\log_2\left(1+\frac{\frac{\bar{P}_1 z_1\zeta}{N_0}}{1+\frac{\bar{P}_2 z_2\zeta}{N_0}}\right)} p_{\mathbf{z}}(z_1,z_2)dz_2dz_1\right) \tag{6.41}$$

$$\mathsf{C}_2(\zeta) = -\frac{\zeta}{\theta_2 T}\log_e\left(\int_0^\infty\int_0^{g(\frac{\bar{P}_2\zeta}{N_0})} e^{-\frac{\theta_2 T}{\zeta}\log_2\left(1+\frac{\bar{P}_2 z_2\zeta}{N_0}\right)} p_{\mathbf{z}}(z_1,z_2)dz_2dz_1 + \int_0^\infty\int_{g(\frac{\bar{P}_2\zeta}{N_0})}^\infty e^{-\frac{\theta_2 T}{\zeta}\log_2\left(1+\frac{\frac{\bar{P}_2 z_2\zeta}{N_0}}{1+\frac{\bar{P}_1 z_1\zeta}{N_0}}\right)} p_{\mathbf{z}}(z_1,z_2)dz_2dz_1\right). \tag{6.42}$$

Then we immediately have the following result.

Theorem 21 *For all $g(SNR_1)$ satisfying the properties in Assumption 1, the minimum bit energies for the two-user multiaccess fading channel in the wideband regime attained through TDMA, superposition coding with fixed decoding order, and superposition decoding with varying decoding order, depend on the individual QoS constraints at the users and are given by*

$$\frac{E_{b,1}}{N_0}_{min} = \frac{-\frac{\theta_1 T\bar{P}_1}{N_0}}{\log_e \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0\log_e 2}z_1}\right\}}, \tag{6.43}$$

$$\frac{E_{b,2}}{N_0}_{min} = \frac{-\frac{\theta_2 T\bar{P}_2}{N_0}}{\log_e \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0\log_e 2}z_2}\right\}}. \tag{6.44}$$

respectively.

Proof: See Appendix I.

Remark: As Theorem 21 shows, the same minimum bit energy is achieved through different transmission strategies (e.g., TDMA or superposition coding) and different

reception schemes (e.g., fixed or variable decoding orders), and therefore TDMA is optimally efficient in the wideband regime as $B \to \infty$.

Remark: A stark difference from the result in Theorem 16 is that the minimum bit energy now varies with the specific QoS constraints at the users. When $\theta = 0$, we can immediately show that the right-hand sides of (6.43) and (6.44) become $\frac{\log_e 2}{\mathbb{E}\{z_1\}}$ and $\frac{\log_e 2}{\mathbb{E}\{z_2\}}$, respectively, which is equivalent to (6.12). For $\theta > 0$, the energy efficiency is now adversely affected by the buffer constraints in the wideband regime.

Similarly as in Section 6.2, we next investigate the wideband slopes in order to quantify the performances and energy efficiencies of different transmission and reception methods in the non-asymptotic regime in which the bandwidth B is large but finite. We have the following results.

Theorem 22 *In the wideband regime, the multiaccess slope region achieved by TDMA is given by*

$$\mathcal{S} = \left\{ (\mathcal{S}_1, \mathcal{S}_2) : 0 \le \mathcal{S}_1 \le \mathcal{S}_1^{up}, \quad 0 \le \mathcal{S}_2 \le \mathcal{S}_2^{up}, \quad \frac{\mathcal{S}_1}{\mathcal{S}_1^{up}} + \frac{\mathcal{S}_2}{\mathcal{S}_2^{up}} \le 1 \right\} \tag{6.45}$$

where

$$\mathcal{S}_1^{up} = 2\left(\frac{N_0 \log_e 2}{\theta_1 T \bar{P}_1}\right)^2 \frac{\mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1}\right\} \left(\log_e \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1}\right\}\right)^2}{\mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} z_1^2\right\}},$$

$$\mathcal{S}_2^{up} = 2\left(\frac{N_0 \log_e 2}{\theta_2 T \bar{P}_2}\right)^2 \frac{\mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2}\right\} \left(\log_e \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2}\right\}\right)^2}{\mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} z_2^2\right\}}.$$

Proof: The second derivatives of (6.37) and (6.38) at $\zeta = 0$ are

$$\ddot{\mathsf{C}}_1(0) = -\frac{1}{\delta \log_e 2}\left(\frac{\bar{P}_1}{N_0}\right)^2 \frac{\mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2}z_1} z_1^2\right\}}{\mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2}z_1}\right\}} \tag{6.46}$$

$$\ddot{\mathsf{C}}_2(0) = -\frac{1}{(1-\delta) \log_e 2}\left(\frac{\bar{P}_2}{N_0}\right)^2 \frac{\mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2}z_2} z_2^2\right\}}{\mathbb{E}_{z_2}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2}z_2}\right\}} \tag{6.47}$$

Using the definition in (1.7), we can express the wideband slopes as

$$\mathcal{S}_1 = 2\delta\left(\frac{N_0 \log_e 2}{\theta_1 T\bar{P}_1}\right)^2 \frac{\mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2}z_1}\right\}\left(\log_e \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2}z_1}\right\}\right)^2}{\mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2}z_1} z_1^2\right\}} \tag{6.48}$$

$$\mathcal{S}_2 = 2(1-\delta)\left(\frac{N_0 \log_e 2}{\theta_2 T\bar{P}_2}\right)^2 \frac{\mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2}z_2}\right\}\left(\log_e \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2}z_2}\right\}\right)^2}{\mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2}z_2} z_2^2\right\}} \tag{6.49}$$

which after simple computation give us the third condition in (6.45). □

Theorem 23 *In the wideband regime, the multiaccess slope region achieved by superposition coding with fixed decoding order is*

$$\begin{aligned}\mathcal{S} = \Bigg\{&(\mathcal{S}_1, \mathcal{S}_2) : 0 \le \mathcal{S}_1 \le \mathcal{S}_1^{up}, \quad 0 \le \mathcal{S}_2 \le \mathcal{S}_2^{up}, \\ &\left(\frac{N_0 \log_e 2}{\theta_1 T}\right)^2 \frac{\left(\log_e \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2}z_1}\right\}\right)^2 \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2}z_1}\right\}}{\bar{P}_1\bar{P}_2\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2}z_1} z_1 z_2\right\}}\left(\frac{1}{\mathcal{S}_1} - \frac{1}{\mathcal{S}_1^{up}}\right) \\ &+ \left(\frac{N_0 \log_e 2}{\theta_2 T}\right)^2 \frac{\left(\log_e \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2}z_2}\right\}\right)^2 \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2}z_2}\right\}}{\bar{P}_1\bar{P}_2\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2}z_2} z_1 z_2\right\}}\left(\frac{1}{\mathcal{S}_2} - \frac{1}{\mathcal{S}_2^{up}}\right) = 1\Bigg\}\end{aligned} \tag{6.50}$$

where $\mathcal{S}_1^{up}$ and $\mathcal{S}_2^{up}$ are defined in Theorem 22.

Proof: Evaluating the second derivatives of (6.39) and (6.40) at $\zeta = 0$ yields

$$\ddot{\mathsf{C}}_1(0) = -\frac{1}{\log_e 2} \frac{\left(\frac{\bar{P}_1}{N_0}\right)^2 \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} z_1^2\right\} + \frac{2(1-\tau)\bar{P}_1\bar{P}_2}{N_0^2}\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} z_1 z_2\right\}}{\mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1}\right\}} \tag{6.51}$$

$$\ddot{\mathsf{C}}_2(0) = -\frac{1}{\log_e 2} \frac{\left(\frac{\bar{P}_2}{N_0}\right)^2 \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} z_2^2\right\} + \frac{2\tau\bar{P}_1\bar{P}_2}{N_0^2}\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} z_1 z_2\right\}}{\mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2}\right\}} \tag{6.52}$$

and as a result, the wideband slopes are given by

$$\mathcal{S}_1 = 2\left(\frac{N_0 \log_e 2}{\theta_1 T}\right)^2 \frac{\left(\log_e \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1}\right\}\right)^2 \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1}\right\}}{\bar{P}_1^2 \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} z_1^2\right\} + 2(1-\tau)\bar{P}_1\bar{P}_2\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} z_1 z_2\right\}} \tag{6.53}$$

$$\mathcal{S}_2 = 2\left(\frac{N_0 \log_e 2}{\theta_2 T}\right)^2 \frac{\left(\log_e \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2}\right\}\right)^2 \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2}\right\}}{\bar{P}_2^2 \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} z_2^2\right\} + 2\tau\bar{P}_1\bar{P}_2\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} z_1 z_2\right\}} \tag{6.54}$$

After solving for τ in (6.53) and (6.54) and subtracting the resulting equations, we have the third condition in (6.50). □

Theorem 24 *For any $g(SNR_1)$ satisfying the properties in Assumption 1, the multi-access slope regions achieved by superposition coding with variable decoding order in*

the wideband regime are different for different decoding orders. The slope region is

$$\mathcal{S} = \bigcup_{\{g(0)\}} \Bigg\{ (\mathcal{S}_1, \mathcal{S}_2) :$$

$$0 \le \mathcal{S}_1 \le 2 \left(\frac{N_0 \log_e 2}{\theta_1 T} \right)^2 \frac{\left(\log_e \mathbb{E}_{z_1} \left\{ e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} \right\} \right)^2 \mathbb{E}_{z_1} \left\{ e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} \right\}}{\bar{P}_1^2 \mathbb{E}_{z_1} \left\{ e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} z_1^2 \right\} + 2 \bar{P}_1 \bar{P}_2 \int_0^\infty \int_0^{g(0)} e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} z_1 z_2 p(z_1, z_2) dz_2 dz_1}$$

$$0 \le \mathcal{S}_2 \le 2 \left(\frac{N_0 \log_e 2}{\theta_2 T} \right)^2 \frac{\left(\log_e \mathbb{E}_{z_2} \left\{ e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} \right\} \right)^2 \mathbb{E}_{z_2} \left\{ e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} \right\}}{\bar{P}_2^2 \mathbb{E}_{z_2} \left\{ e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} z_2^2 \right\} + 2 \bar{P}_1 \bar{P}_2 \int_0^\infty \int_{g(0)}^\infty e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} z_1 z_2 p(z_1, z_2) dz_2 dz_1} \Bigg\} \tag{6.55}$$

if the decoding order is (1,2) when $z_2 < g(z_1, SNR_1)$, and the decoding order is (2,1) when $z_2 > g(z_1, SNR_1)$. The slope region is

$$\mathcal{S} = \bigcup_{\{g(0)\}} \Bigg\{ (\mathcal{S}_1, \mathcal{S}_2) :$$

$$0 \le \mathcal{S}_1 \le 2 \left(\frac{N_0 \log_e 2}{\theta_1 T} \right)^2 \frac{\left(\log_e \mathbb{E}_{z_1} \left\{ e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} \right\} \right)^2 \mathbb{E}_{z_1} \left\{ e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} \right\}}{\bar{P}_1^2 \mathbb{E}_{z_1} \left\{ e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} z_1^2 \right\} + 2 \bar{P}_1 \bar{P}_2 \int_0^\infty \int_{g(0)}^\infty e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} z_1 z_2 p(z_1, z_2) dz_2 dz_1}$$

$$0 \le \mathcal{S}_2 \le 2 \left(\frac{N_0 \log_e 2}{\theta_2 T} \right)^2 \frac{\left(\log_e \mathbb{E}_{z_2} \left\{ e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} \right\} \right)^2 \mathbb{E}_{z_2} \left\{ e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} \right\}}{\bar{P}_2^2 \mathbb{E}_{z_2} \left\{ e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} z_2^2 \right\} + 2 \bar{P}_1 \bar{P}_2 \int_0^\infty \int_0^{g(0)} e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} z_1 z_2 p(z_1, z_2) dz_2 dz_1} \Bigg\} \tag{6.56}$$

if the decoding order is (2,1) when $z_2 < g(z_1, SNR_1)$, and the decoding order is (1,2) when $z_2 > g(z_1, SNR_1)$.

Proof: See Appendix J.

Remark: Unlike previous discussions, we have no closed form expression for the wideband slope region achieved by superposition coding with variable decoding order in the wideband regime. Another observation in the above result is that different

decoding orders can result in different wideband slope regions.

Below we show the superiority of superposition coding with variable decoding compared with fixed decoding order.

Theorem 25 *Superposition coding with variable decoding order achieves better performance in terms of wideband slope region with respect to superposition coding with fixed decoding order.*

Proof: See Appendix K.

In the following, we present the condition under which the suboptimality of TDMA compared with superposition coding with fixed decoding order can be established.

Theorem 26 *If the following is satisfied*

$$\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0\log_e 2}z_1}z_1z_2\right\}\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0\log_e 2}z_2}z_1z_2\right\}\leq\mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0\log_e 2}z_1}z_2^2\right\}\mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0\log_e 2}z_2}z_2^2\right\}, \tag{6.57}$$

then the wideband slope region of TDMA is inside the one attained with superposition coding with fixed decoding order.

Proof: We consider the third conditions in (6.45) and (6.50). Substituting (6.53) and (6.54) into the LHS of the third condition in (6.45), we have

$$\begin{aligned}1-&\frac{2(1-\tau)\bar{P}_2\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0\log_e 2}z_1}z_1z_2\right\}}{\bar{P}_1\mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0\log_e 2}z_1}z_1^2\right\}+2(1-\tau)\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0\log_e 2}z_1}z_1z_2\right\}}\\&+\frac{\bar{P}_2\mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0\log_e 2}z_2}z_2^2\right\}}{\bar{P}_2\mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0\log_e 2}z_2}z_2^2\right\}+2\tau\bar{P}_1\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0\log_e 2}z_2}z_1z_2\right\}}\end{aligned} \tag{6.58}$$

So if the wideband slope region is inside the one attained with superposition coding with fixed decoding order, we must have the above value to be greater than 1 for all $0 \leq \tau \leq 1$. After subtracting 1 from (6.58), we can obtain

$$\begin{aligned}
&\frac{\bar{P}_1}{\bar{P}_1 \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2} z_1} z_1^2\right\} + 2(1-\tau)\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2} z_1} z_1 z_2\right\}} \\
&\frac{\bar{P}_2}{\bar{P}_2 \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2} z_2} z_2^2\right\} + 2\tau\bar{P}_1 \mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2} z_2} z_1 z_2\right\}} \\
&\times \left(4\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2} z_1} z_1 z_2\right\} \mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2} z_2} z_1 z_2\right\} \tau^2 \right. \\
&\qquad - 4\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2} z_1} z_1 z_2\right\} \mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2} z_2} z_1 z_2\right\} \tau \\
&\qquad \left. + \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2} z_1} z_1^2\right\} \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2} z_2} z_2^2\right\} \right) \qquad (6.59)
\end{aligned}$$

The first two terms of the multiplication are positive values. The minimum value of the third term which is a quadratic function of τ is achieved at $\tau = \frac{1}{2}$, and the minimum value is

$$\mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2} z_1} z_1^2\right\} \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2} z_2} z_2^2\right\} - \mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2} z_1} z_1 z_2\right\} \mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2} z_2} z_1 z_2\right\} \qquad (6.60)$$

Thus, we obtain the condition stated in (6.57) for TDMA to be suboptimal. □

Remark: It is interesting that if the condition (6.57) is not satisfied, TDMA can achieve some points outside the wideband slope region attained with superposition coding with fixed decoding order. This tells us that TDMA can be a better choice compared with superposition coding with fixed decoding order in some cases.

In the numerical results, we plot the wideband slope regions for independent Rayleigh fading channels with variances $\mathbb{E}\{z_1\} = \mathbb{E}\{z_2\} = 1$. We assume $\theta_1 = 0.01$,

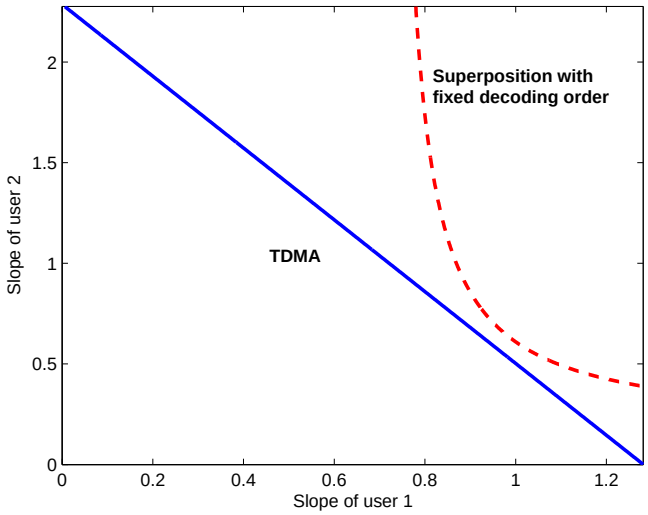

Figure 6.2: The slope regions for independent Rayleigh fading channels.

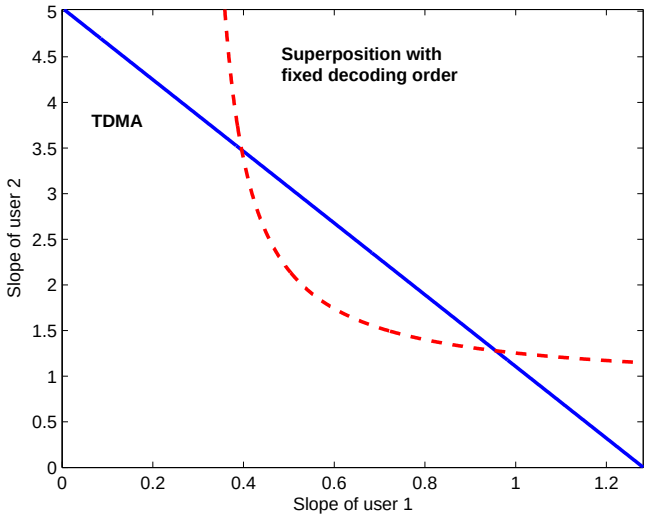

Figure 6.3: The slope regions for independent Rayleigh fading channels.

$\theta_2 = 0.1$, $T = 2$ ms. In Fig. 6.2, we assume $\frac{\bar{P}_1}{N_0} = 2\frac{\bar{P}_2}{N_0} = 10^4$. The left-hand side of (6.57) is 0.1009, while the right-hand side is 0.1283. Hence, the inequality is satisfied. From the figure, we can see that TDMA is suboptimal compared with superposition coding. In Fig. 6.3, we assume $\frac{\bar{P}_1}{N_0} = \frac{1}{2}\frac{\bar{P}_2}{N_0} = 10^4$. The left-hand side of (6.57) is 0.0131, while the right-hand side is 0.006. Hence, the inequality is not satisfied. Confirming the above discussion, we can observe in the figure that TDMA indeed achieves points outside the slope region attained with superposition coding with fixed decoding order.

6.4 Conclusion

In this chapter, we have analyzed the energy efficiency of multiaccess fading channels under QoS constraints by employing the effective capacity as a measure of the maximal throughput under QoS constraints. We have characterized the minimum bit energy and the wideband slope regions for different transmission strategies. We have conducted our analysis in two regimes: low-power regime and wideband regime. Through this analysis, we have shown the impact of QoS constraints on the energy efficiency of multiaccess fading channels. More specifically, we have found that the minimum bit energies are the same for each user when different transmission and reception techniques are employed. While these minimum values are equal those that can be attained in the absence of QoS constraints in the low-power regime, we have shown that strictly higher bit energy values, which depend on the QoS constraints, are needed in the wideband regime. We have also seen that while TDMA is suboptimal in the low-power regime when wideband slope regions are considered, it can outperform superposition coding with fixed decoding order in the wideband regime. Moreover, we have proven in the wideband regime that varying the decoding order can achieve larger slope region when compared with fixed decoding order for superposition coding.

Numerical results validating our results are provided as well.

Chapter 7

Throughput for Two-Hop Communication Systems

In this chapter, we consider two-hop wireless links and investigate the throughput in the presence of QoS constraints by studying the effective capacity. We note that references [68] and [69] have also recently investigated the effective capacity of relay channels. Tang and Zhang in [68] analyzed the power allocation policies in relay networks under the assumption that the relay node has no buffer constraints. Parag and Chamberland in [69] provided a queueing analysis of a butterfly network with constant rate for each link. However, they assumed that there is no congestion at the intermediate nodes. In this work, as a significant departure from previous studies, we assume that both the source and the relay nodes are subject to QoS constraints specified by the QoS exponents θ_1 and θ_2. Now, we face a more challenging scenario in which the buffer constraints at the source and relay interact. Moreover, we consider a general relay channel model in which the fading coefficients for each link can have arbitrary distributions. We concentrate on the decode-and-forward (DF) relaying scheme. Assuming that the relay operates in full-duplex or half-duplex mode, we

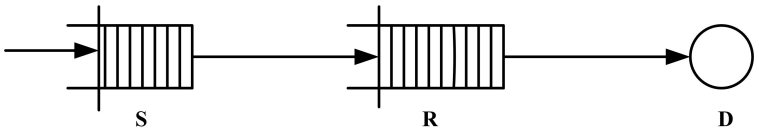

Figure 7.1: The system model.

determine the effective capacity as a function of θ_1 and θ_2. Through this analysis, we characterize the impact of the presence of QoS constraints at the relay and also of half-duplex operation on the throughput of the two-hop link.

7.1 System Model and Preliminaries

7.1.1 System Model

The two-hop communication link is depicted in Figure 7.1. In this model, source **S** is sending information to the destination **D** with the help of the intermediate relay node **R**. We assume that there is no direct link between **S** and **D** (which, for instance, holds, if these nodes are sufficiently far apart in distance). Both the source and the intermediate relay node operate under QoS constraints (i.e., buffer constraints) specified by the QoS exponents θ_1 and θ_2, respectively. Hence, the source and relay buffer violation probabilities should, for some large $Q_{\max}$, satisfy

$$\Pr\{Q_s \geq Q_{\max}\} \approx e^{-\theta_1 Q_{\max}} \tag{7.1}$$

and

$$\Pr\{Q_r \geq Q_{\max}\} \approx e^{-\theta_2 Q_{\max}}, \tag{7.2}$$

respectively. Above, Q_s and Q_r denote the stationary queue lengths at the source and relay, respectively.

We consider both full-duplex and half-duplex relay operation. The full-duplex relay can receive and transmit simultaneously while the half-duplex relay first listens and then transmits. Therefore, reception and transmission at the half-duplex relay occur in non-overlapping intervals.

Next, we identify the discrete-time input and output relationships. In the i^{th} symbol duration, the signal Y_r received at the relay from the source and the signal Y_d received at the destination from the relay can be expressed as

$$Y_r[i] = g_1[i]X_1[i] + n_1[i] \tag{7.3}$$

$$Y_d[i] = g_2[i]X_2[i] + n_2[i] \tag{7.4}$$

where X_j for $j = \{1, 2\}$ denote the inputs for the links $\mathbf{S}-\mathbf{R}$ and $\mathbf{R}-\mathbf{D}$, respectively. More specifically, X_1 is the signal sent from the source and X_2 is sent from the relay. The inputs are subject to individual average energy constraints $\mathbb{E}\{|X_j|^2\} \leq \bar{P}_j/B, j = \{1, 2\}$ where B is the bandwidth. Assuming that the symbol rate is B complex symbols per second, we can easily see that the symbol energy constraint of $\bar{P}_j/B$ implies that the channel input has a power constraint of $\bar{P}_j$. We assume that the fading coefficients $g_j, j = \{1, 2\}$ are jointly stationary and ergodic discrete-time processes, and we denote the magnitude-square of the fading coefficients by $z_j[i] = |g_j[i]|^2$. Above, in the channel input-output relationships, the noise component $n_j[i]$ is a zero-mean, circularly symmetric, complex Gaussian random variable with variance $\mathbb{E}\{|n_j[i]|^2\} = N_j$ for $j = 1, 2$. The additive Gaussian noise samples $\{n_j[i]\}$ are assumed to form an independent and identically distributed (i.i.d.) sequence. We denote the signal-to-noise ratios as $\text{SNR}_j = \frac{\bar{P}_j}{N_j B}$.

7.1.2 Characterization of Effective Capacity

We first state the following result from [18], which identifies the QoS exponent for given arrival and departure processes under certain conditions.

Lemma 1 ([18]) *Consider a queueing system, and suppose that the queue is stable and that both the arrival process $a[n], n = 1, 2, \ldots$ and service process $c[n], n = 1, 2, \ldots$ satisfy the Gärtner-Ellis limit, i.e., for all $\theta \geq 0$, there exists a differentiable asymptotic logarithmic moment generating function (LMGF) $\Lambda_A(\theta)$ defined as*[1]

$$\Lambda_A(\theta) = \lim_{n\to\infty} \frac{\log \mathbb{E}\{e^{\theta \sum_{i=1}^{n} a[i]}\}}{n}, \tag{7.5}$$

and a differentiable asymptotic LMGF $\Lambda_C(\theta)$ defined as

$$\Lambda_C(\theta) = \lim_{n\to\infty} \frac{\log \mathbb{E}\{e^{\theta \sum_{i=1}^{n} c[i]}\}}{n}. \tag{7.6}$$

If there exists a unique $\theta^* > 0$ such that

$$\Lambda_A(\theta^*) + \Lambda_C(-\theta^*) = 0, \tag{7.7}$$

then

$$\lim_{Q_{max}\to\infty} \frac{\log \Pr\{Q > Q_{max}\}}{Q_{max}} = -\theta^*. \tag{7.8}$$

where Q is the stationary queue length. □

Now, we discuss the implications of this result on the two-hop link we study. Assume that the constant arrival rate at the source is $R \geq 0$, and the channels

[1] Throughout the text, logarithm expressed without a base, i.e., $\log(\cdot)$, refers to the natural logarithm $\log_e(\cdot)$.

operate at their capacities. To satisfy the QoS constraint at the source, we should have

$$\tilde{\theta} \geq \theta_1 \tag{7.9}$$

where $\tilde{\theta}$ is the solution to

$$R = -\frac{\Lambda_{sr}(-\tilde{\theta})}{\tilde{\theta}} \tag{7.10}$$

and $\Lambda_{sr}(\theta)$ is the LMGF of the instantaneous capacity of the $\mathbf{S}-\mathbf{R}$ link.

According to [18], the LMGF of the departure process from the source, or equivalently the arrival process to the relay node, is given by

$$\Lambda_r(\theta) = \begin{cases} R\theta, & 0 \leq \theta \leq \tilde{\theta} \\ R\tilde{\theta} + \Lambda_{sr}(\theta - \tilde{\theta}), & \theta > \tilde{\theta} \end{cases}. \tag{7.11}$$

Therefore, in order to satisfy the QoS of the intermediate relay node $\mathbf{R}$, we must have

$$\hat{\theta} \geq \theta_2 \tag{7.12}$$

where $\hat{\theta}$ is the solution to

$$\Lambda_r(\hat{\theta}) + \Lambda_{rd}(-\hat{\theta}) = 0. \tag{7.13}$$

Above, $\Lambda_{rd}(\theta)$ is the LMGF of the instantaneous capacity of the $\mathbf{R}-\mathbf{D}$ link.

After these characterizations, effective capacity of the two-hop communication model can be formulated as follows.

Definition 3 *The effective capacity of the two-hop communication link with the QoS*

constraints specified by θ_1 at the source and θ_2 at the relay node is given by

$$R_E(\theta_1, \theta_2) = \sup_{R \in \mathcal{R}} R \tag{7.14}$$

where $\mathcal{R}$ is the collection of constant arrival rates R for which the solutions $\tilde{\theta}$ and $\hat{\theta}$ of (7.10) and (7.13) satisfy $\tilde{\theta} \geq \theta_1$ and $\hat{\theta} \geq \theta_2$, respectively. Hence, effective capacity is the maximum constant arrival rate that can be supported by the two-hop link in the presence of QoS constraints at both the source and relay nodes.

7.2 Effective Capacity of a Two-Hop Link in Block Fading Channels

We assume that the channel state information of the links $\mathbf{S}-\mathbf{R}$ and $\mathbf{R}-\mathbf{D}$ is available at $\mathbf{S}$ and $\mathbf{R}$, and the channel state information of the link $\mathbf{R}-\mathbf{D}$ is available at $\mathbf{R}$ and $\mathbf{D}$. The transmission power levels at the source and the intermediate-hop node are fixed and hence no power control is employed (i.e., nodes are subject to short-term power constraints). We further assume that the channel capacity for each link can be achieved, i.e., the service processes are equal to the instantaneous Shannon capacities of the links. Moreover, we consider a block fading scenario in which the fading stays constant for a block of T seconds and change independently from one block to another.

7.2.1 Full-Duplex Relay

In this part, we consider the full-duplex relay. The instantaneous capacities of the $\mathbf{S}-\mathbf{R}$ and $\mathbf{R}-\mathbf{D}$ links in each block are given, respectively, by

$$TB\log_2(1+\mathrm{SNR}_1 z_1) \quad \text{and} \quad TB\log_2(1+\mathrm{SNR}_2 z_2) \tag{7.15}$$

in the units of bits per block or equivalently bits per T seconds. These can be regarded as the service processes at the source and relay.

Under the block fading assumption, the logarithmic moment generating functions for the service processes of links $\mathbf{S}-\mathbf{R}$ and $\mathbf{R}-\mathbf{D}$ as functions of θ are given by[2] [35]

$$\Lambda_{sr}(\theta) = \log \mathbb{E}_{z_1}\left\{e^{\theta TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} \tag{7.16}$$

$$\Lambda_{rd}(\theta) = \log \mathbb{E}_{z_2}\left\{e^{\theta TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\} \tag{7.17}$$

and as a result

$$\Lambda_r(\theta) = \begin{cases} R\theta, & 0 \le \theta \le \tilde{\theta} \\ R\tilde{\theta} + \log \mathbb{E}_{z_1}\left\{e^{(\theta-\tilde{\theta})TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}, & \theta > \tilde{\theta} \end{cases}. \tag{7.18}$$

With these formulations for Λ_{sr}, Λ_{rd}, and Λ_r, we can now more explicitly express the equations in (7.10) and (7.13) as

$$R = g(\tilde{\theta}) = -\frac{1}{\tilde{\theta}}\log \mathbb{E}_{z_1}\left\{e^{-\tilde{\theta}TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} \tag{7.19}$$

[2]Due to the assumption that the fading changes independently from one block to another, we can, for instance, simplify (7.5) as $\Lambda_A = \lim_{n\to\infty}\frac{\log \mathbb{E}\{e^{\theta\sum_{i=1}^{n} a[i]}\}}{n} = \lim_{n\to\infty}\frac{\log\prod_{i=1}^{n}\mathbb{E}\{e^{\theta a[i]}\}}{n} = \lim_{n\to\infty}\frac{\sum_{i=1}^{n}\log\mathbb{E}\{e^{\theta a[i]}\}}{n} = \lim_{n\to\infty}\frac{n\log\mathbb{E}\{e^{\theta a[1]}\}}{n} = \log\mathbb{E}\{e^{\theta a[1]}\}$.

and

$$R = h(\tilde{\theta}, \hat{\theta}) = \begin{cases} -\frac{1}{\tilde{\theta}} \log \mathbb{E}_{z_2} \left\{ e^{-\hat{\theta} TB \log_2(1+\mathrm{SNR}_2 z_2)} \right\} & 0 \le \hat{\theta} \le \tilde{\theta} \\ -\frac{1}{\tilde{\theta}} \left(\log \mathbb{E}_{z_2} \left\{ e^{-\hat{\theta} TB \log_2(1+\mathrm{SNR}_2 z_2)} \right\} \right. & \\ \left. + \log \mathbb{E}_{z_1} \left\{ e^{(\hat{\theta}-\tilde{\theta}) TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\} \right) & \hat{\theta} \ge \tilde{\theta} \end{cases}, \tag{7.20}$$

respectively.

We seek to identify the constant arrival rates R that can be supported in the presence of QoS constraints specified by the QoS exponents θ_1 for the $\mathbf{S}-\mathbf{R}$ link and θ_2 for the $\mathbf{R}-\mathbf{D}$ link. In this quest, we have the following characterization. The rates R, which simultaneously satisfy the equations in (7.19) and (7.20) with some $\tilde{\theta} \ge \theta_1$ and $\hat{\theta} \ge \theta_2$, are the arrival rates that can be supported by the two-hop link while having the buffer violation probabilities, for large $Q_{\max}$, behave approximately as $\Pr\{Q_s \ge Q_{\max}\} \approx e^{-\tilde{\theta} Q_{\max}} \le e^{-\theta_1 Q_{\max}}$ and $\Pr\{Q_r \ge Q_{\max}\} \approx e^{-\hat{\theta} Q_{\max}} \le e^{-\theta_2 Q_{\max}}$, where Q_s and Q_r are the stationary queue lengths at the source and relay, respectively. We first establish an upper bound on these arrival rates.

Theorem 27 *The constant arrival rates, which can be supported by the two-hop link in the presence of QoS constraints with QoS exponents θ_1 and θ_2 at the source and relay, respectively, are upper bounded by*

$$R \le \min \left\{ -\frac{1}{\theta_1} \log \mathbb{E}_{z_1} \left\{ e^{-\theta_1 TB \log_2(1+SNR_1 z_1)} \right\}, -\frac{1}{\theta_2} \log \mathbb{E}_{z_2} \left\{ e^{-\theta_2 TB \log_2(1+SNR_2 z_2)} \right\} \right\}. \tag{7.21}$$

Proof: We can see from (7.9) and (7.19) that

$$R = -\frac{1}{\tilde{\theta}} \log \mathbb{E}_{z_1} \left\{ e^{-\tilde{\theta} TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\} \le -\frac{1}{\theta_1} \log \mathbb{E}_{z_1} \left\{ e^{-\theta_1 TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\}. \tag{7.22}$$

Note that the inequality above follows from the assumption that $\tilde{\theta} \geq \theta_1$ and the fact that $-\frac{\Lambda(-\tilde{\theta})}{\tilde{\theta}} = -\frac{1}{\tilde{\theta}} \log \mathbb{E}_{z_1} \left\{ e^{-\tilde{\theta} TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\}$ is a decreasing function of $\tilde{\theta}$ since larger $\tilde{\theta}$ implies a faster decay in the buffer violation probabilities and hence more stringent QoS constraints. Another upper bound can be obtained through the following arguments. Consider the idealistic scenario in which the $\mathbf{S} - \mathbf{R}$ link is deterministic (i.e., there is no fading) and can support any constant arrival rate R (i.e., the capacity of this link is unbounded and $\mathbf{R} - \mathbf{D}$ link is the bottleneck). In such a case, the arriving data can immediately be sent without waiting and consequently there is no need for buffering at the source. Hence, any source QoS constraint can be satisfied. More specifically, if the service rate matches the constant arrival rate, the equation in (7.10) holds for any $\tilde{\theta}$, i.e.,

$$R = -\frac{\Lambda_{sr}(-\tilde{\theta})}{\tilde{\theta}} = -\frac{1}{\tilde{\theta}} \log \mathbb{E} \left\{ e^{-\tilde{\theta} R} \right\} = -\frac{1}{\tilde{\theta}} (-\tilde{\theta} R) = R \tag{7.23}$$

where instantaneous service rate is assumed to be equal to the constant arrival rate R (rather than the random quantity $TB \log_2(1 + \mathrm{SNR}_1 z_1)$ as we have in the fading channel case). Since no buffering is now required at the source, we can freely impose the most strict QoS constraints and assume $\tilde{\theta}$ to be unbounded as well. Then, we have $\hat{\theta} \leq \tilde{\theta}$ for any $\hat{\theta}$. With this, we see from (7.20) that

$$R = -\frac{1}{\hat{\theta}} \log \mathbb{E}_{z_2} \left\{ e^{-\hat{\theta} TB \log_2(1+\mathrm{SNR}_2 z_2)} \right\} \leq -\frac{1}{\theta_2} \log \mathbb{E}_{z_2} \left\{ e^{-\theta_2 TB \log_2(1+\mathrm{SNR}_2 z_2)} \right\} \tag{7.24}$$

where, similarly as before, the inequality is due to the assumption that $\hat{\theta} \geq \theta_2$.

Combining the bounds in (7.22) and (7.24), we can equivalently write

$$R \leq \min\left\{-\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}, -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\}\right\} \tag{7.25}$$

concluding the proof. □

Remark: Note that $-\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}$ is the effective capacity of the $\mathbf{S}-\mathbf{R}$ link with QoS exponent θ_1. Similarly, $-\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\}$ is the effective capacity of the $\mathbf{R}-\mathbf{D}$ link with QoS exponent θ_2. Hence, the arrival rates that can be supported by the two-hop link are upper bounded by the minimum of the effective capacities of the individual links.

Below, we identify, for full-duplex relaying, the effective capacity of the two-hop link, i.e., maximum of the arrival rates that can be supported in the two-hop link in the presence of QoS constraints. According to [18], we know that the queues are not stable if the average transmission rate of link $\mathbf{R}-\mathbf{D}$ is less than the average transmission rate of link $\mathbf{S}-\mathbf{R}$. Therefore, in order to ensure stability, we assume that the condition $\mathbb{E}_{z_1}\{\log_2(1+\mathrm{SNR}_1 z_1)\} < \mathbb{E}_{z_2}\{\log_2(1+\mathrm{SNR}_2 z_2)\}$ is satisfied in the following result.

Theorem 28 *The effective capacity of the two-hop communication system as a function of θ_1 and θ_2 is given by the following:*

***Case I**: If $\theta_1 \geq \theta_2$,*

$$R_E(\theta_1,\theta_2) = \min\left\{-\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\theta_1 TB\log_2(1+SNR_1 z_1)}\right\}, -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+SNR_2 z_2)}\right\}\right\}. \tag{7.26}$$

Case II: *If* $\theta_1 < \theta_2$ *and* $\theta_2 \leq \bar{\theta}$,

$$R_E(\theta_1, \theta_2) = -\frac{1}{\theta_1} \log \mathbb{E}_{z_1} \left\{ e^{-\theta_1 TB \log_2(1+SNR_1 z_1)} \right\} \tag{7.27}$$

where $\bar{\theta}$ *is the unique value of* θ *for which we have the following equality satisfied:*

$$-\frac{1}{\theta_1} \log \mathbb{E}_{z_1} \left\{ e^{-\theta_1 TB \log_2(1+SNR_1 z_1)} \right\} = -\frac{1}{\theta_1} \Big(\log \mathbb{E}_{z_2} \left\{ e^{-\theta TB \log_2(1+SNR_2 z_2)} \right\} + \log \mathbb{E}_{z_1} \left\{ e^{(\theta-\theta_1) TB \log_2(1+SNR_1 z_1)} \right\} \Big). \tag{7.28}$$

Case III: *Assume* $\theta_1 < \theta_2$ *and* $\theta_2 > \bar{\theta}$.

III.a: *If* $-\frac{1}{\theta_2} \log \mathbb{E}_{z_2} \left\{ e^{-\theta_2 TB \log_2(1+SNR_2 z_2)} \right\} \geq -\frac{1}{\theta_2} \log \mathbb{E}_{z_1} \left\{ e^{-\theta_2 TB \log_2(1+SNR_1 z_1)} \right\}$, *then*

$$R_E(\theta_1, \theta_2) = -\frac{1}{\tilde{\theta}^*} \log \mathbb{E}_{z_1} \left\{ e^{-\tilde{\theta}^* TB \log_2(1+SNR_1 z_1)} \right\} \tag{7.29}$$

where $\tilde{\theta}^*$ *is the smallest solution to*

$$-\frac{1}{\tilde{\theta}} \log \mathbb{E}_{z_1} \left\{ e^{-\tilde{\theta} TB \log_2(1+SNR_1 z_1)} \right\} = -\frac{1}{\tilde{\theta}} \Big(\log \mathbb{E}_{z_2} \left\{ e^{-\theta_2 TB \log_2(1+SNR_2 z_2)} \right\} + \log \mathbb{E}_{z_1} \left\{ e^{(\theta_2-\tilde{\theta}) TB \log_2(1+SNR_1 z_1)} \right\} \Big). \tag{7.30}$$

III.b: *If* $-\frac{1}{\theta_2} \log \mathbb{E}_{z_2} \left\{ e^{-\theta_2 TB \log_2(1+SNR_2 z_2)} \right\} < -\frac{1}{\theta_2} \log \mathbb{E}_{z_1} \left\{ e^{-\theta_2 TB \log_2(1+SNR_1 z_1)} \right\}$ *and* $-\frac{1}{\theta_2} \log \mathbb{E}_{z_2} \left\{ e^{-\theta_2 TB \log_2(1+SNR_2 z_2)} \right\} \geq TB \log_2(1 + SNR_1 z_{1,\min})$,

$$R_E(\theta_1, \theta_2) = -\frac{1}{\tilde{\theta}^*} \log \mathbb{E}_{z_1} \left\{ e^{-\tilde{\theta}^* TB \log_2(1+SNR_1 z_1)} \right\} \tag{7.31}$$

where $z_{1,\min}$ is the essential infimum of z_1, and $\tilde{\theta}^*$ is the solution to

$$-\frac{1}{\tilde{\theta}}\log \mathbb{E}_{z_1}\left\{e^{-\tilde{\theta}TB\log_2(1+SNR_1 z_1)}\right\} = -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+SNR_2 z_2)}\right\}. \quad (7.32)$$

***III.c**: Otherwise,*

$$R_E(\theta_1,\theta_2) = -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+SNR_2 z_2)}\right\}. \quad (7.33)$$

Proof: See Appendix L.

Remark: We see that in Case I in which $\theta_1 \geq \theta_2$, the effective capacity upper bound identified in Proposition 27 is attained.

Remark: Note that if $\theta_1 \geq \theta_2$, then the source is operating under more stringent QoS constraints then the relay. In this case, if we have

$$-\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} \leq -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\}, \quad (7.34)$$

then

$$R_E(\theta_1,\theta_2) = -\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}. \quad (7.35)$$

Therefore, under these assumptions, the effective capacity is equal to the effective capacity of the $\mathbf{S}-\mathbf{R}$ link, and the performance is not affected by the presence of the buffer constraints at the relay node $\mathbf{R}$. This is because of the fact that the effective bandwidth of the departure process from the source can be completely supported by the $\mathbf{R}-\mathbf{D}$ link when the QoS exponent imposed at the relay node $\mathbf{R}$ is smaller.

The inequality in (7.34) is, for instance, satisfied when z_1 and z_2 (which are the fading powers in the $\mathbf{S}-\mathbf{R}$ and $\mathbf{R}-\mathbf{D}$ links) have the same distribution, and we

have $\text{SNR}_1 \leq \text{SNR}_2$. We can easily see that

$$-\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\text{SNR}_2 z_2)}\right\} \geq -\frac{1}{\theta_1}\log \mathbb{E}_{z_2}\left\{e^{-\theta_1 TB\log_2(1+\text{SNR}_2 z_2)}\right\} \tag{7.36}$$

$$\geq -\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\theta_1 TB\log_2(1+\text{SNR}_1 z_1)}\right\} \tag{7.37}$$

where (7.36) and (7.37) follow from the facts that $-\frac{1}{\theta}\log \mathbb{E}_z\left\{e^{-\theta TB\log_2(1+\text{SNR}z)}\right\}$ is a decreasing function in θ, and a increasing function in SNR. This discussion also suggests that even if the source operates under more strict buffer constraints, if the fading in the $\mathbf{R}-\mathbf{D}$ link is worse than that in the $\mathbf{S}-\mathbf{R}$ link and/or the signal-to-noise ratio of the relay is smaller, i.e., $\text{SNR}_1 \geq \text{SNR}_2$, then we can have

$$R_E(\theta_1,\theta_2) = \min\left\{ -\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\theta_1 TB\log_2(1+\text{SNR}_1 z_1)}\right\}, \right.$$
$$\left. -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\text{SNR}_2 z_2)}\right\} \right\} \tag{7.38}$$

$$= -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\text{SNR}_2 z_2)}\right\}, \tag{7.39}$$

and hence experience the $\mathbf{R}-\mathbf{D}$ link as the bottleneck.

7.2.2 Half-Duplex Relay

In the case of half-duplex relaying with a fixed time-sharing parameter $\tau \in (0,1)$, we assume that the source first transmits in the τ fraction of the block of T seconds during which the relay listens. Subsequently, in the remaining $(1-\tau)$ fraction of the time, the relay transmits to the destination. Hence, the transmission or service rates (in bits per T seconds) at the source and relay become

$$\tau TB\log_2(1+\text{SNR}_1 z_1) \quad \text{and} \quad (1-\tau)TB\log_2(1+\text{SNR}_2 z_2). \tag{7.40}$$

Now, the logarithmic moment generating functions for the service processes of links $\mathbf{S}-\mathbf{R}$ and $\mathbf{R}-\mathbf{D}$ as functions of θ are given by

$$\Lambda_{sr}(\theta) = \log \mathbb{E}_{z_1}\left\{e^{\tau\theta TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} \tag{7.41}$$

$$\Lambda_{rd}(\theta) = \log \mathbb{E}_{z_2}\left\{e^{(1-\tau)\theta TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\} \tag{7.42}$$

and as a result, we have

$$\Lambda_r(\theta) = \begin{cases} R\theta, & 0 \le \theta \le \tilde{\theta} \\ R\tilde{\theta} + \log \mathbb{E}_{z_1}\left\{e^{\tau(\theta-\tilde{\theta})TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}, & \theta > \tilde{\theta} \end{cases}.$$

With these expressions, equations in (7.10) and (7.13) can be written, for fixed τ, as

$$R = g(\tilde{\theta}) = -\frac{1}{\tilde{\theta}}\log \mathbb{E}_{z_1}\left\{e^{-\tau\tilde{\theta}TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} \tag{7.43}$$

and

$$R = h(\tilde{\theta},\hat{\theta}) = \begin{cases} -\frac{1}{\hat{\theta}}\log \mathbb{E}_{z_2}\left\{e^{-(1-\tau)\hat{\theta}TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\} & 0 \le \hat{\theta} \le \tilde{\theta} \\ -\frac{1}{\hat{\theta}}\Big(\log \mathbb{E}_{z_2}\left\{e^{-(1-\tau)\hat{\theta}TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\} & \\ \quad +\log \mathbb{E}_{z_1}\left\{e^{\tau(\hat{\theta}-\tilde{\theta})TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}\Big) & \hat{\theta} \ge \tilde{\theta} \end{cases}, \tag{7.44}$$

respectively. As in full-duplex relaying, the rates R for which the equations in (7.43) and (7.44) are simultaneously satisfied for some $\tilde{\theta} \ge \theta_1$ and $\hat{\theta} \ge \theta_2$ are the rates that can be supported by the two-hop link in the presence of QoS constraints specified by θ_1 and θ_2. The following result provides the effective capacity, which is defined as the supremum of such rates. Similarly as in full-duplex relaying, we assume that the average transmission rate of the $\mathbf{S}-\mathbf{R}$ link is less than the average transmission rate of the $\mathbf{R}-\mathbf{D}$ link in order to ensure stability in the buffers. Therefore, we

suppose $\mathbb{E}_{z_1}\{\tau \log_2(1+\mathrm{SNR}_1 z_1)\} < \mathbb{E}_{z_2}\{(1-\tau)\log_2(1+\mathrm{SNR}_2 z_2)\}$. Accordingly, in the following result, we assume that the feasible values of τ for half-duplex relaying are upper bounded by

$$\tau < \tau_0 = \frac{\mathbb{E}_{z_2}\{\log_2(1+\mathrm{SNR}_2 z_2)\}}{\mathbb{E}_{z_1}\{\log_2(1+\mathrm{SNR}_1 z_1)\} + \mathbb{E}_{z_2}\{\log_2(1+\mathrm{SNR}_2 z_2)\}}. \tag{7.45}$$

Theorem 29 *In half-duplex relaying, the effective capacity of the two-hop communication link with statistical QoS constraints at the source and the intermediate relay nodes is given by*

$$\textbf{\textit{Case I}}\ \theta_1 \geq \theta_2: \quad R_E(\theta_1,\theta_2) = -\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\tilde{\tau}\theta_1 TB\log_2(1+SNR_1 z_1)}\right\} \tag{7.46}$$

$$\textbf{\textit{Case II}}\ \theta_1 < \theta_2: \quad R_E(\theta_1,\theta_2) = -\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\hat{\tau}\theta_1 TB\log_2(1+SNR_1 z_1)}\right\} \tag{7.47}$$

where $\tilde{\tau} = \min\{\tau_0, \tau^*\}$ *and* τ^* *is the solution to*

$$-\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\tau\theta_1 TB\log_2(1+SNR_1 z_1)}\right\} = -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-(1-\tau)\theta_2 TB\log_2(1+SNR_2 z_2)}\right\} \tag{7.48}$$

and $\hat{\tau} = \min\{\tau_0, \tau'\}$ *and* τ' *is the solution to*

$$\begin{aligned} &-\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\tau\theta_1 TB\log_2(1+SNR_1 z_1)}\right\} \\ &= -\frac{1}{\theta_1}\Big(\log \mathbb{E}_{z_2}\left\{e^{-(1-\tau)\theta_2 TB\log_2(1+SNR_2 z_2)}\right\} + \log \mathbb{E}_{z_1}\left\{e^{\tau(\theta_2-\theta_1) TB\log_2(1+SNR_1 z_1)}\right\}\Big). \end{aligned} \tag{7.49}$$

Proof: See Appendix M.

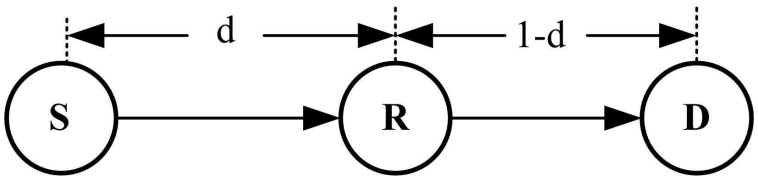

Figure 7.2: The relay model.

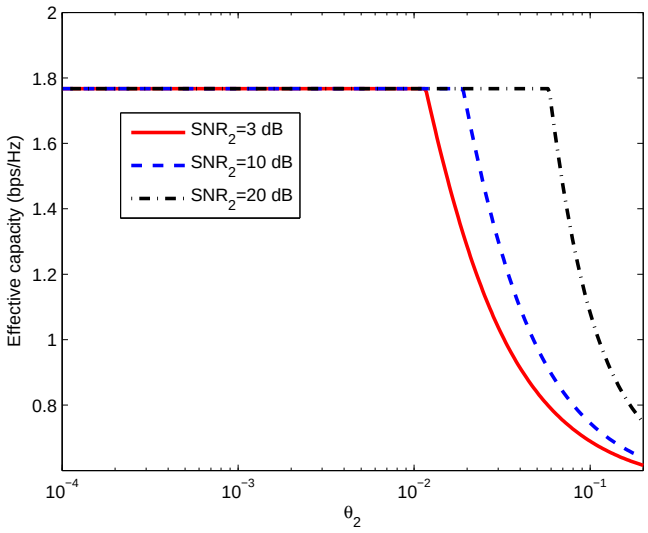

Figure 7.3: The effective capacity as a function of θ_2. $d = 0.5$.

7.2.3 Numerical Results

We consider the relay model depicted in Fig. 7.2. The source, relay, and destination nodes are located on a straight line. The distance between the source and the destination is normalized to 1. Let the distance between the source and the relay node be $d \in (0, 1)$. Then, the distance between the relay and the destination is $1 - d$. We assume the fading distributions for $\mathbf{S} - \mathbf{R}$ and $\mathbf{R} - \mathbf{D}$ links follow independent Rayleigh fading with means $\mathbb{E}\{z_1\} = 1/d^{\alpha}$ and $\mathbb{E}\{z_2\} = 1/(1-d)^{\alpha}$, respectively, where we assume that the path loss $\alpha = 4$. We assume that $\text{SNR}_1 = 0$ dB and $\theta_1 = 0.01$ in the following numerical results.

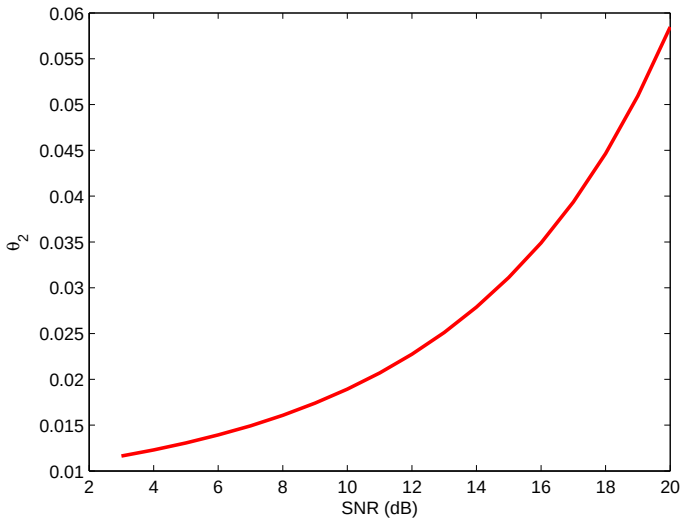

Figure 7.4: θ_2' vs. SNR_2 for $d = 0.5$.

In Fig. 7.3, we plot the effective capacity as a function of the QoS constraints of the full-duplex relay node for different SNR_2 values. We fix $d = 0.5$, in which case the $\mathbf{S}-\mathbf{R}$ and $\mathbf{R}-\mathbf{D}$ links have the same channel conditions. From the figure, we can see that the effective capacity does not decrease for a certain range of θ_2, and this range is increased by increasing SNR_2. Motivated by this observation, we plot the value of θ_2', up to which the effective capacity is unaffected, as a function of SNR_2 in Fig. 7.4. Note that for all values of the pair (SNR, θ_2) below the curve shown in the figure, the QoS constraints of the relay node do not impose any negative effect on the effective capacity. This provides us with useful insight on the design of wireless systems. In Fig. 7.5, we plot the effective capacity as d varies. We assume $\theta_2 = \{0.001, 0.01, 0.05, 0.1\}$. We are interested in the range in which the condition for stable queues (as stated above Theorem 28) is satisfied. More specifically, we note that the optimal d is lower bounded by the value at which we have $\mathbb{E}_{z_1}\{\log_2(1+\text{SNR}_1 z_1)\} = \mathbb{E}_{z_2}\{\log_2(1+\text{SNR}_2 z_2)\}$. We can see from the figure that for small θ_2 (i.e., for $\theta_2 = 0.001$ and $\theta_2 = 0.01$), the

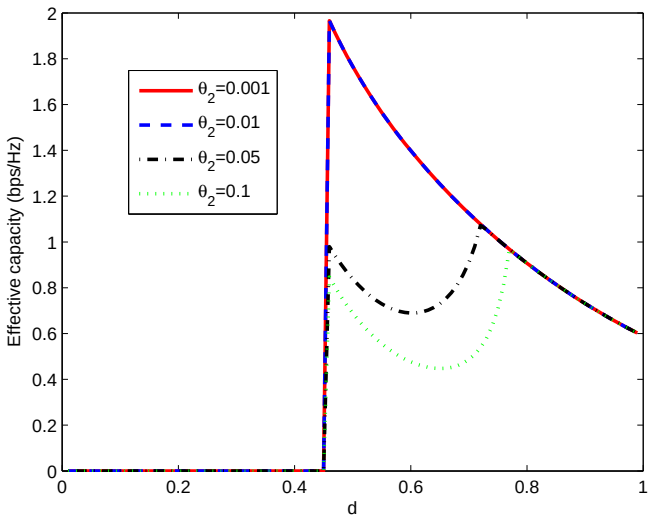

Figure 7.5: The effective capacity as a function of d.

effective capacity curves overlap. In these cases, $\mathbf{S}-\mathbf{R}$ link is the bottleneck and the throughput is determined by the effective capacity of this link. When θ_2 is greater than θ_1 (i.e., when $\theta_2 = 0.05$ or 0.1), it is interesting that the effective capacity decreases first and then increases until the $\mathbf{S}-\mathbf{R}$ link becomes again the bottleneck, in which case the curves overlap. This tells us that with stringent QoS constraints at the relay, having symmetric channel conditions for the links $\mathbf{S}-\mathbf{R}$ and $\mathbf{R}-\mathbf{D}$, i.e., having $d = 0.5$, generally leads to lower performance.

In Fig. 7.6, we plot the effective capacity as a function of θ_2 for half-duplex relaying. We set $d = 0.5$. From the figure, we can find that the effective capacity stays constant for small θ_2, i.e., the QoS constraints at the relay node does not impose any negative effect on the effective capacity of the system. We can also see that as SNR_2 increases, larger QoS constraints at the relay node can be supported while having the effective capacity of the system unaltered. One stark difference from the full-duplex relay is that as SNR_2 increases, the effective capacity of the system increases as well

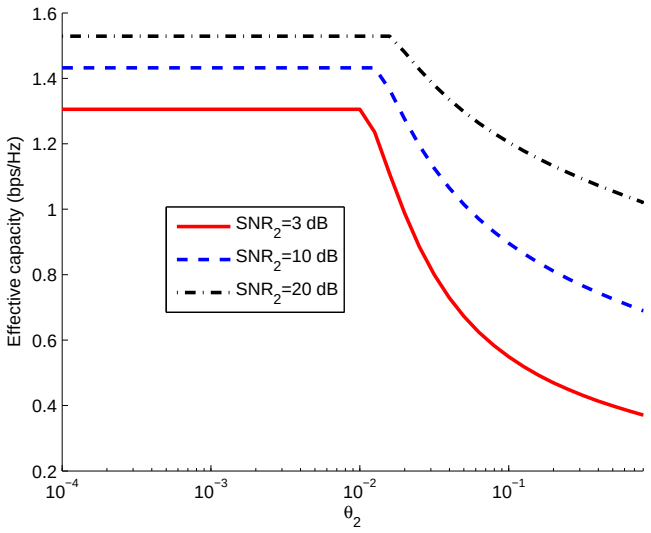

Figure 7.6: The effective capacity as a function of θ_2. $d = 0.5$. $\text{SNR}_2 = \{3, 10, 20\}$ dB.

even for small θ_2. This is due to the nature of the half-duplex operation. As SNR_2 increases, more time can be allocated to the transmission between the source and relay nodes while satisfying (7.45).

In Fig. 7.7, we plot the effective capacity as d and θ_2 varies. We assume $\text{SNR}_2 = 3$ dB. As we can see from the figure, there exists an optimal d that maximizes the effective capacity of the system. Besides, the optimal d increases as θ_2 increases. This is due to the fact that as the QoS constraints at the relay node become more stringent, the effective bandwidth supported by the $\mathbf{R} - \mathbf{D}$ link decreases and this link becomes the bottleneck of the system. In order to counterbalance this negative effect, the channel conditions of the $\mathbf{R} - \mathbf{D}$ link should be improved, which results in a larger d. It is also interesting that the curve is nearly flat for small θ_2 when d is large. So, we plot the effective capacity as d varies for $\theta_2 = \{0.001, 0.01, 0.1\}$ in Fig. 7.8. Confirming the observation in Fig. 7.7, we see that the two curves for $\theta_2 = 0.001$ and $\theta_2 = 0.01$ overlap as d increases. This is because the upperbound for τ specified

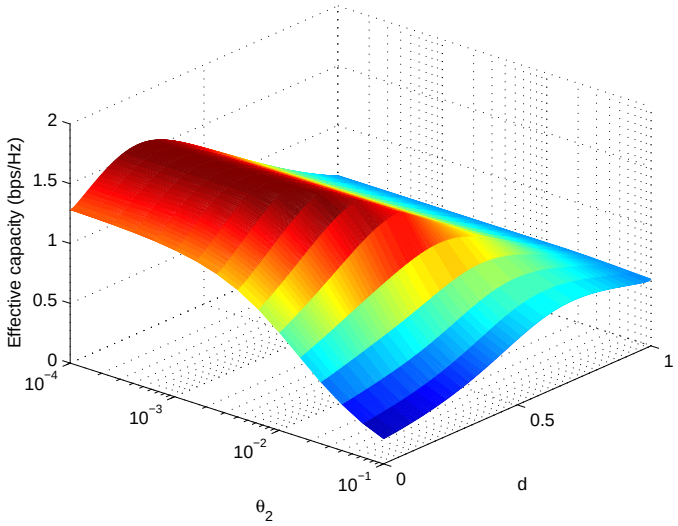

Figure 7.7: The effective capacity v.s. d and θ_2. $\text{SNR}_2 = 3$ dB.

in (7.45) is achieved for both curves.

7.3 Conclusion

In this chapter, we have analyzed the maximum arrival rates that can be supported by a two-hop communication link in which the source and relay nodes are both subject to statistical QoS constraints. We have determined the effective capacity in the block-fading scenario as a function of the signal-to-noise ratio levels SNR_1 and SNR_2 and the QoS exponents θ_1 and θ_2 for both full-duplex and half-duplex relaying. Through this analysis, we have quantified the throughput of a two-hop link operating under buffer constraints. In particular, we have shown that effective capacity can have different characterizations depending on how buffer constraints at the source and relay or more specifically how θ_1 and θ_2 compare. We have noted that if $\theta_1 \geq \theta_2$, the upper bound on the effective capacity is attained. We have also seen that under certain conditions depending on the SNR levels and fading distributions, $\mathbf{S} - \mathbf{R}$ link

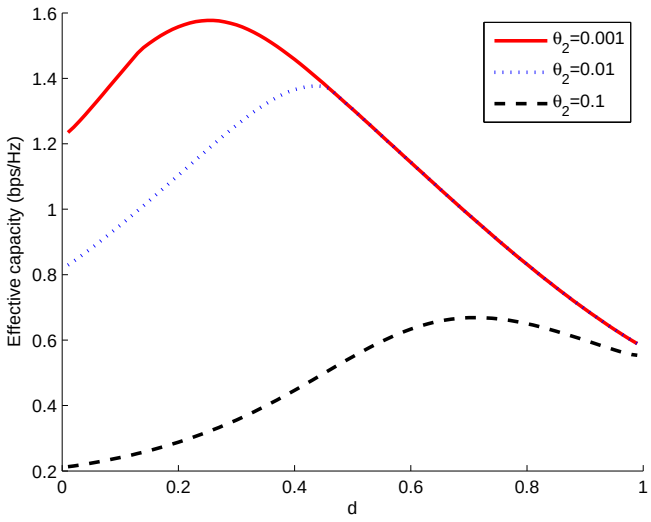

Figure 7.8: The effective capacity as d varies. $\text{SNR}_2 = 3$ dB. $\theta_2 = \{0.001, 0.01, 0.1\}$.

becomes the bottleneck and buffer constraints at the relay do not incur performance losses when the QoS exponent θ_2 is sufficiently small but nonzero. In the numerical results, the threshold for θ_2 above which the effective capacity starts diminishing is identified and is shown to increase with increasing SNR_2. In a simple linear setting, we have numerically investigated the impact of the location of the relay on the effective capacity for different values of the QoS exponents. In half-duplex relaying, we have determined the optimal time-sharing parameter τ. In the numerical results, we have had several interesting observations. We have shown that as the SNR level at the relay node increases, the effective capacity of the system increases for all θ_2. Additionally, as the QoS constraints at the relay node become more stringent, we have observed that the effective capacity of the system can be increased by improving the channel conditions in the $\mathbf{R}-\mathbf{D}$ link through having the relay node approach the destination.

Chapter 8

Throughput for Finite Blocklength Codes

In most prior work, the service rates supported by the wireless channel are assumed to be equal to the instantaneous channel capacity values and no decoding errors are considered, which, from an information theoretic view, is achieved as the coding blocklength grows without bound. On the other hand, practical communication systems employ channel codes with finite blocklengths and operate at rates less than the channel capacity with nonzero probability of decoding error. This is particularly true for systems operating under delay/buffer constraints. For delay sensitive services, using extremely long channel codes can be prohibitive, and hence, results obtained under the idealistic assumption of operation at the channel capacity may not be a faithful representation or prediction of the performance. Therefore, it is of significant interest to study what can be attained with finite-blocklength channel codes in the presence of decoding errors and buffer limitations. Despite their importance, there has only been a handful of work offering insights on such issues. For instance, in [70], Negi and Goel considered the maximization of the joint exponent of the decoding error

and delay violation probability through the appropriate choice of the transmission rate for given delay bound and constant arrival rate. More recently, reference [38] has analyzed the performance of finite blocklength codewords in the presence of statistical QoS constraints. In particular, the effective capacity is formulated by incorporating the recent channel coding results in [71]. However, in [38], coding is assumed to be performed over one coherence block in which the fading stays constant.

In this chapter, we consider a more general setting and assume that codewords are sent over multiple coherence blocks. Hence, each codeword experiences multiple fading realizations. Coding over multiple blocks generally improves the performance since protection against severe fading can be provided as codewords see multiple channel states. At the same time, coding over a large number of blocks may also lead to long delays or buffer overflows especially in the presence of decoding errors and retransmission requirements. Due to these tradeoffs, a throughput analysis of channel coding over multiple coherence blocks in the presence of buffer constraints is called for to identify whether there exists an optimal number of blocks over which coding needs to be performed. With this motivation, we use the following approach and obtain the ensuing original contributions. By making use of the Feinstein's Lemma and employing a Gaussian approximation, we initially derive an approximate lower bound for the instantaneous transmission rate. Then, we regard this lower bound as the service rate and determine, by identifying the effective rate, the arrival rates that can be supported by the channel with certain service guarantees. Subsequently, for both variable- and fixed-rate transmissions, we investigate the interplay between effective rate, coding blocklength, decoding error probabilities, queueing constraints, and signal-to-noise ratio (SNR). In particular, we show that for given coding blocklength, QoS exponent and SNR, the effective rate is maximized at a unique decoding error probability, giving insight on the strength of efficient channel codes for the con-

sidered buffer-constrained systems. We further demonstrate that as the blocklength increases (by coding over a larger number of coherence blocks), effective rate improves if stronger channel codes with lower error probabilities are used. Conversely, we show that if the error probability remains strictly bounded away from zero, effective rate starts diminishing to zero as the blocklength grows without bound. We also investigate the case in which the transmission rate is fixed and error probability varies with the channel fading. In this scenario, the optimal transmission rate that maximizes the throughput is also proven to be unique.

8.1 System Model

We consider a block flat-fading channel, and assume that the fading coefficients stay constant for a coherence block of n symbols and change independently from one block to another. The discrete-time input and output relationship in the l^{th} block is given by

$$y_i = h_l x_i + w_i \quad i = 1, 2, \ldots, n \tag{8.1}$$

where x_i and y_i are the complex-valued channel input and output, respectively, in the i^{th} symbol duration of the l^{th} block, h_l is the channel fading coefficient in the lth block, and w_i is the circularly symmetric complex Gaussian noise with zero mean and variance N_0, i.e., $w_i \sim \mathcal{CN}(0, N_0)$. We assume that the receiver has perfect channel side information (CSI) and hence perfectly knows the realizations of the fading coefficients $\{h_l\}$. On the other hand, we consider both cases of perfect and no CSI at the transmitter.

The channel input is assumed to be subject to $\mathbb{E}\{|x_i|^2\} \leq \mathcal{E}_s$. It is well-known that when the receiver has perfect CSI, the capacity achieving input for the above fading Gaussian channel is Gaussian distributed. Hence, we assume that $x_i \sim \mathcal{CN}(0, \mathcal{E}_s)$.

Since the input and noise are Gaussian distributed, the output is also conditionally Gaussian given the channel fading coefficient, i.e., $y_i|h_l \sim \mathcal{CN}(0, \mathcal{E}_s|h_l|^2 + N_0)$. Moreover, the output is also conditionally Gaussian distributed given h_l and the input x_i, i.e., $y_i|x_i, h_l \sim \mathcal{CN}(h_l x_i, N_0)$. We further assume that the input is independent and identically distributed (i.i.d.) i.e., $p_{x^n} = \prod_{i=1}^{n} p_{x_i}(x_i)$, which implies $p_{y^n|x^n,h_l} = \prod_{i=1}^{n} p_{y_i|x_i,h_l}(y_i|x_i, h_l)$, and $p_{y^n|h_l} = \prod_{i=1}^{n} p_{y_i|h_l}(y_i|h_l)$.

8.2 Mutual Information Density and Channel Coding Rate

As detailed in Section 1.1, effective capacity is determined by specifying the service rate or equivalently the instantaneous transmission rate. We assume that the transmitter performs channel coding over m coherence blocks where $m = 1, 2, \ldots$ Therefore, it sends codewords of length nm and each codeword experiences m independent channel conditions. An upper bound on the maximum decoding error probabilities of random codes of length nm is given by Feinstein's Lemma [72], [73]:

$$\epsilon \leq \Pr\left(\frac{1}{nm} i(x^{nm}; y^{nm}|h_1^m) \leq R + \gamma\right) + \Pr(x^{nm} \notin S_{nm}) + e^{-nm\gamma} \tag{8.2}$$

where $\gamma > 0$ is an arbitrary constant, $S_{nm} = \left\{x^{nm} : \frac{1}{nm}\sum_{i=1}^{nm} |x_i|^2 \leq \mathcal{E}\right\}$ is the constraint set, $i(x^{nm}; y^{nm}|h_1^m)$ is the mutual information density conditioned on the fading coefficients $(h_1, h_2, \ldots, h_m)$ seen in m coherence blocks. The conditional mutual information density is defined as

$$i(x^{nm}; y^{nm}|h_1^m) = \log_2 \frac{p(y^{nm}|x^{nm}, h_1^m)}{p(y^{nm}|h_1^m)}. \tag{8.3}$$

Next, we obtain an expression for the mutual information density of the considered channel and input models (i.e., fading Gaussian channel with Gaussian input), and derive, under certain assumptions, an approximate lower bound on the rates attained by coding over m coherence blocks.

For the system model introduced in Section 8.1, we have

$$\begin{aligned}
&\frac{1}{nm} i(x^{nm}; y^{nm}|h_1^m) \\
&= \frac{1}{nm} \sum_{l=1}^{m} \sum_{i=(l-1)n+1}^{ln} i(x_i; y_i|h_l) && (8.4)\\
&= \frac{1}{nm} \sum_{l=1}^{m} \sum_{i=(l-1)n+1}^{ln} \log_2 \frac{f_{y_i|x_i,h_l}(y_i|x_i,h_l)}{f_{y_i|h_l}(y_i,h_l)} && (8.5)\\
&= \frac{1}{nm} \sum_{l=1}^{m} \sum_{i=(l-1)n+1}^{ln} \left(\log_2 \left(1 + \frac{\mathcal{E}_s|h_l|^2}{N_0}\right) + \frac{|y_i|^2 \log_2 e}{|h_l|^2\mathcal{E}_s + N_0} - \frac{|y_i - h_l x_i|^2 \log_2 e}{N_0} \right) && (8.6)\\
&= \frac{1}{m} \sum_{l=1}^{m} \log_2(1 + \frac{\mathcal{E}_s|h_l|^2}{N_0}) + \frac{\log_2 e}{nm} \sum_{l=1}^{m} \sum_{i=(l-1)n+1}^{ln} \left(\frac{|y_i|^2}{|h_l|^2\mathcal{E}_s + N_0} - \frac{|y_i - h_l x_i|^2}{N_0} \right) && (8.7)
\end{aligned}$$

Denoting $\text{SNR} = \frac{\mathcal{E}_s}{N_0}$ and extending the results in [72] and [73], we can immediately show that $i(x^{nm}; y^{nm}|h_1^m)/(nm)$ has the same distribution as the random variable

$$\frac{1}{m} \sum_{l=1}^{m} \log_2(1 + \text{SNR}|h_l|^2) + \frac{\log_2 e}{nm} \sum_{l=1}^{m} \sqrt{\frac{\text{SNR}|h_l|^2}{1 + \text{SNR}|h_l|^2}} \sum_{i=1}^{n} w_{li} \tag{8.8}$$

where w_{li}'s are i.i.d. Laplace random variables, each with zero mean and variance 2. The sum of nm i.i.d. Laplace random variables has a Bessel-K distribution [72] and generally is difficult to deal with directly. On the other hand, for large enough values of the blocklength nm, the random variable in (8.8) can be well approximated by a Gaussian random variable [73]. Therefore, the mutual information density achieved with the codewords of length nm spreading over m coherence blocks can be regarded

to approximately have the following Gaussian distribution:

$$\frac{1}{nm}i(x^{nm};y^{nm}|h_1^m) \sim \mathcal{CN}\left(\frac{1}{m}\sum_{l=1}^{m}\log_2(1+\mathrm{SNR}z_l), \frac{\log_2^2 e}{m}\sum_{l=1}^{m}\frac{2\mathrm{SNR}z_l}{nm(1+\mathrm{SNR}z_l)}\right) \quad (8.9)$$

where we have defined $z_l = |h_l|^2$. With this approximation, the first probability expression on the right-hand side of (8.2) can be written in terms of the Gaussian Q-function:

$$\Pr\left(\frac{1}{nm}i(x^{nm};y^{nm}|h_1^m) \leq R+\gamma\right) = Q\left(\frac{\frac{1}{m}\sum_{l=1}^{m}\log_2(1+\mathrm{SNR}z_l) - R - \gamma}{\sqrt{\frac{\log_2^2 e}{m}\sum_{l=1}^{m}\frac{2\mathrm{SNR}z_l}{nm(1+\mathrm{SNR}z_l)}}}\right). \quad (8.10)$$

By noting that the Q-function is invertible, we can rewrite the upper bound in (8.2) as a lower bound on the instantaneous rate achieved by coding over m coherence blocks:

$$\begin{aligned} R \geq & \frac{1}{m}\sum_{l=1}^{m}\log_2(1+\mathrm{SNR}z_l) \\ & - \sqrt{\frac{\log_2^2 e}{m}\sum_{l=1}^{m}\frac{2\mathrm{SNR}z_l}{nm(1+\mathrm{SNR}z_l)}}Q^{-1}(\epsilon - \Pr(x^{nm}\notin S_{nm}) - e^{-nm\gamma}) - \gamma \end{aligned} \quad (8.11)$$

for any $\gamma > 0$. Although the above lower bound can also be used in the subsequent analysis, we opt to further simplify it to make the analysis more tractable analytically. We first note that the terms $\Pr(x^{nm}\notin S_{nm})$ and $e^{-nm\gamma}$ decrease exponentially fast with increasing nm and become very small relatively quickly. (For the proof of the exponential decay of $\Pr(x^{nm}\notin S_{nm})$, we refer to Appendix N where a closed-form expression for $\Pr(x^{nm}\notin S_{nm})$ is also given, which can be used to facilitate an analysis with the lower bound in (8.11)). With this observation, we assume that nm is sufficiently large and we neglect these terms for the sake of simplification in the

formulations. Note that these approximations are accurate if $\Pr(x^{nm} \notin S_{nm})$ and $e^{-nm\gamma}$ are much smaller than the error probability ϵ, which we generally observe in the numerical results. Moreover, after the eliminations of these terms, we can see that since the lower bound holds for any $\gamma > 0$, an approximate lower bound for the transmission rate is

$$R \geq R_{l,\epsilon} = \frac{1}{m}\sum_{l=1}^{m}\log_2(1+\mathrm{SNR}z_l) - \sqrt{\frac{\log_2^2 e}{m}\sum_{l=1}^{m}\frac{2\mathrm{SNR}z_l}{nm(1+\mathrm{SNR}z_l)}}Q^{-1}(\epsilon) \tag{8.12}$$

where the notation $R_{l,\epsilon}$ is used to emphasize that this is a lower bound for rates achieved with decoding error probability ϵ. Henceforth, the analysis is based on $R_{l,\epsilon}$.

8.3 Effective Throughput

The rate lower bound in (8.12) gives a characterization of the tradeoffs and interactions between the instantaneous transmission rate, decoding error probability and the fading coefficients when channel coding is performed over multiple coherence blocks using finite blocklength random codes. In particular, we note that $R_{l,e}$ is achieved with probability $1-\epsilon$. With probability ϵ, decoding error occurs. We assume that the receiver reliably detects the errors, and applies a simple ARQ mechanism and sends a negative acknowledgement requesting the retransmission of the message in case of an erroneous reception. Therefore, the data rate is effectively zero when error occurs. Under this assumption, the service rate (in bits per nm symbols) is

$$r_s = \begin{cases} 0, & \text{with probability } \epsilon \\ nmR_{l,\epsilon}, & \text{with probability } 1-\epsilon \end{cases}. \tag{8.13}$$

Similarly as in [38], we obtain the following result on the effective rate by inserting the above service rate formulation into the definition in (1.1) and noting that the service rate varies independently from one sequence of m blocks to another due to the block fading assumption.

Proposition 3 *The effective rate (in bits per channel use) at a given SNR, error probability ϵ, codeword length nm, and QoS exponent θ is*

$$R_E(\theta) = -\frac{1}{\theta nm} \log_e \mathbb{E}_{\mathbf{z}} \left\{ \epsilon + (1-\epsilon) e^{-\theta nm R_{l,\epsilon}} \right\} \tag{8.14}$$

where $R_{l,\epsilon}$ is given in (8.12), and the expectation is with respect to $\mathbf{z} = (z_1, \ldots, z_m)$, which is composed of the magnitude squares of the channel fading coefficients experienced in m blocks.

Proof: The result is immediately obtained through the following steps:

$$\begin{aligned}
R_E(\theta) = -\lim_{t\to\infty} \frac{1}{\theta t} \log_e \mathbb{E}\{e^{-\theta S[t]}\} &= -\lim_{t\to\infty} \frac{1}{\theta t} \log_e \mathbb{E}\{e^{-\theta \sum_{i=1}^{t} r_s[i]}\} && (8.15)\\
&= -\lim_{t\to\infty} \frac{1}{\theta t} \log_e \left(\mathbb{E}\left\{ e^{-\theta r_s} \right\} \right)^t && (8.16)\\
&= -\lim_{t\to\infty} \frac{1}{\theta t} t \log_e \mathbb{E}\left\{ e^{-\theta r_s} \right\} && (8.17)\\
&= -\frac{1}{\theta} \log_e \mathbb{E}\left\{ e^{-\theta r_s} \right\} && (8.18)\\
&= -\frac{1}{\theta} \log_e \mathbb{E}_{\mathbf{z}} \left\{ \epsilon + (1-\epsilon) e^{-\theta mn R_{l,\epsilon}} \right\} && (8.19)
\end{aligned}$$

Above, (8.16) follows from the fact that the service process $r_s(i)$, which depends on the fading vector $\mathbf{z}$, changes independently from one sequence of m blocks to another and has the same distribution for each sequence. This fact is due to the block fading assumption. Hence, $\mathbb{E}\{e^{-\theta \sum_{i=1}^{t} r_s[i]}\} = \mathbb{E}\{\prod_{i=1}^{t} e^{-\theta r_s[i]}\} = \prod_{i=1}^{t} \mathbb{E}\{e^{-\theta r_s[i]}\} = (\mathbb{E}\{e^{-\theta r_s}\})^t$. The term inside the expectation in (8.19) is obtained by evaluating the

expected value of $e^{-\theta r_s}$ for fixed $\mathbf{z}$. Finally, the effective rate expression in (8.14) is obtained by normalizing (8.19) by nm to have the effective rate in the units of bits per channel use. $\square$

Remark: Proposition 3 provides the effective rate achieved with finite block-length codes with possible decoding errors. It is also interesting to consider the idealistic scenario in which the transmission is assumed to be performed at the rate of channel capacity with no decoding errors. In such a case, the transmission rate is $R = \frac{1}{m}\sum_{l=1}^{m}\log_2(1+\text{SNR}z_l)$. Now, the effective rate, again under the block-fading assumption, can be written as

$$\begin{aligned}
\mathsf{R}_{E,\text{ideal}}(\theta) &= -\frac{1}{\theta nm}\log_e \mathbb{E}_{\mathbf{z}}\left\{e^{-\theta nm\frac{1}{m}\sum_{l=1}^{m}\log_2(1+\text{SNR}z_l)}\right\} \\
&= -\frac{1}{\theta nm}\log_e \prod_{l=1}^{m}\mathbb{E}_{z_l}\left\{e^{-\theta n\log_2(1+\text{SNR}z_l)}\right\} \qquad (8.20)\\
&= -\frac{1}{\theta nm}\log_e \left(\mathbb{E}_{z}\left\{e^{-\theta n\log_2(1+\text{SNR}z)}\right\}\right)^{m} \qquad (8.21)\\
&= -\frac{1}{\theta n}\log_e \mathbb{E}_{z}\left\{e^{-\theta n\log_2(1+\text{SNR}z)}\right\} \qquad (8.22)
\end{aligned}$$

where (8.20) is obtained from the assumption of the independence of the fading coefficients for each block, and (8.21) is due to the fact that fading coefficients in different blocks are identically distributed. Interestingly, (8.22) shows us that the effective rate is independent of the number of blocks, m, over which the coding takes place (as long as m is finite). However, we observe in (8.14) that the performance depends on m in the presence of decoding errors. Additionally, we can show using (8.22) that $\lim_{\theta\to 0}\mathsf{R}_{E,\text{ideal}}(\theta) = \mathbb{E}_z\{\log_2(1+\text{SNR}z)\}$. Therefore, in the idealistic case, effective rate becomes equal to the ergodic capacity if queueing constraints are not imposed (i.e., if $\theta = 0$).

The effective rate in (8.14) provides a lower bound on the throughput as a function of SNR, decoding error probability ϵ, fading coefficients, coherence blocklength n, the number of blocks m over which coding is performed, and the QoS exponent θ. Note that using very strong codes and having small error probabilities in the transmission necessitates small transmission rates leading to small throughput. On the other hand, if higher transmission rates with relatively weak channel coding are preferred, then communication reliability degrades and more retransmissions are required again lowering the throughput. With the next result, we show that these effects are balanced when the channel code has a certain strength level specified by its decoding error probability.

Theorem 30 *Given the values of $m > 0, n > 0, \theta > 0$ and SNR > 0, the function*

$$\Psi(\epsilon) = \mathbb{E}_{\mathbf{z}} \left\{ \epsilon + (1-\epsilon) e^{-\theta n m R_{l,\epsilon}} \right\} \tag{8.23}$$

is strictly convex in ϵ, and hence the optimal $\epsilon > 0$ that minimizes $\Psi(\epsilon)$, or equivalently maximizes the effective rate in (8.14), is unique.

Proof: See Appendix O.

Note that the above result holds for $\theta > 0$. If there are no QoS constraints, i.e., $\theta = 0$, we have the following result.

Corollary 6 *When $\theta = 0$, the effective rate becomes*

$$R_E(0) = \lim_{\theta \to 0} R_E(\theta) = (1-\epsilon) \mathbb{E}_{\mathbf{z}} \{ R_{l,\epsilon} \} \tag{8.24}$$

where $R_{l,\epsilon}$ is given by (8.12). $R_E(0)$ is strictly concave in ϵ and hence the optimal ϵ that maximizes $R_E(0)$ is unique.

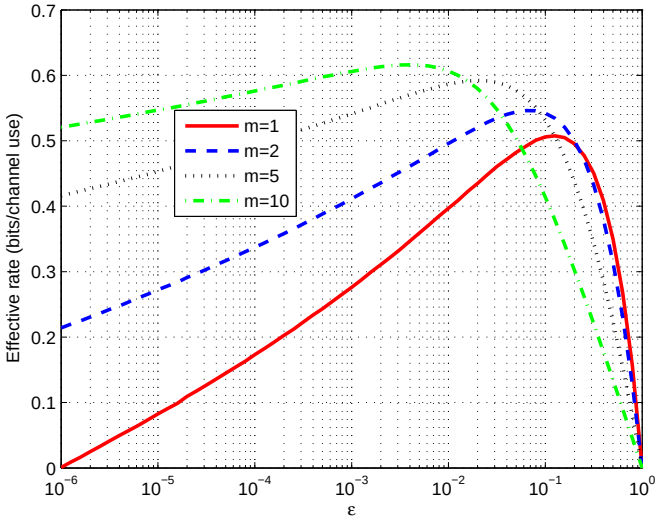

Figure 8.1: The effective rate as a function of ϵ for different values of m. $n = 50$.

The result on the limiting behavior as $\theta \to 0$ can be obtained in a straightforward fashion and the concavity can be shown using similar steps as in the proof of Theorem 30. Due to these, the proof is omitted for brevity.

Note that $\mathsf{R}_E(0)$ is the average transmission rate averaged over all possible channel state vectors, and the result tells us that this arrival rate can be supported in the long term by transmitting over all possible channel states if QoS constraints are not imposed.

Next, we present several numerical results, verifying the theoretical observations and identifying the interplay between some parameters. In Fig. 8.1, we plot the effective rate as a function of ϵ in the Rayleigh fading channel with $\mathbb{E}\{z\} = 1$. Here, we assume SNR $= 0$ dB and $\theta = 0.01$. In Figs. 8.1–8.3, we also assume that $n = 50$. In Fig. 8.1, we provide curves for different values of m. We can see that the effective rate is indeed maximized at a unique ϵ^*, as predicted by Theorem 30. We also observe that as m increases, the optimal ϵ^* decreases, and the effective rate attained at this

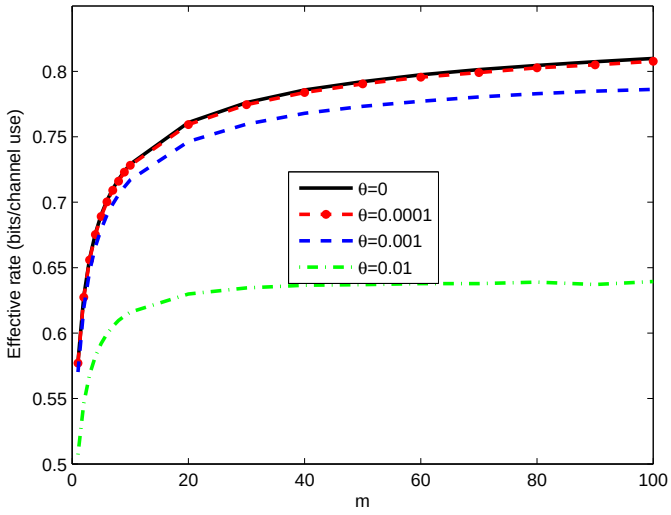

Figure 8.2: The effective rate optimized over ϵ as a function of m. $n = 50$.

optimal ϵ^* is higher. Therefore, coding over an increasing number of coherence blocks is beneficial in terms of increasing the effective rate if the decoding error probability is suitably lowered.

In Fig. 8.2, we plot the effective rate optimized over ϵ as a function of m for different values of θ. From top to bottom, the curves correspond to $\theta = 0,\ 0.0001,\ 0.001,\ 0.01$, respectively. Note that when $\theta = 0$, no queueing constraints are imposed. We see in the figure that all curves increase with m. That is, coding over an increasing number of blocks is always helpful if the decoding error probability is optimized. At the same time, we observe that the rate of increase is smaller for higher values of m, especially if θ is relatively large. Hence, we have diminishing returns as the number of blocks m increases. It is also important to note that as m grows, code complexity and coding and decoding delays increase as well. Hence, these tradeoffs should be carefully considered in the choice of the codeword length.

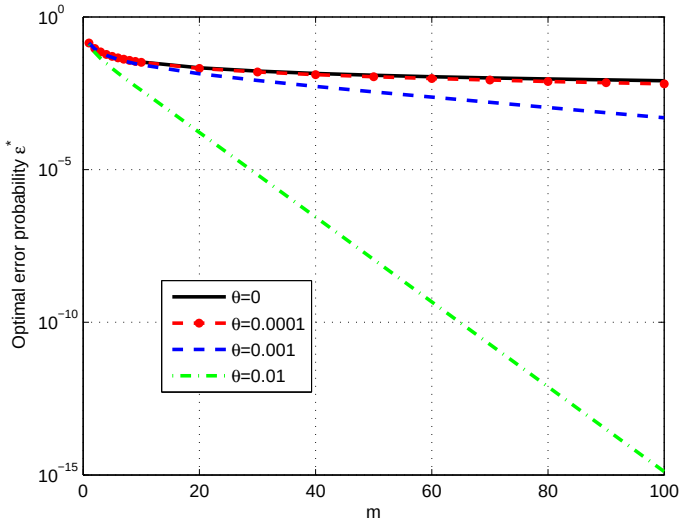

Figure 8.3: The optimal error probability ϵ^* as a function of m. $n = 50$.

In Fig. 8.3, we plot the optimal error probability ϵ^* (at which the effective rate is maximized) as a function of m. As expected, the optimal error probability decreases with increasing m. We also observe that as the QoS constraints become more stringent (i.e., as θ increases), lower decoding error probabilities are required. This can be attributed to the tendency of the more severely buffer-limited system to reduce the number of retransmissions to avoid buffer overflows. The situation is even more critical when m is large and long codewords are transmitted, because when decoding errors occur, the entire long codeword should be retransmitted, and data cannot be cleared from the buffer until successful transmission is achieved. Hence, decoding failures can be quite detrimental under buffer constraints especially for large m. Indeed, in Fig. 8.3, we see that while the error probability curves are relatively close to each other for small values of m, the gap widens as m and θ increase.

8.3.1 Bounded or Fixed error probability

As we have shown in the previous section, coding over multiple blocks is generally beneficial because transmitted codewords experience multiple channel fading realizations and may not get exceedingly affected by severe fading in one block. At the same time, these benefits are realized if the decoding error probabilities are decreased as the number of blocks over which coding is performed increases. Hence, stronger channel codes should be used if codewords are to be transmitted over a larger number of coherence blocks. This is because coding over many blocks with a relatively high frequency of retransmissions may lead to unacceptable delays in systems operating under buffer constraints captured by the QoS exponent θ in this chapter. Hence, we expect to have the optimal error probabilities vanish as $m \to \infty$. Conversely, we can show that if the error probability is bounded away from zero, then $\mathsf{R}_E(\theta)$ approaches zero as m increases without bound.

Theorem 31 *Assume that $\theta > 0$ and the decoding error probability, ϵ_m, which in general depends on m, is lower bounded as $\epsilon_m \geq \epsilon_o > 0$. Then, we have*

$$\lim_{m\to\infty} \mathsf{R}_E(\theta) = 0. \tag{8.25}$$

Proof: See Appendix P.

In order to demonstrate this behavior, we assume in the following numerical results that the error probabilities are kept fixed at a certain level that is strictly greater than zero. In Fig. 8.4, we plot the effective rate as a function of m for different θ values with fixed $\epsilon = 0.01$. In the figure, we observe that the optimal m that maximizes the effective rate under a given ϵ varies with θ. When $\theta = 0$ and therefore there are no buffer constraints, effective rate increases with increasing m. Coding over ever

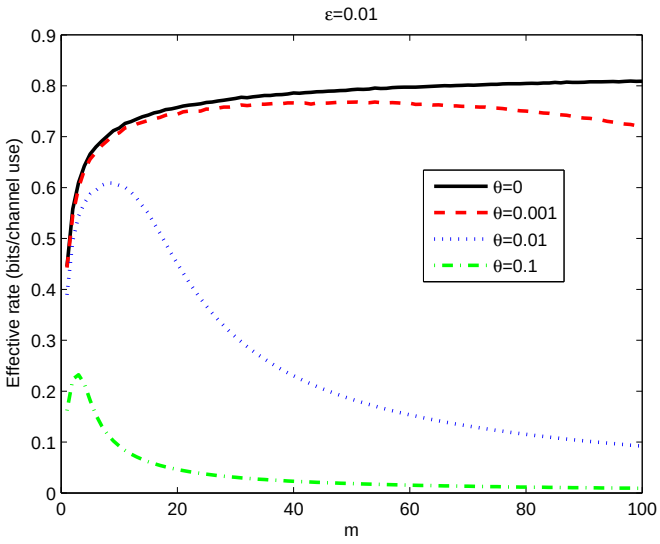

Figure 8.4: The effective rate as a function of m. $n = 50$. $\epsilon = 0.01$.

increasing number of blocks improves the performance. However, we see a strikingly different behavior in the presence of QoS limitations. We note that for $\theta > 0$, effective rate is maximized at a finite value of m, and as predicted by Theorem 31, approaches zero as m grows without bound. Moreover, the optimal value of m diminishes as θ increases. Therefore, coding over fewer blocks should be preferred under stringent buffer limitations and fixed error probability.

In Fig. 8.5, we plot the optimal m that maximizes the effective rate as n varies from 1 to 200 for different $\theta > 0$ values. We again assume $\epsilon = 0.01$. We see that as n increases and hence coherence blocks are larger, smaller m is preferred. Recalling that the codeword length is mn, we here observe that, for fixed error probability, increase in n is being offset by the decrease in m to avoid increases in the codeword length. This is expected in light of the damaging effects of using very long codewords with fixed or bounded decoding error probability as observed in Fig. 8.4. We further see in Fig. 8.5 that the optimal m decreases with increasing θ similarly as in Fig. 8.4.

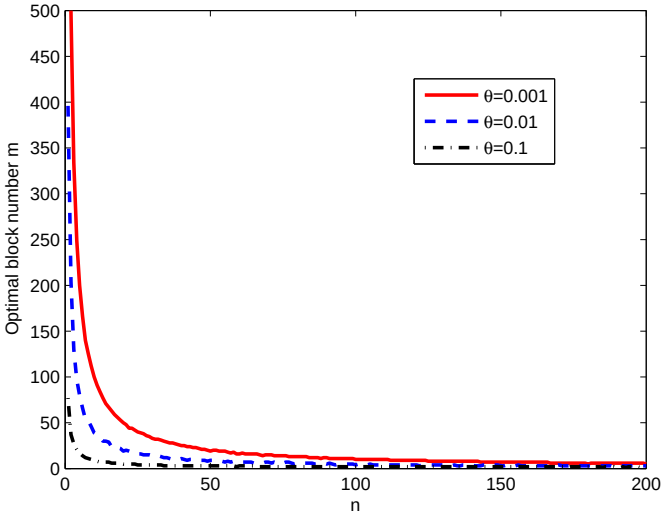

Figure 8.5: The optimal m vs. n. $\epsilon = 0.01$.

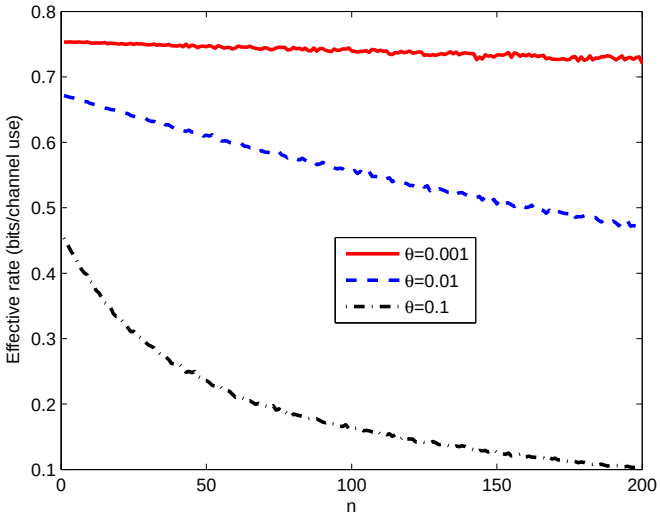

Figure 8.6: The optimal effective rate vs. n. $\epsilon = 0.01$.

Finally, in Fig. 8.6, we plot the optimal effective rate as n varies. It is interesting that while the maximal effective rate does not change much for small θ, it significantly decreases with n for relatively large values of θ. The reason lies in the fact that since n is the coherence duration over which the fading state remains fixed, larger n corresponds to slower fading and slow fading can lead to long durations of deep fading. In such cases, system becomes more conservative and supports lower arrival rates to avoid buffer overflows. While this undermining effect is not as deleterious for loose buffer constraints, it becomes more pronounced with stringent QoS constraints. Overall, we notice that increase in the code blocklength due to increase in coherence duration n has different consequences than that caused by the increase in m.

8.3.2 Fixed Rate Transmissions

Heretofore, we have implicitly assumed that the transmitter has perfect CSI and considered the scenario in which the transmitter employs variable-rate transmissions with rates characterized by $R_{l,\epsilon}$ given in (8.12). Note that in order to transmit at the rate $R_{l,\epsilon}$, the transmitter needs to know the fading coefficients. A more practical scenario is the one in which the transmitter does not know the channel states and send the information at a fixed rate of R. Note that in this case, the decoding error probability varies with the fading coefficients in each set of m blocks. The codeword error probability for a given channel state $\mathbf{z}$ and fixed-transmission rate R is

$$\epsilon(\mathbf{z}, R) = Q\left(\frac{\frac{1}{m}\sum_{l=1}^{m}\log_2(1+\mathrm{SNR}z_l) - R}{\sqrt{\frac{1}{m}\sum_{l=1}^{m}\frac{2\mathrm{SNR}z_l}{nm(1+\mathrm{SNR}z_l)}}\log_2 e}\right) \tag{8.26}$$

obtained by using (8.12). The effective rate is now expressed as

$$\begin{aligned}
&\mathsf{R}_E(\theta, R) \\
&= -\frac{1}{\theta nm} \log_e \mathbb{E}_{\mathbf{z}} \left\{ \epsilon(\mathbf{z}, R) + (1 - \epsilon(\mathbf{z}, R)) e^{-\theta nmR} \right\} \\
&= -\frac{1}{\theta nm} \log_e \mathbb{E}_{\mathbf{z}} \Bigg\{ Q \left(\frac{\frac{1}{m} \sum_{l=1}^{m} \log_2(1 + \mathrm{SNR} z_l) - R}{\sqrt{\frac{1}{m} \sum_{l=1}^{m} \frac{2\mathrm{SNR} z_l}{nm(1+\mathrm{SNR} z_l)}} \log_2 e} \right) \\
&\qquad + \left(1 - Q \left(\frac{\frac{1}{m} \sum_{l=1}^{m} \log_2(1 + \mathrm{SNR} z_l) - R}{\sqrt{\frac{1}{m} \sum_{l=1}^{m} \frac{2\mathrm{SNR} z_l}{nm(1+\mathrm{SNR} z_l)}} \log_2 e} \right) \right) e^{-\theta nmR} \Bigg\}. \qquad (8.27)
\end{aligned}$$

After this formulation, we have the following result that shows that there exists a unique transmission rate that maximizes the throughput for given n, m, $\theta > 0$ and $\mathrm{SNR} > 0$. The reasoning is that very high rates result in frequent errors and retransmissions while very low rates inescapably lead to low throughput even though transmissions are more reliable. Hence, the best performance is attained at a certain transmission rate at which the competing effects are balanced.

Theorem 32 *Assume that the values of n, m, $\theta > 0$ and $SNR > 0$ are fixed, then the function*

$$\begin{aligned}
\Phi(R) = \mathbb{E}_{\mathbf{z}} \Bigg\{ & Q \left(\frac{\frac{1}{m} \sum_{l=1}^{m} \log_2(1 + SNR z_l) - R}{\sqrt{\frac{1}{m} \sum_{l=1}^{m} \frac{2SNR z_l}{nm(1+SNR z_l)}} \log_2 e} \right) \\
& + \left(1 - Q \left(\frac{\frac{1}{m} \sum_{l=1}^{m} \log_2(1 + SNR z_l) - R}{\sqrt{\frac{1}{m} \sum_{l=1}^{m} \frac{2SNR z_l}{nm(1+SNR z_l)}} \log_2 e} \right) \right) e^{-\theta nmR} \Bigg\}
\end{aligned}$$

is minimized at a unique R and hence the optimal R that minimizes $\Phi(R)$ or equivalently maximizes the effective rate in (8.27) is unique.

Proof: See Appendix Q.

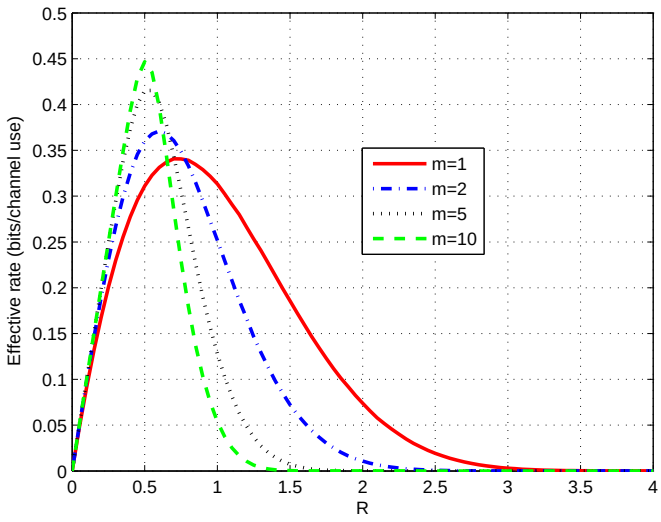

Figure 8.7: The effective rate as a function of R. $n = 50$.

From the proof, we know that the effective rate is maximized at a unique R^*, and $\dot{\Phi}(R^*) = 0$. Numerical methods such as bisection method can be used to determine the optimal R^*.

In Fig. 8.7, we plot the effective rate as a function of the fixed transmission rate R for different m values. We assume that $\theta = 0.01$. It is noted that there is a unique R that maximizes the effective throughput for each m. We can also see in the figure that the maximum effective throughput grows as m increases from 1 to 10, and this improvement is achieved by lowering the transmission rate and consequently the error probabilities. In Fig. 8.8, we plot the optimal effective rate as a function of m for different θ values. We again notice that the effective rate increases with m.

Motivated to see the performance difference between the variable-rate and fixed-rate transmissions, we plot for both transmission schemes the optimal effective rate as a function of θ for different m values in Fig. 8.9. We observe for both transmission schemes that $m = 5$ achieves the highest effective rate for small θ values while $m =$

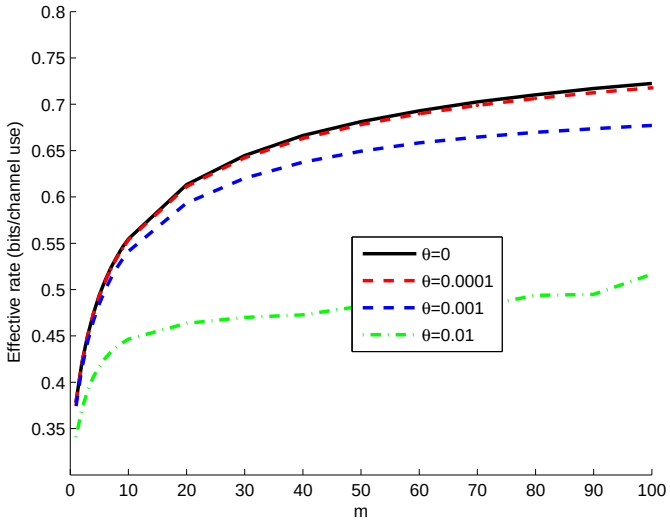

Figure 8.8: The effective rate as a function of m. $n = 50$.

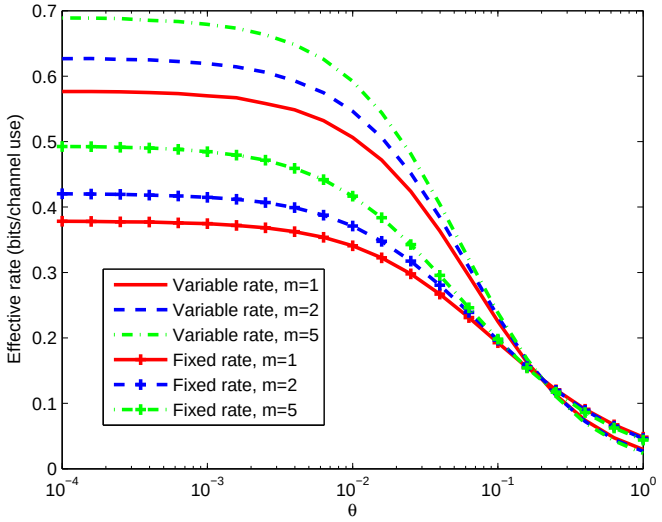

Figure 8.9: The effective rate as a function of θ for different transmission schemes.

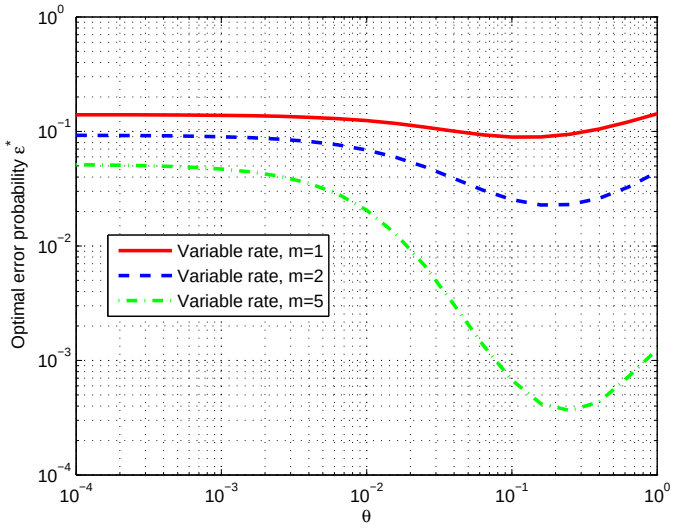

Figure 8.10: The optimal error probability ϵ^* as a function of θ. $n = 50$.

1 begins to perform better as θ increases. Another interesting observation is that fixed-rate transmission outperforms variable-rate transmission for large θ values or equivalently for stringent QoS constraints. We note that the performance of variable-rate schemes are plotted for optimized error probabilities. In Fig. 8.10, we provide the optimal error probability as a function of θ for different m values. We notice in all cases that the optimal error probability initially decreases with θ and then starts to increase after a certain threshold θ value.

8.4 Conclusion

We have analyzed the performance of coding over multiple coherence blocks with possible decoding errors in the presence of queueing constraints. We have characterized the throughput in this scenario by identifying a lower bound on instantaneous transmission rates and determining the effective rate expression. We have investigated two different transmission strategies. For the case of variable-rate transmissions, we have

proven that the optimal error probability that maximizes the effective throughput is unique for given blocklength, QoS exponent θ, and SNR. We have shown that the throughput improves as channel coding is performed over an increasing number of coherence blocks if the probability of decoding errors are lowered accordingly. Additionally, we have remarked that coding over a very large number of blocks can be detrimental if the error probabilities are fixed or cannot be sufficiently decreased. This observation is made for the case in which $\theta > 0$. If, on the other hand, no buffer constraints are imposed i.e., $\theta = 0$, we have noted that using codewords of longer length is always beneficial. Through numerical results, we have further investigated the interplay between effective rate, coherence duration n, QoS exponent θ, and the code blocklength.

In cases in which the transmitter sends the information at a fixed rate R, we have remarked that error probabilities vary with the channel conditions and we have formulated the effective rate. We have shown that the effective rate is maximized at a unique value of R. We have again demonstrated that the performance improves with increasing m but this time if R is suitably lowered. We have further observed interestingly that as QoS constraints become more stringent or equivalently as θ increases, fixed-rate transmission schemes outperform variable-rate ones and moreover coding over a smaller number of coherence blocks starts leading to better performance.

Chapter 9

Conclusions and Future Works

In summary, we have analyzed the energy efficiency in single- and multi- user settings. We have obtained the spectral efficiency–bit energy tradeoff in the low-power and wideband regimes for point-to-point links. We have characterized the minimum bit energy levels and wideband slope regions of fading MAC for different transmission and reception strategies, namely TDMA, superposition coding with fixed decoding order, and superposition coding with variable decoding order. Also, we have studied the effective capacity region of fading MACs, broadcast channels, and secrecy channels under QoS constraints. We have investigated the throughput of two-hop relay channels with QoS constraints at both the source and relay nodes, and the throughput of transmission with finite-blocklength codes.

In the future, we plan to do energy efficiency analysis in network settings, such as broadcast channels and secrecy channels. Practical systems generally operate in multi-user scenario. The energy efficiency in such cases will be of significant interest. We also plan to analyze the throughput of more general relay channels. Our prime results on the throughput of the two-hop relay channels are expected to extend to a wide range of relay channels, such as two-way relay and multiple-relay channels. Also,

our results on the throughput for finite-blocklength codes were carried out in point-to-point links. The performance of finite-block length codes under QoS constraints in the multi-user scenario will be interesting. Some other potential interesting topics may arise in the future as well.

Below is the publication list.

Book Chapters:

1. D. Qiao, M. C. Gursoy, S. Velipasalar, "Green Communications under Quality of Service Constraints," Green Communications: Theoretical Fundamentals, Algorithms and Applications, CRC Press, 2012.

Published/Submitted Journal Papers (refereed):

1. D. Qiao, M. C. Gursoy, S. Velipasalar, "Energy efficiency of multiaccess fading channels under QoS constraints," EURASIP Journal on Wireless Communications and Networking, 2012.

2. D. Qiao, M. C. Gursoy, S. Velipasalar, "Effective capacity of two-hop communication systems," submitted to the IEEE Transactions on Information Theory, July 2011

3. D. Qiao, M. C. Gursoy, S. Velipasalar, "Channel coding over multiple coherence blocks with queueing constraints," submitted to the IEEE Transactions on Information Theory, Dec. 2010.

4. D. Qiao, M. C. Gursoy, S. Velipasalar, "Transmission strategies in multiple access fading channels with statistical QoS constraints," IEEE Transactions on Information Theory, vol. 58, no. 3, March 2012, pp. 1578-1593.

5. D. Qiao, M. C. Gursoy, S. Velipasalar, "Secure wireless communication and optimal power control under statistical queueing constraints," IEEE Transactions on Information Forensics and Security, vol. 6, no. 3, Sep. 2011, pp. 628-639.

6. D. Qiao, M. C. Gursoy, and S. Velipasalar, "Energy efficiency in the low-SNR regime under queueing constraints and channel uncertainty, " IEEE Transactions on Communications, vol. 59, no. 7, July 2011, pp. 2006-2017.

7. D. Qiao, M. C. Gursoy, and S. Velipasalar, "A noncooperative power control game in multi-access fading channels with QoS constraints," Physical Communication (Elsevier), vol. 3, issue 2, pp. 97-104, June 2010

8. D. Qiao, M.C. Gursoy and S. Velipasalar,"The impact of QoS constraints on the energy efficiency of fixed-rate wireless transmissions," IEEE Trans. Wireless Commun.,VOL. 8, NO. 12, Dec. 2009, pp. 5957-5969.

9. M.C. Gursoy, D. Qiao, and S. Velipasalar, "Analysis of energy efficiency in fading channels under QoS constraints," IEEE Trans. Wireless Commun.,VOL. 8, NO. 8, Aug. 2009, pp. 4252 - 4263.

Published/Submitted Conference Papers (refereed):

1. D. Qiao, M. C. Gursoy, and S. Velipasalar, "Throughput regions for fading interference channels under statistical QoS constraints," submitted to 2012 IEEE Global Communications Conference (Globecom).

2. D. Qiao, M. C. Gursoy, and S. Velipasalar, "Energy efficiency of multiaccess fading channels under QoS constraints," to appear in 2012 IEEE International Conference on Communications (ICC).

3. D. Qiao, M. C. Gursoy, and S. Velipasalar, "The impact of half-duplex relaying on the effective capacity of two-hop commuication systems," 46th Conference on Information Sciences and Systems, Princeton, 2012.

4. D. Qiao, M. C. Gursoy, and S. Velipasalar, "Effective capacity region and optimal power control for fading broadcast channels," 2011 IEEE International Symposium on Information Theory (ISIT), St. Petersburg, Russia.

5. D. Qiao, M. C. Gursoy, and S. Velipasalar, "Channel coding over multiple coherence blocks with queueing constraints," 2011 IEEE International Conference on Communications (ICC), Kyoto, Japan.

6. D. Qiao, M. C. Gursoy, and S. Velipasalar, "On the effective capacity of two-hop communication systems," 2011 IEEE International Conference on Communications (ICC), Kyoto, Japan.

7. D. Qiao, M. C. Gursoy, and S. Velipasalar, "Secure broadcasting over fading channels with statistical QoS constraints," Proc. of the 2010 IEEE Global Communications Conference (Globecom), Dec. 2010

8. D. Qiao, M. C. Gursoy, and S. Velipasalar, "Secure communication over fading channels with statistical QoS constraints," Proc. of the 2010 IEEE International Symposium on Information Theory (ISIT), Austin, Texas, June 2010.

9. D. Qiao, M. C. Gursoy, and S. Velipasalar, "On the achievable throughput region of multiple-access fading channels with QoS constraints," Proc. of the 2010 IEEE International Conference on Communications (ICC), Cape Town, South Africa, May 2010.

10. D. Qiao, M. C. Gursoy, and S. Velipasalar, "A noncooperative power control game in multiple-access fading channels with QoS constraints," Proc. of the 2010 IEEE Wireless Communications and Networking Conference (WCNC), Sydney, Australia, Apr. 2010.

11. D. Qiao, M. C. Gursoy and S. Velipasalar, "Energy efficiency of fixed-rate wireless transmissions under queueing constraints and channel uncertainty," Proc. of the IEEE Global Communications Conference (Globecom), Hawaii, Dec. 2009.

12. D. Qiao, M. C. Gursoy, and S. Velipasalar, "Energy efficiency of fixed-rate wireless transmissions under QoS constraints," Proc. of the IEEE International Conference on Communications (ICC), Dresden, Germany, June 2009.

13. D. Qiao, M. C. Gursoy, and S. Velipasalar, "Analysis of energy efficiency in fading channels under QoS constraints," Proc. of the IEEE Global Communications Conference (Globecom), New Orleans, Dec. 2008.

Appendix A

Proof of monotonicity of $\frac{\mathsf{C}_E(\zeta)}{\zeta}$ in ζ

Considering (2.20), we denote

$$C_E(\zeta) = \frac{\mathsf{C}_E(\zeta)}{\zeta} = -\frac{1}{\theta T}\log_e \mathbb{E}\{e^{-\frac{\theta T}{\zeta}\log_2(1+\frac{\bar{P}\zeta}{N_0}z)}\}. \tag{A.1}$$

The first derivative of $C_E(\zeta)$ with respect to ζ is given by

$$\dot{C}_E(\zeta) = -\frac{1}{\zeta^2 \log_e 2}\frac{\mathbb{E}\{e^{-\frac{\theta T}{\zeta}\log_e 2(1+\frac{\bar{P}\zeta}{N_0}z)}\Big[\log_e(1+\frac{\bar{P}\zeta}{N_0}z) - \frac{\frac{\bar{P}\zeta}{N_0}z}{1+\frac{\bar{P}\zeta}{N_0}z}\Big]\}}{\mathbb{E}\{e^{-\frac{\theta T}{\zeta}\log_2(1+\frac{\bar{P}\zeta}{N_0}z)}\}}. \tag{A.2}$$

We let $\nu = \frac{\bar{P}\zeta}{N_0}z \geq 0$, and define $y(\nu) = \log_e(1+\nu) - \frac{\nu}{1+\nu}$, where $y(0) = 0$. It can be easily seen that $\dot{y} = \frac{\nu}{(1+\nu)^2} \geq 0$, so $y(\nu) \geq 0$ holds for all ν. Then, we immediately observe that $\dot{C}_E(\zeta) < 0$ for $\zeta > 0$. Therefore, $\frac{\mathsf{C}_E(\zeta)}{\zeta}$ monotonically increases with *decreasing* ζ. □

Appendix B

Proof of Theorem 6

In [12, Chap. 7, Example 7.2.7], it is shown for Markov modulated processes that

$$\frac{\Lambda(\theta)}{\theta} = \frac{1}{\theta} \log_e sp\left(\mathbf{\Phi}(\theta)\mathbf{P}\right) \tag{B.1}$$

where $sp\left(\mathbf{\Phi}(\theta)\mathbf{P}\right)$ is the spectral radius (i.e., the maximum of the absolute values of the eigenvalues) of the matrix $\mathbf{\Phi}(\theta)\mathbf{P}$, $\mathbf{P}$ is the transition matrix of the underlying Markov process, and $\mathbf{\Phi}(\theta)$ is a diagonal matrix whose j^{th} component, $\phi_j(\theta)$, is the moment generating function of the random process $y_j(t)$ given in this state. Hence, we have $\phi_j(\theta) = E\{e^{\theta y_j(t)}\}$.

The transmission model described for the wideband channel with N subchannels is a Markov-modulated process where the underlying Markov process has $N+1$ states with the transition probabilities given in (3.15). Hence, the transition matrix is given by (B.2) on the next page. Note that the rows of $\mathbf{P}$ are identical due to the fact that the transition probabilities do not depend on the initial state. In each state, the transmission rate is non-random and fixed. Recall that in state j, the transmission rate is equal to $(j-1)rT$. The moment generating function of this deterministic

$$\mathbf{P} = \begin{bmatrix} p_{1,1} & p_{1,2} & . & . & p_{1,N+1} \\ . & & & & . \\ . & & & & . \\ p_{N+1,1} & p_{N+1,2} & . & . & p_{N+1,N+1} \end{bmatrix} = \begin{bmatrix} p_1 & p_2 & . & . & p_{N+1} \\ . & & & & . \\ . & & & & . \\ p_1 & p_2 & . & . & p_{N+1} \end{bmatrix}. \tag{B.2}$$

$$\mathbf{\Phi}(\theta)\mathbf{P} = \underbrace{\begin{bmatrix} 1 & 0 & . & . & 0 \\ 0 & e^{\theta rT} & 0 & . & 0 \\ . & & & & . \\ 0 & 0 & . & 0 & e^{\theta NrT} \end{bmatrix}}_{\mathbf{\Phi}(\theta)} \underbrace{\begin{bmatrix} p_1 & p_2 & . & . & p_{N+1} \\ . & & & & . \\ . & & & & . \\ p_1 & p_2 & . & . & p_{N+1} \end{bmatrix}}_{\mathbf{P}}$$

$$= \begin{bmatrix} p_1 & p_2 & . & . & p_{N+1} \\ p_1 e^{\theta rT} & p_2 e^{\theta rT} & . & . & p_{N+1} e^{\theta rT} \\ . & & & & . \\ p_1 e^{\theta NrT} & p_2 e^{\theta NrT} & . & . & p_{N+1} e^{\theta NrT} \end{bmatrix}. \tag{B.3}$$

process is $\phi_j(\theta) = E\{e^{\theta(j-1)rT}\} = e^{\theta(j-1)rT}$. Therefore, we can express $\mathbf{\Phi}(\theta)\mathbf{P}$ as in (B.3) at the top of this page.

Note that the rows of $\mathbf{\Phi}(\theta)\mathbf{P}$ are multiples of each other, and hence $\mathbf{\Phi}(\theta)\mathbf{P}$ is a matrix of unit rank. This leads to the conclusion that

$$sp\left(\mathbf{\Phi}(\theta)\mathbf{P}\right) = \text{trace}(\mathbf{\Phi}(\theta)\mathbf{P}) = \sum_{j=1}^{N+1} p_j \, e^{\theta(j-1)rT}. \tag{B.4}$$

Therefore, for the wideband channel in consideration, we have

$$\frac{\Lambda(\theta)}{\theta} = \frac{1}{\theta} \log_e sp\left(\mathbf{\Phi}(\theta)\mathbf{P}\right) = \frac{1}{\theta} \log_e \left(\sum_{j=1}^{N+1} p_j \, e^{\theta(j-1)rT} \right). \tag{B.5}$$

Applying the definition

$$\mathsf{R}_E(\text{SNR}, \theta) = \frac{1}{TB} \max_{\substack{r \geq 0 \\ \bar{P}_k \geq 0 \text{ s.t. } \sum \bar{P}_k \leq \bar{P}}} \left\{ -\frac{\Lambda(-\theta)}{\theta} \right\} \tag{B.6}$$

$$\sum_{j=1}^{N+1} p_j \, e^{\theta(j-1)rT} = \sum_{j=1}^{N+1} \left(\begin{array}{c} N \\ j-1 \end{array} \right) \left(P\{z > \alpha\}\right)^{j-1} \left(1 - P\{z > \alpha\}\right)^{N-j+1} \, e^{\theta(j-1)rT} \tag{B.7}$$

$$= \sum_{i=0}^{N} \left(\begin{array}{c} N \\ i \end{array} \right) \left(P\{z > \alpha\} e^{\theta rT}\right)^{i} \left(1 - P\{z > \alpha\}\right)^{N-i} \tag{B.8}$$

$$= \left(1 - P\{z > \alpha\} + P\{z > \alpha\} e^{\theta rT}\right)^{N} \tag{B.9}$$

$$= \left(1 - P\{z > \alpha\}(1 - e^{\theta rT})\right)^{N}. \tag{B.10}$$

---

where we have maximization over the transmission rates and power allocation strategies, we immediately obtain (3.17).

Assume now that $\{z_k\}_{k=1}^{N}$ are identically distributed and therefore p_j is in the binomial form given in (3.16). Then, we can easily obtain (B.7)–(B.10) at the top of this page. Note that (B.8) is obtained by applying a change of variables with $i = j - 1$ and combining the second and fourth terms in the summation in (B.7) to write $\left(P\{z > \alpha\} e^{\theta rT}\right)^{i}$. (B.9) follows from the Binomial Theorem. Now, the expression in (3.18) is readily obtained by noting that $\frac{B}{N} = B_c$. $\square$

Appendix C

Proof of Theorem 7

Assume that the Taylor series expansion of r_{opt} with respect to small $\zeta = \frac{1}{B_c}$ is

$$r_{\text{opt}} = r^*_{\text{opt}} + \dot{r}_{\text{opt}}(0)\zeta + o(\zeta) \tag{C.1}$$

where $r^*_{\text{opt}} = \lim_{\zeta \to 0} r_{\text{opt}}$ and $\dot{r}_{\text{opt}}(0)$ is the first derivative with respect to ζ of r_{opt} evaluated at $\zeta = 0$. From (3.4), we can find that

$$\begin{aligned} \alpha_{\text{opt}} &= \frac{2^{r_{\text{opt}}\zeta} - 1}{\frac{\bar{P}\zeta}{NN_0}} \\ &= \frac{r^*_{\text{opt}} \log_e 2}{\frac{\bar{P}}{NN_0}} + \frac{\dot{r}_{\text{opt}}(0) \log_e 2 + \frac{(r^*_{\text{opt}} \log_e 2)^2}{2}}{\frac{\bar{P}}{NN_0}} \zeta + o(\zeta) \end{aligned} \tag{C.2}$$

from which we have as $\zeta \to 0$ that

$$\alpha^*_{\text{opt}} = \frac{r^*_{\text{opt}} \log_e 2}{\frac{\bar{P}}{NN_0}} \quad \text{and} \tag{C.3}$$

$$\dot{\alpha}_{\text{opt}}(0) = \frac{\dot{r}_{\text{opt}}(0) \log_e 2 + \frac{(r^*_{\text{opt}} \log_e 2)^2}{2}}{\frac{\bar{P}}{NN_0}} \tag{C.4}$$

$$\left.\frac{E_b}{N_0}\right|_{\mathsf{R}_E=0} = \lim_{\zeta\to 0}\frac{\frac{\bar{P}}{NN_0}\zeta}{\mathsf{R}_E(\zeta)} = \frac{\frac{\bar{P}}{NN_0}}{\dot{\mathsf{R}}_E(0)} = \frac{-\frac{\theta T\bar{P}}{NN_0}}{\log_e\left(1-P\{z>\alpha^*_{\text{opt}}\}(1-e^{-\theta T r^*_{\text{opt}}})\right)} = \frac{-\delta\log_e 2}{\log_e \xi} \tag{C.5}$$

---

where $\dot{\alpha}_{\text{opt}}(0)$ is the first derivative with respect to ζ of α_{opt} evaluated at $\zeta = 0$. According to (C.3), $r^*_{\text{opt}} = \frac{\bar{P}\alpha^*_{\text{opt}}}{NN_0\log_e 2}$. We can now derive (C.5) at the top of this page where $\dot{\mathsf{R}}_E(0)$ is the derivative of R_E with respect to ζ at $\zeta = 0$,

$$\delta = \frac{\theta T\bar{P}}{NN_0\log_e 2},$$

and

$$\xi = 1 - P\{z > \alpha^*_{\text{opt}}\}(1 - e^{-\delta\alpha^*_{\text{opt}}}).$$

Since $\frac{E_b}{N_0} = \frac{\frac{\bar{P}}{NN_0}}{\frac{\mathsf{R}_E(\zeta)}{\zeta}}$, the result that $\left.\frac{E_b}{N_0}\right|_{\mathsf{R}_E=0} = \frac{E_b}{N_0}_{\min}$ follows from the fact that $\mathsf{R}_E(\zeta)/\zeta$ monotonically decreases with increasing ζ, and hence achieves its maximum as $\zeta \to 0$. Therefore, we prove (3.20).

The second derivative $\ddot{\mathsf{R}}_E(0)$, required in the computation of the wideband slope $\mathcal{S}_0$, is derived through (C.6)–(C.9) on the next page where $r^*_{\text{opt}} = \frac{\bar{P}\alpha^*_{\text{opt}}}{NN_0\log_e 2}$. Note that (C.8) and (C.9) follow by using L'Hospital's Rule and applying Leibniz Integral Rule [74].

Next, we derive an equality satisfied by α^*_{opt}. Consider the objective function in (3.18)

$$-\frac{1}{\theta TB_c}\log_e\left(1-P\{z>\alpha\}(1-e^{-\theta Tr})\right). \tag{C.10}$$

It can easily be seen that both as $r \to 0$ and $r \to \infty$, this objective function approaches

$$\ddot{\mathsf{R}}_E(0) = \lim_{\zeta\to 0} 2\frac{\mathsf{R}_E(\zeta) - \dot{\mathsf{R}}_E(0)\zeta}{\zeta^2} \tag{C.6}$$

$$= \lim_{\zeta\to 0} 2\frac{1}{\zeta}\Big(-\frac{1}{\theta T}\log_e\Big(1 - P\{z > \alpha_{\text{opt}}\}\Big(1 - e^{-\theta T r_{\text{opt}}}\Big)\Big) + \frac{1}{\theta T}\log_e\Big(1 - P\{z > \alpha^*_{\text{opt}}\}(1 - e^{-\theta T r^*_{\text{opt}}})\Big)\Big) \tag{C.7}$$

$$= \lim_{\zeta\to 0} -\frac{2}{\theta T}\frac{\Big(p_z(\alpha_{\text{opt}})\dot{\alpha}_{\text{opt}}(\zeta)(1 - e^{-\theta T r_{\text{opt}}}) - P\{z > \alpha_{\text{opt}}\}\theta T e^{-\theta T r_{\text{opt}}}\dot{r}_{\text{opt}}(\zeta)\Big)}{1 - P\{z > \alpha_{\text{opt}}\}\Big(1 - e^{-\theta T r_{\text{opt}}}\Big)} \tag{C.8}$$

$$= -\frac{2}{\theta T}\frac{\Big(p_z(\alpha^*_{\text{opt}})\dot{\alpha}_{\text{opt}}(0)(1 - e^{-\theta T r^*_{\text{opt}}}) - P\{z > \alpha^*_{\text{opt}}\}\theta T e^{-\theta T r^*_{\text{opt}}}\dot{r}_{\text{opt}}(0)\Big)}{1 - P\{z > \alpha^*_{\text{opt}}\}\Big(1 - e^{-\theta T r^*_{\text{opt}}}\Big)} \tag{C.9}$$

---

zero[1]. Hence, (C.10) is maximized at a finite and nonzero value of r at which the derivative of (C.10) with respect to r is zero. Differentiating (C.10) with respect to r and making it equal to zero leads to the following equality that needs to be satisfied at the optimal value r_{opt}:

$$\frac{2^{r_{\text{opt}}\zeta} p_z(\alpha_{\text{opt}}) N N_0 \log_e 2}{\bar{P}}(1 - e^{-\theta T r_{\text{opt}}}) = \theta T e^{-\theta T r_{\text{opt}}} P\{z > \alpha_{\text{opt}}\} \tag{C.11}$$

where $\zeta = 1/B_c$. For given θ, as the bandwidth increases (i.e., $\zeta \to 0$), $r_{\text{opt}} \to r^*_{\text{opt}}$. Clearly, $r^*_{\text{opt}} \neq 0$ in the wideband regime. Because, otherwise, if $r_{\text{opt}} \to 0$ and consequently $\alpha_{\text{opt}} \to 0$, the left-hand-side of (C.11) becomes zero, while the right-hand-side is different from zero. So, employing (C.3) and taking the limit of both

[1]Note that α increases without bound with increasing r.

sides of (C.11) as $\zeta \to 0$, we can derive

$$\frac{p_z(\alpha^*_{\text{opt}})NN_0 \log_e 2}{\bar{P}} \left(1 - e^{-\frac{\theta T \bar{P}}{NN_0 \log_e 2}\alpha^*_{\text{opt}}}\right) = \theta T e^{-\frac{\theta T \bar{P}}{NN_0 \log_e 2}\alpha^*_{\text{opt}}} P\{z > \alpha^*_{\text{opt}}\} \tag{C.12}$$

which, after rearranging, yields

$$\frac{\theta T \bar{P}}{NN_0 \log_e 2}\alpha^*_{\text{opt}} = \log_e \left(1 + \frac{\theta T \bar{P}}{NN_0 \log_e 2}\frac{P\{z > \alpha^*_{\text{opt}}\}}{p_z(\alpha^*_{\text{opt}})}\right). \tag{C.13}$$

Denoting $\delta = \frac{\theta T \bar{P}}{NN_0 \log_e 2}$, we obtain the condition (3.22) stated in the theorem.

Combining (C.12) and (C.4) with (C.9) gives us

$$\begin{aligned} \ddot{\mathsf{R}}_E(0) &= -\frac{NN_0 \log_e^2 2}{\theta T \bar{P}} \frac{{r^*_{\text{opt}}}^2 p_z(\alpha^*_{\text{opt}})(1 - e^{-\theta T r^*_{\text{opt}}})}{1 - P\{z > \alpha^*_{\text{opt}}\}\left(1 - e^{-\theta T r^*_{\text{opt}}}\right)} \\ &= -\frac{{r^*_{\text{opt}}}^2 P\{z > \alpha^*_{\text{opt}}\} e^{-\theta T r^*_{\text{opt}}} \log_e 2}{1 - P\{z > \alpha^*_{\text{opt}}\}\left(1 - e^{-\theta T r^*_{\text{opt}}}\right)} \end{aligned} \tag{C.14}$$

Substituting (C.14) and the expression for $\dot{\mathsf{R}}_E(0)$ in (C.5) into (3.12), we obtain (3.21).

□

Appendix D

Proof of Theorem 8

We first consider the Taylor series expansion of r_{opt} in the low-SNR regime:

$$r_{\text{opt}} = a\text{SNR} + b\text{SNR}^2 + o(\text{SNR}^2) \tag{D.1}$$

where a and b are real-valued constants. Substituting (D.1) into (3.4), we obtain the Taylor series expansion for α_{opt}:

$$\alpha_{\text{opt}} = \frac{a \log_e 2}{B} + \left(\frac{b \log_e 2}{B} + \frac{a^2 \log_e^2 2}{2B^2} \right) \text{SNR} + o(\text{SNR}). \tag{D.2}$$

From (D.2), we note that in the limit as $\text{SNR} \to 0$, we have

$$\alpha^*_{\text{opt}} = \frac{a \log_e 2}{B}. \tag{D.3}$$

Next, we obtain the Taylor series expansion with respect to SNR for $P\{z > \alpha_{\text{opt}}\}$ using the Leibniz Integral Rule [74] as in (D.4) on the next page.

Using (D.1), (D.2), and (D.4), we find the series expansion for R_E given in (3.10) as in (D.5) on the next page. Then, using (D.3), we immediately derive from (D.5)

$$P\{z > \alpha_{\rm opt}\} = P\{z > \alpha^*_{\rm opt}\} - \left(\frac{b\log_e 2}{B} + \frac{a^2\log_e^2 2}{2B^2}\right) p_z(\alpha^*_{\rm opt}){\rm SNR} + o({\rm SNR}). \qquad \text{(D.4)}$$

$$\begin{aligned}
\mathsf{R}_E({\rm SNR}) &= -\frac{1}{\theta TB}\log_e\Bigg[1 - \Big(P\{z > \alpha^*_{\rm opt}\} - \Big(\frac{b\log_e 2}{B} + \frac{a^2\log_e^2 2}{2B^2}\Big)p_z(\alpha^*_{\rm opt}){\rm SNR} + o({\rm SNR})\Big) \\
&\qquad\qquad \times \big(\theta Ta{\rm SNR} + (\theta Tb - \frac{(\theta Ta)^2}{2}){\rm SNR}^2 + o({\rm SNR}^2)\big)\Bigg] \\
&= \frac{aP\{z > \alpha^*_{\rm opt}\}}{B}{\rm SNR} + \frac{1}{B}\Big(-\frac{\theta Ta^2}{2}P\{z > \alpha^*_{\rm opt}\} - \frac{a^3 p_z(\alpha^*_{\rm opt})\log_e^2 2}{2B^2} \\
&\qquad + \frac{\theta T(P\{z > \alpha^*_{\rm opt}\}a)^2}{2}\Big){\rm SNR}^2 + o({\rm SNR}^2). \qquad \text{(D.5)}
\end{aligned}$$

---

that

$$\dot{\mathsf{R}}_E(0) = \frac{\alpha^*_{\rm opt}P\{z > \alpha^*_{\rm opt}\}}{\log_e 2}, \qquad \text{(D.6)}$$

$$\begin{aligned}
\ddot{\mathsf{R}}_E(0) &= -\frac{{\alpha^*_{\rm opt}}^3 p_z\{\alpha^*_{\rm opt}\}}{\log_e 2} \\
&\qquad - \frac{\theta TB{\alpha^*_{\rm opt}}^2}{\log_e^2 2}P\{z > \alpha^*_{\rm opt}\}(1 - P\{z > \alpha^*_{\rm opt}\}). \qquad \text{(D.7)}
\end{aligned}$$

Similarly as in the discussion in the proof of Theorem 7 in section 3.3, the optimal fixed-rate $r_{\rm opt}$, akin to (C.11), should satisfy

$$\frac{2^{r_{\rm opt}/B}p_z(\alpha_{\rm opt})\log_e 2}{B{\rm SNR}}(1 - e^{-\theta Tr_{\rm opt}}) = \theta Te^{-\theta Tr_{\rm opt}}P\{z > \alpha_{\rm opt}\}. \qquad \text{(D.8)}$$

Taking the limits of both sides of (D.8) as ${\rm SNR} \to 0$ and employing (D.1), we obtain

$$\frac{ap_z(\alpha^*_{\rm opt})\log_e 2}{B} = P\{z > \alpha^*_{\rm opt}\}. \qquad \text{(D.9)}$$

From (D.3), (D.9) simplifies to

$$\alpha^*_{\text{opt}} p_z(\alpha^*_{\text{opt}}) = P\{z > \alpha^*_{\text{opt}}\}, \tag{D.10}$$

proving the condition in (3.26). Moreover, using (D.10), the first term in the expression for $\ddot{\mathsf{R}}_E(0)$ in (D.7) becomes $-\frac{{\alpha^*_{\text{opt}}}^2 P\{z \geq \alpha^*_{\text{opt}}\}}{\log_e 2}$. Together with this change, evaluating the expressions in (3.12) with the results in (D.6) and (D.7), we obtain (3.24) and (3.25). □

Appendix E

Proof of Proposition 2

From the maximization problem above (4.10) and the definition of α in (4.8), we can easily see that for fixed r, the only term in the objective function in this maximization that depends on ρ is α. Moreover, α has this dependency through SNR_{eff}. Therefore, ρ_{opt} that maximizes the objective function can be found by minimizing α, or equivalently maximizing SNR_{eff}. Substituting the definitions in (4.2) and the expressions for $\sigma^2_{h_{est}}$ and $\sigma^2_{h_{err}}$ into (4.6), we have

$$\text{SNR}_{\text{eff}} = \frac{\mathcal{E}_s \sigma^2_{h_{est}}}{\sigma^2_{h_{err}} \mathcal{E}_s + N_0} = \frac{\rho(1-\rho)\gamma^2 T^2 B^2 \text{SNR}^2}{\rho\gamma TB(TB-2)\text{SNR} + \gamma TB\text{SNR} + TB - 1} \tag{E.1}$$

where $\text{SNR} = \frac{\bar{P}}{N_0 B}$. Evaluating the derivative of SNR_{eff} with respect to ρ and making it equal to zero leads to the expression in (4.12). Clearly, ρ_{opt} is independent of θ and r.

Above, we have implicitly assumed that the maximization is performed with respect to first ρ and then r. However, the result will not alter if the order of the maximization is changed. Note that the objective function in the maximization above

(4.10),

$$g(\mathrm{SNR}_{\mathrm{eff}}, r) = -\frac{1}{\theta TB} \log_e \left(1 - e^{-\frac{2^{\frac{rT}{TB-1}}-1}{\mathrm{SNR}_{\mathrm{eff}}}} (1 - e^{-\theta T r}) \right), \tag{E.2}$$

is a monotonically increasing function of $\mathrm{SNR}_{\mathrm{eff}}$ for all r. It can be easily verified that maximization does not affect the monotonicity of g, and hence $\max_{r \geq 0} g(\mathrm{SNR}_{\mathrm{eff}}, r)$ is still a monotonically increasing function of $\mathrm{SNR}_{\mathrm{eff}}$. Therefore, in the outer maximization with respect to ρ, the choice of ρ that maximizes $\mathrm{SNR}_{\mathrm{eff}}$ will also maximize $\max_{r \geq 0} g(\mathrm{SNR}_{\mathrm{eff}}, r)$, and the optimal value of ρ is again given by (4.12). □

Appendix F

Proof of Theorem 10

Note that as $\mathrm{SNR} \to 0$, transmission rates also approach zero and therefore we have $r_{\mathrm{opt}} \to 0$. Using this fact, it can be shown that the derivative of R_E in (4.15) with respect to SNR at $\mathrm{SNR} = 0$ is

$$\dot{\mathsf{R}}_E(0) = \lim_{\mathrm{SNR}\to 0} \frac{1}{B} e^{-\alpha_{\mathrm{opt}}}\, \dot{r}_{\mathrm{opt}}\, e^{-\theta T r_{\mathrm{opt}}} - \frac{1}{\theta T B}\, \dot{\alpha}_{\mathrm{opt}}\, e^{-\alpha_{\mathrm{opt}}} (1 - e^{-\theta T r_{\mathrm{opt}}}) \tag{F.1}$$

where $\dot{r}_{\mathrm{opt}}$ and $\dot{\alpha}_{\mathrm{opt}}$ are the derivatives of r_{opt} and α_{opt}, respectively, with respect to SNR, and $\alpha_{\mathrm{opt}} = \frac{2^{\frac{r_{\mathrm{opt}} T}{TB-1}} - 1}{\mathrm{SNR}_{\mathrm{eff,opt}}}$. Next, we investigate how $\mathrm{SNR}_{\mathrm{eff,opt}}$ scales as SNR vanishes. Note that as $\mathrm{SNR} \to 0$, $\eta \to \infty$, $\rho_{\mathrm{opt}} \to 1/2$, and hence $\phi(\mathrm{SNR}) \to 1/4\gamma^2 T^2 B^2$. Then, we have

$$\mathrm{SNR}_{\mathrm{eff,opt}} = \frac{\gamma^2 T^2 B^2}{4(TB-1)} \mathrm{SNR}^2 + o(\mathrm{SNR}^2). \tag{F.2}$$

Therefore, $\mathrm{SNR}_{\mathrm{eff,opt}}$ decreases as SNR^2 as SNR diminishes to zero. Now, we consider the behavior of r_{opt} at low SNRs. If r_{opt} diminishes slower than SNR^2 (for instance, if r_{opt} decreases as SNR^a where $0 < a < 2$), then it can be verified that $\alpha_{\mathrm{opt}} \to \infty$ as $\mathrm{SNR} \to 0$ from which we can immediately see that $\dot{\mathsf{R}}_E(0) = 0$ due to exponentially

decreasing term $e^{-\alpha_{\text{opt}}}$. On the other hand, if r_{opt} reduces to zero faster than or as SNR^2 (e.g., as SNR^a where $a \geq 2$), α_{opt} approaches a finite value. However in this case, we can show that $\dot{r}_{\text{opt}} \to 0$ and $\dot{\alpha}_{\text{opt}}(1 - e^{-\theta T r_{\text{opt}}}) \to 0$ as $\text{SNR} \to 0$, leading again the conclusion that $\dot{\mathsf{R}}_E(0) = 0$. $\square$

Appendix G

Proof of Theorem 11

We define $\zeta = \frac{1}{B_c}$. Recall that in the scenario considered in Theorem 11, B_c grows linearly with B while N is kept fixed. Therefore, we have $\zeta \to 0$ as $B \to \infty$. According to the expression of SNR_{eff} given in the line below (4.26), we have the following result similar to (4.12) for ρ_{opt} in this case:

$$\rho_{\text{opt}} = \sqrt{\eta(\eta+1)} - \eta \tag{G.1}$$

where

$$\eta = \frac{\gamma T B_c \text{SNR} + TB_c - 1}{\gamma T B_c (TB_c - 2)\text{SNR}} \quad \text{and} \quad \text{SNR} = \frac{\bar{P}}{N_0 B}. \tag{G.2}$$

We first derive the following asymptotic expansion for the optimal fraction ρ_{opt}

$$\rho_{\text{opt}} = \rho_{\text{opt}}^* + \dot{\rho_{\text{opt}}}(0)\zeta + o(\zeta) \tag{G.3}$$

where ρ_{opt}^* is the asymptotic value of ρ_{opt} attained as $\zeta \to 0$, and $\dot{\rho_{\text{opt}}}(0)$ is the first

$$\omega = \frac{\frac{\gamma^2 P^2 T}{(NN_0)^2}}{1+\frac{\rho^*_{\rm opt}\gamma\bar{P}T}{NN_0}}\left(\dot{\rho_{\rm opt}}(0)(1-2\rho^*_{\rm opt}) - \frac{(1-2\rho^*_{\rm opt})\frac{\gamma\bar{P}}{NN_0}+\dot{\rho_{\rm opt}}(0)\frac{\gamma\bar{P}T}{NN_0}-\frac{1}{T}}{1+\frac{\rho^*_{\rm opt}\gamma\bar{P}T}{NN_0}}\rho^*_{\rm opt}(1-\rho^*_{\rm opt})\right)$$
$$= -\frac{\gamma\bar{P}}{NN_0T}\left(\sqrt{1+\frac{NN_0}{\gamma\bar{P}T}}-\sqrt{\frac{NN_0}{\gamma\bar{P}T}}\right)^2\left(\sqrt{1+\frac{\gamma\bar{P}T}{NN_0}}-2\right). \quad \text{(G.8)}$$

---

derivative of $\rho_{\rm opt}$ evaluated at $\zeta = 0$. We can easily find that

$$\rho^*_{\rm opt} = \sqrt{\frac{NN_0}{\gamma\bar{P}T}\left(1+\frac{NN_0}{\gamma\bar{P}T}\right)} - \frac{NN_0}{\gamma\bar{P}T} \quad \text{(G.4)}$$

and

$$\dot{\rho_{\rm opt}}(0) = \frac{1}{2T}\sqrt{1+\frac{\gamma\bar{P}T}{NN_0}}\left(\sqrt{1+\frac{NN_0}{\gamma\bar{P}T}}-\sqrt{\frac{NN_0}{\gamma\bar{P}T}}\right)^2. \quad \text{(G.5)}$$

Furthermore, $\mathrm{SNR}_{\rm eff,opt}$ defined in the line below (4.26) satisfies

$$\mathrm{SNR}_{\rm eff,opt} = \varphi\zeta + \omega\zeta^2 + o(\zeta^2) \quad \text{(G.6)}$$

where

$$\varphi = \frac{\rho^*_{\rm opt}(1-\rho^*_{\rm opt})\frac{\gamma^2\bar{P}^2T}{(NN_0)^2}}{1+\frac{\rho^*_{\rm opt}\gamma\bar{P}T}{NN_0}} = \frac{\gamma\bar{P}}{NN_0}\left(\sqrt{1+\frac{NN_0}{\gamma\bar{P}T}}-\sqrt{\frac{NN_0}{\gamma\bar{P}T}}\right)^2 \quad \text{(G.7)}$$

and ω is given by (G.8) at the top of this page.

Now, assume that the Taylor series expansion of $r_{\rm opt}$ with respect to small ζ is

$$r_{\rm opt} = r^*_{\rm opt} + \dot{r}_{\rm opt}(0)\zeta + o(\zeta) \quad \text{(G.9)}$$

where $r^*_{\rm opt} = \lim_{\zeta\to 0} r_{\rm opt}$ and $\dot{r}_{\rm opt}(0)$ is the first derivative with respect to ζ of $r_{\rm opt}$ evaluated at $\zeta = 0$. From (4.8) and the above asymptotic expansions of $\rho_{\rm opt}$ and $\mathrm{SNR}_{\rm eff,opt}$, we can find $\alpha_{\rm opt}$ as (G.10) in the next page, from which we have as $\zeta \to 0$

$$\begin{aligned}\alpha_{\text{opt}} &= \frac{2^{\frac{r_{\text{opt}}\zeta}{1-\zeta/T}}-1}{\text{SNR}_{\text{eff,opt}}} \\ &= \frac{r^*_{\text{opt}}\log_e 2 + \left[\left(\frac{r^*_{\text{opt}}}{T}+\dot{r}_{\text{opt}}(0)\right)\log_e 2 + \frac{(r^*_{\text{opt}}\log_e 2)^2}{2}\right]\zeta + o(\zeta)}{\varphi+\omega\zeta+o(\zeta)} \\ &= \frac{r^*_{\text{opt}}\log_e 2}{\varphi} + \left(\frac{\dot{r}_{\text{opt}}(0)\log_e 2}{\varphi} + \frac{r^*_{\text{opt}}\log_e 2}{\varphi}\left(\frac{1}{T}-\frac{\omega}{\varphi}\right) + \frac{(r^*_{\text{opt}}\log_e 2)^2}{2\varphi}\right)\zeta + o(\zeta)\end{aligned} \tag{G.10}$$

---

that

$$\alpha^*_{\text{opt}} = \frac{r^*_{\text{opt}}\log_e 2}{\varphi} \tag{G.11}$$

and that

$$\dot{\alpha}_{\text{opt}}(0) = \frac{\dot{r}_{\text{opt}}(0)\log_e 2}{\varphi} + \frac{r^*_{\text{opt}}\log_e 2}{\varphi}\left(\frac{1}{T}-\frac{\omega}{\varphi}\right) + \frac{(r^*_{\text{opt}}\log_e 2)^2}{2\varphi} \tag{G.12}$$

where $\dot{\alpha}_{\text{opt}}(0)$ is the first derivative with respect to ζ of α_{opt} evaluated at $\zeta = 0$. Note also that we have $r^*_{\text{opt}} = \frac{\varphi\alpha^*_{\text{opt}}}{\log_e 2}$ according to (G.11).

Note that the derivative with respect to r of the objective function in the maximization in (4.26) is zero at the optimal value $r = r_{\text{opt}}$. Combining (G.6) and (G.11) and letting $\zeta \to 0$ in this derivative expression at $r = r_{\text{opt}}$, we obtain

$$\frac{\log_e 2}{\varphi}\left(1 - e^{-\frac{\theta T\varphi\alpha^*_{\text{opt}}}{\log_e 2}}\right) - \theta T e^{-\theta T r^*_{\text{opt}}} = 0 \tag{G.13}$$

from which we get

$$\alpha^*_{\text{opt}} = \frac{\log_e 2}{\theta T\varphi}\log_e\left(1+\frac{\theta T\varphi}{\log_e 2}\right). \tag{G.14}$$

Since $\frac{E_b}{N_0} = \frac{\frac{\bar{P}}{NN_0}}{\frac{\mathsf{R}_E(\zeta)}{\zeta}}$, the result that $\left.\frac{E_b}{N_0}\right|_{\mathsf{R}_E=0} = \frac{E_b}{N_0\,\min}$ follows from the fact that $\mathsf{R}_E(\zeta)/\zeta$ monotonically decreases with increasing ζ, and hence achieves its maximum

$$\begin{aligned}
\ddot{\mathsf{R}}_E(0) &= \lim_{\zeta\to 0} 2\frac{\mathsf{R}_E(\zeta)-\dot{\mathsf{R}}_E(0)\zeta}{\zeta^2} \\
&= \lim_{\zeta\to 0} 2\frac{1}{\zeta}\Big(-\frac{1}{\theta T}\log_e\Big(1-P\{|w|^2\geq\alpha_{\rm opt}\}\Big(1-e^{-\theta T r_{\rm opt}}\Big)\Big) \\
&\qquad +\frac{1}{\theta T}\log_e\Big(1-P\{|w|^2\geq\alpha^*_{\rm opt}\}(1-e^{-\theta T r^*_{\rm opt}})\Big)\Big) \\
&= \lim_{\zeta\to 0} -\frac{2e^{-\alpha_{\rm opt}}}{\theta T\left(1-P\{|w|^2\geq\alpha_{\rm opt}\}\left(1-e^{-\theta T r_{\rm opt}}\right)\right)} \\
&\qquad \times\Big(\dot\alpha_{\rm opt}(\zeta)(1-e^{-\theta T r_{\rm opt}})-\theta T e^{-\theta T r_{\rm opt}}\dot r_{\rm opt}(\zeta)\Big) \qquad \text{(G.16)} \\
&= -\frac{2e^{-\alpha^*_{\rm opt}}}{\theta T\Big(1-P\{|w|^2\geq\alpha^*_{\rm opt}\}\Big(1-e^{-\theta T r^*_{\rm opt}}\Big)\Big)} \\
&\qquad \times\Big(\dot\alpha_{\rm opt}(0)(1-e^{-\theta T r^*_{\rm opt}})-\theta T e^{-\theta T r^*_{\rm opt}}\dot r_{\rm opt}(0)\Big) \qquad \text{(G.17)}
\end{aligned}$$

---

as $\zeta\to 0$. We now have

$$\begin{aligned}
\frac{E_b}{N_0\,\min} &= \lim_{\zeta\to 0}\frac{\frac{\bar P}{NN_0}\zeta}{\mathsf{R}_E(\zeta)} = \frac{-\frac{\theta T\bar P}{NN_0}}{\log_e\Big(1-P\{|w|^2\geq\alpha^*_{\rm opt}\}(1-e^{-\theta T r^*_{\rm opt}})\Big)} \\
&= \frac{-\delta}{\log_e\xi} = \frac{\frac{\bar P}{NN_0}}{\dot{\mathsf{R}}_E(0)} \qquad \text{(G.15)}
\end{aligned}$$

where $\dot{\mathsf{R}}_E(0)$ is the derivative of R_E with respect to ζ at $\zeta=0$, $\delta=\frac{\theta T\bar P}{NN_0}$, and $\xi=1-P\{|w|^2\geq\alpha^*_{\rm opt}\}(1-e^{-\frac{\theta T\varphi\alpha^*_{\rm opt}}{\log_e 2}})$. Obviously, (G.15) provides (4.27).

Note that the second derivative $\ddot{\mathsf{R}}_E(0)$, required in the computation of the wide-band slope $\mathcal{S}_0$, can be obtained from (G.16)(G.17) at the top of this page, where $r^*_{\rm opt}=\frac{\bar P\alpha^*_{\rm opt}}{NN_0\log_e 2}$. (G.16) and (G.17) follow by using L'Hospital's Rule and applying Leibniz Integral Rule.

Meanwhile, substituting (G.13) and (G.12) into (G.17) gives us

$$\begin{aligned}
\ddot{\mathsf{R}}_E(0) &= -\frac{2e^{-\alpha^*_{\text{opt}}}}{\theta T\left(1-P\{|w|^2\geq\alpha^*_{\text{opt}}\}\left(1-e^{-\theta T r^*_{\text{opt}}}\right)\right)} \\
&\quad\times\alpha^*_{\text{opt}}(1-e^{-\theta T r^*_{\text{opt}}})\left(\frac{1}{T}-\frac{\omega}{\varphi}+\frac{\varphi\alpha^*_{\text{opt}}}{2}\right) \\
&= -\frac{2(1-\xi)\alpha^*_{\text{opt}}}{\theta T\xi}\left(\frac{1}{T}-\frac{\omega}{\varphi}+\frac{\varphi\alpha^*_{\text{opt}}}{2}\right) \\
&= -\frac{2(1-\xi)\alpha^*_{\text{opt}}}{\theta T\xi}\left(\frac{1}{T}\left(\sqrt{1+\frac{\gamma\bar{P}T}{NN_0}}-1\right)+\frac{\varphi\alpha^*_{\text{opt}}}{2}\right)
\end{aligned} \tag{G.18}$$

Combining (G.18) and (G.15), we can prove (4.28) using the wideband slope formula in (3.12). □

Appendix H

Proof for Theorem 19

We need to consider the wideband slopes for different decoding order assignments. Due to the complex expressions involved, we here state the derivation for $\mathcal{S}_1$ for the case in which the decoding order is (1,2) when $z_2 < g(z_1, \mathrm{SNR}_1)$, and the decoding order is (2,1) when $z_2 > g(z_1, \mathrm{SNR}_1)$. Taking the second derivative of (6.9), we have

$$\ddot{\mathsf{C}}_1(\mathrm{SNR}_1) = -\frac{\ddot{\phi}_1\phi_1 - (\dot{\phi}_1)^2}{\beta_1\phi_1^2 \log_e 2} \tag{H.1}$$

where $\dot{\phi}_1$ is provided in (6.19) and $\ddot{\phi}_1$ is given by

$$\begin{aligned}
\ddot{\phi}_1 &= \int_0^\infty \left(1+\frac{\mathrm{SNR}_1 z_1}{1+\mathrm{SNR}_1 g(\mathrm{SNR}_1)/\lambda}\right)^{-\beta_1} p(z_1, g(\mathrm{SNR}_1))\ddot{g}(\mathrm{SNR}_1)dz_1 \\
&- 2\beta_1 \int_0^\infty \left(1+\frac{\mathrm{SNR}_1 z_1}{1+\mathrm{SNR}_1 g(\mathrm{SNR}_1)/\lambda}\right)^{-\beta_1-1} \frac{z_1}{(1+\mathrm{SNR}_1 g(\mathrm{SNR}_1)/\lambda)^2} \\
&\times p(z_1, g(\mathrm{SNR}_1))\dot{g}(\mathrm{SNR}_1)dz_1 \\
&+ \int_0^\infty \left(1+\frac{\mathrm{SNR}_1 z_1}{1+\mathrm{SNR}_1 g(\mathrm{SNR}_1)/\lambda}\right)^{-\beta_1} \dot{p}(z_1, g(\mathrm{SNR}_1)) \left(\dot{g}(\mathrm{SNR}_1)\right)^2 dz_1 \\
&+ \beta_1(\beta_1+1)\int_0^\infty \int_0^\infty \left(1+\frac{\mathrm{SNR}_1 z_1}{1+\mathrm{SNR}_1 z_1/\lambda}\right)^{-\beta_1-2} \frac{z_1^2}{(1+\mathrm{SNR}_1 g(\mathrm{SNR}_1)/\lambda)^4} p(z_1, z_2)dz_2dz_1 \\
&+ \frac{2\beta_1}{\lambda}\int_0^\infty \int_0^{g(\mathrm{SNR}_1)} \left(1+\frac{\mathrm{SNR}_1 z_1}{1+\mathrm{SNR}_1 z_1/\lambda}\right)^{-\beta_1-1} \frac{z_1 z_2}{(1+\mathrm{SNR}_1 g(\mathrm{SNR}_1)/\lambda)^3} p(z_1, z_2)dz_2dz_1 \\
&- \int_0^\infty (1+\mathrm{SNR}_1 z_1)^{-\beta_1} p(z_1, g(\mathrm{SNR}_1))\ddot{g}(\mathrm{SNR}_1)dz_1 \\
&+ 2\beta_1 \int_0^\infty (1+\mathrm{SNR}_1 z_1)^{-\beta_1-1} z_1 p(z_1, g(\mathrm{SNR}_1))\dot{g}(\mathrm{SNR}_1)dz_1 \\
&- \int_0^\infty (1+\mathrm{SNR}_1 z_1)^{-\beta_1} \dot{p}(z_1, g(\mathrm{SNR}_1)) \left(\dot{g}(\mathrm{SNR}_1)\right)^2 dz_1 \\
&+ \beta_1(\beta_1+1)\int_0^\infty \int_{g(\mathrm{SNR}_1)}^\infty (1+\mathrm{SNR}_1 z_1)^{-\beta_1-2} z_1^2 p(z_1, z_2)dz_2dz_1. \qquad \text{(H.2)}
\end{aligned}$$

Letting $\mathrm{SNR}_1 = 0$ and supposing that $g(0)$, $\dot{g}(0)$, and $\ddot{g}(0)$ are finite, we have

$$\ddot{\mathsf{C}}_1(0) = -\frac{1}{\log_e 2}\left(\beta_1\left(\mathbb{E}\{z_1^2\} - (\mathbb{E}\{z_1\})^2\right) + \mathbb{E}\{z_1^2\} + \frac{2}{\lambda}\int_0^\infty \int_0^{g(0)} z_1 z_2 p(z_1, z_2)dz_2dz_1\right) \qquad \text{(H.3)}$$

Substituting (H.3) and (6.21) into (1.7), we obtain

$$\mathcal{S}_1 = \frac{2(\mathbb{E}\{z_1\})^2}{\beta_1\left(\mathbb{E}\{z_1^2\} - (\mathbb{E}\{z_1\})^2\right) + \mathbb{E}\{z_1^2\} + \frac{2}{\lambda}\int_0^\infty \int_0^{g(0)} z_1 z_2 p(z_1, z_2)dz_2dz_1}. \qquad \text{(H.4)}$$

Similarly, we can derive

$$\mathcal{S}_2 = \frac{2(\mathbb{E}\{z_2\})^2}{\beta_2\left(\mathbb{E}\{z_2^2\} - (\mathbb{E}\{z_2\})^2\right) + \mathbb{E}\{z_2^2\} + 2\lambda \int_0^\infty \int_{g(0)}^\infty z_1 z_2 p(z_1, z_2) dz_2 dz_1}. \tag{H.5}$$

If the decoding order is (2,1) when $z_2 < g(z_1, \mathrm{SNR}_1)$, and is (1,2) when $z_2 > g(z_1, \mathrm{SNR}_1)$, following the steps described above, we can obtain

$$\mathcal{S}_1 = \frac{2(\mathbb{E}\{z_1\})^2}{\beta_1\left(\mathbb{E}\{z_1^2\} - (\mathbb{E}\{z_1\})^2\right) + \mathbb{E}\{z_1^2\} + \frac{2}{\lambda} \int_0^\infty \int_{g(0)}^\infty z_1 z_2 p(z_1, z_2) dz_2 dz_1} \tag{H.6}$$

$$\mathcal{S}_2 = \frac{2(\mathbb{E}\{z_2\})^2}{\beta_2\left(\mathbb{E}\{z_2^2\} - (\mathbb{E}\{z_2\})^2\right) + \mathbb{E}\{z_2^2\} + 2\lambda \int_0^\infty \int_0^{g(0)} z_1 z_2 p(z_1, z_2) dz_2 dz_1}. \tag{H.7}$$

Combining (H.4) and (H.5) and eliminating $g(0)$, we can obtain the third condition in (6.32). It is interesting that combining (H.6) and (H.7) and eliminating $g(0)$, we still get the same third condition stated in (6.32). This shows us that the slope regions for different decoding order assignments overlap. □

Appendix I

Proof of Theorem 21

Taking the first derivatives of (6.37) and (6.38) and letting $\zeta = 0$, we obtain

$$\dot{\mathsf{C}}_1(0) = -\frac{1}{\theta_1 T} \log_e \mathbb{E}_{z_1} \left\{ e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} \right\} \tag{I.1}$$

$$\dot{\mathsf{C}}_2(0) = -\frac{1}{\theta_2 T} \log_e \mathbb{E}_{z_2} \left\{ e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} \right\}. \tag{I.2}$$

Substituting (I.1) and (I.2) into (6.36), we get the results in (6.43) and (6.44).

Next, we consider the superposition coding with fixed decoding. Evaluating the first derivative of (6.39) and (6.40) at $\zeta = 0$, we again get

$$\dot{\mathsf{C}}_1(0) = -\frac{1}{\theta_1 T} \log_e \mathbb{E}_{z_1} \left\{ e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} \right\} \tag{I.3}$$

$$\dot{\mathsf{C}}_2(0) = -\frac{1}{\theta_2 T} \log_e \mathbb{E}_{z_2} \left\{ e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} \right\}. \tag{I.4}$$

which imply the results in (6.43) and (6.44).

We can also prove the results for the variable decoding case similarly as in the proof of Theorem 16. Consider (6.41) and (6.42) with the associated decoding order.

The first derivative of (6.41) can be expressed as

$$\dot{\mathsf{C}}_1(\zeta) = -\frac{1}{\theta_1 T}\log_e \phi_1 - \frac{\zeta\dot{\phi}_1}{\theta_1 T\phi_1} \tag{I.5}$$

where ϕ_1 is

$$\begin{aligned}\phi_1 = &\int_0^\infty \int_{g(\frac{\bar{P}_1\zeta}{N_0})}^\infty e^{-\frac{\theta_1 T}{\zeta}\log_2\left(1+\frac{\bar{P}_1 z_1\zeta}{N_0}\right)} p_{\mathbf{z}}(z_1,z_2)dz_2dz_1 \\ &+\int_0^\infty \int_0^{g(\frac{\bar{P}_1\zeta}{N_0})} e^{-\frac{\theta_1 T}{\zeta}\log_2\left(1+\frac{\frac{\bar{P}_1 z_1\zeta}{N_0}}{1+\frac{\bar{P}_2 z_2\zeta}{N_0}}\right)} p_{\mathbf{z}}(z_1,z_2)dz_2dz_1\end{aligned} \tag{I.6}$$

and $\dot{\phi}_1$ is

$$\begin{aligned}\dot{\phi}_1 = &-\int_0^\infty \dot{g}\left(\frac{\bar{P}\zeta}{N_0}\right)\frac{\bar{P}_1}{N_0}e^{-\frac{\theta_1 T}{\zeta}\log_2\left(1+\frac{\bar{P}_1 z_1\zeta}{N_0}\right)} p_{\mathbf{z}}(z_1, g(\bar{P}_1\zeta/N_0))dz_1 \\ &+\int_0^\infty \int_{g(\frac{\bar{P}_1\zeta}{N_0})}^\infty e^{-\frac{\theta_1 T}{\zeta}\log_2\left(1+\frac{\bar{P}_1 z_1\zeta}{N_0}\right)}\left(\frac{\theta_1 T}{\zeta^2}\log_2(1+\frac{\bar{P}_1 z_1\zeta}{N_0}) - \frac{\theta_1 T}{\zeta}\frac{\frac{\bar{P}_1 z_1}{N_0\log_e 2}}{1+\frac{\bar{P}_1 z_1\zeta}{N_0}}\right) p_{\mathbf{z}}(z_1,z_2)dz_2dz_1 \\ &+\int_0^\infty \dot{g}\left(\frac{\bar{P}_1\zeta}{N_0}\right)\frac{\bar{P}_1}{N_0}e^{-\frac{\theta_1 T}{\zeta}\log_2\left(1+\frac{\frac{\bar{P}_1 z_1\zeta}{N_0}}{1+\frac{\bar{P}_2 g(\bar{P}_1\zeta/N_0)\zeta}{N_0}}\right)} p_{\mathbf{z}}(z_1, g(\bar{P}_1\zeta/N_0))dz_1 \\ &+\int_0^\infty \int_0^{g(\frac{\bar{P}_1\zeta}{N_0})} e^{-\frac{\theta_1 T}{\zeta}\log_2\left(1+\frac{\frac{\bar{P}_1 z_1\zeta}{N_0}}{1+\frac{\bar{P}_2 z_2\zeta}{N_0}}\right)}\left(\frac{\theta_1 T}{\zeta^2}\log_2\left(1+\frac{\frac{\bar{P}_1 z_1\zeta}{N_0}}{1+\frac{\bar{P}_2 z_2\zeta}{N_0}}\right)\right. \\ &\left.-\frac{\theta_1 T}{\zeta}\frac{\frac{\bar{P}_1 z_1}{N_0\log_e 2}}{(1+\frac{\bar{P}_2 z_2\zeta}{N_0})(1+\frac{\bar{P}_1 z_1\zeta}{N_0}+\frac{\bar{P}_2 z_2\zeta}{N_0})}\right) p_{\mathbf{z}}(z_1,z_2)dz_2dz_1.\end{aligned} \tag{I.7}$$

If we define $f(\zeta) = \frac{\theta_1 T}{\zeta^2}\log_2(1+\frac{\bar{P}_1 z_1 \zeta}{N_0}) - \frac{\theta_1 T}{\zeta}\frac{\frac{\bar{P}_1 z_1}{N_0 \log_e 2}}{1+\frac{\bar{P}_1 z_1 \zeta}{N_0}}$, we can show that

$$\begin{aligned}
\lim_{\zeta\to 0} f(\zeta) &= \theta_1 T \lim_{\zeta\to 0} \frac{\frac{\log_2(1+\frac{\bar{P}_1 z_1 \zeta}{N_0})}{\zeta} - \frac{\frac{\bar{P}_1 z_1}{N_0 \log_e 2}}{1+\frac{\bar{P}_1 z_1 \zeta}{N_0}}}{\zeta} \\
&= \theta_1 T \lim_{\zeta\to 0} \left(-\frac{1}{\zeta^2}\log_2(1+\frac{\bar{P}_1 z_1 \zeta}{N_0}) + \frac{1}{\zeta}\frac{\frac{\bar{P}_1 z_1}{N_0 \log_e 2}}{1+\frac{\bar{P}_1 z_1 \zeta}{N_0}} + \frac{\left(\frac{\bar{P}_1 z_1}{N_0}\right)^2}{(1+\frac{\bar{P}_1 z_1 \zeta}{N_0})^2 \log_e 2} \right) \\
&= -\lim_{\zeta\to 0} f(\zeta) + \frac{\theta_1 T}{\log_e 2}\left(\frac{\bar{P}_1 z_1}{N_0}\right)^2
\end{aligned} \tag{I.8}$$

which gives us that

$$\lim_{\zeta\to 0} f(\zeta) = \frac{\theta_1 T}{2\log_e 2}\left(\frac{\bar{P}_1 z_1}{N_0}\right)^2. \tag{I.9}$$

Similarly, we can show that

$$\begin{aligned}
&\lim_{\zeta\to 0}\left(\frac{\theta_1 T}{\zeta^2}\log_2\left(1+\frac{\frac{\bar{P}_1 z_1 \zeta}{N_0}}{1+\frac{\bar{P}_2 z_2 \zeta}{N_0}}\right) - \frac{\theta_1 T}{\zeta}\frac{\frac{\bar{P}_1 z_1}{N_0 \log_e 2}}{(1+\frac{\bar{P}_2 z_2 \zeta}{N_0})(1+\frac{\bar{P}_1 z_1 \zeta}{N_0}+\frac{\bar{P}_2 z_2 \zeta}{N_0})}\right) \\
&= \frac{\theta_1 T}{2\log_e 2}\left(\frac{\bar{P}_1 z_1}{N_0}\right)^2 + \frac{\theta_1 T \bar{P}_1 \bar{P}_2 z_1 z_2}{N_0^2 \log_e 2}.
\end{aligned} \tag{I.10}$$

With (I.9) and (I.10) in mind, we can obtain

$$\begin{aligned}
\lim_{\zeta\to 0}\dot{\phi}_1 &= \frac{\theta_1 T}{2\log_e 2}\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1}\left(\frac{\bar{P}_1 z_1}{N_0}\right)^2\right\} \\
&\quad + \frac{\theta_1 T}{\log_e 2}\int_0^\infty \int_0^{g(0)} e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1}\frac{\bar{P}_1 \bar{P}_2 z_1 z_2}{N_0^2} p(z_1, z_2) dz_2 dz_1
\end{aligned} \tag{I.11}$$

and hence

$$\dot{\mathsf{C}}_1(0) = -\frac{1}{\theta_1 T}\log_e \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1}\right\}. \tag{I.12}$$

Similarly, taking the derivative of (6.42) and letting $\zeta = 0$, we have

$$\dot{\mathsf{C}}_2(0) = -\frac{1}{\theta_2 T} \log_e \mathbb{E}_{z_2} \left\{ e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} \right\} \tag{I.13}$$

which, after incorporating (6.36), again gives us the results in (6.43) and (6.44). For the reverse decoding order assignment, following similar steps, we still get the results in (6.43) and (6.44). □

Appendix J

Proof of Theorem 24

Similar to Theorem 19, we here present the derivation for $\mathcal{S}_1$ for the case when the decoding order is (1,2) when $z_2 < g(z1, \text{SNR}_1)$, and the decoding order is (2,1) when $z_2 > g(z_1, \text{SNR}_1)$. The second derivative of (6.41) is

$$\ddot{\mathsf{C}}_1(\zeta) = -\frac{2\dot{\phi}_1}{\theta_1 T \phi_1} - \frac{\zeta\left(\ddot{\phi}_1\phi_1 - \dot{\phi}_1^2\right)}{\theta_1 T \phi_1^2} \tag{J.1}$$

where ϕ_1 and $\dot{\phi}_1$ are (I.6) and (I.7), respectively, and $\ddot{\phi}_1$ is given by

$$
\begin{aligned}
\ddot{\phi}_1 = &- \int_0^\infty \ddot{g}\left(\frac{\bar{P}_1\zeta}{N_0}\right)\left(\frac{\bar{P}_1}{N_0}\right)^2 e^{-\frac{\theta_1 T}{\zeta}\log_2\left(1+\frac{\bar{P}_1 z_1\zeta}{N_0}\right)} p(z_1, g(\bar{P}_1\zeta/N_0))dz_1 \\
&- 2\int_0^\infty \dot{g}\left(\frac{\bar{P}_1\zeta}{N_0}\right)\frac{\bar{P}_1}{N_0} e^{-\frac{\theta_1 T}{\zeta}\log_2\left(1+\frac{\bar{P}_1 z_1\zeta}{N_0}\right)}\left(\frac{\theta_1 T}{\zeta^2}\log_2\left(1+\frac{\bar{P}_1 z_1\zeta}{N_0}\right) - \frac{\theta_1 T}{\zeta}\frac{\frac{\bar{P}_1 z_1}{N_0\log_e 2}}{1+\frac{\bar{P}_1 z_1\zeta}{N_0}}\right) \\
&\qquad\qquad p(z_1, g(\bar{P}_1\zeta/N_0))dz_1 \\
&- \int_0^\infty \left(\dot{g}\left(\frac{\bar{P}_1\zeta}{N_0}\right)\frac{\bar{P}_1}{N_0}\right)^2 e^{-\frac{\theta_1 T}{\zeta}\log_2\left(1+\frac{\bar{P}_1 z_1\zeta}{N_0}\right)}\left(\frac{\theta_1 T}{\zeta^2}\log_2\left(1+\frac{\bar{P}_1 z_1\zeta}{N_0}\right) - \frac{\theta_1 T}{\zeta}\frac{\frac{\bar{P}_1 z_1}{N_0\log_e 2}}{1+\frac{\bar{P}_1 z_1\zeta}{N_0}}\right) \\
&\qquad\qquad \dot{p}(z_1, g(\bar{P}_1\zeta/N_0))dz_1 \\
&+ \int_0^\infty \int_{g\left(\frac{\bar{P}_1\zeta}{N_0}\right)}^\infty e^{-\frac{\theta_1 T}{\zeta}\log_2\left(1+\frac{\bar{P}_1 z_1\zeta}{N_0}\right)}\left(\left(\frac{\theta_1 T}{\zeta^2}\log_2\left(1+\frac{\bar{P}_1 z_1\zeta}{N_0}\right) - \frac{\theta_1 T}{\zeta}\frac{\frac{\bar{P}_1 z_1}{N_0\log_e 2}}{1+\frac{\bar{P}_1 z_1\zeta}{N_0}}\right)^2\right. \\
&\left. - \frac{2}{\zeta}\left(\frac{\theta_1 T}{\zeta^2}\log_2\left(1+\frac{\bar{P}_1 z_1\zeta}{N_0}\right) - \frac{\theta_1 T}{\zeta}\frac{\frac{\bar{P}_1 z_1}{N_0\log_e 2}}{1+\frac{\bar{P}_1 z_1\zeta}{N_0}}\right) + \frac{\theta_1 T}{\zeta\log_e 2}\left(\frac{\frac{\bar{P}_1 z_1}{N_0}}{1+\frac{\bar{P}_1 z_1\zeta}{N_0}}\right)^2\right) p(z_1, z_2)dz_2 dz_1 \\
&+ \int_0^\infty \ddot{g}\left(\frac{\bar{P}_1\zeta}{N_0}\right)\left(\frac{\bar{P}_1}{N_0}\right)^2 e^{-\frac{\theta_1 T}{\zeta}\log_2\left(1+\frac{\frac{\bar{P}_1 z_1\zeta}{N_0}}{1+\frac{\bar{P}_2 g(P_1\zeta/N_0)\zeta}{N_0}}\right)} p(z_1, g(\bar{P}_1\zeta/N_0))dz_1 \\
&+ 2\int_0^\infty \dot{g}\left(\frac{\bar{P}_1\zeta}{N_0}\right)\frac{\bar{P}_1}{N_0} e^{-\frac{\theta_1 T}{\zeta}\log_2\left(1+\frac{\frac{\bar{P}_1 z_1\zeta}{N_0}}{1+\frac{\bar{P}_2 g(P_1\zeta/N_0)\zeta}{N_0}}\right)}\left(\frac{\theta_1 T}{\zeta^2}\log_2\left(1+\frac{\frac{\bar{P}_1 z_1\zeta}{N_0}}{1+\frac{\bar{P}_2 g(\bar{P}_1\zeta/N_0)\zeta}{N_0}}\right)\right. \\
&\left. - \frac{\theta_1 T}{\zeta}\frac{\frac{\bar{P}_1 z_1}{N_0\log_e 2}}{\left(1+\frac{\bar{P}_2 g(\bar{P}_1\zeta/N_0)\zeta}{N_0}\right)\left(1+\frac{\bar{P}_1 z_1\zeta}{N_0}+\frac{\bar{P}_2 g(\bar{P}_1\zeta/N_0)\zeta}{N_0}\right)}\right) p(z_1, g(\bar{P}_1\zeta/N_0))dz_1 \\
&+ \int_0^\infty \left(\dot{g}\left(\frac{\bar{P}_1\zeta}{N_0}\right)\frac{\bar{P}_1}{N_0}\right)^2 e^{-\frac{\theta_1 T}{\zeta}\log_2\left(1+\frac{\frac{\bar{P}_1 z_1\zeta}{N_0}}{1+\frac{\bar{P}_2 g(P_1\zeta/N_0)\zeta}{N_0}}\right)} \dot{p}(z_1, g(\bar{P}_1\zeta/N_0))dz_1 \\
&+ \int_0^\infty \int_0^{g\left(\frac{\bar{P}_1\zeta}{N_0}\right)} e^{-\frac{\theta_1 T}{\zeta}\log_2\left(1+\frac{\frac{\bar{P}_1 z_1\zeta}{N_0}}{1+\frac{\bar{P}_2 z_2\zeta}{N_0}}\right)} \\
&\qquad \times \left(\left(\frac{\theta_1 T}{\zeta^2}\log_2\left(1+\frac{\frac{\bar{P}_1 z_1\zeta}{N_0}}{1+\frac{\bar{P}_2 z_2\zeta}{N_0}}\right) - \frac{\theta_1 T}{\zeta}\frac{\frac{\bar{P}_1 z_1}{N_0\log_e 2}}{\left(1+\frac{\bar{P}_2 z_2\zeta}{N_0}\right)\left(1+\frac{\bar{P}_1 z_1\zeta}{N_0}+\frac{\bar{P}_2 z_2\zeta}{N_0}\right)}\right)^2\right. \\
&\left. - \frac{2}{\zeta}\left(\frac{\theta_1 T}{\zeta^2}\log_2\left(1+\frac{\frac{\bar{P}_1 z_1\zeta}{N_0}}{1+\frac{\bar{P}_2 z_2\zeta}{N_0}}\right) - \frac{\theta_1 T}{\zeta}\frac{\frac{\bar{P}_1 z_1}{N_0\log_e 2}}{\left(1+\frac{\bar{P}_2 z_2\zeta}{N_0}\right)\left(1+\frac{\bar{P}_1 z_1\zeta}{N_0}+\frac{\bar{P}_2 z_2\zeta}{N_0}\right)}\right)\right. \\
&\left. + \frac{\theta_1 T}{\zeta}\frac{\frac{\bar{P}_1 z_1}{N_0\log_e 2}\left(\frac{\bar{P}_2 z_2}{N_0}\left(1+\frac{\bar{P}_1 z_1\zeta}{N_0}+\frac{\bar{P}_2 z_2\zeta}{N_0}\right)+\left(\frac{\bar{P}_1 z_1}{N_0}+\frac{\bar{P}_2 z_2}{N_0}\right)\left(1+\frac{\bar{P}_2 z_2\zeta}{N_0}\right)\right)}{\left(1+\frac{\bar{P}_2 z_2\zeta}{N_0}\right)^2\left(1+\frac{\bar{P}_1 z_1\zeta}{N_0}+\frac{\bar{P}_2 z_2\zeta}{N_0}\right)^2}\right) p(z_1, z_2)dz_2 dz_1
\end{aligned}
\tag{J.2}
$$

By letting $\zeta = 0$ and recalling (I.9) and (I.10), we can show that

$$\ddot{\mathsf{C}}_1(0) = -\frac{1}{\log_e 2} \frac{\left(\frac{\bar{P}_1}{N_0}\right)^2 \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} z_1^2\right\} + \frac{2\bar{P}_1\bar{P}_2}{N_0^2} \int_0^\infty \int_0^{g(0)} e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} z_1 z_2 p(z_1, z_2) dz_2 dz_1}{\mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1}\right\}} \tag{J.3}$$

Combining (I.12) and (J.3) with (1.7), we have

$$\mathcal{S}_1 = 2\left(\frac{N_0 \log_e 2}{\theta_1 T}\right)^2 \frac{\left(\log_e \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1}\right\}\right)^2 \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1}\right\}}{\bar{P}_1^2 \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} z_1^2\right\} + 2\bar{P}_1\bar{P}_2 \int_0^\infty \int_0^{g(0)} e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} z_1 z_2 p(z_1, z_2) dz_2 dz_1}. \tag{J.4}$$

Following similar steps, we can derive that

$$\mathcal{S}_2 = 2\left(\frac{N_0 \log_e 2}{\theta_2 T}\right)^2 \frac{\left(\log_e \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2}\right\}\right)^2 \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2}\right\}}{\bar{P}_2^2 \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} z_2^2\right\} + 2\bar{P}_1\bar{P}_2 \int_0^\infty \int_{g(0)}^\infty e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} z_1 z_2 p(z_1, z_2) dz_2 dz_1}. \tag{J.5}$$

If the decoding order is (2,1) when $z_2 < g(z_1, \mathrm{SNR}_1)$, and is (1,2) when $z_2 > g(z_1, \mathrm{SNR}_1)$, following the steps described above, we can obtain

$$\mathcal{S}_1 = 2\left(\frac{N_0 \log_e 2}{\theta_1 T}\right)^2 \frac{\left(\log_e \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1}\right\}\right)^2 \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1}\right\}}{\bar{P}_1^2 \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} z_1^2\right\} + 2\bar{P}_1\bar{P}_2 \int_0^\infty \int_{g(0)}^\infty e^{-\frac{\theta_1 T \bar{P}_1}{N_0 \log_e 2} z_1} z_1 z_2 p(z_1, z_2) dz_2 dz_1}, \tag{J.6}$$

$$\mathcal{S}_2 = 2\left(\frac{N_0 \log_e 2}{\theta_2 T}\right)^2 \frac{\left(\log_e \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2}\right\}\right)^2 \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2}\right\}}{\bar{P}_2^2 \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} z_2^2\right\} + 2\bar{P}_1\bar{P}_2 \int_0^\infty \int_0^{g(0)} e^{-\frac{\theta_2 T \bar{P}_2}{N_0 \log_e 2} z_2} z_1 z_2 p(z_1, z_2) dz_2 dz_1}. \tag{J.7}$$

Also note that the wideband slopes have non-negative values and we have the inequalities in (6.55) and (6.56). □

Appendix K

Proof of Theorem 25

We need to compare the upper bound of the slope region in (6.50) with the upper bounds of both (6.55) and (6.56).

By moving the term with $g(0)$ to the LHS of the equation, we can rewrite (J.4) and (J.5) as

$$\frac{\int_0^{\infty}\int_0^{g(0)} e^{-\frac{\theta_1 T\bar{P}_1}{N_0\log_e 2}z_1} z_1 z_2 p(z_1,z_2)dz_2dz_1}{\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0\log_e 2}z_1} z_1 z_2\right\}} = \left(\frac{N_0\log_e 2}{\theta_1 T}\right)^2 \frac{\left(\log_e \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0\log_e 2}z_1}\right\}\right)^2 \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0\log_e 2}z_1}\right\}}{\bar{P}_1\bar{P}_2\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0\log_e 2}z_1} z_1 z_2\right\}}\left(\frac{1}{\mathcal{S}_1}-\frac{1}{\mathcal{S}_1^{up}}\right) \quad \text{(K.1)}$$

and

$$\frac{\int_0^{\infty}\int_{g(0)}^{\infty} e^{-\frac{\theta_2 T\bar{P}_2}{N_0\log_e 2}z_2} z_1 z_2 p(z_1,z_2)dz_2dz_1}{\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0\log_e 2}z_2} z_1 z_2\right\}} = \left(\frac{N_0\log_e 2}{\theta_2 T}\right)^2 \frac{\left(\log_e \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0\log_e 2}z_2}\right\}\right)^2 \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0\log_e 2}z_1}\right\}}{\bar{P}_1\bar{P}_2\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0\log_e 2}z_2} z_1 z_2\right\}}\left(\frac{1}{\mathcal{S}_2}-\frac{1}{\mathcal{S}_2^{up}}\right) \quad \text{(K.2)}$$

Denote

$$\gamma_1 = \frac{\int_0^{\infty}\int_0^{g(0)} e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2} z_1} z_1 z_2 p(z_1, z_2) dz_2 dz_1}{\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2} z_1} z_1 z_2\right\}} \tag{K.3}$$

$$\gamma_2 = \frac{\int_0^{\infty}\int_{g(0)}^{\infty} e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2} z_2} z_1 z_2 p(z_1, z_2) dz_2 dz_1}{\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2} z_2} z_1 z_2\right\}} \tag{K.4}$$

We know that $0 \leq \gamma_1 \leq 1$ and $0 \leq \gamma_2 \leq 1$ vary with different $g(0)$. Substitute (K.1) and (K.2) into the third condition of (6.50), we can obtain

$$\begin{aligned}
&\left(\frac{N_0 \log_e 2}{\theta_1 T}\right)^2 \frac{\left(\log_e \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2} z_1}\right\}\right)^2 \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2} z_1}\right\}}{\bar{P}_1\bar{P}_2\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2} z_1} z_1 z_2\right\}} \left(\frac{1}{\mathcal{S}_1} - \frac{1}{\mathcal{S}_1^{up}}\right) \\
&+ \left(\frac{N_0 \log_e 2}{\theta_2 T}\right)^2 \frac{\left(\log_e \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2} z_2}\right\}\right)^2 \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2} z_1}\right\}}{\bar{P}_1\bar{P}_2\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2} z_2} z_1 z_2\right\}} \left(\frac{1}{\mathcal{S}_2} - \frac{1}{\mathcal{S}_2^{up}}\right) \\
&= \gamma_1 + \gamma_2
\end{aligned} \tag{K.5}$$

Following similar steps, we can get from (J.6) and (J.7)

$$\begin{aligned}
&\left(\frac{N_0 \log_e 2}{\theta_1 T}\right)^2 \frac{\left(\log_e \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2} z_1}\right\}\right)^2 \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2} z_1}\right\}}{\bar{P}_1\bar{P}_2\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_1 T\bar{P}_1}{N_0 \log_e 2} z_1} z_1 z_2\right\}} \left(\frac{1}{\mathcal{S}_1} - \frac{1}{\mathcal{S}_1^{up}}\right) \\
&+ \left(\frac{N_0 \log_e 2}{\theta_2 T}\right)^2 \frac{\left(\log_e \mathbb{E}_{z_2}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2} z_2}\right\}\right)^2 \mathbb{E}_{z_1}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2} z_1}\right\}}{\bar{P}_1\bar{P}_2\mathbb{E}_{\mathbf{z}}\left\{e^{-\frac{\theta_2 T\bar{P}_2}{N_0 \log_e 2} z_2} z_1 z_2\right\}} \left(\frac{1}{\mathcal{S}_2} - \frac{1}{\mathcal{S}_2^{up}}\right) \\
&= 2 - \gamma_1 - \gamma_2
\end{aligned} \tag{K.6}$$

Considering (K.5) and (K.6), we know that either $\gamma_1 + \gamma_2$ or $2 - \gamma_1 - \gamma_2$ must be less than 1, which implies that variable decoding order achieves points outside the

region attained with fixed decoding order, proving the theorem. □

Appendix L

Proof of Theorem 28

Case I $\theta_1 \geq \theta_2$:

For this case, we can show that the upper bound in (7.21) can be attained. First assume that

$$-\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\} \leq -\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}. \qquad \text{(L.1)}$$

Hence, the second term on the right-hand side of (7.21) is the minimum one. Now, set $\hat{\theta} = \theta_2$ in (7.20). Assume that $\tilde{\theta} \geq \hat{\theta} = \theta_2$ where $\tilde{\theta}$ is the solution to (7.19). The validity of this assumption will be shown later below. Under these assumptions, we see from (7.20) that

$$R = h(\tilde{\theta}, \theta_2) = -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\} \quad \text{for all } \tilde{\theta} \geq \hat{\theta} = \theta_2. \qquad \text{(L.2)}$$

Now, in order to show that this rate can be supported, we have to prove that the

equation in (7.19) is also satisfied for this choice of R, i.e., we should have

$$R = -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\} = g(\tilde{\theta}) = -\frac{1}{\tilde{\theta}}\log \mathbb{E}_{z_1}\left\{e^{-\tilde{\theta} TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} \tag{L.3}$$

for some $\tilde{\theta}$ satisfying $\tilde{\theta} \geq \theta_1$ and $\tilde{\theta} \geq \hat{\theta} = \theta_2$. From (L.1) and (L.2), we have

$$R \leq -\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}. \tag{L.4}$$

Since $-\frac{1}{\theta}\log \mathbb{E}_{z_1}\left\{e^{-\theta TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}$ is a decreasing function of θ, (L.4) implies that there exists a $\tilde{\theta} \geq \theta_1$ such that

$$R = -\frac{1}{\tilde{\theta}}\log \mathbb{E}_{z_1}\left\{e^{-\tilde{\theta} TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} \leq -\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} \tag{L.5}$$

showing that (L.3) holds. Note that in Case I, the original assumption is that $\theta_1 \geq \theta_2$. Then, we have $\tilde{\theta} \geq \theta_1 \geq \hat{\theta} = \theta_2$. Hence, in case I, we satisfy $\tilde{\theta} \geq \hat{\theta} = \theta_2$, verifying the earlier assumption. In summary, we have shown that (7.19) and (7.20) simultaneously hold for $\tilde{\theta} \geq \theta_1$ and $\hat{\theta} = \theta_2$ when we have

$$R = \min\left\{-\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}, -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\}\right\} \tag{L.6}$$

$$= -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\}. \tag{L.7}$$

Hence, the upper bound in (7.21) can be achieved and this is the effective capacity.

Above, we have assumed that the second term in (7.21) is the minimum one. On

the other hand, if we have

$$-\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} \leq -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\}, \qquad \text{(L.8)}$$

similar arguments follow. In particular, we can choose $\tilde{\theta} = \theta_1$ in this case, and have from (7.19)

$$R = g(\theta_1) = -\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}. \qquad \text{(L.9)}$$

Through a similar approach as above, we can show that (7.20) can be satisfied with $\hat{\theta} \geq \theta_2$ for this choice of R and establish that the upper bound in (7.21) is again attained.

Case II: $\theta_1 < \theta_2$ and $\theta_2 \leq \bar{\theta}$:

Suppose that the effective capacity is decided by the $\mathbf{S}-\mathsf{R}$ link and $\tilde{\theta} = \theta_1$ returns the highest R. Hence, we set $\tilde{\theta} = \theta_1$ in (7.19) and have

$$R = -\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}. \qquad \text{(L.10)}$$

Clearly, this rate can be supported by the $\mathbf{S}-\mathsf{R}$ link while the QoS constraint at the source is satisfied. In order to prove that this rate is viable for the two-hop link in the presence of the QoS constraint at the relay, we have to show that the equality in (7.20) is satisfied as well for some $\hat{\theta} \geq \theta_2$. Note that the assumption in Case II is $\tilde{\theta} = \theta_1 < \theta_2$. Then, having $\hat{\theta} \geq \theta_2$ implies that $\hat{\theta} > \tilde{\theta} = \theta_1$. Consequently, in order to

satisfy (7.20), we should have

$$R = -\frac{1}{\theta_1}\left(\log \mathbb{E}_{z_2}\left\{e^{-\hat{\theta}TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\} + \log \mathbb{E}_{z_1}\left\{e^{(\hat{\theta}-\theta_1)TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}\right) \quad \text{(L.11)}$$

where we have used the assumption that $\tilde{\theta} = \theta_1$. Our goal is to see whether (L.10) and (L.11) for some $\hat{\theta} \geq \theta_2$ can be satisfied simultaneously. In this quest, we first show several properties of the function on the right-hand side of (L.11).

Lemma 2 *Consider the function*

$$f(\theta) = -\frac{1}{\theta_1}\left(\log \mathbb{E}\left\{e^{-\theta TB\log_2(1+SNR_2 z_2)}\right\} + \log \mathbb{E}\left\{e^{(\theta-\theta_1)TB\log_2(1+SNR_1 z_1)}\right\}\right) \quad \textit{for } \theta \geq 0. \quad \text{(L.12)}$$

This function has the following properties:

a) $f(\theta)$ *is a continuous function of* θ*.*

b) $f(0) = -\frac{1}{\theta_1}\log \mathbb{E}\left\{e^{-\theta_1 TB\log_2(1+SNR_1 z_1)}\right\}$*.*

c) The first derivative of $f(\theta)$ *with respect to* θ *at* $\theta = 0$ *is positive, i.e.,* $\dot{f}(0) > 0$*. Hence,* $f(\theta)$ *is initially an increasing function in the vicinity of the origin as* θ *increases.*

d) $f(\theta)$ *is a concave function of* θ*.*

e) If $TB\log_2(1 + SNR_1 z_{1,\max}) > TB\log_2(1 + SNR_2 z_{2,\min})$ *where* $z_{1,\max}$ *is the essential supremum of the random variable* z_1 *and* $z_{2,\min}$ *is the essential infimum of* z_2*, then there exists a* $\theta^* > 0$ *such that* $f(\theta^*) = 0$*.*

Proof:

a) The continuity can be shown by noting the continuity of the logarithm and exponential functions and employing the Dominated Convergence Theorem and Monotone Convergence Theorem for the justification of the interchange of the limit and expectations. For the first expectation in (L.12), we can apply the Dominated Convergence Theorem by observing that we have $|e^{-\theta TB\log_2(1+\mathrm{SNR}_2 z_2)}| \leq 1$ for all $\theta \geq 0$ and the bounding function is integrable, i.e., $\mathbb{E}\{1\} = 1 < \infty$. For the second expectation, we immediately note that $e^{(\theta-\theta_1)TB\log_2(1+\mathrm{SNR}_1 z_1)}$ is nonnegative and increases with increasing θ, and consequently we can use the Monotone Convergence Theorem to justify the interchange of limit and expectation.

b) This property can be readily seen by evaluating the function at $\theta = 0$.

c) The first derivative of f with respect to θ can be evaluated as

$$\dot{f}(\theta) = -\frac{1}{\theta_1}\left(\frac{-\mathbb{E}_{z_2}\left\{e^{-\theta TB\log_2(1+\mathrm{SNR}_2 z_2)}TB\log_2(1+\mathrm{SNR}_2 z_2)\right\}}{\mathbb{E}_{z_2}\left\{e^{-\theta TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\}} + \frac{\mathbb{E}_{z_1}\left\{e^{(\theta-\theta_1)TB\log_2(1+\mathrm{SNR}_1 z_1)}TB\log_2(1+\mathrm{SNR}_1 z_1)\right\}}{\mathbb{E}_{z_1}\left\{e^{(\theta-\theta_1)TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}}\right). \tag{L.13}$$

Then, $\dot{f}(0)$ can be written as

$$\dot{f}(0) = \frac{TB}{\theta_1}\left(\mathbb{E}_{z_2}\{\log_2(1+\mathrm{SNR}_2 z_2)\} - \frac{\mathbb{E}_{z_1}\{e^{-\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\log_2(1+\mathrm{SNR}_1 z_1)\}}{\mathbb{E}_{z_1}\{e^{-\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\}}\right). \tag{L.14}$$

Let us define

$$\alpha(\theta_1) = \mathbb{E}_{z_2}\{\log_2(1+\mathrm{SNR}_2 z_2)\} - \frac{\mathbb{E}_{z_1}\{e^{-\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\log_2(1+\mathrm{SNR}_1 z_1)\}}{\mathbb{E}_{z_1}\{e^{-\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\}}. \tag{L.15}$$

We can see that $\alpha(0) = \mathbb{E}_{z_2}\{\log_2(1 + \text{SNR}_2 z_2)\} - \mathbb{E}_{z_1}\{\log_2(1 + \text{SNR}_1 z_1)\} > 0$ (due to our original assumption to ensure stability). The first derivative of $\alpha(\theta_1)$ with respect to θ_1 is

$$\begin{aligned}\dot{\alpha}(\theta_1) = TB & \frac{1}{\left(\mathbb{E}_{z_1}\{e^{-\theta_1 TB\log_2(1+\text{SNR}_1 z_1)}\}\right)^2} \\ & \times \Bigg(\mathbb{E}_{z_1}\{e^{-\theta_1 TB\log_2(1+\text{SNR}_1 z_1)}\left(\log_2(1+\text{SNR}_1 z_1)\right)^2\}\mathbb{E}_{z_1}\{e^{-\theta_1 TB\log_2(1+\text{SNR}_1 z_1)}\} \\ & \quad - \left(\mathbb{E}_{z_1}\{e^{-\theta_1 TB\log_2(1+\text{SNR}_1 z_1)}\log_2(1+\text{SNR}_1 z_1)\}\right)^2\Bigg) \end{aligned} \tag{L.16}$$

By Cauchy-Schwarz inequality, we know that $\mathbb{E}\{X^2\}\mathbb{E}\{Y^2\} \geq \left(\mathbb{E}\{XY\}\right)^2$. Then, denoting

$X = \sqrt{e^{-\theta_1 TB\log_2(1+\text{SNR}_1 z_1)}\left(\log_2(1+\text{SNR}_1 z_1)\right)^2}$ and $Y = \sqrt{e^{-\theta_1 TB\log_2(1+\text{SNR}_1 z_1)}}$, we easily see that $\dot{\alpha}(\theta_1) \geq 0$ for all θ_1. Thus, $\alpha(\theta_1)$ is an increasing function and we have $\alpha(\theta_1) \geq \alpha(0) > 0$. Hence, $\dot{f}(0) > 0$.

d) The second derivative of f with respect to θ can be expressed as

$$\begin{aligned}
\ddot{f}(\theta) = -\frac{1}{\theta_1}\Bigg(& \frac{1}{\left(\mathbb{E}_{z_2}\left\{e^{-\theta TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\}\right)^2} \\
& \times \Bigg(\mathbb{E}_{z_2}\left\{e^{-\theta TB\log_2(1+\mathrm{SNR}_2 z_2)}\left(TB\log_2(1+\mathrm{SNR}_2 z_2)\right)^2\right\}\mathbb{E}_{z_2}\left\{e^{-\theta TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\} \\
& \quad - \left(\mathbb{E}_{z_2}\left\{e^{-\theta TB\log_2(1+\mathrm{SNR}_2 z_2)} TB\log_2(1+\mathrm{SNR}_2 z_2)\right\}\right)^2\Bigg) \\
& + \frac{1}{\left(\mathbb{E}_{z_1}\left\{e^{(\theta-\theta_1)TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}\right)^2} \\
& \times \Bigg(\mathbb{E}_{z_1}\left\{e^{(\theta-\theta_1)TB\log_2(1+\mathrm{SNR}_1 z_1)}\left(TB\log_2(1+\mathrm{SNR}_1 z_1)\right)^2\right\} \\
& \quad \times \mathbb{E}_{z_1}\left\{e^{(\theta-\theta_1)TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} \\
& \quad - \left(\mathbb{E}_{z_1}\left\{e^{(\theta-\theta_1)TB\log_2(1+\mathrm{SNR}_1 z_1)} TB\log_2(1+\mathrm{SNR}_1 z_1)\right\}\right)^2\Bigg)\Bigg) \qquad \text{(L.17)} \\
& \leq 0 \qquad \text{(L.18)}
\end{aligned}$$

where Cauchy-Schwarz inequality is used again. With this characterization, we establish that f is a concave function of θ.

e) We first express $f(\theta)$ in the following form:

$$f(\theta) = -\frac{1}{\theta_1}\left(\log\mathbb{E}_{z_2}\left\{e^{-\theta TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\} + \log\mathbb{E}_{z_1}\left\{e^{(\theta-\theta_1)TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}\right) \qquad \text{(L.19)}$$

$$\begin{aligned}
&= \frac{\theta}{\theta_1}\Bigg(-\frac{1}{\theta}\log\mathbb{E}_{z_2}\left\{e^{-\theta TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\} \\
&\qquad - \left(1-\frac{\theta_1}{\theta}\right)\frac{1}{\theta-\theta_1}\log\mathbb{E}_{z_1}\left\{e^{(\theta-\theta_1)TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}\Bigg) \\
&= \frac{\theta}{\theta_1}\left(E_C(\theta) - E_B(\theta-\theta_1)\right) \qquad \text{(L.20)}
\end{aligned}$$

where

$$E_C(\theta) = -\frac{1}{\theta}\log \mathbb{E}_{z_2}\left\{e^{-\theta TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\} \tag{L.21}$$

is the virtual effective capacity with respect to θ, and

$$E_B(\theta-\theta_1) = \left(1-\frac{\theta_1}{\theta}\right)\frac{1}{\theta-\theta_1}\log \mathbb{E}_{z_1}\left\{e^{(\theta-\theta_1)TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}$$

is the virtual effective bandwidth with respect to $\theta-\theta_1$. Similar to the discussion in [8], we know that $E_C(\theta)$ is decreasing in θ. Moreover, when $\theta = 0$, we have $E_C(0) = \mathbb{E}_{z_2}\{TB\log_2(1+\mathrm{SNR}_2 z_2)\}$, and as $\theta\to\infty$, $E_C(\theta)$ approaches the delay limited capacity [19], i.e., $E_C(\theta)\to TB\log_2(1+\mathrm{SNR}_2 z_{2,\min})$ where $z_{2,\min}$ is the essential infimum of the random variable z_2. Furthermore, $E_B(\theta-\theta_1)$ is an increasing function of θ. For $\theta<\theta_1$, $E_B(\theta-\theta_1)$ has a negative value. At $\theta=\theta_1$, we have $E_B(\theta_1-\theta_1) = E_B(0) = 0$. As $\theta\to\infty$, $E_B(\theta-\theta_1)$ approaches the highest rate of the $\mathbf{S}-\mathsf{R}$ link, i.e., $E_B(\theta-\theta_1)\to TB\log_2(1+\mathrm{SNR}_1 z_{1,\max})$ where $z_{1,\max}$ is the essential supremum of the random variable z_1. Therefore, as long as $TB\log_2(1+\mathrm{SNR}_1 z_{1,\max}) > TB\log_2(1+\mathrm{SNR}_2 z_{2,\min})$, the decreasing curve $E_C(\theta)$ and increasing curve $E_B(\theta-\theta_1)$ will meet at some point $\theta=\theta^*>0$ at which we have $f(\theta^*) = \frac{\theta^*}{\theta_1}\left(E_C(\theta^*)-E_B(\theta^*-\theta_1)\right) = 0$.

A numerical result provides a visualization of the above discussion. In Fig. L.1, we plot the virtual effective capacity and virtual effective bandwidth normalized by TB as a function of θ in the Rayleigh fading channel. We assume that $T = 2$ ms, $B = 10^5$ Hz, $\theta_1 = 0.01$, $\mathrm{SNR}_1 = 0$ dB, and $\mathrm{SNR}_2 = 10$ dB. Note that we have $z_{1,\max} = \infty$ and $z_{2,\min} = 0$ in the Rayleigh fading model. □

Recall that we are seeking to establish whether (L.10) and (L.11) can simultane-

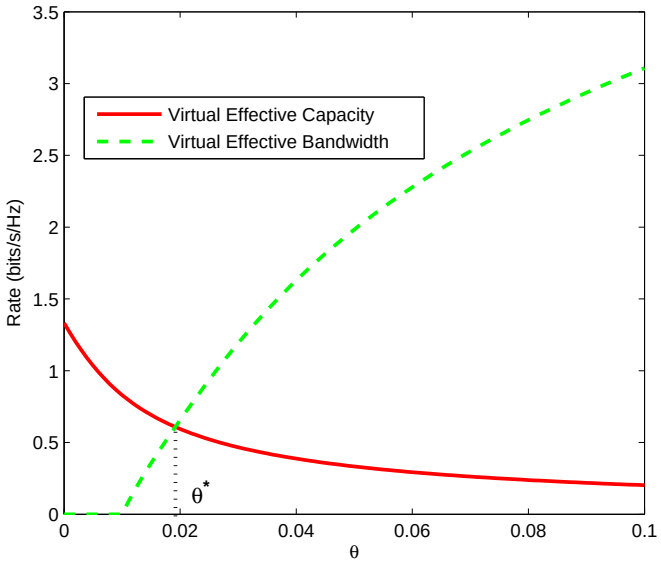

Figure L.1: The virtual effective capacity and virtual effective bandwidth as a function of θ in Rayleigh fading channels with full-duplex relay. $\mathbb{E}\{z_1\} = \mathbb{E}\{z_2\} = 1$.

ously be satisfied for some $\hat{\theta} \geq \theta_2$. With the definition of the function $f(\cdot)$ whose properties are delineated in Lemma 2, the equations in (L.10) and (L.11) can be combined to write

$$f(\hat{\theta}) = -\frac{1}{\theta_1} \log \mathbb{E}_{z_1} \left\{ e^{-\theta_1 TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\}. \tag{L.22}$$

Hence, our goal is to see whether the equation in (L.22) can be satisfied for some $\hat{\theta} \geq \theta_2$. In Lemma 2, we have noted that the function $f(\theta)$ is equal to the right-hand side of (L.22) at $\theta = 0$, and then it increases. At some point, $f(\theta)$ approaches zero. Since it is a concave function, we immediately see that $f(\theta)$ is a function that initially increases, hits a peak value, and then starts decreasing. This leads us to conclude that $f(\theta)$ becomes equal to the right-hand side of (L.22) once again at some unique

$\theta > 0$. Let us denote this unique point as $\bar{\theta}$. Hence,

$$f(\bar{\theta}) = -\frac{1}{\theta_1} \log \mathbb{E}_{z_1} \left\{ e^{-\theta_1 TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\}. \tag{L.23}$$

If $\bar{\theta} \geq \theta_2$, then (L.22) is satisfied for $\hat{\theta} = \bar{\theta} \geq \theta_2$. Therefore, (L.10) and (L.11) are satisfied simultaneously. Hence, the arrival rate

$$R = -\frac{1}{\theta_1} \log \mathbb{E}_{z_1} \left\{ e^{-\theta_1 TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\} \tag{L.24}$$

can be supported by the two-hop link. Since this rate is an upper bound on the arrival rates as proved in Proposition 27, this arrival rate is the effective capacity, proving (7.27) in Theorem 28.

It is important to note that the above result implicitly assumes that $TB \log_2(1 + \mathrm{SNR}_1 z_{1,\max}) > TB \log_2(1 + \mathrm{SNR}_2 z_{2,\min})$ which is a condition in part e) of Lemma 2. Note that if this condition does not hold, then it means that the maximum service rate from the source is equal to or lower than the minimum service rate from the relay. Hence, the relay can immediately support any arrival rate without requiring any buffering. The bottleneck is the $\mathbf{S} - \mathbf{R}$ link and arrival rates are limited by the effective capacity of this link. Therefore, we again have effective capacity of the two-hop link given by (7.27).

Case III: Assume $\theta_1 < \theta_2$ and $\theta_2 > \bar{\theta}$:

Above, we have discussed the case in which $\bar{\theta} \geq \theta_2$. If, on the other hand, $\bar{\theta} < \theta_2$, then (L.22) and consequently (L.11) cannot be satisfied for some $\hat{\theta} \geq \theta_2$. Hence, the arrival rate in (L.24) cannot be supported by the two-hop link, and we need to

consider possibly smaller rates, i.e.,

$$R = g(\tilde{\theta}) = -\frac{1}{\tilde{\theta}} \log \mathbb{E}_{z_1} \left\{ e^{-\tilde{\theta} TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\} \tag{L.25}$$

for some $\tilde{\theta} \geq \theta_1$. The rate given above is supported by the two-hop link if the equation

$$g(\tilde{\theta}) = h(\tilde{\theta}, \hat{\theta}) \tag{L.26}$$

is satisfied for some $\hat{\theta} \geq \theta_2$ and $\tilde{\theta} \geq \theta_1$. Recall that the function h is defined in (7.20) as

$$h(\tilde{\theta}, \hat{\theta}) = \begin{cases} -\frac{1}{\hat{\theta}} \log \mathbb{E}_{z_2} \left\{ e^{-\hat{\theta} TB \log_2(1+\mathrm{SNR}_2 z_2)} \right\} & 0 \leq \hat{\theta} \leq \tilde{\theta} \\ -\frac{1}{\hat{\theta}} \Bigg(\log \mathbb{E}_{z_2} \left\{ e^{-\hat{\theta} TB \log_2(1+\mathrm{SNR}_2 z_2)} \right\} & \\ \quad + \log \mathbb{E}_{z_1} \left\{ e^{(\hat{\theta}-\tilde{\theta}) TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\} \Bigg) & \hat{\theta} \geq \tilde{\theta} \end{cases} . \tag{L.27}$$

We first note that for fixed $\tilde{\theta}$, $h(\tilde{\theta}, \hat{\theta})$ is a decreasing function of $\hat{\theta}$ because as $\hat{\theta}$ increases, the QoS constraints at the relay become more stringent and consequently lower rates can be supported by the relay. Therefore, in order to identify the highest arrival rates R, we consider the smallest allowed value of $\hat{\theta}$ and set $\hat{\theta} = \theta_2$. We now consider the equation

$$g(\tilde{\theta}) = h(\tilde{\theta}, \theta_2) \tag{L.28}$$

and seek whether this equation is satisfied for some $\tilde{\theta} \geq \theta_1$. At $\tilde{\theta} = \theta_1$, the left-hand side of (L.28) becomes

$$g(\theta_1) = -\frac{1}{\theta_1} \log \mathbb{E}_{z_1} \left\{ e^{-\theta_1 TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\} \tag{L.29}$$

while the right-hand side is

$$h(\theta_1, \theta_2) = -\frac{1}{\theta_1}\left(\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\} + \log \mathbb{E}_{z_1}\left\{e^{(\theta_2-\theta_1)TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}\right) \tag{L.30}$$

$$= f(\theta_2) \tag{L.31}$$

where $f(\cdot)$ is the function defined in Lemma 2. Note that our assumption in this case is $\theta_2 > \bar{\theta}$. Recalling (L.23), we know that

$$f(\bar{\theta}) = -\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} = g(\theta_1). \tag{L.32}$$

Then, from the properties of f and the assumption that $\theta_2 > \bar{\theta}$, we immediately see that

$$f(\theta_2) = h(\theta_1, \theta_2) \leq -\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} = g(\theta_1). \tag{L.33}$$

Therefore, at $\tilde{\theta} = \theta_1$, the left-hand side of (L.28) is larger than the value at the right-hand side.

Now, let us consider the values at $\tilde{\theta} = \theta_2$. The left-hand and right-hand sides of (L.28) become, respectively,

$$g(\theta_2) = -\frac{1}{\theta_2}\log \mathbb{E}_{z_1}\left\{e^{-\theta_2 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} \tag{L.34}$$

and

$$h(\theta_2, \theta_2) = -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\} \tag{L.35}$$

If we have

$$g(\theta_2) = -\frac{1}{\theta_2}\log \mathbb{E}_{z_1}\left\{e^{-\theta_2 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}$$
$$\leq h(\theta_2,\theta_2) = -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\}, \tag{L.36}$$

then the left-hand side of (L.28) is smaller that the value of the right-hand side at θ_2. Therefore, being continuous functions, $g(\tilde{\theta})$ and $h(\tilde{\theta},\theta_2)$ meet at some $\theta_1 \leq \tilde{\theta} \leq \theta_2$. Denote the smallest value of $\tilde{\theta}$ for which we have $g(\tilde{\theta}) = h(\tilde{\theta},\theta_2)$ as $\tilde{\theta}^*$. Then, the highest rate that can be supported by the two-hop link is

$$R = g(\tilde{\theta}^*) = -\frac{1}{\tilde{\theta}^*}\log \mathbb{E}_{z_1}\left\{e^{-\tilde{\theta}^* TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} \tag{L.37}$$

Above result is obtained under the assumption that $g(\theta_2) \leq h(\theta_2,\theta_2)$. Let us now consider the other possibility in which $g(\theta_2) > h(\theta_2,\theta_2)$. For this case, we first have the following lemma.

Lemma 3 *Assume that $g(\theta_2) > h(\theta_2,\theta_2)$. Then, $h(\tilde{\theta},\theta_2)$ is an increasing function of $\tilde{\theta}$ for $\tilde{\theta} \leq \theta_2$.*

Proof: For $\tilde{\theta} \leq \theta_2$, we can express

$$h(\tilde{\theta},\theta_2) = -\frac{1}{\tilde{\theta}}\left(\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\} + \log \mathbb{E}_{z_1}\left\{e^{(\theta_2-\tilde{\theta})TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}\right). \tag{L.38}$$

The first derivative of $h(\tilde{\theta}, \theta_2)$ with respect to $\tilde{\theta}$ is

$$\dot{h}(\tilde{\theta}, \theta_2) = \frac{1}{\tilde{\theta}^2}\left(\tilde{\theta}\frac{\mathbb{E}_{z_1}\left\{e^{(\theta_2-\tilde{\theta})TB\log_2(1+\mathrm{SNR}_1 z_1)}TB\log_2(1+\mathrm{SNR}_1 z_1)\right\}}{e^{(\theta_2-\tilde{\theta})TB\log_2(1+\mathrm{SNR}_1 z_1)}}\right.$$
$$\left. + \log\mathbb{E}_{z_1}\left\{e^{(\theta_2-\tilde{\theta})TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} + \log\mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\}\right). \quad \text{(L.39)}$$

Let us define

$$\beta(\tilde{\theta}) = \tilde{\theta}\frac{\mathbb{E}_{z_1}\left\{e^{(\theta_2-\tilde{\theta})TB\log_2(1+\mathrm{SNR}_1 z_1)}TB\log_2(1+\mathrm{SNR}_1 z_1)\right\}}{\mathbb{E}_{z_1}\left\{e^{(\theta_2-\tilde{\theta})TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}}$$
$$+ \log\mathbb{E}_{z_1}\left\{e^{(\theta_2-\tilde{\theta})TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} + \log\mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\}. \quad \text{(L.40)}$$

We can show that $\beta(\tilde{\theta})$ is nonnegative.

The first derivative of $\beta(\tilde{\theta})$ with respect to $\tilde{\theta}$ is

$$\dot{\beta}(\tilde{\theta}) = \frac{\tilde{\theta}}{\left(\mathbb{E}_{z_1}\left\{e^{(\theta_2-\tilde{\theta})TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}\right)^2}$$
$$\times\left(-\mathbb{E}_{z_1}\left\{e^{(\theta_2-\tilde{\theta})TB\log_2(1+\mathrm{SNR}_1 z_1)}\left(TB\log_2(1+\mathrm{SNR}_1 z_1)\right)^2\right\}\mathbb{E}_{z_1}\left\{e^{(\theta_2-\tilde{\theta})TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}\right.$$
$$\left.+\left(\mathbb{E}_{z_1}\left\{e^{(\theta_2-\tilde{\theta})TB\log_2(1+\mathrm{SNR}_1 z_1)}TB\log_2(1+\mathrm{SNR}_1 z_1)\right\}\right)^2\right) \quad \text{(L.41)}$$
$$\leq 0 \quad \text{(L.42)}$$

where Cauchy-Schwarz inequality is used for (L.42). Therefore, $\beta(\tilde{\theta})$ is a decreasing

function of $\tilde{\theta}$, and hence for $\tilde{\theta} \leq \theta_2$ we have

$$\beta(\tilde{\theta}) \geq \beta(\theta_2) = \theta_2 \mathbb{E}_{z_1}\left\{TB\log_2(1+\text{SNR}_1 z_1)\right\} + \log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\text{SNR}_2 z_2)}\right\} \tag{L.43}$$

$$= -\theta_2\left(-TB\mathbb{E}_{z_1}\{\log_2(1+\text{SNR}_1 z_1)\} - \frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\text{SNR}_2 z_2)}\right\}\right) \tag{L.44}$$

Note that our assumption is that

$$\begin{aligned} g(\theta_2) &= -\frac{1}{\theta_2}\log \mathbb{E}_{z_1}\left\{e^{-\theta_2 TB\log_2(1+\text{SNR}_1 z_1)}\right\} \\ &> h(\theta_2,\theta_2) = -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\text{SNR}_2 z_2)}\right\}. \end{aligned} \tag{L.45}$$

Since $TB\mathbb{E}_{z_1}\{\log_2(1+\text{SNR}_1 z_1)\} \geq -\frac{1}{\theta_2}\log \mathbb{E}_{z_1}\left\{e^{-\theta_2 TB\log_2(1+\text{SNR}_1 z_1)}\right\}$, the above inequality implies that

$$TB\mathbb{E}_{z_1}\{\log_2(1+\text{SNR}_1 z_1)\} > -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\text{SNR}_2 z_2)}\right\} \tag{L.46}$$

which further implies that $\beta(\theta_2) > 0$. Finally, we immediately see that

$$\dot{h}(\tilde{\theta},\theta_2) = \frac{1}{\tilde{\theta}^2}\beta(\tilde{\theta}) \geq \frac{1}{\tilde{\theta}^2}\beta(\theta_2) > 0 \tag{L.47}$$

proving that $h(\tilde{\theta},\theta_2)$ is an increasing function of $\tilde{\theta}$ for $\tilde{\theta} \leq \theta_2$. □

In effect, we have shown that if $h(\theta_2,\theta_2) < g(\theta_2)$, then $h(\tilde{\theta},\theta_2) < g(\theta_2)$ for all $\tilde{\theta} \leq \theta_2$. Note that since $g(\tilde{\theta})$ is a decreasing function, $g(\theta_2) \leq g(\tilde{\theta})$ for all $\tilde{\theta} \leq \theta_2$.

Combining these, we observe that

$$h(\tilde{\theta}, \theta_2) < g(\theta_2) \leq g(\tilde{\theta}) \quad \forall \tilde{\theta} \leq \theta_2. \tag{L.48}$$

Therefore, the equality $g(\tilde{\theta}) = h(\tilde{\theta}, \theta_2)$ cannot be satisfied for any $\theta_1 \leq \tilde{\theta} \leq \theta_2$. Hence, we should have $\tilde{\theta} > \theta_2$. Note that for $\tilde{\theta} > \theta_2$, $h(\tilde{\theta}, \theta_2)$, which can be expressed as

$$h(\tilde{\theta}, \theta_2) = -\frac{1}{\theta_2} \log \mathbb{E}_{z_2} \left\{ e^{-\theta_2 TB \log_2(1+\mathrm{SNR}_2 z_2)} \right\}, \tag{L.49}$$

is a constant for given θ_2. On the other hand,

$$g(\tilde{\theta}) = -\frac{1}{\tilde{\theta}} \log \mathbb{E}_{z_1} \left\{ e^{-\tilde{\theta} TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\} \tag{L.50}$$

is a decreasing function with minimum value given by

$$\lim_{\tilde{\theta} \to \infty} g(\tilde{\theta}) = TB \log_2(1 + \mathrm{SNR}_1 z_{1,\min}) \tag{L.51}$$

where $z_{1,\min}$ is the essential infimum of z_1. Hence, if

$$TB \log_2(1 + \mathrm{SNR}_1 z_{1,\min}) \leq h(\tilde{\theta}, \theta_2) = -\frac{1}{\theta_2} \log \mathbb{E}_{z_2} \left\{ e^{-\theta_2 TB \log_2(1+\mathrm{SNR}_2 z_2)} \right\}, \tag{L.52}$$

then the equation $g(\tilde{\theta}) = h(\tilde{\theta}, \theta_2)$ can be satisfied at some $\tilde{\theta} = \tilde{\theta}^* \geq \theta_2$, and the maximum arrival rate is given by

$$R = g(\tilde{\theta}^*) = -\frac{1}{\tilde{\theta}^*} \log \mathbb{E}_{z_1} \left\{ e^{-\tilde{\theta}^* TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\}. \tag{L.53}$$

If, on the other hand,

$$TB\log_2(1+\mathrm{SNR}_1 z_{1,\min}) > h(\tilde{\theta},\theta_2) = -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\}, \quad \text{(L.54)}$$

the bottleneck is the $\mathbf{R}-\mathbf{D}$ link, and the highest arrival rate that can be supported by the two-hop link is

$$R = -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\}. \quad \text{(L.55)}$$

Note that this arrival rate is smaller than the smallest possible transmission rate of the source and hence no buffering is needed at the source in this extreme case. $\square$

Appendix M

Proof of Theorem 29

We first identify the following upper bound on the rates that can be supported with half-duplex relaying in the two-hop link:

$$\begin{aligned} R \leq \sup_{\tau\in[0,\tau_0)} \min\Bigg\{ & -\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\tau\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}, \\ & -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-(1-\tau)\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\}\Bigg\} \qquad \text{(M.1)} \\ = & -\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\tilde{\tau}\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} \qquad \text{(M.2)} \end{aligned}$$

where $\tilde{\tau} = \min\{\tau_0, \tau^*\}$ and τ^* is the solution to

$$-\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\tau\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} = -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-(1-\tau)\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\} \quad \text{(M.3)}$$

and τ_0, as defined in (7.45), is the upper bound on the time-sharing parameter τ. Above, (M.1) can be easily obtained by using a similar approach as in the proof of Proposition 27. (M.2) follows from the fact that the first term inside the minimization in (M.1) is an increasing function of τ while the second term is a decreasing function. Hence, the upper bound in (M.1) is maximized at τ^* at which the two terms inside the

minimization are equal to each other. If $\tau^* < \tau_0$, the optimal value of τ is selected as τ^*. If, on the other hand, τ^* exceeds the upper bound, i.e., $\tau^* \geq \tau_0$, then the optimal value is τ_0.

Case I $\theta_1 \geq \theta_2$:

In this case in which the QoS constraint at the source is more stringent, we can show that the upper bound in (M.2) can be achieved or be approached arbitrarily closely. Let us set $\tilde{\theta} = \theta_1$, $\hat{\theta} = \theta_2$, and choose the time-sharing parameter as $\tau = \tilde{\tau} = \min\{\tau_0, \tau^*\}$. Now, the equation in (7.43) becomes

$$R = g(\theta_1) = -\frac{1}{\theta_1} \log \mathbb{E}_{z_1} \left\{ e^{-\tilde{\tau}\theta_1 TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\}. \tag{M.4}$$

Since $\hat{\theta} = \theta_2 \leq \tilde{\theta} = \theta_1$ by our assumption in Case I, (7.44) reduces to

$$R = h(\theta_1, \theta_2) = -\frac{1}{\theta_2} \log \mathbb{E}_{z_2} \left\{ e^{-(1-\tilde{\tau})\theta_2 TB \log_2(1+\mathrm{SNR}_2 z_2)} \right\}. \tag{M.5}$$

Now, first assume that $\tilde{\tau} = \tau^*$. As seen in (M.3), we have, by the definition of τ^*, that the right-hand sides of (M.4) and (M.5) are equal and therefore these equations are simultaneously satisfied.

Next, consider the other possibility in which $\tilde{\tau} = \min\{\tau_0, \tau^*\} = \tau_0$ which implies that $\tau_0 \leq \tau^*$. Note again that τ^* is the value of τ at which the functions

$$-\frac{1}{\theta_1} \log \mathbb{E}_{z_1} \left\{ e^{-\tau\theta_1 TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\} \tag{M.6}$$

and

$$-\frac{1}{\theta_2} \log \mathbb{E}_{z_2} \left\{ e^{-(1-\tau)\theta_2 TB \log_2(1+\mathrm{SNR}_2 z_2)} \right\} \tag{M.7}$$

are equal. Note that the function in (M.6) increases with increasing τ while the function in (M.7) decreases. They meet at τ^*. Therefore, at $\tau = \tau_0 \leq \tau^*$, we have

$$-\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\tau_0\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} \leq -\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-(1-\tau_0)\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\}. \tag{M.8}$$

Hence, the rate

$$R = -\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\tau_0\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\} \tag{M.9}$$

can be supported. More specifically, the equations in (7.43) and (7.44) can simultaneously be satisfied by setting $\tilde{\theta} = \theta_1$, $\tau = \tau_0$, and also by choosing $\hat{\theta} > \theta_2$ so that the right-hand side of (7.44) becomes smaller than $-\frac{1}{\theta_2}\log \mathbb{E}_{z_2}\left\{e^{-(1-\tau_0)\theta_2 TB\log_2(1+\mathrm{SNR}_2 z_2)}\right\}$ and matches $-\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\tau_0\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}$.

One subtlety in the above argument is the following. Note that we have the strict inequality $\tau < \tau_0$. Hence, we cannot actually set $\tau = \tau_0$ but we can select a value of τ that is arbitrarily close to τ_0. Therefore, since the function in (M.6) increases with increasing τ, we can approach the maximum rate $-\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\tau_0\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}$ arbitrarily closely. Because the effective capacity is defined as the supremum of rates (see e.g., (7.14)), $R = -\frac{1}{\theta_1}\log \mathbb{E}_{z_1}\left\{e^{-\tau_0\theta_1 TB\log_2(1+\mathrm{SNR}_1 z_1)}\right\}$ is indeed the effective capacity.

Case II $\theta_1 < \theta_2$:

We now consider the scenario in which the relay node is subject to a more stringent QoS constraint. In this case, the approach behind the proof is identical to the one employed in Case I. Again, we set $\tilde{\theta} = \theta_1$ and $\hat{\theta} = \theta_2$. Because, otherwise if we have $\tilde{\theta} > \theta_1$ and/or $\hat{\theta} > \theta_2$, we impose more strict QoS constraints than necessary and hence end up supporting only lower arrival rates. Now, for fixed τ, the equations in

(7.43) and (7.44) become

$$R = g(\theta_1) = -\frac{1}{\theta_1} \log \mathbb{E}_{z_1} \left\{ e^{-\tau \theta_1 TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\} \tag{M.10}$$

and

$$\begin{aligned} R = h(\theta_1, \theta_2) = -\frac{1}{\theta_1} \Bigg(& \log \mathbb{E}_{z_2} \left\{ e^{-(1-\tau)\theta_2 TB \log_2(1+\mathrm{SNR}_2 z_2)} \right\} \\ & + \log \mathbb{E}_{z_1} \left\{ e^{\tau(\theta_2-\theta_1) TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\} \Bigg), \end{aligned} \tag{M.11}$$

respectively. Note that (M.11) follows from (7.44) by noting that $\hat{\theta} = \theta_2 > \theta_1 = \tilde{\theta}$ in this case. Similarly as before, the right-hand side of (M.10) is an increasing function of τ while the right-hand side of (M.11) is a decreasing function. Therefore, the equations in (M.10) and (M.11) can simultaneously be satisfied by choosing $\tau = \tau'$ where τ' is solution to

$$\begin{aligned} & -\frac{1}{\theta_1} \log \mathbb{E}_{z_1} \left\{ e^{-\tau \theta_1 TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\} \\ & \quad = -\frac{1}{\theta_1} \left(\log \mathbb{E}_{z_2} \left\{ e^{-(1-\tau)\theta_2 TB \log_2(1+\mathrm{SNR}_2 z_2)} \right\} + \log \mathbb{E}_{z_1} \left\{ e^{\tau(\theta_2-\theta_1) TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\} \right). \end{aligned} \tag{M.12}$$

Choosing values other than $\tilde{\theta} = \theta_1$, $\hat{\theta} = \theta_2$, and $\tau = \tau'$ will lead to smaller arrival rates. Hence, the effective capacity is given by

$$R_E(\theta_1, \theta_2) = -\frac{1}{\theta_1} \log \mathbb{E}_{z_1} \left\{ e^{-\tau' \theta_1 TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\}. \tag{M.13}$$

Above discussion implicitly assumes that $\tau' < \tau_0$. If τ' exceeds the threshold τ_0, then the optimal value of the time-sharing parameter is set to $\tau = \tau_0$. Using similar

ideas as in Case I, we can show that the effective capacity in this case is

$$R_E(\theta_1, \theta_2) = -\frac{1}{\theta_1} \log \mathbb{E}_{z_1} \left\{ e^{-\tau_0 \theta_1 TB \log_2(1+\mathrm{SNR}_1 z_1)} \right\}. \tag{M.14}$$

□

Appendix N

Exponential Decay of $\Pr(x^{nm} \notin S_{nm})$ with nm

Assume that the codewords $x^{nm} = (x_1, x_2, \ldots, x_{nm})$ are generated randomly with each component independent and identically distributed according to $x_i \sim \mathcal{CN}(0, \mathcal{E} - \delta)$ for some arbitrarily small $\delta > 0$. Now, we have

$$\Pr(x^{nm} \notin S_{nm}) = \Pr\left(\frac{1}{nm}\sum_{i=1}^{nm} |x_i|^2 > \mathcal{E}\right) \tag{N.1}$$

$$= \Pr\left(\sum_{i=1}^{nm} |x_i|^2 > nm\mathcal{E}\right) \tag{N.2}$$

$$= 1 - \frac{\gamma\left(nm, nm\frac{\mathcal{E}}{\mathcal{E}-\delta}\right)}{(nm-1)!}. \tag{N.3}$$

The expression in (N.3) is obtained by noting that $\sum_{i=1}^{nm} |x_i|^2$ is a central chi-square random variable with $2nm$ degrees of freedom and $\mathbb{E}\{|x_i|^2\} = \mathcal{E} - \delta$ for all i, and observing that the probability in (N.2) is the complementary cumulative distribution function of this chi-square random variable. In (N.3), $\gamma(\cdot, \cdot)$ is the lower incomplete Gamma function defined as

$$\gamma(s,y) = \int_0^y t^{s-1} e^{-t}\, dt. \tag{N.4}$$

For positive integer nm, the lower incomplete Gamma function has the following equivalent expression [75, Section 8.352]:

$$\gamma(nm,y) = (nm-1)!\left(1 - e^{-y}\sum_{i=1}^{nm-1}\frac{y^i}{i!}\right) \tag{N.5}$$

Making use of this expression, we can write

$$\Pr(x^{nm} \notin S_{nm}) = e^{-anm}\sum_{i=1}^{nm-1}\frac{a^i(nm)^i}{i!} \tag{N.6}$$

$$\leq e^{-anm}a^{nm-1}\sum_{i=1}^{nm-1}\frac{(nm)^i}{i!} \tag{N.7}$$

$$\leq e^{-anm}a^{nm-1}\sum_{i=1}^{\infty}\frac{(nm)^i}{i!} \tag{N.8}$$

$$= e^{-anm}a^{nm-1}e^{nm} \tag{N.9}$$

$$= e^{-(a-1-\log_e a)nm-\log_e a} \tag{N.10}$$

where we have defined $a = \frac{\mathcal{E}}{\mathcal{E}-\delta} > 1$. Above, the upper bound in (N.7) is obtained by noting that for $a > 1$, we have $a^i \leq a^{nm-1}$ for all $i = 1, \ldots, nm-1$. (N.8) follows by increasing the upper limit of the summation to infinity. (N.9) follows from the fact that the sum expression in (N.8) is the power series of the exponential function and is equal to e^{nm}. (N.10) is obtained by expressing $a^{nm-1} = e^{(nm-1)\log_e a}$ and combining the exponential functions. Finally, noting the fact that $a - 1 - \log_e a > 0$ for $a > 1$, we immediately see from the upper bound in (N.10) that $\Pr(x^{nm} \notin S_{nm})$ decreases exponentially fast with increasing nm. $\square$

Appendix O

Proof of Theorem 30

We first prove the following proposition whose proof uses some techniques also employed in [76].

Proposition 4 *The function*

$$f(\epsilon) = (1-\epsilon)e^{-\theta nmR_{l,\epsilon}} \tag{O.1}$$

is strictly convex in ϵ.

Proof: Denote

$$-\theta nmR_{l,\epsilon} = aQ^{-1}(\epsilon) + b \tag{O.2}$$

where we, from (8.12), define

$$a = \theta\sqrt{\sum_{l=1}^{m}\frac{2n\mathrm{SNR}z_l}{1+\mathrm{SNR}z_l}}\log_2 e \quad \text{and} \quad b = -\theta n\sum_{l=1}^{m}\log_2(1+\mathrm{SNR}z_l). \tag{O.3}$$

Note that $a > 0$ since, by our assumption, SNR > 0 and $\theta > 0$ and also we also consider the interesting case in which at least one fading coefficient is strictly greater than zero[1]. Then, we can rewrite (O.1) as

$$f(\epsilon) = (1-\epsilon)e^{aQ^{-1}(\epsilon)+b}. \tag{O.4}$$

The first and second derivative of $f(\epsilon)$ with respect to ϵ are

$$\dot{f}(\epsilon) = \left(a\dot{Q}^{-1}(\epsilon)(1-\epsilon) - 1\right) e^{aQ^{-1}(\epsilon)+b} \tag{O.5}$$

$$\ddot{f}(\epsilon) = \left(a(1-\epsilon)\left(\dot{Q}^{-1}(\epsilon)\right)^2 - 2\dot{Q}^{-1}(\epsilon) + (1-\epsilon)\ddot{Q}^{-1}(\epsilon)\right) ae^{aQ^{-1}(\epsilon)+b} \tag{O.6}$$

where $\dot{Q}^{-1}(\epsilon)$ and $\ddot{Q}^{-1}(\epsilon)$ denote the first and second derivatives of $Q^{-1}(\epsilon)$ with respect to ϵ. Note that for an invertible and differentiable function g, we have $g(g^{-1}(x)) = x$. Taking the derivative of both sides, we have

$$\dot{g}(g^{-1}(x))\dot{g}^{-1}(x) = 1 \Rightarrow \dot{g}^{-1}(x) = \frac{1}{\dot{g}(g^{-1}(x))} \tag{O.7}$$

where $\dot{g}(g^{-1}(x))$ denotes the first derivative of g evaluated at $g^{-1}(x)$, and $\dot{g}^{-1}(x)$ is the derivative of g^{-1} with respect to x. Noting that

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad \text{and} \quad \dot{Q}(x) = -\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \tag{O.8}$$

we can derive the following

$$\dot{Q}^{-1}(\epsilon) = -\sqrt{2\pi} e^{\frac{(Q^{-1}(\epsilon))^2}{2}}. \tag{O.9}$$

[1]If $z_l = 0$ for all l, then $R_{l,\epsilon} = 0$ and consequently $f(\epsilon) = (1-\epsilon)$ is a linear function of ϵ. Hence, the strict convexity will not be affected by this linear behavior when the expectation is taken over all possible values of $\mathbf{z}$ as will also be more explicitly discussed at the end of Appendix O.

Note that $\dot{Q}^{-1}(\epsilon) < 0$ for $0 \leq \epsilon \leq 1$. Differentiating $\dot{Q}^{-1}(\epsilon)$ with respect to ϵ, we have

$$\ddot{Q}^{-1}(\epsilon) = 2\pi Q^{-1}(\epsilon) e^{(Q^{-1}(\epsilon))^2}. \tag{O.10}$$

Next, we consider the following two cases:

1) $\epsilon < \frac{1}{2}$: We have $Q^{-1}(\epsilon) > 0$ for this case and hence $\ddot{Q}^{-1}(\epsilon) > 0$. Together with the fact that $\dot{Q}^{-1}(\epsilon) < 0$, we can immediately see that $\ddot{f}(\epsilon) > 0$ for $\epsilon < \frac{1}{2}$.

2) $\epsilon > \frac{1}{2}$: We have $Q^{-1}(\epsilon) < 0$ for this case. Substituting (O.9) and (O.10) into (O.6) and denoting $x = Q^{-1}(\epsilon)$, the expression inside the parentheses on the right-hand side of (O.6) can be written as

$$a(1-\epsilon)\left(\dot{Q}^{-1}(\epsilon)\right)^2 - 2\dot{Q}^{-1}(\epsilon) + (1-\epsilon)\ddot{Q}^{-1}(\epsilon) \tag{O.11}$$

$$= a(1-\epsilon)2\pi e^{(Q^{-1}(\epsilon))^2} + 2\sqrt{2\pi} e^{\frac{(Q^{-1}(\epsilon))^2}{2}} + (1-\epsilon)2\pi Q^{-1}(\epsilon) e^{Q^{-1}(\epsilon))^2} \tag{O.12}$$

$$= a(1-Q(x))2\pi e^{\frac{x^2}{2}} + 2\sqrt{2\pi} e^{\frac{x^2}{2}} + (1-Q(x))2\pi x e^{x^2} \tag{O.13}$$

$$= e^{\frac{x^2}{2}} \left(2\pi(1-Q(x))(x+a)e^{\frac{x^2}{2}} + 2\sqrt{2\pi}\right) \tag{O.14}$$

$$\geq e^{\frac{x^2}{2}} \left(2\pi(1-Q(x))x e^{\frac{x^2}{2}} + 2\sqrt{2\pi}\right) \tag{O.15}$$

$$\geq e^{\frac{x^2}{2}} \left(2\pi \frac{1}{\sqrt{2\pi}(-x)} e^{-\frac{x^2}{2}} x e^{\frac{x^2}{2}} + 2\sqrt{2\pi}\right) \tag{O.16}$$

$$= e^{\frac{x^2}{2}}(-\sqrt{2\pi} + 2\sqrt{2\pi}) = e^{\frac{x^2}{2}}\sqrt{2\pi} > 0 \tag{O.17}$$

where (O.15) follows from the facts that $a > 0$ and hence $x + a > x$. (O.16) is obtained from the following upper bound

$$1 - Q(x) = Q(-x) \leq \frac{1}{\sqrt{2\pi}(-x)} e^{-\frac{x^2}{2}} \text{ for } x < 0 \tag{O.18}$$

and the facts that $x = Q^{-1}(\epsilon) < 0$ for this case and as a result $x(1 - Q(x))$ can be lower bounded by $x\frac{1}{\sqrt{2\pi}(-x)}e^{-\frac{x^2}{2}}$. Therefore, $\ddot{f}(\epsilon) > 0$ for $\epsilon > \frac{1}{2}$.

Also note that $\epsilon = \frac{1}{2}$ means $Q^{-1}(\epsilon) = 0$, so we have

$$a(1-\epsilon)\left(\dot{Q}^{-1}(\epsilon)\right)^2 - 2\dot{Q}^{-1}(\epsilon) + (1-\epsilon)\ddot{Q}^{-1}(\epsilon) \tag{O.19}$$

$$= a(1-\epsilon)2\pi e^{(Q^{-1}(\epsilon))^2} + 2\sqrt{2\pi}e^{\frac{(Q^{-1}(\epsilon))^2}{2}} + (1-\epsilon)2\pi Q^{-1}(\epsilon)e^{Q^{-1}(\epsilon))^2} \tag{O.20}$$

$$= a\pi + 2\sqrt{2\pi} > 0 \tag{O.21}$$

and as a result $\ddot{f}(\epsilon) > 0$.

From the above discussion, we can find that $\ddot{f}(\epsilon) > 0$ for all $\epsilon \in [0, 1]$. $f(\epsilon)$ is strictly convex in ϵ. □

Now, let $\psi(\epsilon) = \epsilon + (1-\epsilon)e^{-\theta nmR_{l,\epsilon}} = \epsilon + f(\epsilon)$. We have $\ddot{\psi}(\epsilon) = \ddot{f}(\epsilon) > 0$. Hence, $\psi(\epsilon)$ is also strictly convex. Moreover, since the nonnegative weighted sum of strictly convex functions is strictly convex [53] and the addition of a constant does not alter the strict convexity (note that in the case in which $z_l = 0$ for all l, we have $R_{l,\epsilon} = 0$ and $\psi(\epsilon) = 1$), we can conclude that $\Psi(\epsilon) = \mathbb{E}\{\psi(\epsilon)\}$ is strictly convex in ϵ. □

Appendix P

Proof of Theorem 31

We first write the the effective rate expression

$$\mathsf{R}_E(\theta) = -\frac{1}{\theta nm} \log_e \mathbb{E}_{\mathbf{z}} \left\{ \epsilon_m + (1-\epsilon_m) e^{-\theta nm R_{l,\epsilon_m}} \right\} \tag{P.1}$$

where the subscript m in ϵ_m is used to explicitly indicate the dependence of the decoding error probability on m. Recall that we assume $\epsilon_m \geq \epsilon_o > 0$ for all m. Under this assumption, we first show the boundedness of the function inside the expectation in (P.1).

Lemma 4 *Assume that $\epsilon_m \geq \epsilon_o > 0$ for all m. Then, there exists an integer M such that for all $m \geq M$, we have*

$$\epsilon_o \leq \epsilon_m + (1-\epsilon_m) e^{-\theta nm R_{l,\epsilon_m}} \leq 1. \tag{P.2}$$

Proof: The lower bound is immediate as $\epsilon_o \leq \epsilon_m \leq 1$ and $(1-\epsilon_m) e^{-\theta nm R_{l,\epsilon_m}} \geq 0$, and actually holds for all m. Additionally, it is easy to see that the upper bound in (P.2) holds when $R_{l,\epsilon_m} \geq 0$. Being a lower bound, the nonnegativity of R_{l,ϵ_m} in (8.12)

is not guaranteed for any given ϵ_m. Note that for arbitrarily small ϵ_m, $Q^{-1}(\epsilon_m)$ is arbitrarily large, which can lead to negative values for R_{l,ϵ_m}. However, we show below that when ϵ_m is lower bounded by $\epsilon_o > 0$, R_{l,ϵ_m} is nonnegative for sufficiently large values of m. We first establish the following lower bound:

$$R_{l,\epsilon_m} = \frac{1}{m}\sum_{l=1}^{m}\log_2(1+\mathrm{SNR}z_l) - \sqrt{\frac{\log_2^2 e}{m}\sum_{l=1}^{m}\frac{2\mathrm{SNR}z_l}{nm(1+\mathrm{SNR}z_l)}}Q^{-1}(\epsilon_m) \tag{P.3}$$

$$\geq \frac{1}{m}\sum_{l=1}^{m}\log_2(1+\mathrm{SNR}z_l) - \sqrt{\frac{\log_2^2 e}{m}\sum_{l=1}^{m}\frac{2\mathrm{SNR}z_l}{nm(1+\mathrm{SNR}z_l)}}Q^{-1}(\epsilon_o) \tag{P.4}$$

$$\geq \frac{1}{m}\sum_{l=1}^{m}\log_2(1+\mathrm{SNR}z_l) - \sqrt{\frac{2\log_2^2 e}{nm^2}\sum_{l=1}^{m}\log_e(1+\mathrm{SNR}z_l)}Q^{-1}(\epsilon_o) \tag{P.5}$$

$$= \sqrt{\frac{1}{m}\sum_{l=1}^{m}\log_2(1+\mathrm{SNR}z_l)}\left(\sqrt{\frac{1}{m}\sum_{l=1}^{m}\log_2(1+\mathrm{SNR}z_l)} - \sqrt{\frac{2\log_2 e}{nm}}Q^{-1}(\epsilon_o)\right) \tag{P.6}$$

where (P.4) is due to the observation that $Q^{-1}(\epsilon_o) \geq Q^{-1}(\epsilon_m)$ for $\epsilon_m \geq \epsilon_o$, and (P.5) follows from the fact that $\log_e(1+x) \geq \frac{x}{1+x}$ for all $x \geq 0$. By the law of large numbers, we know that $\frac{1}{m}\sum_{l=1}^{m}\log_2(1+\mathrm{SNR}z_l) \to \mathbb{E}\{\log_2(1+\mathrm{SNR}z)\}$ as $m \to \infty$. On the other hand, $\sqrt{\frac{2\log_2 e}{nm}}Q^{-1}(\epsilon_o) \to 0$ as m increases. Hence, from the lower bound in (P.6), we conclude that there exists M such that for all $m \geq M$, $R_{l,\epsilon_m} \geq 0$, proving the Lemma. □

The result of Lemma 4 implies that for sufficiently large m, we have

$$\epsilon^* \leq \mathbb{E}_{\mathbf{z}}\left\{\epsilon_m + (1-\epsilon_m)e^{-\theta nmR_{l,\epsilon_m}}\right\} \leq 1 \tag{P.7}$$

and hence

$$0 \leq \lim_{m\to\infty} -\log_e \mathbb{E}_{\mathbf{z}}\left\{\epsilon_m + (1-\epsilon_m)e^{-\theta nmR_{l,\epsilon_m}}\right\} \leq -\log_e \epsilon_o < \infty \tag{P.8}$$

showing that the numerator in (P.1) approaches a finite value with increasing m. On the other hand, the denominator in (P.1) increases linearly with m. Therefore, $\lim_{m\to\infty} \mathsf{R}_E(\theta) = 0$. □

Appendix Q

Proof of Theorem 32

First, for any given channel state pair $\mathbf{z} = (z_1, z_2, \ldots, z_m)$, we define

$$\mu = \frac{1}{m}\sum_{l=1}^{m} \log_2(1 + \mathrm{SNR} z_l), \tag{Q.1}$$

$$\delta = \sqrt{\frac{1}{m}\sum_{l=1}^{m} \frac{2\mathrm{SNR} z_l}{nm(1 + \mathrm{SNR} z_l)}} \log_2 e \tag{Q.2}$$

and note that $\mu > 0$, $\delta > 0$. We can find that $\Phi(0) = 1$, $\Phi(\infty) = 1$, and $\Phi(R) < 1$ for all $R \in (0, \infty)$. Note that

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt. \tag{Q.3}$$

The first and second derivatives of $\Phi(R)$ in R are given by

$$\dot{\Phi}(R) = \mathbb{E}_{\mathbf{z}}\left\{\frac{1}{\sqrt{2\pi}\delta}e^{-\frac{(\mu-R)^2}{2\delta^2}}\right\}\left(1-e^{-\theta nmR}\right) - \theta nm\left(1-\mathbb{E}_{\mathbf{z}}\left\{Q\left(\frac{\mu-R}{\delta}\right)\right\}\right)e^{-\theta nmR}, \tag{Q.4}$$

$$\begin{aligned}\ddot{\Phi}(R) = {} & \mathbb{E}_{\mathbf{z}}\left\{\frac{1}{\sqrt{2\pi}\delta}e^{-\frac{(\mu-R)^2}{2\delta^2}}\frac{\mu-R}{\delta^2}\right\}\left(1-e^{-\theta nmR}\right) \\ & + \theta nme^{-\theta nmR}\left(\mathbb{E}_{\mathbf{z}}\left\{\frac{2}{\sqrt{2\pi}\delta}e^{-\frac{(\mu-R)^2}{2\delta^2}}\right\} + \theta nm\left(1-\mathbb{E}_{\mathbf{z}}\left\{Q\left(\frac{\mu-R}{\delta}\right)\right\}\right)\right).\end{aligned} \tag{Q.5}$$

Now, we need the following result.

Proposition 5 $\ddot{\Phi}(R) = 0$ *has only one solution.*

Proof: Obviously, $\ddot{\Phi}(0) > 0$. Letting $\ddot{\Phi}(R) = 0$ and performing a simple computation, we have

$$-\frac{\mathbb{E}_{\mathbf{z}}\left\{\frac{1}{\sqrt{2\pi}\delta}e^{-\frac{(\mu-R)^2}{2\delta^2}}\frac{\mu-R}{\delta^2}\right\}}{\mathbb{E}_{\mathbf{z}}\left\{\frac{1}{\sqrt{2\pi}\delta}e^{-\frac{(\mu-R)^2}{2\delta^2}}\right\}} = \theta nm\left(2+\theta nm\frac{1-\mathbb{E}_{\mathbf{z}}\left\{Q\left(\frac{\mu-R}{\delta}\right)\right\}}{\mathbb{E}_{\mathbf{z}}\left\{\frac{1}{\sqrt{2\pi}\delta}e^{-\frac{(\mu-R)^2}{2\delta^2}}\right\}}\right)\frac{e^{-\theta nmR}}{1-e^{-\theta nmR}}. \tag{Q.6}$$

First, we can show that the left-hand side (LHS) of (Q.6) is a nondecreasing function in R. Let

$$g(R) = \mathbb{E}_{\mathbf{z}}\left\{\frac{1}{\sqrt{2\pi}\delta}e^{-\frac{(\mu-R)^2}{2\delta^2}}\right\}. \tag{Q.7}$$

$\frac{1}{\sqrt{2\pi}\delta}e^{-\frac{(\mu-R)^2}{2\delta^2}}$ is a log-concave function in R for all $\mathbf{z}$, and since integration over $\mathbf{z}$ does not change the log-concavity, $g(R)$ is also a log-concave function [53]. And hence

$-\log_e g(R)$ is convex. Note that

$$\text{LHS} = \frac{d}{dR}\left(-\log_e g(R)\right) \tag{Q.8}$$

thus the derivative of LHS of (Q.6) is greater than or equal to 0, and as a result it is nondecreasing in R.

Next, we can prove that the right-hand side (RHS) of (Q.6) is a strictly decreasing function in R. Note that $\frac{e^{-\theta nmR}}{1-e^{-\theta nmR}}$ is strictly decreasing with increasing R. Let

$$1 - \mathbb{E}_{\mathbf{z}}\left\{Q\left(\frac{\mu - R}{\delta}\right)\right\} = \mathbb{E}_{\mathbf{z}}\left\{\int_{-\infty}^{\frac{\mu-R}{\delta}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt\right\} = f(u(R)), \tag{Q.9}$$

where $f(x) = \mathbb{E}_{\mathbf{z}}\left\{\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt\right\}$ and $u(R) = \frac{\mu-R}{\delta}$. We know that $f(x)$ is a log-concave function [53], and from [53, Eq. 3.10], we can see that $\log_e f$ is concave and nondecreasing, and u is concave (actually linear) in R, and hence $\log_e f(u(R))$ is a concave function in R directly. Now that

$$\frac{\mathbb{E}_{\mathbf{z}}\left\{\frac{1}{\sqrt{2\pi}\delta} e^{-\frac{(\mu-R)^2}{2\delta^2}}\right\}}{1 - \mathbb{E}_{\mathbf{z}}\left\{Q\left(\frac{\mu-R}{\delta}\right)\right\}} = -\frac{\dot{f}(u(R))}{f(u(R))} = \frac{d}{dR}(-\log_e f(u(R))) \tag{Q.10}$$

and $-\log_e f(u(R))$ is a convex function. So, $\frac{\mathbb{E}_{\mathbf{z}}\left\{\frac{1}{\sqrt{2\pi}\delta} e^{-\frac{(\mu-R)^2}{2\delta^2}}\right\}}{1-\mathbb{E}_{\mathbf{z}}\left\{Q\left(\frac{\mu-R}{\delta}\right)\right\}}$ is a nondecreasing function, i.e., $\frac{1-\mathbb{E}_{\mathbf{z}}\left\{Q\left(\frac{\mu-R}{\delta}\right)\right\}}{\mathbb{E}_{\mathbf{z}}\left\{\frac{1}{\sqrt{2\pi}\delta} e^{-\frac{(\mu-R)^2}{2\delta^2}}\right\}}$ is a nonincreasing function in R. Thus, due to the strictly decreasing behavior of $\frac{e^{-\theta nmR}}{1-e^{-\theta nmR}}$ and the facts that $\theta nm\left(2 + \theta nm \frac{1-\mathbb{E}_{\mathbf{z}}\left\{Q\left(\frac{\mu-R}{\delta}\right)\right\}}{\mathbb{E}_{\mathbf{z}}\left\{\frac{1}{\sqrt{2\pi}\delta} e^{-\frac{(\mu-R)^2}{2\delta^2}}\right\}}\right) > 0$ and $\frac{e^{-\theta nmR}}{1-e^{-\theta nmR}} > 0$ for $n, m, \theta > 0$, the RHS of (Q.6) is strictly decreasing in R, and hence (Q.6) has only one solution. □

Denote the unique solution to $\ddot{\Phi}(R) = 0$ as R'. We know that $\ddot{\Phi}(R) > 0$ for all $R < R'$, or $\dot{\Phi}(R)$ is increasing equivalently, and $\ddot{\Phi}(R) < 0$ for all $R > R'$, or $\dot{\Phi}(R)$ is decreasing equivalently. Note here that $\int_0^\infty \dot{\Phi}(R)dR = \Phi(\infty) - \Phi(0) = 0$, $\dot{\Phi}(0) = -\theta nm(1 - \mathbb{E}_{\mathbf{z}}\left\{Q(\frac{\mu}{\delta})\right\}) < 0$, so $\dot{\Phi}(R') > 0$. Otherwise, $\dot{\Phi}(R)$ is decreasing for $R > R'$, and hence $\dot{\Phi}(R) \leq 0$, $\int_0^\infty \dot{\Phi}(R)dR < 0$, leading to a contradiction. Also note that $\dot{\Phi}(\infty) = 0$, so $\dot{\Phi}(R) > 0$ for $R > R'$. Thus, there is only one solution to $\dot{\Phi}(R) = 0$. This solution is in the range $R \in (0, R')$, and $\Phi(R)$ is minimized at this value. $\square$

Bibliography

[1] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005. 1, 5.1.1

[2] A. Ephremides and B. Hajek, "Information theory and communication networks: an unconsummated union," *Information Theory, IEEE Transactions on*, vol. 44, no. 6, pp. 2416 –2434, Oct. 1998. 1

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, D. L. Schilling, Ed. Wiley, 1991, vol. 6. 1, 1.3

[4] L. Ozarow, S. Shamai, and A. Wyner, "Information theoretic considerations for cellular mobile radio," *Vehicular Technology, IEEE Transactions on*, vol. 43, no. 2, pp. 359 –378, May 1994. 1

[5] S. Hanly and D. Tse, "Multiaccess fading channels. ii. delay-limited capacities," *Information Theory, IEEE Transactions on*, vol. 44, no. 7, pp. 2816 –2831, Nov. 1998. 1

[6] R. Berry and R. Gallager, "Communication over fading channels with delay constraints," *Information Theory, IEEE Transactions on*, vol. 48, no. 5, pp. 1135 –1149, May 2002. 1

[7] M. Neely, "Optimal energy and delay tradeoffs for multiuser wireless downlinks," *Information Theory, IEEE Transactions on*, vol. 53, no. 9, pp. 3095 –3113, Sept. 2007. 1

[8] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *Wireless Communications, IEEE Transactions on*, vol. 2, no. 4, pp. 630 – 643, July 2003. 1, 1.1, e

[9] F. Kelly, "Notes on effective bandwidths," 1996. 1

[10] C.-S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *Automatic Control, IEEE Transactions on*, vol. 39, no. 5, pp. 913 –931, May 1994. 1, 3.2

[11] C.-S. Chang and J. Thomas, "Effective bandwidth in high-speed digital networks," *Selected Areas in Communications, IEEE Journal on*, vol. 13, no. 6, pp. 1091 –1100, Aug. 1995. 1

[12] C.-S. Chang, *Performance Guarantees in Communication Networks.* London, UK: Springer-Verlag, 2000. 1, 3.2, 3.2, 2, B

[13] D. Wu and R. Negi, "Effective capacity-based quality of service measures for wireless networks," in *Broadband Networks, 2004. BroadNets 2004. Proceedings. First International Conference on*, Oct. 2004, pp. 527 – 536. 1.1

[14] ——, "Utilizing multiuser diversity for efficient support of quality of service over a fading channel," *Vehicular Technology, IEEE Transactions on*, vol. 54, no. 3, pp. 1198 – 1206, May 2005. 1.1

[15] ——, "Downlink scheduling in a cellular network for quality-of-service assurance," *Vehicular Technology, IEEE Transactions on*, vol. 53, no. 5, pp. 1547 – 1557, Sept. 2004. 1.1

[16] J. Tang and X. Zhang, "Cross-layer modeling for quality of service guarantees over wireless links," *Wireless Communications, IEEE Transactions on*, vol. 6, no. 12, pp. 4504 –4512, Dec. 2007. 1.1, 1.4

[17] L. Liu and J.-F. Chamberland, "On the effective capacities of multiple-antenna gaussian channels," in *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*, July 2008, pp. 2583 –2587. 1.1, 1.4

[18] C.-S. Chang and T. Zajic, "Effective bandwidths of departure processes from queues with time varying capacities," in *INFOCOM '95. Fourteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Bringing Information to People. Proceedings. IEEE*, Apr. 1995, pp. 1001 –1009 vol.3. 1.1, 7.1.2, 1, 7.1.2, 7.2.1

[19] M. Gursoy, D. Qiao, and S. Velipasalar, "Analysis of energy efficiency in fading channels under QoS constraints," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 8, pp. 4252 –4263, Aug. 2009. 1.1, 1.4, e

[20] V. K. Garg, *Wireless Communications & Networking*. Elsevier, 2007. 1.2

[21] S. Verdu, "Spectral efficiency in the wideband regime," *Information Theory, IEEE Transactions on*, vol. 48, no. 6, pp. 1319 –1343, June 2002. 1.2, 1.2, 1.2, 1, 3.2, 3.3, 6

[22] R. G. Gallager, *An inequality on the capacity region of multiaccess fading channels*. Kluwer, 1994, pp. 129–139. 1.3, 5, 5.1.1

[23] D. Tse and S. Hanly, "Multiaccess fading channels. I. polymatroid structure, optimal resource allocation and throughput capacities," *Information Theory, IEEE Transactions on*, vol. 44, no. 7, pp. 2796 –2815, Nov. 1998. 1.3, 5.1.1, 5.1.2, 5.2

[24] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Communications, 1995. ICC '95 Seattle, 'Gateway to Globalization', 1995 IEEE International Conference on*, vol. 1, June 1995, pp. 331 –335 vol.1. 1.3

[25] G. Gupta and S. Toumpis, "Power allocation over parallel gaussian multiple access and broadcast channels," *Information Theory, IEEE Transactions on*, vol. 52, no. 7, pp. 3274 – 3282, July 2006. 1.3

[26] S. Vishwanath, S. Jafar, and A. Goldsmith, "Optimum power and rate allocation strategies for multiple access fading channels," in *Vehicular Technology Conference, 2001. VTC 2001 Spring. IEEE VTS 53rd*, vol. 4, 2001, pp. 2888 –2892 vol.4. 1.3, 5, 5.1.1, 5.3.3

[27] G. Caire, D. Tuninetti, and S. Verdu, "Suboptimality of TDMA in the low-power regime," *Information Theory, IEEE Transactions on*, vol. 50, no. 4, pp. 608 – 620, Apr. 2004. 1.3, 5.3.3, 5.3.3, 6, 6.2

[28] E. M. Yeh and A. S. Cohen, "Information theory, queueing, and resource allocation in multi-user fading communications," in *2004 Conference on Information Sciences and Systems*, 2004, pp. 1396–1401. 1.3

[29] E. Yeh and A. Cohen, "Throughput optimal power and rate control for queued multiaccess and broadcast communications," in *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*, June-2 July 2004, p. 112. 1.3

[30] ——, "Throughput and delay optimal resource allocation in multiaccess fading channels," in *Information Theory, 2003. Proceedings. IEEE International Symposium on*, June-4 July 2003, p. 245. 1.3

[31] E. M. Yeh and A. S. Cohen, "Delay optimal rate allocation in multiaccess fading communications," in *Allerton Conference on Communication, Control, and Computing*, 2004, pp. 140–149. 1.3

[32] N. Ehsan and T. Javidi, "Delay optimal transmission policy in a wireless multiaccess channel," *Information Theory, IEEE Transactions on*, vol. 54, no. 8, pp. 3745 –3751, Aug. 2008. 1.3

[33] J. Yang and S. Ulukus, "Delay-minimal transmission for average power constrained multi-access communications," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 9, pp. 2754 –2767, Sept. 2010. 1.3

[34] L. Liu, P. Parag, and J.-F. Chamberland, "Quality of service analysis for wireless user-cooperation networks," *Information Theory, IEEE Transactions on*, vol. 53, no. 10, pp. 3833 –3842, Oct. 2007. 1.3, 1.4

[35] J. Tang and X. Zhang, "Quality-of-service driven power and rate adaptation over wireless links," *Wireless Communications, IEEE Transactions on*, vol. 6, no. 8, pp. 3058 –3068, Aug. 2007. 1.4, 2.1, 2.1, 5.4.1, 5.4.1, 5.4.2, 7.2.1

[36] ——, "Quality-of-service driven power and rate adaptation for multichannel communications over wireless links," *Wireless Communications, IEEE Transactions on*, vol. 6, no. 12, pp. 4349 –4360, Dec. 2007. 1.4

[37] ——, "Cross-layer modeling for quality of service guarantees over wireless links," *Wireless Communications, IEEE Transactions on*, vol. 6, no. 12, pp. 4504 –4512, Dec. 2007. 1.4

[38] L. Liu, P. Parag, J. Tang, W.-Y. Chen, and J.-F. Chamberland, "Resource allocation and quality of service evaluation for wireless communication systems using fluid models," *Information Theory, IEEE Transactions on*, vol. 53, no. 5, pp. 1767 –1777, May 2007. 1.4, 3.1, 8, 8.3

[39] A. Balasubramanian and S. Miller, "The effective capacity of a time division downlink scheduling system," *Communications, IEEE Transactions on*, vol. 58, no. 1, pp. 73 –78, Jan. 2010. 1.4

[40] E. Jorswieck, R. Mochaourab, and M. Mittelbach, "Effective capacity maximization in multi-antenna channels with covariance feedback," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 10, pp. 2988 –2993, Oct. 2010. 1.4

[41] A. Balasubramanian and S. Miller, "On optimal scheduling for time-division systems with quality of service constraints," in *Information Sciences and Systems, 2009. CISS 2009. 43rd Annual Conference on*, Mar. 2009, pp. 719 –722. 1.4, 5.3.2.1

[42] D. Qiao, M. Gursoy, and S. Velipasalar, "The impact of QoS constraints on the energy efficiency of fixed-rate wireless transmissions," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 12, pp. 5957 –5969, Dec. 2009. 1.4, 4.5.1, 4.5.2

[43] ——, "Energy efficiency in the low-SNR regime under queueing constraints and channel uncertainty," *Communications, IEEE Transactions on*, vol. 59, no. 7, pp. 2006 –2017, July 2011. 1.4

[44] ——, "Transmission strategies in multiple access fading channels with statistical QoS constraints," *Information Theory, IEEE Transactions on*, vol. 58, no. 3, pp. 1578 –1593, Mar. 2012. 1.4

[45] ——, "Energy efficiency in multiaccess fading channels under QoS constraints," *EURASIP Journal on Wireless Communications and Networking*, 2012. 1.4

[46] ——, "Effective capacity of two-hop wireless communication systems," *IEEE Transactions on Information Theory*, submitted in 2011. 1.4

[47] ——, "Channel coding over multiple coherence blocks with queueing constraints," in *Communications (ICC), 2011 IEEE International Conference on*, June 2011, pp. 1 –5. 1.4, 5.3.1

[48] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *Information Theory, IEEE Transactions on*, vol. 44, no. 6, pp. 2619 –2692, Oct. 1998. 2

[49] *LT Codes*, ser. FOCS '02. Washington, DC, USA: IEEE Computer Society, 2002. [Online]. Available: http://dl.acm.org/citation.cfm?id=645413.652135 2.1

[50] A. Shokrollahi, "Raptor codes," *Information Theory, IEEE Transactions on*, vol. 52, no. 6, pp. 2551 –2567, June 2006. 2.1

[51] J. Castura and Y. Mao, "Rateless coding and relay networks," *Signal Processing Magazine, IEEE*, vol. 24, no. 5, pp. 27 –35, Sept. 2007. 2.1

[52] ——, "Rateless coding over fading channels," *Communications Letters, IEEE*, vol. 10, no. 1, pp. 46 – 48, Jan. 2006. 2.1

[53] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004. 2.1, 5.3.1, O, Q, Q

[54] S. Shamai and S. Verdu, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *Information Theory, IEEE Transactions on*, vol. 47, no. 4, pp. 1302 –1327, May 2001. 2.2.2, 2.2.2

[55] S. Borade and L. Zheng, "Wideband fading channels with feedback," *Information Theory, IEEE Transactions on*, vol. 56, no. 12, pp. 6058 –6065, Dec. 2010. 1

[56] P. Sadeghi and P. Rapajic, "Capacity analysis for finite-state Markov mapping of flat-fading channels," *Communications, IEEE Transactions on*, vol. 53, no. 5, pp. 833 – 840, May 2005. 3.1

[57] D. Porrat, D. Tse, and S. Nacu, "Channel uncertainty in ultra-wideband communication systems," *Information Theory, IEEE Transactions on*, vol. 53, no. 1, pp. 194 –208, Jan. 2007. 3.3, 2, 4.5.2

[58] V. Raghavan, G. Hariharan, and A. Sayeed, "Capacity of sparse multipath channels in the ultra-wideband regime," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, no. 3, pp. 357 –371, Oct. 2007. 3.3, 2, 4.5.2

[59] I. Telatar and R. Gallager, "Combining queueing theory with information theory for multiaccess," *Selected Areas in Communications, IEEE Journal on*, vol. 13, no. 6, pp. 963 –969, Aug. 1995. 3.3, 2a

[60] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *Information Theory, IEEE Transactions on*, vol. 49, no. 4, pp. 951 – 963, Apr. 2003. 4.2.1, 4.2.2

[61] M. Gursoy, "On the capacity and energy efficiency of training-based transmissions over fading channels," *Information Theory, IEEE Transactions on*, vol. 55, no. 10, pp. 4543 –4567, Oct. 2009. 4.2.1, 4.2.2

[62] S. Shakkottai, "Effective capacity and QoS for wireless scheduling," *Automatic Control, IEEE Transactions on*, vol. 53, no. 3, pp. 749 –761, Apr. 2008. 5.2

[63] V. Venkataramanan and X. Lin, "On wireless scheduling algorithms for minimizing the queue-overflow probability," *Networking, IEEE/ACM Transactions on*, vol. 18, no. 3, pp. 788 –801, June 2010. 5.2

[64] L. Ying, R. Srikant, A. Eryilmaz, and G. Dullerud, "A large deviations analysis of scheduling in wireless networks," *Information Theory, IEEE Transactions on*, vol. 52, no. 11, pp. 5088 –5098, Nov. 2006. 5.2

[65] M. Gursoy, "Throughput analysis of buffer-constrained wireless systems in the finite blocklength regime," in *Communications (ICC), 2011 IEEE International Conference on*, June 2011, pp. 1 –5. 5.3.1

[66] G. Arfken and H. Weber, *Mathematical methods for physicists*. Elsevier, 2005. 5.3.2.1

[67] L. Lai and H. El Gamal, "The water-filling game in fading multiple-access channels," *Information Theory, IEEE Transactions on*, vol. 54, no. 5, pp. 2110 –2122, May 2008. 27

[68] J. Tang and X. Zhang, "Cross-layer resource allocation over wireless relay networks for quality of service provisioning," *Selected Areas in Communications, IEEE Journal on*, vol. 25, no. 4, pp. 645 –656, May 2007. 7

[69] P. Parag and J.-F. Chamberland, "Queueing analysis of a butterfly network for comparing network coding to classical routing," *Information Theory, IEEE Transactions on*, vol. 56, no. 4, pp. 1890 –1908, Apr. 2010. 7

[70] R. Negi and S. Goel, "An information-theoretic approach to queuing in wireless channels with large delay bounds," in *Global Telecommunications Conference, 2004. GLOBECOM '04. IEEE*, vol. 1, Nov.-3 Dec. 2004, pp. 116 – 122 Vol.1. 8

[71] Y. Polyanskiy, H. Poor, and S. Verdu, "Channel coding rate in the finite block-length regime," *Information Theory, IEEE Transactions on*, vol. 56, no. 5, pp. 2307 –2359, May 2010. 8

[72] J. N. Laneman, "On the distribution of mutual information," in *Workshop on Information Theory and its Applications (ITA)*, 2006. 8.2, 8.2, 8.2

[73] D. Buckingham and M. Valenti, "The information-outage probability of finite-length codes over AWGN channels," in *Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on*, Mar. 2008, pp. 390 –395. 8.2, 8.2, 8.2

[74] M. Protter and C. Morrey, *A first course in real analysis*. Springer-Verlag, 1977. C, D

[75] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, 7th ed. Elsevier/Academic Press, Amsterdam, 2007. N

[76] P. Wu and N. Jindal, "Coding versus arq in fading channels: How reliable should the PHY be?" *Communications, IEEE Transactions on*, vol. 59, no. 12, pp. 3363 –3374, Dec. 2011. O