ADAPTIVE RECEIVER DESIGN AND OPTIMAL RESOURCE ALLOCATION STRATEGIES FOR FADING CHANNELS

Thesis by

Lifang Li

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2000
(Submitted November 17, 1999)

Acknowledgements

I am very grateful to many individuals that have helped and supported me through my doctoral study at Caltech. First I would like to thank my thesis advisor, Professor Andrea Goldsmith, for giving me the great opportunity to be a member of her research group and for providing constant support and encouragement in my research. I believe that a large part of my accomplishment has been due to her diligent work and appropriate guidance. Her broad knowledge and deep insights have been an important resource to me over the past three and half years. Her constructive criticisms and detailed comments have greatly improved my technical writing and oral presentations.

I am grateful to my friend Song Wang in Mathematics Department for many useful discussions about my research problems from a mathematical point of view. Another friend who always took his time to listen to me and gave me useful suggestions is Srinivas Aji.

I would like to thank Dr. Dariush Divsalar at JPL and Professor Ghassan Kawas Kaleh for spending time with me discussing my work and for their helpful comments. I am also grateful to Professors Robert J. McEliece, Michelle Effros, Joel N. Franklin, Ali Hajimiri and Dariush Divsalar for serving as committee members on my candidacy and/or defense examinations, and for their various suggestions regarding my research.

I appreciate the assistance with paperwork provided by our secretaries Lilian Porter, Linda Dozsa, Julie Ponce, etc., and the help on software maintenance from our former and current system administrators, Robort Freeman and Joseph Chiu.

I owe a lot to my friend Xin An and his wife Xiaolei Xu for helping me settle down at Caltech and get used to the life here during the first year. Later, friends such as Wen Chen, Hui Jin, Jing Liu, Yang Song, Hongxing Tang, Shaoyu Wang, Tianxiang Weng, Qian Zhao, etc., have helped me generously on various things. I would also like to thank my former and current lab-mates, Srinivas Aji, Mohamed-Slim Alouini, Jung-Fu Chen, Gavin Horn, Syed Jafar, Hui Jin, Aamod Khandekar, Wei Lin, Kim Lumbard, Neelesh B. Mehta, Xiaoyi Tang, Hai-Xin Tie, Sriram Vishwanath, Meina Xu, and Zhong Yu, for sharing their knowledge with me and for their friendship and collaboration. In addition, I would like to thank many other students, professors, and staffs in different groups and departments that have made my staying at Caltech an enjoyable experience. A special thanks goes to my friend and now husband Andrew for his great patience, companionship, support, and care.

Finally, I cannot thank my sister Yanfang and brother-in-law Xing Zhou too much. They have been supporting me in every way through all my undergraduate and graduate years. I am also eternally indebted to my father for his meticulous private tutoring till high school. Although he did not even make it to see me enter college, he had taught me the spirit of hard-working and independence, which benefits me forever.

Abstract

The mobile wireless environment has been a challenge to reliable communications because of the time-varying nature of the channel. Detrimental effects such as path loss, shadowing, and multipath fading can greatly attenuate the transmitted signal. Therefore, adaptive channel estimation and data detecting algorithms must be designed for such channels. Moreover, in a multi-user system, dynamic resource allocation is an important means to transmit information efficiently through the varying channel.

In this thesis we first propose two adaptive feedback maximum-likelihood detection techniques, a decision-feedback decoder and an output-feedback decoder, for coded signals transmitted over channels with correlated fading. Both analysis and simulation results demonstrate that they have far better performance than the conventional decoder. We also propose a simple improvement to conventional decoders by using a weighted metric. The BER performance of all these decoders is analyzed through a sliding window decoding method.

Next we derive the ergodic (Shannon) capacity region and optimal dynamic resource allocation for an M-user fading broadcast channel under code-division with and without successive decoding, time-division, and frequency-division. For this channel we also derive the outage and zero-outage capacity regions and the corresponding optimal resource allocation strategies under different spectrum-sharing techniques. We obtain the outage capacity region implicitly by deriving the minimum common outage probability or the outage probability region for a given rate vector. The corresponding optimal power allocation scheme is a multi-user generalization of the single-user threshold-decision rule.

Finally, we obtain the outage capacity region and optimal power allocation for fading multiple access channels. Successive decoding is proved to be optimal and iterative algorithms are proposed to obtain the optimal decoding order and power allocation in each fading state under the average power and outage probability constraints of each user. We also obtain the average power regions that can support a rate vector with the given average outage probability of each user satisfied.

Contents

Acknowledgements iii

Abstract v

1 Introduction 1

1.1 Background 1

1.2 Thesis Summary 1

1.3 Thesis Outline 2

2 Fading Channel Characterization 5

2.1 Introduction 5

2.2 Single-User Fading Channels 5

2.2.1 Path Loss and Shadowing 5

2.2.2 Small-Scale Fading 6

2.2.3 Statistical Modeling for Amplitude Fading 8

2.3 Multi-User Fading Channels 10

2.3.1 Fading Broadcast Channels 10

2.3.2 Fading Multiple-Access Channels 11

3 Feedback Maximum-Likelihood Detection of Coded Signals over Flat-Fading Channels 12

3.1 Introduction 12

3.1.1 Pilot Symbol-Aided Technique 12

3.1.2 Statistics-Dependent Techniques 13

3.1.3 Statistics-Independent Technique 14

3.1.4 Our FSMC Model-Based Technique 14

3.2 FSMC Model of Slowly Fading Channels 17

3.3 System Model 20

3.3.1 Perfect Phase Compensation 20

3.3.2 Differential Encoding/Decoding of the Signal Phase 22

3.4 Decoder Structures 22

3.4.1 Decision-Feedback Decoder 23

3.4.2 Output-Feedback Decoder 25

3.4.3 Conventional Decoding with Weighting 26

3.4.4 Maximum-Likelihood Decoding with Perfect CSI 26

3.5 Performance Analysis 27

3.6 Numerical Results 32

3.7 Conclusions 40

4 Ergodic Capacity and Optimal Resource Allocation for Fading Broadcast Channels 42

4.1 Introduction 42

4.2 The Fading Broadcast Channel 43

4.3 Ergodic Capacity Regions 45

4.3.1 CD 45

4.3.2 TD 53

4.3.3 Sub-Optimal TD Policy 65

4.3.4 CD without Successive Decoding 66

4.4 Numerical Results 69

4.5 Conclusions 72

5 Outage Capacities and Optimal Resource Allocation for Fading Broadcast Channels 74

5.1 Introduction 74

5.2 The Fading Broadcast Channel 76

5.3 Zero-Outage Capacity Region 76

5.3.1 CD 77

5.3.2 CD without Successive Decoding 79

5.3.3 TD 81

5.3.4 The Limiting Zero-Outage Capacity Region for Nakagami Fading . . 83

5.4 Outage Capacities and Minimum Outage Probability 86

5.4.1 Minimum Common Outage Probability 88

5.4.2 Outage Probability Region 89

5.4.3 Multimedia Outage Probability Region 101

5.5 Numerical Results 102

5.6 Conclusions 114

6 Outage Capacities and Optimal Power Allocation for Fading Multiple-Access Channels 115

6.1 Introduction 115

6.2 The Fading Multiple-Access Channel 117

6.3 Definitions and Notations 118

6.4 Simultaneous Outage Declaration 120

6.4.1 Power Allocation Policy 120

6.4.2 Minimum Common Outage Probability 121

6.4.3 Common Outage Transmission Policy 125

6.4.4 Average Power Region 126

6.5 Independent Outage Declaration 131

6.5.1 Power Allocation Policy 131

6.5.2 Outage Probability Region 132

6.5.3 Independent Outage Transmission Policy 138

6.5.4 Average Power Region 141

6.6 Iterative Algorithms for Power Allocation 147

6.7 Auxiliary Constraints on Transmit Power 155

6.8 Conclusions 157

7 Summary of Contributions 158

A Equivalence of the Conventional Detector with Weighting and the Output-Feedback Detector with Hard-Decision in Chapter 3 161

B Proofs in Chapter 4 163

B.1 Proof of Theorem 4.1 163

B.1.1 Achievability of the Capacity Region 163

B.1.2 Converse 167

B.1.3 Convexity of the Capacity Region 170

B.1.4 Capacity of a Probabilistic Broadcast Channel with CSI 171

B.2 Interpretation of the "Water-Filling" Procedure for CD with Theorem 4.2 . 180

B.3 Proof of Theorem 4.3 181

B.4 Proof of Lemma 4.1 184

B.5 Proof of Theorem 4.4 185

B.6 Decision Region Comparison for Optimal and Sub-Optimal TD Schemes 187

B.7 Proof of Lemma 4.2 189

C Proofs in Chapter 5 **191**

C.1 Proof of Theorem 5.1 191

C.2 Necessary and Sufficient Condition for Equations in (5.8) to Have Positive Solutions 199

C.3 Proof of Lemma 5.1 200

C.4 Proof of Lemma 5.2 201

C.5 Proof of Lemma 5.3 203

C.6 Proof of Theorem 5.3 204

D Proofs in Chapter 6 **208**

D.1 Proof of Lemma 6.1 208

D.2 Proof of Lemma 6.2 209

D.3 Proof of Lemma 6.3 211

D.4 The HT Algorithm 215

D.5 Proof of Lemma 6.4 216

D.6 Proof of Lemma 6.5 217

D.7 The HT^* Algorithm 218

D.8 The HT^{**} Algorithm 220

D.9 Determination of Users with Zero-Outage 222

D.10 Derivation of $\mathbf{Pr}(n)$ in Algorithm 6.3 224

D.11 Proof of Theorem 6.2 225

Bibliography **231**

List of Figures

2.1 A broadcast channel. . . . 10

2.2 A multiple-access channel. . . . 11

3.1 FSMC modeling of the fading channel. . . . 17

3.2 The FSMC model. . . . 18

3.3 Baseband system model. . . . 20

3.4 Decision-feedback decoder. . . . 23

3.5 Output-feedback decoder. . . . 25

3.6 The encoder structure of the $(2, 1, 1)$ convolutional code. . . . 27

3.7 Sliding window decoding for the $(2, 1, 1)$ convolutional code. . . . 28

3.8 Two-state fading channel. . . . 33

3.9 Analytical and simulated BER for QPSK and two-state convolutional coding (two-state fading channel). . . . 34

3.10 Analytical and simulated BER for TC-4AM (Rayleigh fading). . . . 35

3.11 Simulated BER for QPSK and 64-state convolutional coding (two-state fading channel). . . . 36

3.12 Simulated BER for TC-8PSK (two-state fading channel). . . . 37

3.13 Simulated BER for TC-16QAM (Rayleigh fading). . . . 38

3.14 Simulated BER for differentially phase-encoded TC-16QAM (Rayleigh fading). 39

4.1 An M-user fading broadcast channel model. . . . 44

4.2 An equivalent M-user fading broadcast channel model. . . . 44

4.3 The functions $u_1(z)$ and $u_2(z)$ when $\frac{\mu_1}{n_1} > \frac{\mu_2}{n_2}$ and $\lambda < \frac{\mu_2 - \mu_1}{n_2 B - n_1 B}$. . . . 51

4.4 The functions $u_1(z)$ and $u_2(z)$ when $\frac{\mu_1}{n_1} > \frac{\mu_2}{n_2}$ and $\lambda \geq \frac{\mu_2 - \mu_1}{n_2 B - n_1 B}$. . . . 52

4.5 The functions $f_1(P)$, $f_2(P)$ and $J(P)$ when $\frac{\mu_1}{n_1} > \frac{\mu_2}{n_2}$. . . . 57

4.6 The functions $J(P)$ and $f_j(P) = \mu_j \ln \left[1 + \frac{P}{n_j B}\right]$ $(1 \leq j \leq 4)$ which satisfy (4.34) and (4.35). . . . 60

4.7 Two-user ergodic capacity region comparison: 3 dB SNR difference. . . . 70

4.8 Two-user throughput capacity region comparison: 20 dB SNR difference with strong signal power. 71
4.9 Two-user throughput capacity region comparison: 20 dB SNR difference with weak signal power. 72

5.1 Power $P(\mathbf{n})$ vs. Reward $J(P(\mathbf{n}))$ for $N = 3$. 94
5.2 Power $P(\mathbf{n})$ vs. Reward $J(P(\mathbf{n}))$ for three remaining sets of users. 95
5.3 Two-user zero-outage capacity region in Nakagami fading: 20 dB SNR difference. 102
5.4 Two-user zero-outage capacity region in Nakagami fading: 3 dB SNR difference. 103
5.5 Two-user capacity region for a given common outage probability in Nakagami fading using CD: 3 dB SNR difference. 104
5.6 Two-user capacity region for a given common outage probability in Nakagami fading using CD: 20 dB SNR difference. 105
5.7 Two-user capacity region comparison of different common outage probabilities in Nakagami fading using CD: 3 dB SNR difference. 106
5.8 Minimum common outage probability for a given rate vector vs. average transmit power in Nakagami fading using CD. 106
5.9 Two-user capacity region for a given common outage probability in Nakagami fading using CDWO: 3 dB SNR difference. 107
5.10 Two-user capacity region for a given common outage probability in Nakagami fading using CDWO: 20 dB SNR difference. 108
5.11 Two-user capacity region comparison for a given common outage probability in Nakagami fading using CD, TD and CDWO: 3 dB SNR difference. 109
5.12 Two-user capacity region for a given outage probability vector in Nakagami fading using CD: 3 dB SNR difference. 110
5.13 Two-user capacity region for a given outage probability vector in Nakagami fading using CD: 20 dB SNR difference. 110
5.14 Two-user outage probability region comparison for different average transmit power in Nakagami fading using CD: 3 dB SNR difference. 111
5.15 Two-user capacity region comparison for different outage probability vectors in Nakagami fading using CD: 3 dB SNR difference. 112

5.16 Two-user capacity regions with a common outage probability and with an outage probability vector in Nakagami fading using CD: 3 dB SNR difference. 113

5.17 Two-user capacity regions with a common outage probability and with an outage probability vector in Nakagami fading using CD: 20 dB SNR difference. 113

6.1 The iterative procedure of *Algorithm 6.2*. 152

6.2 The iterative procedure of *Algorithm 6.3*. 155

B.1 Probabilistic broadcast channel: Channel 1 in operation with probability p_c, and Channel 2 in operation with probability $\bar{p}_c$. 172

D.1 Four possible cases of the two-user system: Case (1) and (2). 213

D.2 Four possible cases of the two-user system: Case (3) and (4). 214

Chapter 1 Introduction

1.1 Background

The mobile wireless environment has been a challenge to reliable communications because of the time-varying nature of the channel. The received radio signal fluctuates with the changing channel due to path loss, shadowing, multipath fading, and mobility of the users and surrounding objects. Since we have to transmit information in the face of a constantly changing channel, adaptive channel estimation and data detecting algorithms must be designed for the receiver. Moreover, in a multi-user system, dynamic allocation of resources such as the transmit power, bandwidth, and rate is an important means to deal with the time-varying nature of the channel. Indeed, adaptability in a wireless system's architecture and algorithms is critical for effective communications. The terminal devices, the base station, and even the network topology are all expected to be adaptive in order to transmit information efficiently and reliably through the wireless network [1]. In existing wireless systems, common uses of adaptability include adaptive power control enabling a terminal close to the base station to transmit at lower power, and adaptive channel allocation where the selection of a channel by the terminal and the base station is based on local interference and channel conditions.

1.2 Thesis Summary

This thesis focuses on two subjects. First, based on a finite-state Markov chain model of the slowly time-varying channel, we develop adaptive maximum-likelihood decoding algorithms for coded signals transmitted over slowly flat fading channels and examine their bit-error-rate performance. These decoding algorithms outperform the conventional decoder without decreasing the transmit information rate or increasing the signal bandwidth. In addition, the complexity of these algorithms is relatively low compared to the existing algorithms.

Secondly, in a multi-user system, assuming that both the transmitter(s) and receiver(s) have perfect channel side information (CSI) and dynamic resource allocation is allowed, we obtain three types of capacity regions for fading broadcast channels, one type of capacity region for fading multiple-access channels, and their corresponding optimal resource allocation strategies. The three types of capacities are: ergodic (Shannon) capacity, zero-outage capacity, and outage capacity. The capacity type investigated for fading multiple-access channels is the last one, since the ergodic capacity region and the zero-outage capacity region for fading multiple-access channels are obtained in [2] and [3], respectively[1]. For fading broadcast channels, the capacity regions are derived for four different spectrum-sharing techniques: code-division (CD) with and without successive decoding, time-division (TD), and frequency-division (FD).

1.3 Thesis Outline

The rest of the thesis is organized as follows. We start in Chapter 2 with a brief introduction to the main characteristics and modeling of single-user fading channels. The concepts such as fast and slow fading, frequency-flat and frequency-selective fading, and several statistical models for amplitude fading, are reviewed, which will be used in the subsequent chapters. A short description of fading broadcast channels and fading multiple-access channels is also provided.

In Chapter 3, we first give a rather thorough review of various existing techniques on channel estimation and data detection, including several pilot symbol-aided techniques, numerous statistics-dependent techniques, and some statistics-independent techniques. We then introduce our feedback maximum-likelihood decoding techniques for coded signals transmitted over channels with correlated slow fading. These techniques are based on a finite-state Markov model for the fading channel. The system model and the structures of five decoders, i.e., our proposed decision-feedback decoder and output-feedback decoder, the conventional decoder with and without weighting, and the decoder with perfect channel state information, are described. The bit-error-rate performances of the five decoders are then investigated through both analysis and simulation.

In Chapters 4-5, we study three types of capacity regions for fading broadcast channels and obtain their corresponding optimal resource allocation strategies: the ergodic capacity region, the zero-outage capacity region, and the outage capacity region with nonzero outage.

[1]The ergodic (Shannon) capacity of a fading channel is called "throughput capacity" in [2], and the zero-outage capacity is called "delay-limited capacity" in [3].

In Chapter 4, we first give a discrete-time M-user fading broadcast channel model. We then derive the ergodic capacity region of fading broadcast channels for CD, TD, and FD, assuming that both the transmitter and the receivers have perfect CSI. Optimal resource allocation policies that achieve the capacity region boundaries are obtained for CD with successive decoding and for TD using the Lagrangian method. A simple sub-optimal policy is also proposed for TD. It is shown that CD without successive decoding has the same ergodic capacity region as TD and FD. Numerical results for a two-user system are provided for these different spectrum-sharing techniques under different amplitude-fading models.

In Chapter 5, we derive zero-outage capacity regions and outage capacity regions of fading broadcast channels for CD, TD, and FD, assuming that both the transmitter and the receivers have perfect CSI. The zero-outage capacity is a special case of the outage capacity, with the allowed outage probability equal to zero. For convenience we derive the zero-outage capacity regions for different spectrum-sharing techniques first. We then show that in an M-user broadcast system, the outage capacity region is implicitly obtained by deriving the minimum common outage probability or the outage probability region for a given rate vector. Therefore, given the required rate of each user, we derive the minimum common outage probability and the corresponding optimal power allocation strategy under the assumption that the broadcast channel is either not used at all when fading is severe or is used simultaneously for all users. When each user can declare an outage independently, we derive the outage probability region boundary and the corresponding optimal power allocation scheme for a given rate vector. Numerical results for the different outage capacity regions are obtained for the Nakagami-m fading model.

We derive the outage capacity region of fading multiple-access channels in Chapter 6, assuming that both the receiver and all the transmitters have perfect CSI. In this chapter we first describe a discrete-time M-user fading multiple-access channel model. Then under similar assumptions about whether the outage declaration from each user is simultaneous or independent, we argue that the outage capacity region can be implicitly obtained by deriving the minimum common outage probability or the outage probability region for a given rate vector. Given a required rate and average power constraint for each user, we find a successive decoding strategy and a power allocation policy that minimize the common outage probability or bound the outage probability region. Iterative algorithms are proposed for obtaining the optimal decoding order and power allocation in each fading state under a given power constraint for each user. Given the outage probability constraint of each user, we also obtain the average power region that can support a given rate vector with all outage probability constraints satisfied. At the end of this chapter we address the outage capacity region problem when additional peak-power constraints are imposed on the M-users.

Finally, in Chapter 7, we conclude the thesis with a summary of contributions.

Chapter 2 Fading Channel Characterization

2.1 Introduction

In this chapter we first describe the main characteristics and models of a single-user fading channel. We then briefly introduce the fading channels of a multi-user system, including the broadcast channel and the multiple-access channel.

2.2 Single-User Fading Channels

In a single-user system, there is only one transmitter and one receiver. The received radio signal from the transmitter typically consists of a line-of-sight (LOS) component as well as multipath components that arrive at the receiver delayed in time and shifted in phase relative to the LOS path. These signal components may be obstructed by buildings, foliage, or other objects. This radio channel is time-varying due to mobility of either the transmitter or the receiver, or even the surrounding objects. Mobile channel characterization and modeling are dealt with in depth in many textbooks [4]-[8]. Here we give a brief review for some concepts that will be used in the subsequent chapters.

2.2.1 Path Loss and Shadowing

Path loss is the decrease of the signal strength (average power) with the distance d between the transmitter and receiver due to multipath and shadowing. Path loss is often modeled as being proportional to d^γ and the radio wave frequency f, where γ is some value between two to six, depending on the specific environment [8]. For example, assuming a single unobstructed LOS transmission path, $\gamma = 2$. In a two-ray model where both the direct path and a ground reflected propagation path are considered, $\gamma = 4$ for d sufficiently large. In most environments, given the same transmitter-receiver separation distance, the actual received average signal power in different surroundings will deviate from the predicted one due to obstructions from buildings and other objects. The variations due to the obstructions, as measured in decibels (dBs), can be modeled as Gaussian or normal distributed. This log-normal distribution of the path loss describes the random shadowing effects which occur over a large number of measurement locations [7].

2.2.2 Small-Scale Fading

Small-scale fading, or simply *fading*, refers to the rapid fluctuation of the received signal over a short distance (a few wavelengths) due to constructive and destructive combining of the multipath components. If the transmitter or receiver is moving, then this results in rapid fluctuation of the received signal over time. The small-scale fading characteristics are determined by the multipath environment (from reflection, diffraction, and scattering) as well as the speed of the mobile and surrounding objects and the signal bandwidth.

Fast and Slow fading

It is well-known that whenever there is relative motion between a transmitter and a receiver, the received carrier frequency is shifted relative to the transmitted carrier frequency. This shift of frequency is called the Doppler frequency shift. Assuming that the relative speed between the transmitter and a receiver is v_m and a pure sinusoid with frequency f_c is transmitted, then the Doppler frequency shift f_d is:

$$f_d = \frac{v_m}{c} f_c,$$

where c is the speed of light. In reality the received signal arrives from multiple paths, and the velocity of movement in the direction of each path is usually different from that of another path. Therefore, the received signal of a transmitted sinusoid will have a spectrum composed of frequencies in the range $f_c - f_d$ to $f_c + f_d$, and this spectrum is referred to as the Doppler spectrum.

Doppler spectrum is a measurement of the spectral broadening caused by the rate of change in the mobile radio channel. The coherence time $T_c \approx \frac{1}{f_d}$ of the channel is a statistical measure of the time duration over which the channel impulse response is essentially invariant. Depending on how rapidly the transmitted baseband signal changes as compared to the changing rate of the channel, a channel may be identified as either *fast fading* or *slow fading*. If the coherence time of the channel is smaller than the symbol duration of the transmitted signal, then the channel impulse response will change rapidly within each symbol period and we call such channel a fast fading channel. Otherwise, if the channel coherence time is much larger than the symbol duration, then the channel impulse response will change much slower than the symbol rate, and we call such channel a slow fading channel.

Frequency-Flat Fading and Frequency-Selective Fading

Since the same transmitted signal may arrive at the receiver through different paths with different delays, we characterize the time-dispersive nature of the multipath channel using its multipath delay spread, defined as the overall span of path delays (i.e., earliest arrival to latest arrival). However, different channels with the same multipath delay spread can exhibit very different signal intensity profiles over the delay span. Therefore, the root-mean-square (rms) delay spread τ_{rms} is often considered,

$$\tau_{rms} = \sqrt{\overline{\tau^2} - (\bar{\tau})^2},$$

where, given L propagation paths,

$$\overline{\tau^n} = \frac{\sum_{i=1}^{L} \tau_i^n |\beta_i|^2}{\sum_{i=1}^{L} |\beta_i|^2}, \quad n = 1, 2,$$

and $|\beta_i|$ represents the amplitude of the ith path arriving with delay τ_i.

In the frequency domain, the coherence bandwidth refers to the range of frequencies over which all spectrum components have roughly the same gain and linear phase. If the coherence bandwidth B_c is defined as the bandwidth over which the frequency correlation function is above 0.9, then

$$B_c \approx \frac{1}{50\tau_{rms}}.$$

Therefore, a channel exhibits "flat fading" if the signal bandwidth is much smaller than the coherence bandwidth B_c of the channel, and is "frequency-selective" if the signal bandwidth is larger than B_c. In flat fading, the spectral characteristics of the transmitted signal are preserved at the receiver. However, the strength of the received signal changes with time due to fluctuation in the channel gain caused by multipath. In frequency-selective fading, the received signal includes multiple versions of the transmitted waveform which are attenuated and delayed in time. Thus the channel induces intersymbol interference. In the frequency domain, some spectral components of the received signal have greater gains than others, i.e., the fading in the channel becomes frequency selective.

2.2.3 Statistical Modeling for Amplitude Fading

In mobile radio channels, several distributions are used to describe the received envelope of the flat fading signal, or the envelope of an individual multipath component. We now provide a brief introduction for the Rayleigh, Rician, and Nakagami-m distributions that will be used in subsequent chapters. These distributions are applicable to different multipath environments, as described in more detail below. More distributions and their corresponding applications can be found in [9, 10].

Rayleigh

The Rayleigh distribution typically models signal fading when there is no direct LOS multipath component. This distribution is commonly used to model terrestrial mobile communication channels since it represents the worst-case scenario and is also relatively easy to manipulate mathematically. The Rayleigh probability density function (p.d.f.) for the amplitude α is given by

$$p_\alpha(\alpha) = \frac{\alpha}{\Omega} \exp\left(-\frac{\alpha^2}{2\Omega}\right); \qquad \alpha \geq 0, \tag{2.1}$$

which is described by a single parameter Ω. The mean and variance of the Rayleigh distributed random variable are given by $\sqrt{\Omega\pi/2}$ and $(2-\pi/2)\Omega$, respectively.

The square of the magnitude of a Rayleigh distributed random variable represents the signal power and has an exponential distribution. That is, if we let $\gamma = \alpha^2$, the p.d.f. of γ is:

$$p_r(\gamma) = \frac{1}{\overline{\gamma}} \exp\left(-\frac{\gamma}{\overline{\gamma}}\right); \qquad \gamma \geq 0,$$

where $\overline{\gamma}$ is the average received power. We will use this distribution in several places in Chapters 3-4.

Rician

The Rician distribution is commonly used to model the amplitude variation in the presence of a strong stationary (nonfading) LOS path, such as in satellite communication systems. Its p.d.f. is given by

$$p_\alpha(\alpha) = \frac{\alpha}{\Omega} \exp\left(-\frac{\alpha^2+\nu^2}{2\Omega}\right) I_0\left(\frac{\alpha\nu}{\Omega}\right), \quad \alpha \geq 0,$$

where Ω represents the variance of the random component of the received signal, ν is the amplitude of the LOS component, and $I_0(\cdot)$ is the zero-order modified Bessel function of the first kind. The parameter $K \triangleq \nu^2/\Omega$ is the ratio between the deterministic LOS signal power and the variance of the multipath. Usually K is used in terms of dB as the parameter identifying the Rician distribution function,

$$K \;\; (dB) = 10 \log \frac{\nu^2}{\Omega} \;\; dB.$$

As $\nu^2 \to 0$, $K \to -\infty$, and as the LOS dominant path decreases in amplitude, the Rician distribution degenerates to a Rayleigh distribution. We will use the Rician distribution in Chapter 4.

Nakagami-*m*

The Nakagami-m distribution often gives the best description of the amplitude fading in land-mobile [11, 12, 13] and indoor-mobile [14] communication systems. Its p.d.f. is given by [15]

$$p_\alpha(\alpha) = \frac{2\, m^m\, \alpha^{2m-1}}{\Omega^m\, \Gamma(m)} \exp\left(-\frac{m\, \alpha^2}{\Omega}\right), \quad \alpha \geq 0,$$

where Ω is the mean-square value of the random variable α, $\Gamma(\cdot)$ is the gamma function, and m is the Nakagami-m fading parameter which ranges from $1/2$ to ∞. The square of the random variable α represents the signal power and has a gamma distribution. That is, if we let $\gamma = \alpha^2$, the p.d.f. of γ is:

$$p_r(\gamma) = \frac{m^m\, \gamma^{m-1}}{\overline{\gamma}^m\, \Gamma(m)} \exp\left(-\frac{m\, \gamma}{\overline{\gamma}}\right), \quad \gamma \geq 0,$$

where $\overline{\gamma}$ is the average received power. Note that for $m = 1$, the Nakagami-m distribution reduces to Rayleigh, and for $m = 1/2$, it is a one-sided Gaussian distribution. With proper adjustment of the parameters it can also fit Rician and lognormal distributions very tightly. In the limit as $m \longrightarrow +\infty$, the Nakagami-$m$ fading channel converges to an additive white Gaussian noise (AWGN) channel. We will use the Nakagami-m fading distribution in several instances in Chapter 5.

2.3 Multi-User Fading Channels

2.3.1 Fading Broadcast Channels

In a multi-user communication system, as shown in Figure 2.1, the broadcast channel refers to the channel through which a centralized transmitter (or base station) sends information to multiple receivers (users). It is also known as the downlink channel. systems are typical

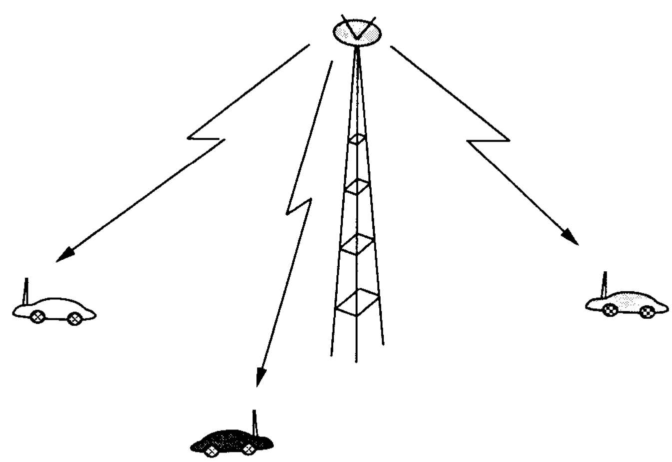


Figure 2.1: A broadcast channel.

broadcast systems. In a mobile broadcast system, the subchannel between the transmitter and each individual receiver is time-varying due to user mobility and the changing environment. Since the receivers may be located in various places, the same transmitted signal may pass through different fading subchannels before reaching those receivers. A mathematical model of the fading broadcast channel will be given in Chapters 4-5.

2.3.2 Fading Multiple-Access Channels

As shown in Figure 2.2, the multiple-access channel refers to the channel through which all the users in the system send information to a centralized receiver or base station. It is also known as the uplink channel, which is the reverse of the downlink channel. Since each

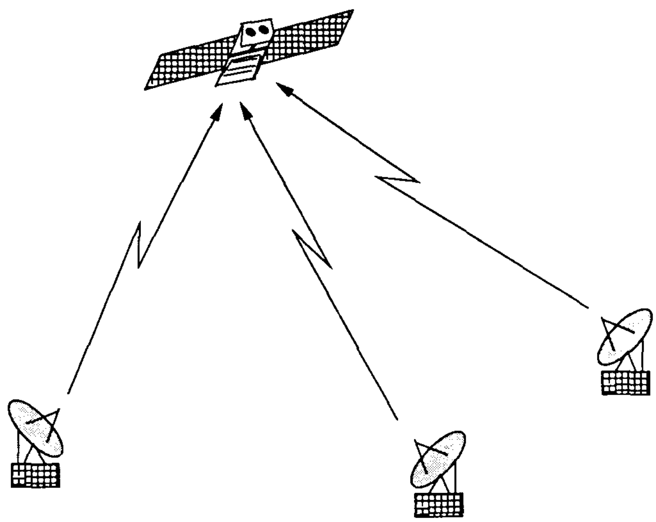


Figure 2.2: A multiple-access channel.

user has a transmitter, the multiple access channel consists of many transmitters sending information to one receiver. The transmitted signal from each user is distinct and the received signal at the base station is a sum of all the transmitted signals. In a mobile multiple-access system, the subchannels are time-varying and the transmitted signals from different users typically reach the base station through quite different paths. Therefore, the subchannels are usually different from user to user. A mathematical model of the fading multiple-access channel will be given in Chapter 6.

Chapter 3 Feedback Maximum-Likelihood Detection of Coded Signals over Flat-Fading Channels

3.1 Introduction

It is well known that the transmission link of many mobile communication systems can be characterized as a frequency-flat fading channel. Such channel exhibits memory due to correlation of the multiplicative fading distortion [4]. The channel memory should be considered in optimal detection of both coded and uncoded signals transmitted over fading channels and much work has been done in this area. In order to implement coherent detection, the multiplicative fading channel distortion (fading gains) must be estimated explicitly. In the following subsections we discuss various methods for obtaining these estimates.

3.1.1 Pilot Symbol-Aided Technique

Pilot symbol-aided channel estimation is proposed in [16] for the Rician fading channel, where one known symbol is sent for each data block and linear interpolation is used to estimate the complex fading gains for the unknown data symbols, based on the fact that the fading gains are correlated. This scheme is later extended to the Rayleigh fading channel with uncoded M-ary quadrature amplitude modulation (QAM, MQAM) in [17, 18], and with coded 16QAM in [19]. Theoretical analyses are given in [20] for uncoded systems. The idea that fading gains can be estimated based on both pilot symbols and data symbols is proposed in [21] by using tentative decisions and a noise smoothing filter for coded phase shift keying (PSK) signal. This method is improved in [22] with an adaptive linear predictor. A similar scheme that uses tentative decisions and a smoothing filter is proposed in [23] for uncoded MQAM systems. In [24], the fading gains are estimated from both pilot and data symbols based on the principle of per-survivor processing [25].

Although pilot symbol-aided detection is considered superior to the pilot-tone assisted method [26] and it compares favorably with differential detection in bit-error-rate (BER) performance, it requires an increase in the power or bandwidth of the transmitted signal, or a decrease in information rate, and thus reduces the transmission efficiency. Therefore, in some applications transmission without pilot symbols may be desirable. The following subsections describe techniques belonging to this category.

3.1.2 Statistics-Dependent Techniques

Assuming that the channel fading correlation function is known, the minimum symbol error probability receiver and the minimum sequence error probability receiver are derived for uncoded PSK modulation in [27] and [28], respectively. However, both receivers are unrealizable, since their complexity grows exponentially with the length of the received symbol sequence. These papers therefore propose suboptimal implementations, and these implementations have similar structures, i.e., they both consist of a finite-memory decision-directed estimator of the fading distortion and a detector that uses these estimates. For coded transmission, the fading distortion is similarly estimated by the estimator before deinterleaving and decoding based on tentative decisions [28, 29]. Note that because of the large deinterleaving and/or decoding delay, the more reliable symbol decisions from the error-correcting decoder are not available in these receivers.

By assuming a finite memory length of the fading channel, a practical linear predictive receiver structure based on optimal maximum-likelihood (ML) sequence estimation (MLSE) is derived in [30] for uncoded continuous phase modulation (CPM) with Rayleigh fading. This receiver is shown to be a special case of a general innovations-based MLSE receiver that uses the Viterbi algorithm and is applicable to any modulation type [31]. An extended Kalman filter is applied as a near-optimal solution to the estimation of the fading distortion in [32] when additional channel parameters such as the received signal power, frequency offset and Doppler spread are to be estimated at the same time.

In [33], an optimal ML receiver structure is developed for uncoded M-ary differential PSK (M-DPSK) reception. An approximate ML decoder employing the Viterbi algorithm based on the principle of per-survivor processing has been designed for coded and uncoded M-ary PSK (MPSK) [34] as well as coded and uncoded M-DPSK [35]. This method is extended in [36] to uncoded QAM signaling using adaptive predictive filters. For interleaved coded MPSK systems, a two-stage estimator-detector is also proposed in [34]. A different two-stage receiver based on joint maximum *a posteriori* (MAP) estimation of the fading distortion and the coded data sequence is proposed in [37]. For uninterleaved coded PSK or QAM systems, an ML sequence estimator is derived in [38] for fast Rayleigh or Rician fading channels.

3.1.3 Statistics-Independent Technique

All of the above techniques assume exact knowledge of the fading channel correlation function. Estimating second-order statistics requires an additional training sequence [33]. Moreover, since fading channel statistics may vary with time, incorrect estimation of the statistics will lead to suboptimal performance. A robust receiver structure that does not use any knowledge of the fading correlation function is proposed in [39]. Double-filtering receivers based on linearly time-varying fading models are proposed in [40] and [41] for PSK and DPSK modulation, respectively, and they use blind algorithms that do not require information on the fading statistics.

3.1.4 Our FSMC Model-Based Technique

Since many error-correcting codes are designed for additive white Gaussian noise (AWGN) channels where errors tend to occur randomly, when these codes are used for the fading channels, interleaving/deinterleaving is usually employed to break up long deep fades due to slow fading. We propose a decision-feedback detection algorithm and an output-feedback detection algorithm for interleaved coded signals transmitted over channels with slow, flat fading. The two-stage decision-feedback decoder structure was originally proposed in [42] for the finite-state Markov chain (FSMC) channel model to achieve the finite-state Markov channel capacity. We now give a brief description of our techniques, which will be described in more detail in the rest of this chapter.

Based on an FSMC model of the slowly fading channel, both feedback decoders consist of a recursive channel state estimator and an ML sequence detector, assuming that convolutional codes or trellis codes are used. In the decision-feedback decoder, the channel state estimator makes a soft decision about the current fade level (channel state) based on past channel output observations and feedback decisions from the ML sequence detector. In the output-feedback detector, the channel state estimate is based only on past output observations. Since the decision-feedback decoder uses feedback decisions for its state estimate, it suffers from error propagation when bits are decoded in error, similar to the error propagation exhibited by decision-feedback equalizers. The output-feedback decoder does not suffer from error propagation, so it has slightly better performance than the decision-feedback decoder on channels with a low signal-to-noise ratio (SNR).

Since both feedback decoders are based on an FSMC model of the slowly fading channel, which is obtained from the channel correlation function, they are statistics-dependent. However, due to the fact that the channel state estimate is computed recursively, the complexity of these feedback detectors is independent of channel memory, and is only slightly higher than that of conventional decoding techniques developed for the AWGN channels [43]. This is different from those statistics-dependent techniques discussed in Section 3.1.2, most of which are developed for fast fading channels, since their complexity usually grows exponentially with the channel memory length.

The FSMC model of the fading channel is a generalization of the Gilbert-Elliott channel model [44]. By partitioning the range of the received SNR into a finite number of intervals, the first-order FSMC model is constructed for the slowly Rayleigh fading channel in [45]. It is then extended to the more general Nakagame-m fading channel in [46], where the state transition probabilities are obtained from the second-order fading statistics. For slowly time-varying channels, the first-order FSMC model is shown to be enough [47]. Higher-order Markov modeling for fast-fading channels is proposed in [48, 49].

Note that since phase distortion caused by fading is not included in the FSMC channel model, perfect phase compensation is assumed in our two feedback decoders, which is equivalent to perfect carrier recovery at the receiver [50, 51]. However, in reality it is not easy to implement idea carrier recovery, and an effective amplitude-phase FSMC model of the fading channel will be too complex for practical implementation. In our simulations we compensate for the effect of phase distortion by differential decoding and encoding the phase of the transmit signal over slowly fading channels and compare the results to that of perfect phase compensation.

Differential encoding/decoding of PSK signals is often used in practice for its simplicity, since it does not require carrier phase acquisition and tracking at the receiver. By using a nonlinear transformation named multilag high-order instantaneous moment (ml-HIM), the idea of differential encoding/decoding has been generalized to higher orders and extended to nonconstant-envelope modulation such as QAM [52]. The ml-HIM method is shown to be capable of removing the effects of phase ambiguity, Doppler frequency shift, and even higher order phase distortions. Differential encoding of the QAM signals discussed in [38, 53] is a special case of the ml-HIM encoding method. In [53], the differentially phase-encoded QAM signals and its differentially coherent detection are proposed to provide a practical solution to the effect of phase discontinuities in a frequency-hopping system. For the Rayleigh fading channel and nonconstant-envelope modulation, we differentially encode and decode the phase of the transmit signals to compensate for the effect of phase distortion due to slow fading. The amplitude distortion will then be estimated in the two feedback decoders prior to sequence detection of the coded signals.

For comparison, we also obtain the performance of two other decoders: an ML decoder with perfect channel state information (CSI) and a conventional Viterbi decoder (VD) with interleaving, which ignores the channel memory. The conventional decoder has extremely poor performance; we therefore propose a simple improvement for its design which weights the decision metric by the fading distribution. This technique is less complex than the feedback decoders, since the recursive calculation of the state estimate is eliminated.

It is difficult to analyze the performance of our proposed decoders directly, due to the complexity of their path metrics in the ML sequence detection. Therefore, we approximate their performance with that of a sliding window decoder (SWD) which uses the same path metrics. The SWD was originally proposed to calculate the BER of convolutional codes on very noisy AWGN channels [54]. For our analyses, we generalize the SWD to channels with fading and either convolutional or trellis codes. We analyze two simple cases using the SWD approximation and the analytical results are compared to the simulated BER performance of the decoders with very good agreement of the two. Our analysis is applicable to more complex code structures as well, but the computational complexity becomes prohibitive. Therefore, we investigate the performance of all decoders for more complex codes by simulations alone.

The remainder of this chapter is organized as follows. In Section 3.2, we briefly introduce the FSMC Model of slowly fading channels. In Section 3.3 we present the system model. The decision-feedback decoder, output-feedback decoder, and conventional decoder with and without weighting are described in Section 3.4. In Section 3.5 we present our performance analysis of these decoders. The numerical results from analysis and simulation along with the coding, modulation, and channel models are shown in Section 3.6. Our conclusions are given in Section 3.7.

3.2 FSMC Model of Slowly Fading Channels

In our two feedback decoders, an FSMC model of the fading amplitude is used. We now give a brief review on how to obtain this FSMC model for the channel fade level (fading amplitude) α with a given first-order distribution $p_\alpha(a)$ and known second-order statistics of the fading process. Detailed discussions can be found in [45, 46].

Let S denote the dynamic range of the fade level α. For example, $S = (0, \infty)$ if α is Rayleigh distributed. Divide S into a finite number of disjoint sets $(S_1, \cdots, S_K)$ with the thresholds $a_1, \cdots, a_{K-1}$ as shown in Figure 3.1, where each set corresponds to a channel state. That is, by denoting $a_0 = 0$ and $a_K = \infty$, the fade level $\alpha \in S_i$ if $a_{i-1} \leq \alpha < a_i$,

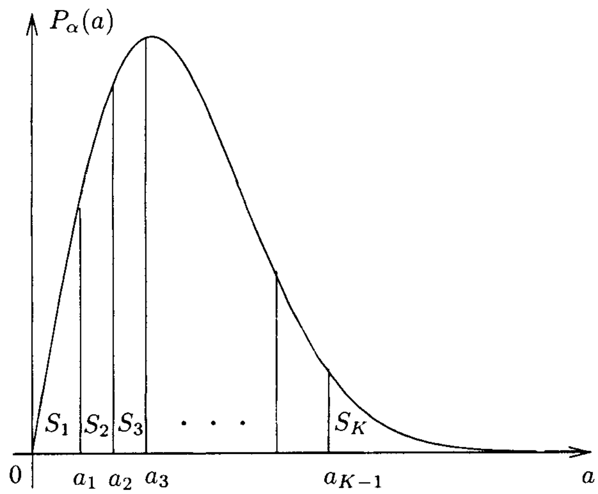


Figure 3.1: FSMC modeling of the fading channel.

$1 \leq i \leq K$. The K channel states $S_1, \cdots, S_K$ will form a Markov chain as shown in Figure 3.2, where the transition probability from state S_i to state S_j is denoted as P_{ij}, $i, j = 1, \cdots, K$. Given the two-dimensional joint probability density function $p(\gamma_{t1}, \gamma_{t2})$ of the fading process, P_{ij} can be calculated as [46]:

$$P_{ij} \quad = \quad P(\gamma_{t2} \in S_j | \gamma_{t1} \in S_i) \tag{3.1}$$

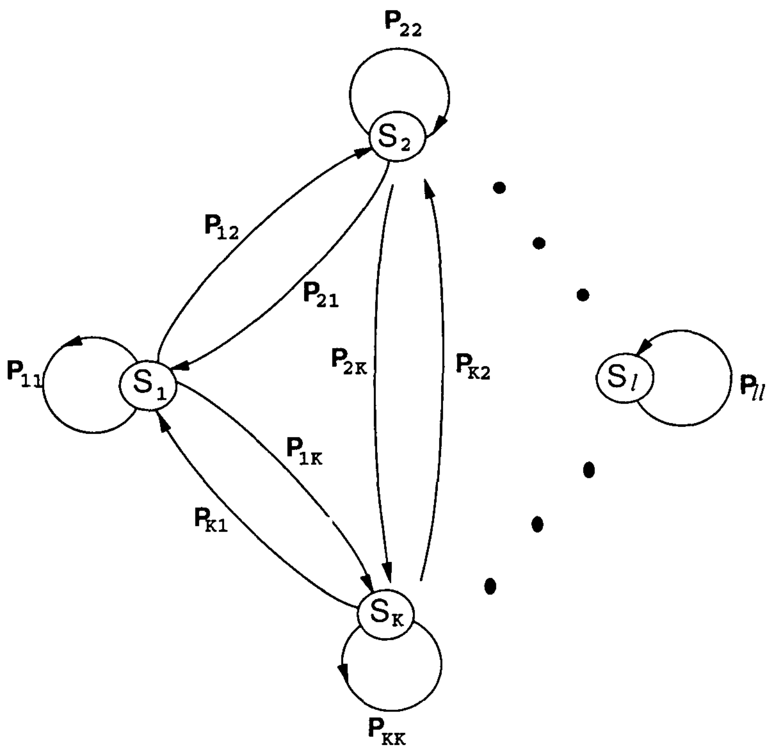


Figure 3.2: The FSMC model.

$$= \frac{\int_{a_{i-1}}^{a_i} \int_{a_{j-1}}^{a_j} p(\gamma_{t1}, \gamma_{t2}) d\gamma_{t1} d\gamma_{t2}}{P(S_i)},$$

where $P(S_i)$ is the stationary probability of the channel state S_i, i.e.,

$$P(S_i) = \int_{a_{i-1}}^{a_i} p_\alpha(a) da, \quad i = 1, \cdots, K.$$

In situations where $p(\gamma_{t1}, \gamma_{t2})$ is unknown, Monte Carlo simulation can be used to construct the FSMC model from experimental data.

For the Rayleigh fading channel, $p_\alpha(a)$ given in (2.1) is parameterized by Ω and $p(\gamma_{t1}, \gamma_{t2})$ is [4]:

$$\begin{aligned} p(\gamma_{t1}, \gamma_{t2}) &= \frac{\gamma_{t1}\gamma_{t2}}{\sqrt{|B|}} \exp\left\{ -\frac{\Omega(\gamma_{t1}^2 + \gamma_{t2}^2)}{2\sqrt{|B|}} \right\} \\ &\cdot I_0\left(\frac{\gamma_{t1}\gamma_{t2}}{\sqrt{|B|}} R(t2 - t1) \right), \quad \gamma_{t1}, \gamma_{t2} \geq 0, \end{aligned} \quad (3.2)$$

where $\sqrt{|B|} = \Omega^2 - [R(t2-t1)]^2$, $I_0(\cdot)$ is the zero-order modified Bessel function of the first kind, and $R(\cdot)$ is the autocorrelation function of the underlying complex Gaussian process. That is,

$$R(\tau) = J_0^2[2\pi f_d \tau],$$

where $J_0[\cdot]$ is the zero-order Bessel function of the first kind, and f_d is the Doppler frequency shift due to user mobility.

For slow Rayleigh fading, it is shown in [45] that the transition probabilities between channel states can be well approximated based on level-crossing rates. Specifically, since the channel changes slowly, transitions are assumed to occur only between neighboring states, i.e.,

$$P_{ij} = 0, \quad \forall |i-j| > 1.$$

Let $A_0 \cdots, A_K$ be the thresholds of the received SNR corresponding to the thresholds $a_0, \cdots, a_K$. Then the expected rate that the received SNR passes downward across the threshold A_i is [4]:

$$N_i = \sqrt{\frac{\pi A_i}{\Omega}} f_d \cdot \exp\left\{-\frac{A_i}{2\Omega}\right\}, \quad 1 \leq i \leq K-1.$$

Therefore, assuming that the transmission symbol rate is R_s, the transition probabilities are [45]:

$$P_{i,i+1} \approx \frac{N_i}{R_s P(S_i)}, \quad 1 \leq i \leq K-1,$$

$$P_{i,i-1} \approx \frac{N_{i-1}}{R_s P(S_i)}, \quad 2 \leq i \leq K,$$

and

$$P_{11} = 1 - P_{12},$$

$$P_{K,K} = 1 - P_{K,K-1},$$

$$P_{ii} = 1 - P_{i,i+1} - P_{i,i-1}, \quad 2 \leq i \leq K-1.$$

In our analyses and simulations we will use this simple method to obtain the FSMC model of the slowly Rayleigh fading channel and, as in [45], the thresholds are chosen such that the stationary probability of each state is the same.

3.3 System Model

3.3.1 Perfect Phase Compensation

Our baseband system model under the assumption of idea carrier recovery (perfect phase compensation) is shown in Figure 3.3. At the transmitter, the input binary data is first passed through a convolutional encoder. The encoded code words are then interleaved by a $J \times L$ block interleaver and mapped to a PSK or QAM signal constellation. The resulting sequence $\{x_n\}$ is fed to a square-root Nyquist filter, the output of which is the baseband transmit signal $s(t)$.

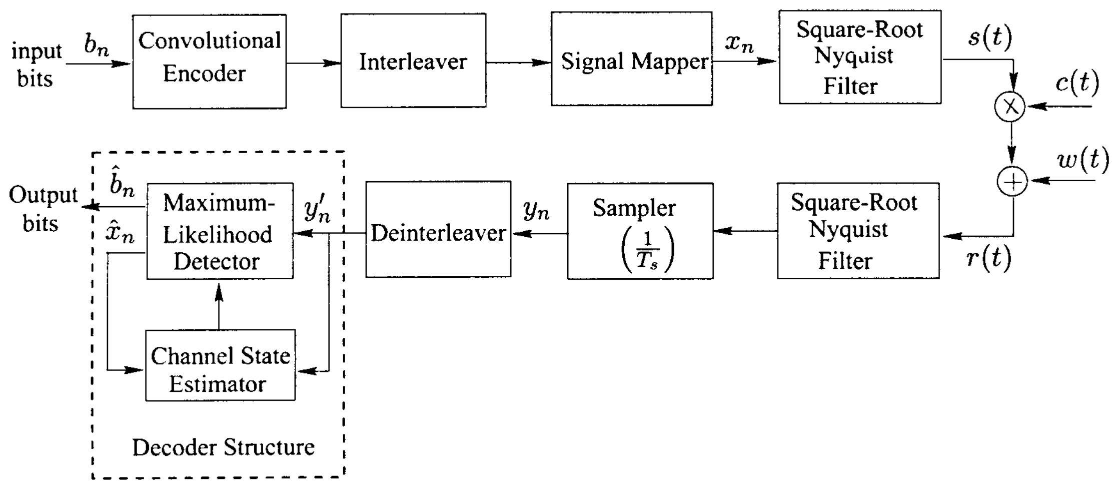


Figure 3.3: Baseband system model.

The transmit signal $s(t)$ passes through a fading channel with zero-mean additive Gaussian noise $w(t)$, where the complex multiplicative distortion process of the channel is denoted by $c(t)$. With Rayleigh fading $c(t)$ can be modeled as a zero-mean Gaussian low-pass process, the bandwidth of which equals the Doppler spread due to user mobility. We assume that the power of $c(t)$ is normalized to unity.

The received baseband signal $r(t)$ is first passed through a matched filter and then sampled at the symbol rate $1/T_s$. Generally, assuming that $c(t)$ varies slowly with respect to the symbol duration T_s and perfect timing is available (no intersymbol interference), the received sample sequence $\{y_n\}$ without phase compensation will be:

$$y_n = c_n x_n + w_n, \quad n = 0, 1, 2, \cdots, \tag{3.3}$$

where c_n is the n-th sample of $c(t)$, with amplitude α_n and phase θ_n, i.e., $c_n = \alpha_n e^{j\theta_n}$, and $\{w_n\}$ is a set of statistically independent complex Gaussian variables with zero-mean and variance σ^2 in each dimension. The variance σ^2 is:

$$\sigma^2 = \frac{1}{2E_s/N_0},$$

where E_s is the average signal energy per symbol and $N_0/2$ is the power spectral density of the noise.

If we assume perfect carrier recovery at the receiver, then the phase distortion caused by the fading process $c(t)$ will be completely compensated for. In this case, (3.3) can be simplified to:

$$y_n = \alpha_n x_n + w_n, \quad n = 0, 1, 2, \cdots,$$

and only the fade lever α_n is to be estimated for each received symbol. After deinterleaving, the symbol sequence is denoted as $\{y'_n\}$. Since the fading channel has finite memory, the symbols within any row of the deinterleaver become independent as the interleaving depth J becomes large. However, the symbols within any column of the deinterleaver are received from consecutive channel uses, and are thus dependent due to channel correlation. This dependence will be used in the decision-feedback decoder and the output-feedback decoder for channel state estimation.

In each of the two feedback decoders, the channel state estimator computes an estimate of the fade level distribution conditioned on past channel outputs. For the decision-feedback decoder, this conditional fade distribution at time n is computed from past received symbols $(y_1, \ldots, y_{n-1})$ and the feedback decisions $(\hat{x}_1, \ldots, \hat{x}_{n-1})$ which are obtained by re-encoding the bit decisions $(\hat{b}_1, \ldots, \hat{b}_{n-1})$ from the ML sequence detector. For the output-feedback decoder, only the past received symbols $(y_1, \ldots, y_{n-1})$ are used to compute the conditional distribution. In the case of perfect CSI, the state estimator output at time n is assumed to be the exact fade level α_n. More details on the state estimators and the corresponding ML sequence detectors will be given in Section 3.4.

3.3.2 Differential Encoding/Decoding of the Signal Phase

Our baseband system model for transmitting and receiving differentially phase-encoded nonconstant-envelope signals is similar to that of Figure 3.3, except that at the transmitter, a differential phase-encoder is added after the signal mapper, and at the receiver, a differential phase-decoder is added after the sampler. That is, at the transmitter, if we still let x_n denote the signal input to the square-root Nyquist filter at time n while denoting the output of the signal mapper as

$$v_n = A_n e^{j\phi_n},$$

with A_n and ϕ_n representing the amplitude and phase of a signal point on the signal constellation, respectively, then

$$x_n = A_n e^{j\psi_n},$$

where $\psi_n = \psi_{n-1} + \phi_n$.

Similarly, at the receiver, if we still denote the input to the deinterleaver as y_n while denoting the output of the sampler as

$$z_n = B_n e^{j\Phi_n},$$

with B_n and Φ_n representing the amplitude and phase of the received symbol at time n, respectively, then

$$y_n = B_n e^{j\Psi_n},$$

where $\Psi_n = \Phi_n - \Phi_{n-1}$.

3.4 Decoder Structures

In this section we describe the two feedback decoder structures for ML sequence detection of coded signals over fading channels. We also describe the conventional decoding technique for fading channels, and propose an improvement to this method which weights the decoding metric by the stationary channel distribution.

3.4.1 Decision-Feedback Decoder

A detailed block diagram of the decision-feedback decoder is shown in Figure 3.4. To make use of the channel memory, this decoder approximates the channel variation with a stationary FSMC model. That is, as shown in Section 3.2, the range S of the channel fade level α_n is divided into a finite number of disjoint sets $(S_1, \ldots, S_K)$, corresponding to a finite number of channel states. The channel varies over these states according to a Markov process. Based on this channel model, the decoder is composed of a channel state estimator and a soft-decision ML sequence detector. The state estimator computes the conditional distribution $\pi_n(l)$ of the channel state S_l based on past channel output observations and feedback decisions.

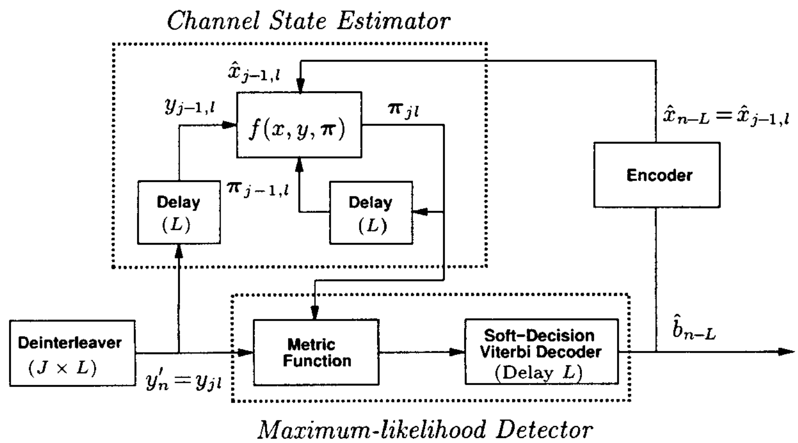


Figure 3.4: Decision-feedback decoder.

As shown in Figure 3.3, $\{y_n\}$ denotes the received symbol sequence after sampling at rate $\frac{1}{T_s}$. The corresponding transmitted symbol sequence is $\{x_n\}$. Let $y^n = (y_1, \ldots, y_n)$ and $x^n = (x_1, \ldots, x_n)$. Then the conditional state distribution based on past channel outputs and perfect feedback decisions $(\hat{x}_n = x_n, \forall n)$ is

$$\pi_n(l) = p(\alpha_n \in S_l | x^{n-1}, y^{n-1}). \tag{3.4}$$

For an FSMC channel model, $\pi_n(l)$ can be computed recursively according to [42]

$$\pi_{n+1}(l) = \frac{\sum_{j=1}^{K} p(y_n|x_n, \alpha_n \in S_j) p(\alpha_n \in S_j | x^{n-1}, y^{n-1}) P_{jl}}{\sum_{k=1}^{K} p(y_n|x_n, \alpha_n \in S_k) p(\alpha_n \in S_k | x^{n-1}, y^{n-1})},$$

where P is the matrix of transition probabilities for the channel states with $P_{jl} = p(\alpha_{n+1} \in S_l | \alpha_n \in S_j)$. This recursion can be written in vector form

$$\boldsymbol{\pi}_{n+1} = \frac{\boldsymbol{\pi}_n D(x_n, y_n) P}{\boldsymbol{\pi}_n D(x_n, y_n) \underline{\mathbf{1}}} \triangleq f(x_n, y_n, \boldsymbol{\pi}_n), \tag{3.5}$$

where

$$\boldsymbol{\pi}_n = (\pi_n(1), \ldots, \pi_n(K)), \tag{3.6}$$

$D(x_n, y_n)$ is a diagonal $K \times K$ matrix with the kth diagonal term $p(y_n|x_n, \alpha_n \in S_k)$, and $\underline{\mathbf{1}} = (1, \ldots, 1)^T$ is a K-dimensional vector. The value of $f(x_n, y_n, \boldsymbol{\pi}_n)$ is calculated recursively in the state estimator, as shown in Figure 3.4. Note that although $y_1, \ldots, y_n$ are received subsequently over the channel, the deinterleaver delays subsequent channel outputs by L (the number of interleaver columns). Therefore, the calculation of $f(x_n, y_n, \boldsymbol{\pi}_n)$ in our state estimator requires a delay of L in the deinterleaver output y'_n, and a corresponding delay in the state estimator output $\boldsymbol{\pi}_n$ and feedback decision $\hat{x}_n = x_n$. These delays are indicated in Figure 3.4, where $y_{jl} \triangleq y'_n$ explicitly denotes that y'_n is in the jth row and lth column of the deinterleaver, and $\boldsymbol{\pi}_{jl} \triangleq \boldsymbol{\pi}_n$ and $x_{jl} \triangleq x_n$ denote, respectively, the state estimate and feedback decision corresponding to y_{jl}. With this notation the recursive calculation (3.5) becomes

$$\boldsymbol{\pi}_{jl} = \frac{\boldsymbol{\pi}_{j-1,l} D(x_{j-1,l}, y_{j-1,l}) P}{\boldsymbol{\pi}_{j-1,l} D(x_{j-1,l}, y_{j-1,l}) \underline{\mathbf{1}}}. \tag{3.7}$$

It is shown in [42] that, assuming perfect feedback decisions and with asymptotically deep interleaving, the conditional distribution $\boldsymbol{\pi}_n$ is a sufficient statistic for the channel state given all past channel outputs and corresponding feedback decisions. The metric used in the ML sequence detector, $m(x^n, y^n) = \sum_{j=1}^{n} m(x_j, y_j)$, is updated at each n, where

$$m(x_j, y_j) = -\log \left[\sum_k p(y_j|x_j, \alpha_j \in S_k) \pi_j(k)\right] \tag{3.8}$$

and

$$p(y_j|x_j, \alpha_j \in S_k) = \frac{\int_{S_k} p(y_j|x_j, \gamma) p_\alpha(\gamma) d\gamma}{\int_{S_k} p_\alpha(\gamma) d\gamma}. \tag{3.9}$$

In (3.9), $p_\alpha(\cdot)$ is the stationary distribution of the channel fading. The Viterbi decoder, after a decoding delay $\delta = L$, outputs the bit decisions $\hat{b}_n$ corresponding to the path with the smallest metric, which are then re-encoded to obtain the feedback decisions $\hat{x}_n$.

3.4.2 Output-Feedback Decoder

The structure of the output-feedback decoder, shown in Figure 3.5, is similar to that of the decision-feedback decoder, except that the state estimator computes the conditional distribution of the channel state based only on past channel output observations: $\rho_n(l) = p(\alpha_n \in S_l|y^{n-1})$. Since no feedback decisions are used by the state estimator, the estimator can operate on received symbols prior to deinterleaving.

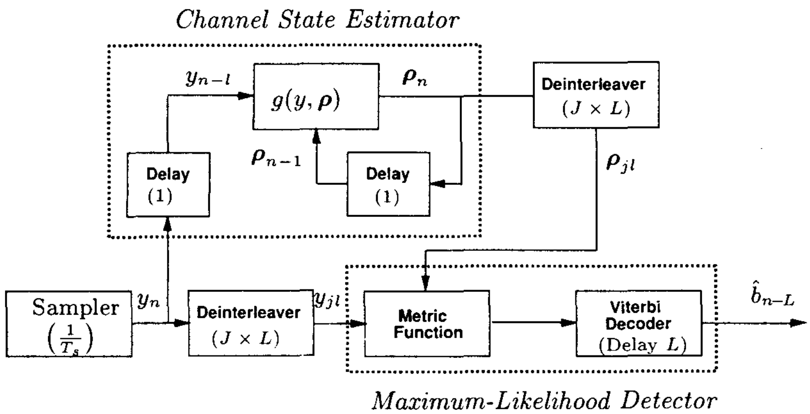


Figure 3.5: Output-feedback decoder.

If we denote the channel input alphabet by χ and the distribution of x_n by $p(x_n)$, ρ_n can also be calculated recursively as [42]

$$\rho_{n+1} = \frac{\rho_n B(y_n) P}{\rho_n B(y_n) \underline{1}} \triangleq g(y_n, \rho_n), \tag{3.10}$$

where $B(y_n)$ is a diagonal $K \times K$ matrix with kth diagonal term

$$p(y_n|\alpha_n \in S_k) = \frac{\sum_{x_n \in \chi} \int_{S_k} p(y_n|x_n, \gamma) p_\alpha(\gamma) p(x_n) d\gamma}{\int_{S_k} p_\alpha(\gamma) d\gamma}. \tag{3.11}$$

The metric function in this decoder is similar to the one used for the decision-feedback decoder, with the sufficient statistic π_j in (3.8) replaced by ρ_j:

$$m(x_j, y_j) = -\log\left[\sum_k p(y_j|\alpha_j \in S_k)\rho_j(k)\right]. \tag{3.12}$$

3.4.3 Conventional Decoding with Weighting

The conventional decoder does not use the fading statistics of the channel in its ML sequence detector. That is, the decoder processes the deinterleaved symbols using a metric

$$m(y_j, x_j) = -\log p(y_j|x_j), \tag{3.13}$$

where $p(y_j|x_j)$ is based on a memoryless AWGN channel with the same SNR as the fading channel.

We can improve the performance of this decoder by using the stationary fading distribution in the metric calculation. This will not be as accurate as using the conditional distribution computed by either the decision-feedback or the output-feedback decoder, but will use at least some information about the channel statistics in the decoding process. This weighted metric is given by

$$m(y_j, x_j) = -\log \int_{-\infty}^{\infty} p(y_j|x_j, \gamma)p_\alpha(\gamma)d\gamma. \tag{3.14}$$

Note that the conventional decoder with weighting is equivalent to the feedback decoders designed for a single channel state S, where S is the range of fading values. Also, it is proved in Appendix A that when hard decision is used, the conventional decoder with weighting and the output-feedback decoder are equivalent for any type of modulation with a decision rule independent of the fade level, such as MPSK modulation.

3.4.4 Maximum-Likelihood Decoding with Perfect CSI

When the channel fade level α_j is known at the receiver [55], the decoder processes the deinterleaved symbols using a metric

$$m(y_j, x_j) = -\log p(y_j|x_j, \alpha_j). \tag{3.15}$$

The performance of this idea decoder yields a lower bound on the BER performance of the feedback decoders, which only know the conditional distribution of the channel state instead of its actual value.

3.5 Performance Analysis

There is much previous work estimating the BER of convolutional codes and trellis codes on fading channels ([56]-[62] and the references therein). All of these papers evaluate the decoder performance directly, typically using Chernoff upper bounds or exact expressions for the pairwise error probabilities of the trellis codes and then using the well-known transfer function method [56] to bound the total BER of a decoder with or without CSI. It is difficult to obtain similar bounds or expressions for our feedback decoders and for the conventional decoder with channel weighting, since these decoders have relatively complex decision metrics. The sliding window decoding technique is a method of approximating the BER performance of a VD. The SWD approaches an optimal ML symbol decoder as its window size increases. An optimal ML symbol decoder and an optimal ML sequence decoder have similar performance at high SNRs. Thus, we can well-approximate the performance of our feedback decoders from that of a SWD with a large window size and large SNR. Since the BER of the SWD is easily obtained using a Markov model for the decoding process, this approximation is an effective method to calculate the BER of an ML sequence decoder with a complex metric.

Let (n, k, m) denote a convolutional code which encodes $m+1$ blocks of k-tuple input bits into n coded bits. In the following we briefly illustrate the operation of a SWD with a $(2,1,1)$ convolutional code [54], the encoder structure of which is given in Figure 3.6. The encoded bits are mapped to the QPSK signal constellation through Gray encoding.

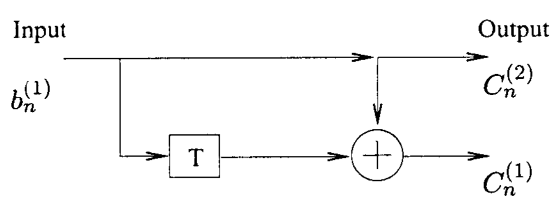


Figure 3.6: The encoder structure of the $(2,1,1)$ convolutional code.

Figure 3.7 shows the tree structure of this code, with the output symbol indicated on each branch. This figure is similar to Fig. 1 in [54]. Since there are only two states in this code, starting from a given state, the decoder will either remain in that state or transfer to the other state as the decoding process proceeds. The state transition depends on L_w symbols of the received sequence, where L_w is the sliding window size.

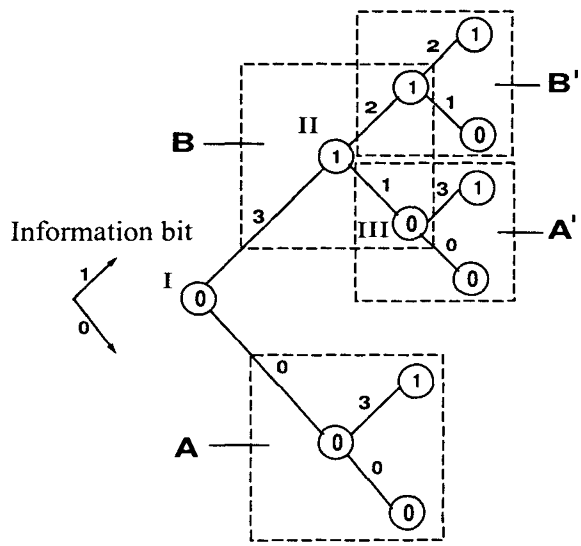


Figure 3.7: Sliding window decoding for the (2, 1, 1) convolutional code.

Consider an example with sliding window size $L_w = 2$. Starting from an initial state of 0 (labeled as node I in Figure 3.7), there are four possible output symbol sequences of length L_w: $\mathbf{v_1} = (0,0)$, $\mathbf{v_2} = (0,3)$, $\mathbf{v_3} = (3,1)$, $\mathbf{v_4} = (3,2)$. Let $\mathbf{A} = \{\mathbf{v_1}, \mathbf{v_2}\}$ and $\mathbf{B} = \{\mathbf{v_3}, \mathbf{v_4}\}$, as indicated in Figure 3.7. Denote the received symbol sequence as $\mathbf{y} = (y_0, y_1)$ and denote the distance of $\mathbf{y}$ to sets $\mathbf{A}$ and $\mathbf{B}$ as $d(\mathbf{y}, \mathbf{A})$ and $d(\mathbf{y}, \mathbf{B})$, respectively. That is,

$$d(\mathbf{y}, \mathbf{A}) = \min_{\mathbf{v_i} \in \mathbf{A}} \{d(\mathbf{y}, \mathbf{v_i})\},$$

$$d(\mathbf{y}, \mathbf{B}) = \min_{\mathbf{v_i} \in \mathbf{B}} \{d(\mathbf{y}, \mathbf{v_i})\},$$

where $d(\cdot, \cdot)$ is the distance between two sequences according to a specific decision metric. Specifically, for sequences $\mathbf{y}$ and $\mathbf{v}$,

$$d(\mathbf{y}, \mathbf{v}) = \sum_{l=0}^{L_w-1} m(v_l, y_l), \tag{3.16}$$

where $m(\cdot,\cdot)$ is the metric of the specific decoder under consideration (e.g., (3.8), (3.12), (3.13), (3.14) or (3.15)). If $\mathbf{y}$ is closer to set $\mathbf{A}$ than to set $\mathbf{B}$, i.e., $d(\mathbf{y},\mathbf{A}) < d(\mathbf{y},\mathbf{B})$, then the decoder starting from state 0 will remain in state 0 and the input information bit will be determined as 0. Otherwise, the decoder will go to state 1 and the input bit will be decoded as 1. Assuming that state 1 is chosen as the next starting node (labeled as node II), then the four possible output symbol sequences of length L_w ($L_w = 2$) are: $\mathbf{v}'_1 = (1,0)$, $\mathbf{v}'_2 = (1,3)$, $\mathbf{v}'_3 = (2,1)$, $\mathbf{v}'_4 = (2,2)$. Note that $\mathbf{A}' = \{\mathbf{v}'_1, \mathbf{v}'_2\}$ and $\mathbf{B}' = \{\mathbf{v}'_3, \mathbf{v}'_4\}$. Now $\mathbf{y} = (y_1, y_2)$. If $\mathbf{y}$ is closer to set $\mathbf{A}'$ than to set $\mathbf{B}'$, then the decoder starting from state 1 will go to state 0 (labeled as node III) and the input bit will be decoded as 0. Otherwise, the decoder will remain in state 1 and the input will be determined as 1. It is obvious that the BER of this SWD decreases as L_w becomes large, since a longer symbol sequence is used to decode each bit. Moreover, as L_w approaches infinity, this SWD becomes the optimal ML symbol decoder since the entire sequence is used to decode each bit.

The decoding process of a SWD for N-state convolutional codes or trellis codes can be modeled as an N-state Markov chain. Let q_{ij} be the probability that the decoder will go from state i to state j, and q_i be the stationary probability of being in state i, for $i,j = 0,1,2,\cdots,N-1$. Assume that the all-zero codeword sequence is transmitted and let $\mathbf{x}_{ij}$ be the k-tuple input bits associated with the transition from state i to state j in the decoding of an (n,k,m) convolutional code. The decoded BER will then be [54]

$$P_b = \frac{1}{k} \sum_{i=0}^{N-1} q_i \sum_{j=0}^{N-1} q_{ij} w[\mathbf{x}_{ij}], \tag{3.17}$$

where $w[\mathbf{x}_{ij}]$ denotes the Hamming weight of $\mathbf{x}_{ij}$. Note that $\forall 0 \le i \le N-1$, $\sum_{j=0}^{N-1} q_{ij} = 1$, $\sum_{i=0}^{N-1} q_i = 1$, and the stationary probabilities can be derived in terms of the transition probabilities.

For the $(2,1,1)$ code, since $w[\mathbf{x}_{00}] = 0$, $w[\mathbf{x}_{01}] = 1$, $w[\mathbf{x}_{10}] = 0$, $w[\mathbf{x}_{11}] = 1$, we have

$$P_b = q_0 q_{01} + q_1 q_{11} \tag{3.18}$$

and

$$q_{01} = \sum_{\mathbf{y}} 1[\mathbf{y} : d(\mathbf{y},\mathbf{A}) > d(\mathbf{y},\mathbf{B})]\; p(\mathbf{y}|\mathbf{x} = \mathbf{0}), \tag{3.19}$$

$$q_{11} = \sum_{\mathbf{y}} 1[\mathbf{y} : d(\mathbf{y}, \mathbf{A}') > d(\mathbf{y}, \mathbf{B}')] \; p(\mathbf{y}|\mathbf{x} = \mathbf{0}), \tag{3.20}$$

where $1[\cdot]$ is the indicator function.

For an FSMC channel with channel states $S_1, \ldots, S_K$, the derivation of q_{01} and q_{11} will be modified as:

$$q_{01} = \sum_{\mathbf{S}} \sum_{\mathbf{y}} 1[\mathbf{y} : d(\mathbf{y}, \mathbf{A}) > d(\mathbf{y}, \mathbf{B})| \; \mathbf{S}] \; p(\mathbf{y}|\mathbf{x} = \mathbf{0}, \mathbf{S}) \; p(\mathbf{S}), \tag{3.21}$$

$$q_{11} = \sum_{\mathbf{S}} \sum_{\mathbf{y}} 1[\mathbf{y} : d(\mathbf{y}, \mathbf{A}') > d(\mathbf{y}, \mathbf{B}')| \; \mathbf{S}] \; p(\mathbf{y}|\mathbf{x} = \mathbf{0}, \mathbf{S}) \; p(\mathbf{S}), \tag{3.22}$$

where $\mathbf{S} = (S^{(0)}, \ldots, S^{(L_w-1)})$ is a length-L_w state vector, with each of its elements being one of the K channel states $S_1, \ldots, S_K$. There are K^{L_w} such vectors. In addition, since we assume ideal interleaving, the adjacent channel states and received symbols after deinterleaving are independent. Therefore, in (3.21) and (3.22),

$$p(\mathbf{S}) = \prod_{l=0}^{L_w-1} p(S^{(l)}), \tag{3.23}$$

$$p(\mathbf{y}|\mathbf{x} = \mathbf{0}, \mathbf{S}) = \prod_{l=0}^{L_w-1} p(y_l|x_l = 0, S^{(l)}), \tag{3.24}$$

where $p(S^{(l)})$ is the stationary probability of the channel being in state $S^{(l)}$.

For conventional decoders with and without channel weighting, their metrics (3.14) and (3.13) do not depend on channel states. For the ideal decoder with perfect CSI, its metric (3.15) is determined once the channel state is given. Thus, conditioned on channel states $\mathbf{S} = (S^{(0)}, \ldots, S^{(L_w-1)})$, the distances $d(\mathbf{y}, \mathbf{A}), d(\mathbf{y}, \mathbf{B}), d(\mathbf{y}, \mathbf{A}')$ and $d(\mathbf{y}, \mathbf{B}')$ in (3.21) and (3.22) are straightforward to calculate for these three decoders using their corresponding metrics. Consequently, the transition probabilities (3.21) and (3.22) of their SWDs are calculated and the BER (3.18) is easily obtained. However, for the decision-feedback decoder, its metric (3.8) is a function of the channel state distribution, which is calculated recursively from all past channel observations and which is therefore changing from symbol to symbol with time. The same is true for the metric (3.12) of the output-feedback decoder. Therefore, the indicator functions in (3.21) and (3.22) require further analysis.

In particular, we derive the expressions of q_{01} and q_{11} in (3.21) and (3.22) for the feedback decoders using the limiting probabilities of the channel state distributions, which will be discussed further shortly. For the decision-feedback decoder,

$$1[\mathbf{y}: d(\mathbf{y},\mathbf{A}) > d(\mathbf{y},\mathbf{B})|\ \mathbf{S}] = \sum_{\underline{\boldsymbol{\pi}}} 1[\mathbf{y}: d(\mathbf{y},\mathbf{A}) > d(\mathbf{y},\mathbf{B})|\ \underline{\boldsymbol{\pi}}]\ p(\underline{\boldsymbol{\pi}}|\mathbf{S}), \tag{3.25}$$

$$1[\mathbf{y}: d(\mathbf{y},\mathbf{A}') > d(\mathbf{y},\mathbf{B}')|\ \mathbf{S}] = \sum_{\underline{\boldsymbol{\pi}}} 1[\mathbf{y}: d(\mathbf{y},\mathbf{A}') > d(\mathbf{y},\mathbf{B}')|\ \underline{\boldsymbol{\pi}}]\ p(\underline{\boldsymbol{\pi}}|\mathbf{S}), \tag{3.26}$$

where $\underline{\boldsymbol{\pi}} = (\boldsymbol{\pi}^{(0)}, \ldots, \boldsymbol{\pi}^{(L_w-1)})$, with $\boldsymbol{\pi}^{(l)}$, $l = 0, \ldots, L_w - 1$, denoting the channel state distributions, and

$$p(\underline{\boldsymbol{\pi}}|\mathbf{S}) = \prod_{l=0}^{L_w-1} p(\boldsymbol{\pi}^{(l)}|S^{(l)}). \tag{3.27}$$

In (3.27), $p(\boldsymbol{\pi}^{(l)}|S^{(l)})$ is the limiting conditional probability that the distribution for the channel state will be vector $\boldsymbol{\pi}^{(l)}$ given the current channel being in state $S^{(l)}$. In [42] it is proved that for independent identically distributed (i.i.d) inputs, the channel state distribution $\boldsymbol{\pi}_n$ defined in (3.6) is a Markov chain. Moreover, it converges very fast in distribution to a limit which is independent of the initial channel state. That is, as n approaches infinity, the limit of $p(\boldsymbol{\pi}_n|s_0 = S_i)$ exists and it is equal for all i, $i = 1, \ldots, K$. In addition, we find that there is also a limit for $p(\boldsymbol{\pi}_n|s_n = S_i)$, $i = 1, \ldots, K$, which is different for different i, and we use this limit to calculate the indicator functions in (3.25) and (3.26). This limiting conditional probability can be derived iteratively with the following recursive relation between $p(\boldsymbol{\pi}_{n+1}|s_{n+1})$ and $p(\boldsymbol{\pi}_n|s_n)$:

$$\begin{aligned} p(\boldsymbol{\pi}_{n+1}|s_{n+1}) &= \sum_{s_n}\sum_{\boldsymbol{\pi}_n} p(\boldsymbol{\pi}_{n+1}, \boldsymbol{\pi}_n, s_n|s_{n+1}) \\ &= \sum_{s_n}\sum_{\boldsymbol{\pi}_n} \{p(\boldsymbol{\pi}_{n+1}|\boldsymbol{\pi}_n, s_n)p(\boldsymbol{\pi}_n|s_n)p(s_{n+1}|s_n)p(s_n)\}/p(s_{n+1}) \\ &= \sum_{s_n}\sum_{\boldsymbol{\pi}_n}\sum_{x_n}\sum_{y_n} 1[(x_n, y_n): f(x_n, y_n, \boldsymbol{\pi}_n) = \boldsymbol{\pi}_{n+1}]\ p(y_n|x_n, s_n)p(x_n) \\ &\quad \cdot p(\boldsymbol{\pi}_n|s_n)p(s_{n+1}|s_n)p(s_n)/p(s_{n+1}). \end{aligned} \tag{3.28}$$

Similarly, for the output-feedback decoder, the limiting conditional probability of the channel state distribution can be obtained from the recursion between $p(\boldsymbol{\rho}_{n+1}|s_{n+1})$ and $p(\boldsymbol{\rho}_n|s_n)$:

$$p(\boldsymbol{\rho}_{n+1}|s_{n+1}) = \sum_{s_n}\sum_{\boldsymbol{\rho}_n} p(\boldsymbol{\rho}_{n+1}, \boldsymbol{\rho}_n, s_n|s_{n+1})$$

$$= \sum_{s_n}\sum_{\rho_n}\left\{p(\rho_{n+1}|\rho_n, s_n)p(\rho_n|s_n)p(s_{n+1}|s_n)p(s_n)\right\}/p(s_{n+1})$$
$$= \sum_{s_n}\sum_{\rho_n}\sum_{y_n} 1[y_n : g(y_n, \rho_n) = \rho_{n+1}]\; p(y_n|s_n)$$
$$\cdot\, p(\rho_n|s_n)p(s_{n+1}|s_n)p(s_n)/p(s_{n+1}). \quad (3.29)$$

Now by substituting (3.25) into (3.21) and (3.26) into (3.22), we get the expressions of q_{01} and q_{11} for the decision-feedback decoder:

$$q_{01} = \sum_{\mathbf{S}}\sum_{\mathbf{y}}\sum_{\underline{\pi}} 1[\mathbf{y} : d(\mathbf{y}, \mathbf{A}) > d(\mathbf{y}, \mathbf{B})|\; \underline{\pi}]\; p(\underline{\pi}|\mathbf{S})\; p(\mathbf{y}|\mathbf{x} = \mathbf{0}, \mathbf{S})\; p(\mathbf{S}), \quad (3.30)$$

$$q_{11} = \sum_{\mathbf{S}}\sum_{\mathbf{y}}\sum_{\underline{\pi}} 1[\mathbf{y} : d(\mathbf{y}, \mathbf{A}') > d(\mathbf{y}, \mathbf{B}')|\; \underline{\pi}]\; p(\underline{\pi}|\mathbf{S})\; p(\mathbf{y}|\mathbf{x} = \mathbf{0}, \mathbf{S})\; p(\mathbf{S}). \quad (3.31)$$

Using a similar derivation, we find that for the output-feedback decoder, its q_{01} and q_{11} in (3.18) can be calculated simply by substituting vector $\underline{\pi}$ in (3.30) and (3.31) with vector $\underline{\rho} = (\rho^{(0)}, \ldots, \rho^{(L_w-1)})$. Note that in (3.28), perfect feedback decisions are assumed, since it is too complicated to consider error propagations in the BER analysis for the decision-feedback decoder.

3.6 Numerical Results

In this section we describe our analytical and simulation results for the five decoders: decision-feedback, output-feedback, conventional, conventional with weighting, and ML with perfect CSI. A simulation for each of the decoders was built using COSSAP [63].

Due to computational complexity in the analysis of codes with complicated structures, our analytical results are given for two simple codes. In order to compare the analytical results with the corresponding simulated results, we assume perfect feedback decisions in both the analysis and simulation for the decision-feedback decoder. When actual decisions from the VD are used, some performance degradation of the decision-feedback decoder is expected. This degradation will be most pronounced for simple codes with relatively high BERs; for powerful codes which sharply reduce the BER, we expect the effect of error propagation to be small. In our simulations for complex codes, we investigate the performance of the decision-feedback decoder using the actual decisions from the VD. Therefore, in these simulations, the true performance of the decision-feedback decoder is obtained and compared to that of other decoders.

We first determine the performance of all the decoders for the $(2,1,1)$ convolutional code with quadrature PSK (QPSK) modulation on a two-state fading channel. This channel, shown in Figure 3.8, consists of two states, a "good" state G and a "bad" state B, where each state is a memoryless AWGN channel with spectral noise density $N_0/2$ and fade level α_G for the good state and α_B for the bad state. The channel state varies between the good and bad states according to a stationary Markov process, with transition probabilities as shown in Figure 3.8. The stationary distribution of this channel is $p(G) = g/(g+b)$ and $p(B) = b/(g+b)$.

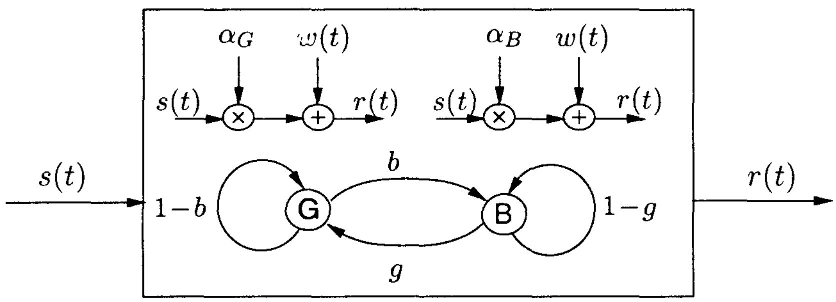


Figure 3.8: Two-state fading channel.

In our analysis and simulation we assume that the channel SNR is 10dB for the good state and −5dB for the bad state, and that perfect decisions are made in the decision-feedback decoder. The BER performance of the various decoders, shown in Figure 3.9, is analyzed and simulated for $b = 0.001$ fixed and a range of g values. As g increases, the channel spends more time in the good channel state, and the BER decreases accordingly. The dashed curves in Figure 3.9 correspond to analytical results and the solid lines are simulated ones. They agree with each other quite well for all the decoders. In our analysis, the sliding window size is $L_v = 3$ and we quantize the received signals into four levels in each dimension. The simulations are run under the same quantizing strategy for all decoders, except that two-level hard decision is used for the conventional decoder without weighting, since for this decoder, soft decision results in much worse performance than hard decision. Obviously the BER will decrease with an increase in the number of quantization levels. But we see that even with a two-bit quantizer, the decision-feedback decoder with perfect feedback decisions performs almost as well as the ideal decoder with exact CSI.

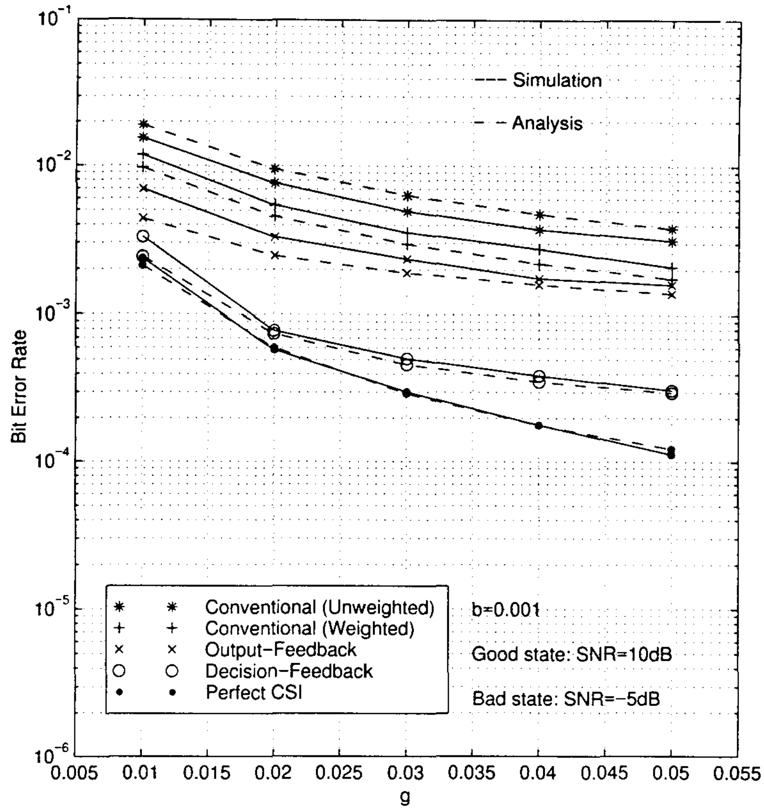


Figure 3.9: Analytical and simulated BER for QPSK and two-state convolutional coding (two-state fading channel).

The output-feedback decoder and the decoder with a weighted metric both have better performance than the best conventional decoder which uses hard decision and no metric weighting.

Figure 3.10 shows the analytical and simulation results for the four-state TC-4AM code [64] on a Rayleigh fading channel with normalized fading rate $f_dT = 0.001$, where f_d is the Doppler frequency shift and T is the symbol duration. In these calculations we again assume perfect decisions in the decision-feedback decoder. As in the previous figure, the dashed curves represent analytical results and the solid ones are simulations. Based on (3.17), the BER of this code can be derived from the state transition diagram of its corresponding SWD, and the simplified expression is

$$P_b = \frac{1 + p_{13}/p_{32}}{2 + p_{13}/p_{32} + p_{20}/p_{01}}.$$

A simple two-state Markov chain model of the Rayleigh fading channel is adopted in both

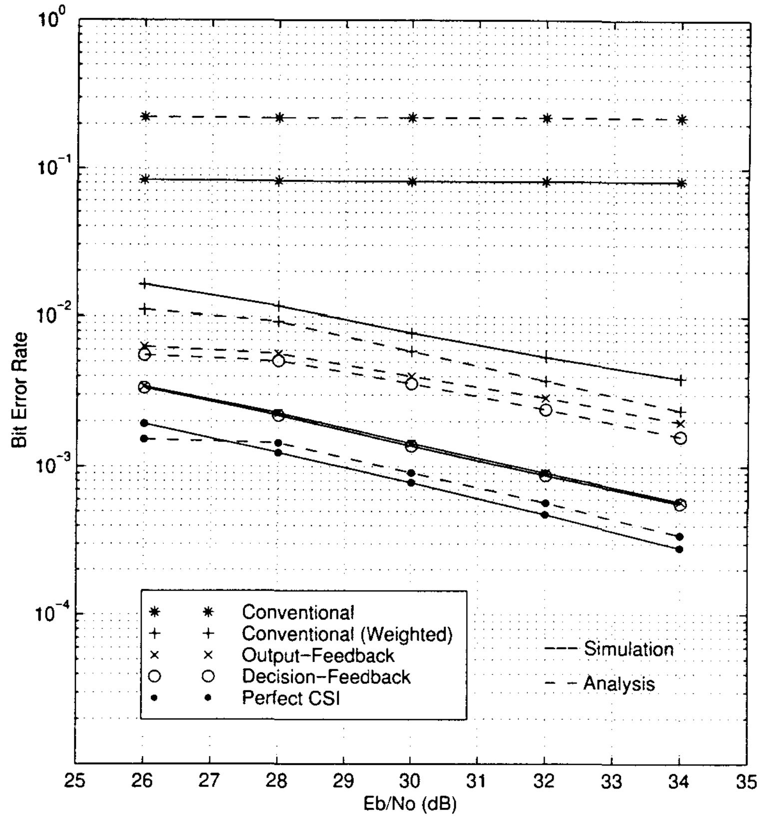


Figure 3.10: Analytical and simulated BER for TC-4AM (Rayleigh fading).

analysis and simulation of the two feedback decoders to keep the computation complexity reasonable. For each state, its probability is obtained by averaging the corresponding fading range over the Rayleigh distribution as shown in Section 3.2. The received signals are quantized to ten levels and the sliding window size remains to be $L_w = 3$. In this case the output-feedback decoder and the perfect decision-feedback decoder have very similar performance, due to the fact that the two-state Markov chain is a very rough approximation for the Rayleigh fading variation. Nevertheless, the feedback decoders still exhibit great performance advantages compared to the conventional decoder, which uses a quantized Euclidean distance metric and has a BER of around 0.08 for a large range of E_b/N_0. Note that in both Figure 3.9 and Figure 3.10, the decision-feedback decoder always outperforms the output-feedback decoder since feedback decisions are assumed perfect.

We investigate the performance of the five decoders for more complicated codes using simulation alone, due to the computational complexity of the analysis. As mentioned before, in these simulations we use the actual feedback decisions of the decoder, so the decoder will exhibit error propagation. Figure 3.11 shows the BER of these decoders for a 64-state maximum free distance $(2,1,6)$ code [43] with QPSK modulation on the two-state fading channel of Figure 3.8. The solid lines in this figure now correspond to hard decision of the

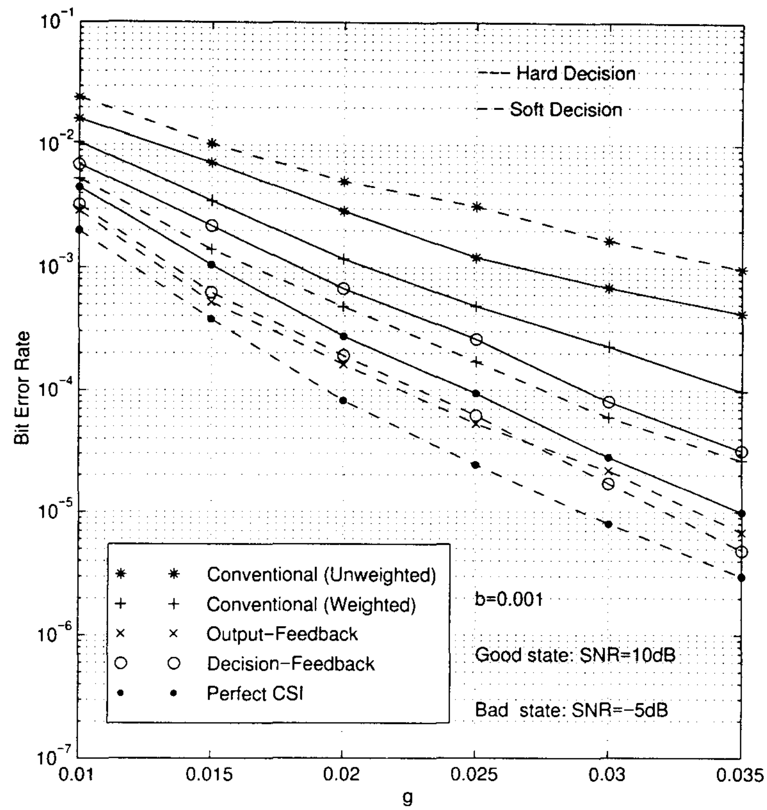


Figure 3.11: Simulated BER for QPSK and 64-state convolutional coding (two-state fading channel).

received symbols, so the in-phase and quadrature components are quantized to ± 1 before decoding. The dashed curves correspond to soft decision, so the unquantized received symbols are used in decoding. For hard decision, the solid line showing the performance of the conventional decoder with weighting also represents the performance of the output-feedback decoder, since the two decoders are equivalent when hard decision is used, as mentioned in Section 3.4.3.

We see that for soft decision, the two feedback decoders have similar performance: the output-feedback decoder does slightly better than the decision-feedback decoder on poor channels (g small), since it does not suffer from error propagation, and slightly worse on good channels. The effects of hard decision are quite dramatic. Hard decision increases the

BER by almost an order of magnitude in some cases, except for the conventional decoder without weighting, where hard decision actually decreases BER. Note that adding channel weighting to the conventional decoder decreases its BER significantly for both hard and soft decisions.

Figure 3.12 shows the simulated BER for eight-state TC-8PSK modulation, as described in [65], applied to the two-state fading channel of Figure 3.8. Soft decisions are used in each

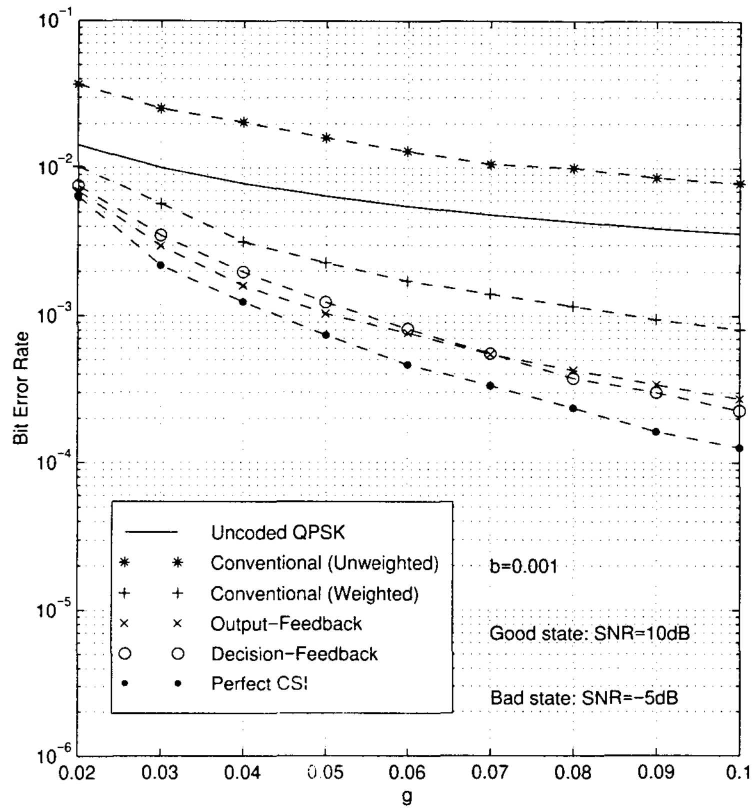


Figure 3.12: Simulated BER for TC-8PSK (two-state fading channel).

of the simulations in Figure 3.12. The solid line shows the BER for uncoded QPSK, which is lower than the BER of a conventional decoder applied to the TC-8PSK. As in the previous figure, the decision-feedback and output-feedback decoders have similar performance, and the output-feedback decoder performs slightly better than the decision-feedback decoder on poor channels and slightly worse on good channels. Using channel weighting in the conventional decoder decreases its BER by almost an order of magnitude.

The simulated BER for eight-state TC-16QAM applied to the Rayleigh fading channel with perfect phase compensation is shown in Figure 3.13. In these simulations, the normal-

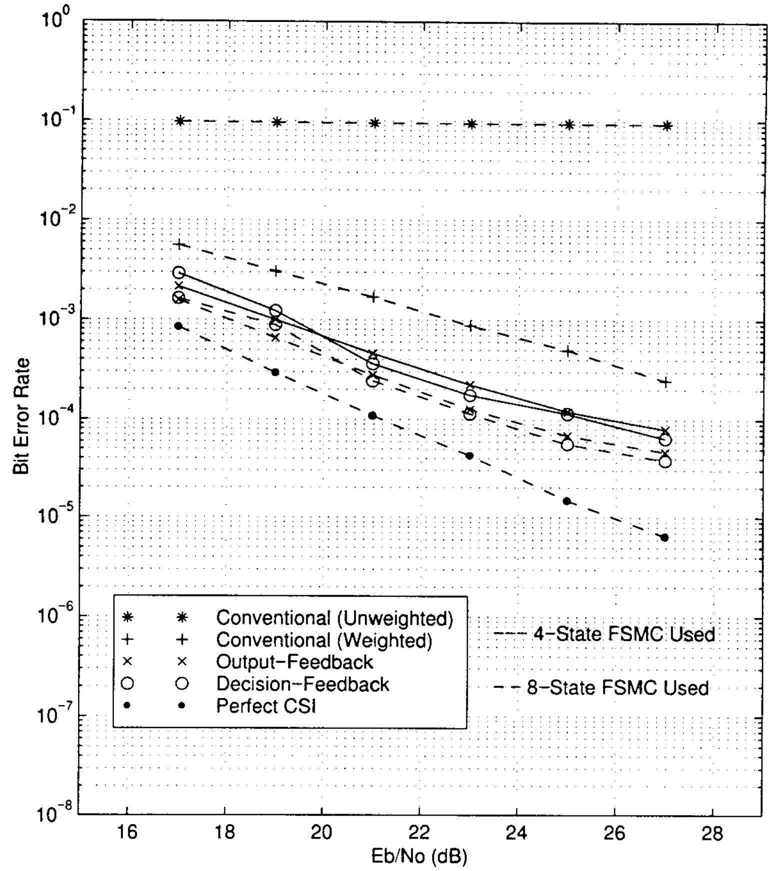


Figure 3.13: Simulated BER for TC-16QAM (Rayleigh fading).

ized fading rate is $f_d T = 0.001$ and soft-decisions are used in each decoder. The trellis-code, described in more detail in [66], is designed specifically for fading channels. The decision-feedback and output-feedback decoders assume an FSMC model of the Rayleigh channel and our BER results are reported for two different FSMC approximations. The solid lines correspond to a four-state Markov model and the dashed lines correspond to an eight-state model, with fading regions and transition probabilities computed as shown in Section 3.2. The conventional decoders with and without weighting and the ML decoder with perfect CSI do not use this channel approximation, so their BER performance is independent of the FSMC model.

We see that the BER of the conventional decoder in Rayleigh fading is approximately 0.1 over a large range of E_b/N_0 values: these results match those of [67], as do our simulation results for the ML decoder with perfect CSI. The two feedback decoders have similar performance for this channel as well, and their performance is not that sensitive to the channel approximation: a four-state channel model yields almost the same BER as the eight-state model. We found that increasing the number of channel states above eight has a negligible impact on the BER. Note that the feedback decoders outperform the conventional decoder without weighting by three orders of magnitude at high E_b/N_0 values. In this SNR range, the ML decoder with perfect CSI outperforms the feedback decoders by roughly an order of magnitude, with a smaller performance gap at lower SNRs. Weighting improves the BER performance of the conventional decoder by two orders of magnitude, so the performance improvement on the Rayleigh fading channel is significantly higher than on the two-state fading channel.

Figure 3.14 shows the simulated BER for differentially phase-encoded eight-state TC-16QAM signals applied to the same Rayleigh fading channel as discussed in Figure 3.13. Since in this case, phase distortion caused by the fading channel is not totally compensated

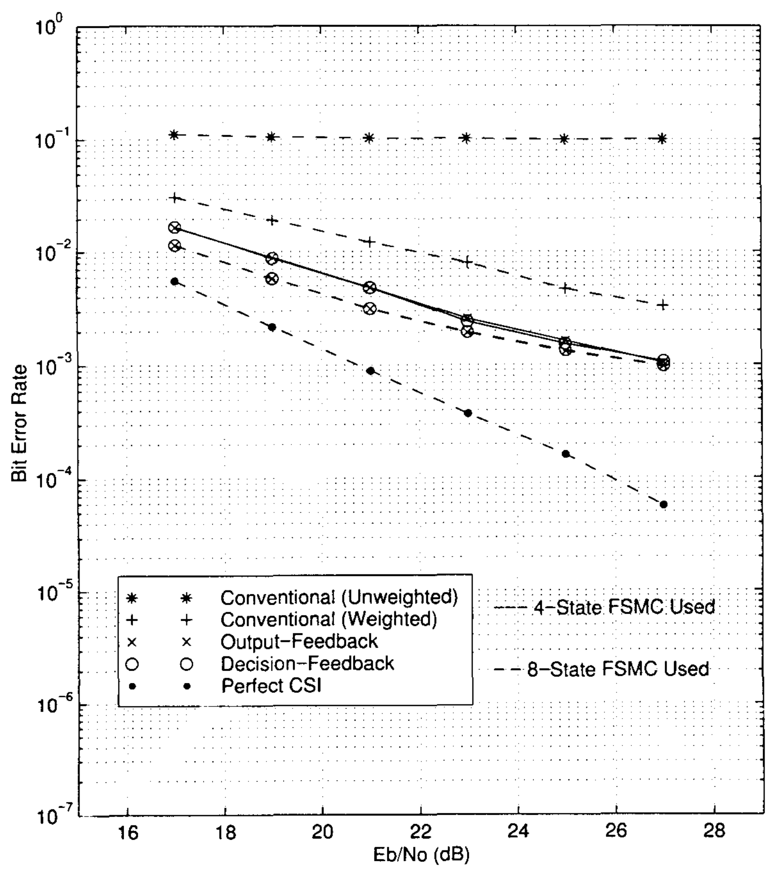


Figure 3.14: Simulated BER for differentially phase-encoded TC-16QAM (Rayleigh fading).

for by differential decoding of the signal phase, the BERs of all decoders except the conventional decoder have increased by around one order of magnitude compared to the case of perfect phase compensation. However, the relative behaviors of the five decoders are similar to that of the perfect phase compensation case in Figure 3.13. Although the two feedback decoders have similar performance in both cases, it is found in our simulations that the channel state estimator in the decision-feedback decoder may occasionally diverge (i.e., lose track of the fading channel) due to a burst of decision errors and whenever this happens, the simulated BER will be around 0.5. To prevent this divergence, we use a header of 3000 data bits at the beginning of each simulation. In practice, a known header may be required for each interleaving block of data in order to guarantee reliable communications. Since no feedback decisions are used in the output-feedback decoder, it does not have this divergent behavior and therefore does not need any header.

3.7 Conclusions

We propose two feedback decoders, a decision-feedback decoder and an output-feedback decoder, for coded signals transmitted over channels with correlated slow fading. These decoder structures recursively compute the channel fade distribution conditioned on past received symbols and, for the decision-feedback decoder, on past feedback decisions as well. We also propose a simple improvement to conventional decoders by using a weighted metric. The performance of each decoder is investigated by both analysis and simulation.

In our simulations, hard-decision as well as soft-decision on received symbols are considered and it is shown that all decoders exhibit a significant performance penalty when hard-decision is used, except for the conventional decoder, where hard-decision actually improves performance. The BER performance of all these decoders is analyzed through a sliding window decoding method. Both analysis and simulation demonstrate that the two feedback decoders have far better performance than the conventional decoder. Assuming ideal phase compensation for the Rayleigh fading channel, our results indicate that the BER of both feedback decoders comes within an order of magnitude of the performance of the ML sequence detector with perfect CSI, though the decision-feedback decoder does slightly worse on channels with low SNR due to error propagation, and slightly better on channels with high SNR. Moreover, these decoders reduce the BER relative to conventional techniques by several orders of magnitude. The weighting technique improves the BER performance of the conventional decoder by up to two orders of magnitude, with only a minor increase in complexity. With simple practical phase compensation method such as differential decoding, the two feedback decoders and the conventional decoder with weighting also yield significant performance improvement over the conventional decoder. The two feedback decoders have very similar performance. However, the decision-feedback decoder may not be reliable without a header during severe fading, while the output-feedback decoder is robust.

Chapter 4 Ergodic Capacity and Optimal Resource Allocation for Fading Broadcast Channels

4.1 Introduction

As discussed in Chapter 2, the wireless communication channel for both point-to-point and broadcast communications varies with time due to user mobility, which induces time-varying path loss, shadowing and multipath-fading in the received signal power. For these time-varying channels, dynamic allocation of resources such as power, rate, and bandwidth can result in better performance than fixed resource allocation strategies [68]-[71]. Indeed, adaptive techniques are currently used in both wireless and wireline systems and are being proposed as standards for next-generation cellular systems.

By using an optimal dynamic power and rate allocation strategy, the ergodic (Shannon) capacity of a single-user fading channel with channel side information (CSI) at both the transmitter and the receiver is obtained in [72]. The corresponding optimal power allocation strategy is a water-filling over time or, equivalently, over the fading states. This capacity corresponds to the maximum long-term achievable rate averaged over all states of the time-varying channel. In [2], assuming perfect CSI at the receiver and at all transmitters, the ergodic capacity[1] region of a fading multiple access channel (MAC) with Gaussian noise and the corresponding optimal power and rate allocation are obtained using the polymatroidal structure of the region. The optimal power allocation for the MAC is a generalization of water-filling for the single-user channel, and both the optimal rate and optimal power allocation in each joint fading state are derived via a greedy algorithm. The boundary of the ergodic capacity region can be achieved by successive decoding, as for the MAC with additive white Gaussian noise (AWGN).

In this chapter, we derive the ergodic capacity of an M-user fading broadcast channel with transmitter and receiver CSI and obtain the corresponding optimal resource allocation

[1]The Shannon capacity of a fading channel is called "throughput capacity" in [2].

strategy for code-division (CD) with and without successive decoding, time-division (TD), and frequency-division (FD). The optimal power allocation that achieves the boundary of the ergodic capacity region is derived by solving an optimization problem over a set of time-invariant AWGN broadcast channels with a total average transmit power constraint. For CD with successive decoding, the optimization problem is similar to that of the parallel Gaussian broadcast channels discussed in [73, 74], and is solved directly by applying the results therein. We solve the optimization problem for TD and show that one of the non-unique optimal power allocation strategies is also optimal for CD without successive decoding, and the ergodic capacity regions for these two techniques are the same. TD and FD are equivalent in the sense that they have the same ergodic capacity region and the optimal power allocation for one of them can be directly obtained from that of the other [75]. Thus, we obtain the optimal resource allocation for FD as well. For TD and CD without successive decoding we also propose a simple sub-optimal power allocation strategy that results in an ergodic rate region close to their capacity region.

The remainder of this chapter is organized as follows: the discrete-time fading broadcast channel model is presented in Section 4.2. The ergodic capacity regions and the optimal resource allocation for CD with and without successive decoding, TD, and FD, as well as the sub-optimal power and time allocations for TD are obtained in Section 4.3. Section 4.4 shows various numerical results, followed by our conclusions in the last section.

Notation. The prime ($'$) is used to denote the derivative of a function throughout this chapter and Appendix B except in the proofs about the convexity of a capacity region in Appendix B.1 and B.3, where the prime or double prime of a symbol just denotes another symbol.

4.2 The Fading Broadcast Channel

We consider a discrete-time M-user broadcast channel with fading as shown in Figure 4.1. In this model, the signal source $X[i]$ is composed of M independent information sources and the broadcast channel consists of M independent fading sub-channels. The time-varying sub-channel gains are denoted as $\sqrt{g_1[i]}, \sqrt{g_2[i]}, \ldots, \sqrt{g_M[i]}$, and the Gaussian noises of these sub-channels are denoted as $z_1[i], z_2[i], \ldots, z_M[i]$. Let $\bar{P}$ be the total average transmit power, B the received signal bandwidth, and ν_j the noise density of $z_j[i]$, $j = 1, 2, \ldots, M$. Since the time-varying received SNR $\gamma_j[i] = \bar{P}g_j[i]/(\nu_j B)$, $j = 1, 2, \ldots, M$, if we define[2] $n_j[i] = \nu_j/g_j[i]$ we have $\gamma_j[i] = \bar{P}/(n_j[i]B)$.

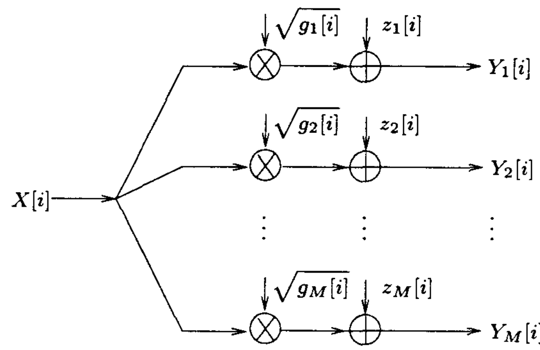


Figure 4.1: An M-user fading broadcast channel model.

Therefore, for slowly time-varying broadcast channel, we obtain an equivalent channel model, which is shown in Figure 4.2. In this model, the noise density of $z_j[i]/\sqrt{g_j[i]}$ is

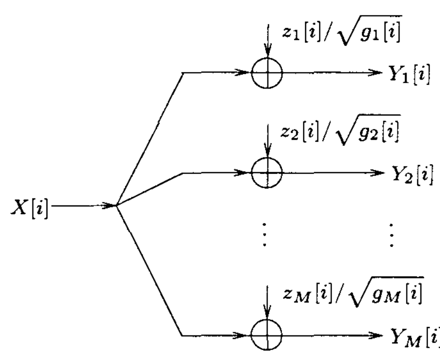


Figure 4.2: An equivalent M-user fading broadcast channel model.

$n_j[i]$, $j = 1, 2, \ldots, M$. We assume that $\{n_j[i]\}_{j=1}^{M}$ are known to the transmitter and all the M receivers at time i. Thus, the transmitter can vary the transmit power $P_j[i]$ for each user relative to the noise density vector $\mathbf{n}[i] = (n_1[i], n_2[i], \ldots, n_M[i])$, subject only to the average power constraint $\bar{P}$. For TD or FD, it can also vary the fraction of transmission time or bandwidth $\tau_j[i]$ assigned to each user, subject to the constraint $\sum_{j=1}^{M} \tau_j[i] = 1$ for all i. For CD, the superposition code can be varied at each transmission. Since every receiver

[2]Note that when $g_j[i] = 0$ for some j, $n_j[i] = \infty$ and no information can be transmitted through the jth sub-channel.

knows the noise density vector $\mathbf{n}[i]$, they can decode their individual signals by successive decoding based on the known resource allocation strategy given the M noise densities. In practice, it is necessary to send the transmitter strategy to each receiver through either a header on the transmitted data or a pilot tone. We call $\mathbf{n}[i]$ the joint fading process and denote $\mathcal{N}$ as the set of all possible joint fading states. $F(\mathbf{n})$ denotes a given cumulative distribution function (c.d.f.) on $\mathcal{N}$.

4.3 Ergodic Capacity Regions

Under the assumption that both the transmitters and the receiver have perfect CSI, the ergodic capacity region of a fading MAC is derived in [2] by exploiting its special polymatroidal structure. In that work the optimal resource allocation scheme is obtained by solving a family of optimization problems over a set of parallel Gaussian MACs, one for each fading state. In this section we derive the ergodic capacity region for the fading broadcast channel under the assumption that the transmitter and all receivers have perfect CSI. The corresponding optimal resource allocation strategy is obtained by optimizing over a set of parallel Gaussian broadcast channels for CD with and without successive decoding and for TD. For FD it is shown that the ergodic capacity region is the same as for TD and the corresponding optimal power and bandwidth allocation policy can be derived directly from that of TD [75].

4.3.1 CD

We first consider superposition coding and successive decoding where, in each joint fading state, the M-user broadcast channel can be viewed as a degraded Gaussian broadcast channel with noise densities $n_1[i], n_2[i], \ldots, n_M[i]$ and the multiresolution signal constellation is optimized relative to these instantaneous noise densities. Given a power allocation policy $\mathcal{P}$, let $P_j(\mathbf{n})$ be the transmit power allocated to User j for the joint fading state $\mathbf{n} = (n_1, n_2, \ldots, n_M)$ and denote $\mathcal{F}$ as the set of all possible power policies satisfying the average power constraint $E_{\mathbf{n}}\left[\sum_{j=1}^{M} P_j(\mathbf{n})\right] \leq \bar{P}$, where $E[\cdot]$ denotes the expectation function. For simplicity, assume that the stationary distributions of the fading processes have continuous densities[3], i.e., $Pr\{n_i = n_j\} = 0$, $\forall i \neq j$.

[3]If $Pr\{n_i = n_j\} \neq 0$ for some i, j then, in state $\mathbf{n}$, User i and User j can be viewed as a single user and superposition coding and successive decoding are applied to $M-1$ users. The information for User i and

Theorem 4.1 *The ergodic capacity region for the fading broadcast channel when the transmitter and all the receivers know the current channel state is given by:*

$$\mathcal{C}(\bar{P}) = \bigcup_{\mathcal{P}\in\mathcal{F}} \mathcal{C}_{CD}(\mathcal{P}), \tag{4.1}$$

where

$$\mathcal{C}_{CD}(\mathcal{P}) = \left\{ R_j \leq E_{\mathbf{n}}\left[B\log\left(1 + \frac{P_j(\mathbf{n})}{n_jB + \sum_{i=1}^{M} P_i(\mathbf{n})\mathbf{1}[n_j > n_i]}\right)\right],\ 1 \leq j \leq M \right\}, \tag{4.2}$$

and $\mathbf{1}[\cdot]$ *denotes the indicator function* ($\mathbf{1}[x] = 1$ *if* x *is true and zero otherwise). Moreover, the capacity region* $\mathcal{C}(\bar{P})$ *is convex.*

Proof: See Appendix B.1. □

Since this capacity region is convex, $\forall \boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_M) \in \Re_+^M$ with $\sum_{i=1}^M \mu_i = 1$, if a rate vector $\mathbf{R} = (R_1, R_2, \ldots, R_M)$ is a solution to the following maximization problem, it will be on the boundary surface of $\mathcal{C}(\bar{P})$ in (4.1):

$$\max_{\mathbf{R}\in\mathcal{C}(\bar{P})} \boldsymbol{\mu}\boldsymbol{R}. \tag{4.3}$$

The maximization problem in (4.3) is equivalent to

$$\begin{cases} \max\limits_{\mathbf{P}(\mathbf{n})} E_{\mathbf{n}}\left[J_0(\mathbf{P}(\mathbf{n}))\right] \\ \text{subject to: } E_{\mathbf{n}}\left[\sum_{j=1}^M P_j(\mathbf{n})\right] = \bar{P}, \end{cases} \tag{4.4}$$

where $\mathbf{P}(\mathbf{n}) = [P_1(\mathbf{n}), P_2(\mathbf{n}), \ldots, P_M(\mathbf{n})]$ and the objective function

$$J_0(\mathbf{P}(\mathbf{n})) = \sum_{j=1}^{M} \mu_j \ln\left(1 + \frac{P_j(\mathbf{n})}{n_jB + \sum_{i=1}^{M} P_i(\mathbf{n})\mathbf{1}[n_j > n_i]}\right) - \lambda \sum_{j=1}^{M} P_j(\mathbf{n}). \tag{4.5}$$

In (4.5), λ is the Lagrangian multiplier and μ_i can be viewed as a weighting parameter proportional to the priority of User i ($i = 1, 2, \ldots, M$). The problem in (4.4) is quite similar to that of the parallel AWGN broadcast channels discussed in [73, 74] and its solution is obtained by applying the results therein. For each fading state $\mathbf{n} = (n_1, n_2, \cdots, n_M)$, let the permutation $\pi(i)$ be defined such that $n_{\pi(1)} < n_{\pi(2)} < \cdots < n_{\pi(M)}$. The optimal power

User j are then transmitted by time-sharing the channel.

allocation procedure for state $\mathbf{n}$ as derived in [73] is essentially water-filling, which will be discussed later in this section in the context of a greedy algorithm. We now describe the optimal power allocation procedure from [73]:

Initialization: Do not assign power to any user $\pi(i)$ for which $\exists j$, $i < j \leq M$ with $\frac{\mu_{\pi(i)}}{n_{\pi(i)}} \leq \frac{\mu_{\pi(j)}}{n_{\pi(j)}}$. Remove these users from further consideration.

Step 1: Denote the number of remaining users as K and define the permutation $\rho(\cdot)$ such that

$$n_{\rho(1)} < n_{\rho(2)} < \cdots < n_{\rho(K)}.$$

Then, due to the removal criterion, we have

$$\frac{\mu_{\rho(1)}}{n_{\rho(1)}} > \frac{\mu_{\rho(2)}}{n_{\rho(2)}} > \cdots > \frac{\mu_{\rho(K)}}{n_{\rho(K)}}, \tag{4.6}$$

i.e.,

$$\frac{n_{\rho(1)}}{\mu_{\rho(1)}} < \frac{n_{\rho(2)}}{\mu_{\rho(2)}} < \cdots < \frac{n_{\rho(K)}}{\mu_{\rho(K)}}.$$

Step 2: Define

$$P^*_{\rho(1)}(\mathbf{n}) = \min\left\{\left[\frac{\mu_{\rho(1)}}{\lambda} - n_{\rho(1)}B\right]_+, \min_{\mu_{\rho(i)} > \mu_{\rho(1)}} \frac{\mu_{\rho(1)}n_{\rho(i)}B - \mu_{\rho(i)}n_{\rho(1)}B}{\mu_{\rho(i)} - \mu_{\rho(1)}}\right\}, \tag{4.7}$$

where $[x]_+ \stackrel{\triangle}{=} \max(x, 0)$ and assign power $P^*_{\rho(1)}(\mathbf{n})$ to User $\rho(1)$. If

$$P^*_{\rho(1)}(\mathbf{n}) = \left[\frac{\mu_{\rho(1)}}{\lambda} - n_{\rho(1)}B\right]_+,$$

the total power for state $\mathbf{n}$ has been allocated. If not, only the power for User $\rho(1)$ has been allocated. In this case, increase the noises $n_{\rho(i)}B$ $(1 < i \leq K)$ by $P^*_{\rho(1)}(\mathbf{n})$ and do not assign power to any user $\rho(i)$ for which $\exists j$ such that $i < j \leq K$ with $\frac{\mu_{\rho(i)}}{n_{\rho(i)}} \leq \frac{\mu_{\rho(j)}}{n_{\rho(j)}}$. Remove these users from further consideration. Also remove User $\rho(1)$ and return to Step 1. □

In the above procedure, $s^* = \frac{1}{\lambda}$ is the water-filling power level that satisfies the total power constraint in (4.4). In each iteration, this water-filling procedure consists of selecting the best receiver according to a modified noise criterion using the weighting parameter μ_i

for each user i, and adding power to the corresponding subchannel until a predetermined power is achieved. Note that in each iteration, some subchannels will be identified to hold no power. This optimal power allocation procedure can be found using a greedy algorithm [74]. In order to show this, we first require the following theorem, which is similar to *Theorem 3.14* in [2]:

Theorem 4.2 $\forall \boldsymbol{\xi} \in \Re_+^M$, $\mathbf{y} \in \Re_+^M$, $\lambda > 0$, *consider the maximization problem*

$$\max_{\mathbf{y}} \sum_{j=1}^{M} \xi_j x_j(\mathbf{y}) - \lambda \sum_{j=1}^{M} y_j, \tag{4.8}$$

where, for a set of non-decreasing concave functions $\{g_j(\cdot)\}_{j=1}^M$ *with* $\{g_j'(\cdot)\}_{j=1}^M$ *either concave or convex,*

$$\begin{cases} x_1(\mathbf{y}) = g_1(y_1) - g_1(0), \\ x_j(\mathbf{y}) = g_j(\sum_{i=1}^{j} y_i) - g_j(\sum_{i=1}^{j-1} y_i), \quad 2 \le j \le M. \end{cases} \tag{4.9}$$

Define the utility functions

$$\begin{aligned} u_j(z) &= \xi_j g_j'(z) - \lambda, \quad 1 \le j \le M. \\ u^*(z) &= \left[\max_{1 \le j \le M} u_j(z)\right]_+ . \end{aligned}$$

Then the solution to (4.8) is given by $\int_0^\infty u^*(z)dz$. *Moreover, let* $z_0 = 0$ *and let* z_k *(*$1 \le k \le K-1$*) denote the point where* $u_{i_k}(z)$ *intersects* $u_{i_{k+1}}(z)$ *for some* i_k *and* i_{k+1} *so that*[4]

$$u^*(z) = \begin{cases} u_{i_k}(z) & \text{if } z_{k-1} \le z \le z_k, \forall 1 \le k \le K, \\ 0 & \text{if } z \ge z_K. \end{cases} \tag{4.10}$$

Then the optimizing point $\mathbf{y}^*$ *is*

$$y_j^* = \begin{cases} z_k - z_{k-1}, & \text{if } j = i_k \text{ for some } k, 1 \le k \le K \\ 0, & \text{otherwise.} \end{cases} \tag{4.11}$$

[4]From (4.10) we see that z_K is defined as the smallest z for which $u^*(z) = 0$; if no such point exists, then $z_K = \infty$.

Proof: Let $y_0 \triangleq 0$ and by substituting (4.9) into (4.8), we have:

$$\begin{aligned}
& \max_{\mathbf{y}} \sum_{j=1}^{M} \xi_j x_j(\mathbf{y}) - \lambda \sum_{j=1}^{M} y_j \\
= & \max_{\mathbf{y}} \sum_{j=1}^{M} \xi_j \left[g_j(\sum_{k=0}^{j} y_k) - g_j(\sum_{k=0}^{j-1} y_k) \right] - \lambda \sum_{j=1}^{M} y_j \\
= & \max_{\mathbf{y}} \sum_{j=1}^{M} \int_{\sum_{k=0}^{j-1} y_k}^{\sum_{k=0}^{j} y_k} u_j(z) dz \\
\leq & \int_0^{\infty} u^*(z) dz. \qquad (4.12)
\end{aligned}$$

Note that since the functions $\{g_j(\cdot)\}_{j=1}^M$ are concave, $\forall 1 \leq j \leq M$, $g_j'(\cdot)$ is monotonically decreasing and hence so are $\{u_j(\cdot)\}_{j=1}^M$ and $u^*(\cdot)$. Moreover, since $\{g_j'(\cdot)\}_{j=1}^M$ are either concave or convex, for any i, j, two curves $u_i(\cdot)$ and $u_j(\cdot)$ can cross each other at most once. Thus, the set of cross points $\{z_k\}_{k=1}^K$ for the curves $\{u_j(\cdot)\}_{j=1}^M$ satisfying the condition that if and only if $z \in [z_{k-1}, z_k]$, $u^*(z) = u_{i_k}(z)$ for some i_k $(1 \leq k \leq K)$ is unique and it is easy to verify that point $\mathbf{y}^*$ given in (4.11) achieves the upper bound in (4.12). □

We now apply *Theorem 4.2* to the maximization problem (4.4). Recall that in a given state $\mathbf{n}$, we have defined the permutation $\pi(\cdot)$ such that $n_{\pi(1)} < n_{\pi(2)} < \cdots < n_{\pi(M)}$. Now for $1 \leq j \leq M$, let $y_j \triangleq P_{\pi(j)}(\mathbf{n})$, $\xi_j \triangleq \mu_{\pi(j)}$, and

$$g_j(z) \triangleq \ln(n_{\pi(j)} B + z). \qquad (4.13)$$

With these definitions, $x_j(\mathbf{y})$ $(1 \leq j \leq M)$ in *Theorem 4.2* becomes

$$\begin{aligned}
x_1(\mathbf{y}) &= g_1(y_1) - g_1(0) \\
&= \ln\left[1 + \frac{P_{\pi(1)}(\mathbf{n})}{n_{\pi(1)} B}\right],
\end{aligned}$$

$$\begin{aligned}
x_j(\mathbf{y}) &= g_j(\sum_{i=1}^{j} y_i) - g_j(\sum_{i=1}^{j-1} y_i) \\
&= \ln\left[1 + \frac{P_{\pi(j)}(\mathbf{n})}{n_{\pi(j)} B + \sum_{i=1}^{j-1} P_{\pi(i)}(\mathbf{n})}\right], \quad 2 \leq j \leq M, \qquad (4.14)
\end{aligned}$$

and the utility function $u_j(z)$ $(1 \leq j \leq M)$ becomes

$$\begin{aligned} u_j(z) &= \xi_j g_j'(z) - \lambda \\ &= \frac{\mu_{\pi(j)}}{n_{\pi(j)}B + z} - \lambda, \quad 1 \leq j \leq M. \end{aligned} \tag{4.15}$$

Thus the maximization problem (4.8) is equivalent to the maximization problem (4.4). Consequently, by *Theorem 4.2*, the solution y_j^* to (4.8) is the power allocation for User $\pi(j)$ that maximizes (4.4) for state $\mathbf{n}$. That is,

$$P^*_{\pi(j)}(\mathbf{n}) = \begin{cases} z_k - z_{k-1}, & \text{if } j = i_k \text{ for some } k, 1 \leq k \leq K \\ 0, & \text{otherwise,} \end{cases}$$

where z_k $(0 \leq k \leq K)$ is as defined in *Theorem 4.2*. This optimal solution to (4.4) can be interpreted as follows.

Assuming that in fading state $\mathbf{n}$, the interference to each user $\pi(j)$ $(1 \leq j \leq M)$ from other users is z $(z \geq 0)$, i.e.,

$$z = \sum_{i=1}^{j-1} P_{\pi(i)}(\mathbf{n}),$$

then the objective function $J_0(\mathbf{P}(\mathbf{n}))$ in (4.4) becomes

$$J_0(\mathbf{P}(\mathbf{n})) = \sum_{j=1}^{M} \mu_{\pi(j)} \ln\left(1 + \frac{P_{\pi(j)}(\mathbf{n})}{n_{\pi(j)}B + z}\right) - \lambda \sum_{j=1}^{M} P_{\pi(j)}(\mathbf{n}). \tag{4.16}$$

Therefore, when a marginal power δp is allocated to User $\pi(j)$ at interference level z, the marginal weighted rate increase in the objective function $J_0(\mathbf{P}(\mathbf{n}))$ is

$$\begin{aligned} & \left.\frac{\partial J_0(\mathbf{P}(\mathbf{n}))}{\partial P_{\pi(j)}(\mathbf{n})}\right|_{P_{\pi(j)}(\mathbf{n})=0} \cdot \delta p \\ = & \left\{ \left.\frac{\mu_{\pi(j)}}{n_{\pi(j)}B + z + P_{\pi(j)}(\mathbf{n})}\right|_{P_{\pi(j)}(\mathbf{n})=0} - \lambda \right\} \cdot \delta p \\ = & \; u_j(z) \cdot \delta p, \end{aligned}$$

where the utility function $u_j(z)$ is given in (4.15). Thus, *Theorem 4.2* indicates that the optimal solution to (4.4) can be obtained greedily by allocating power δp to the User i^* with the largest positive marginal weighted rate increase $u^*(z) \cdot \delta p = \left(\frac{\mu_{i^*}}{n_{i^*}B+z} - \lambda\right) \cdot \delta p$

at each interference level z ($z \geq 0$). This power allocation process continues until no user obtains a positive marginal weighted rate increase with the addition of power δp (i.e., $u^*(z) = \left[\max_{1\leq j\leq M} u_j(z)\right]_+ = 0$), in which case we allocate no more power to state $\mathbf{n}$.

For example, in a two-user system, let $u_j(z) \triangleq \frac{\mu_j}{n_jB+z} - \lambda$, $j = 1,2$. In order that $\max_{j\in\{1,2\}}\{u_j(z)\} > 0$, z must be no larger than $\max_{j\in\{1,2\}}\{\frac{\mu_j}{\lambda} - n_jB\}$. Assuming that $\mu_1 < \mu_2$, we have:

1. if $\frac{\mu_1}{n_1} \leq \frac{\mu_2}{n_2}$, then $\forall z > 0$, $u_1(z) < u_2(z)$, which means that no power should be assigned to User 1. Therefore, the optimal power $\mathbf{P}^*(\mathbf{n})$ is:

$$\begin{cases} P_1^*(\mathbf{n}) = 0, \\ P_2^*(\mathbf{n}) = \left[\frac{\mu_2}{\lambda} - n_2B\right]_+; \end{cases}$$

2. if $\frac{\mu_1}{n_1} > \frac{\mu_2}{n_2}$, $u_1(z)$ and $u_2(z)$ will cross each other once at $z_c \triangleq \frac{\mu_1 n_2 B - \mu_2 n_1 B}{\mu_2 - \mu_1}$. In this case,

   (a) if $u_1(z_c) = u_2(z_c) > 0$, i.e., $\lambda < \frac{\mu_2-\mu_1}{n_2B-n_1B}$, then as shown in Figure 4.3,

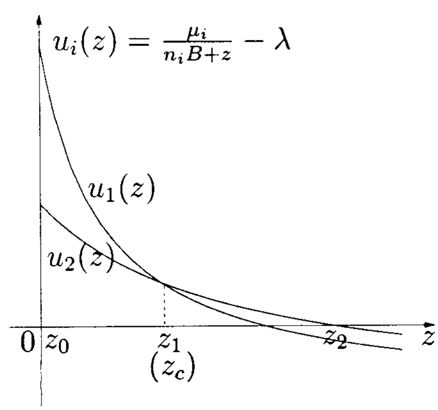


Figure 4.3: The functions $u_1(z)$ and $u_2(z)$ when $\frac{\mu_1}{n_1} > \frac{\mu_2}{n_2}$ and $\lambda < \frac{\mu_2-\mu_1}{n_2B-n_1B}$.

$$\begin{cases} u_1(z) > u_2(z), & \text{for } z_0 < z < z_1, \\ u_1(z) < u_2(z), & \text{for } z_1 < z < z_2, \end{cases}$$

where $z_0 = 0$, $z_1 = z_c$, $z_2 = \frac{\mu_2}{\lambda} - n_2 B$. Thus,

$$u^*(z) = \begin{cases} u_1(z), & \text{for } z_0 < z < z_1, \\ u_2(z), & \text{for } z_1 < z < z_2, \\ 0, & \text{for } z \geq z_2. \end{cases}$$

Therefore, according to *Theorem 4.2*,

$$\begin{cases} P_1^*(\mathbf{n}) = z_1 - z_0 = \frac{\mu_1 n_2 B - \mu_2 n_1 B}{\mu_2 - \mu_1}, \\ P_2^*(\mathbf{n}) = z_2 - z_1 = \left[\frac{\mu_2}{\lambda} - n_2 B - \frac{\mu_1 n_2 B - \mu_2 n_1 B}{\mu_2 - \mu_1}\right]_+ . \end{cases}$$

(b) if $u_1(z_c) = u_2(z_c) \leq 0$, i.e., $\lambda \geq \frac{\mu_2 - \mu_1}{n_2 B - n_1 B}$, then as shown in Figure 4.4,

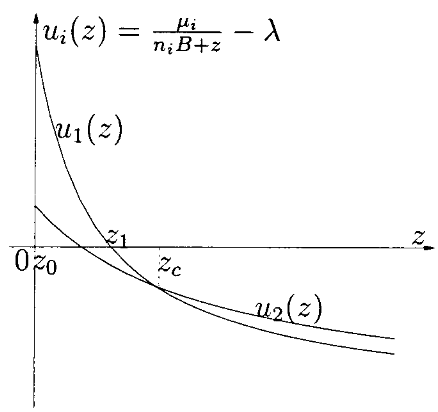


Figure 4.4: The functions $u_1(z)$ and $u_2(z)$ when $\frac{\mu_1}{n_1} > \frac{\mu_2}{n_2}$ and $\lambda \geq \frac{\mu_2 - \mu_1}{n_2 B - n_1 B}$.

$$u_1(z) > u_2(z), \qquad \text{for } z_0 < z < z_1,$$

where $z_0 = 0$ and $z_1 = \frac{\mu_1}{\lambda} - n_1 B$. Thus,

$$u^*(z) = \begin{cases} u_1(z), & \text{for } z_0 < z < z_1, \\ 0, & \text{for } z \geq z_1. \end{cases}$$

and

$$\begin{cases} P_1^*(\mathbf{n}) = z_1 - z_0 = \left[\frac{\mu_1}{\lambda} - n_1 B\right]_+ , \\ P_2^*(\mathbf{n}) = 0. \end{cases}$$

When $\mu_1 > \mu_2$, the power policy for User 1 and User 2 is similarly obtained using appropriate substitutions for all subscripts. When $\mu_1 = \mu_2$, the power policy is simplified to the following: if $n_1 \leq n_2$, then $P_1^*(\mathbf{n}) = \left[\frac{\mu_1}{\lambda} - n_1 B\right]_+$, $P_2^*(\mathbf{n}) = 0$; otherwise $P_1^*(\mathbf{n}) = 0$, $P_2^*(\mathbf{n}) = \left[\frac{\mu_2}{\lambda} - n_2 B\right]_+$.

Therefore, from the discussion for the two-user system, one can easily see that the two-step water-filling procedure described earlier in this section is a concise summarization of the M-user optimal power allocation for CD derived from *Theorem 4.2*. That is, in each iteration, according to (4.6), User $\rho(1)$ has the largest positive marginal weighted rate increase and the procedure starts to assign power to User $\rho(1)$ by water-filling until the total power allocated to him satisfies (4.7), in which case if

$$P_{\rho(1)}^*(\mathbf{n}) = \min_{\mu_{\rho(i)} > \mu_{\rho(1)}} \frac{\mu_{\rho(1)} n_{\rho(i)} B - \mu_{\rho(i)} n_{\rho(1)} B}{\mu_{\rho(i)} - \mu_{\rho(1)}},$$

then some user other than User $\rho(1)$ will begin to have the largest positive marginal weighted rate increase and the available power will be allocated to him in the next iteration; if $P_{\rho(1)}^*(\mathbf{n}) = \left[\frac{\mu_{\rho(1)}}{\lambda} - n_{\rho(1)} B\right]_+$, then no user will have a positive marginal weighted rate increase with any additional power and no more power will be allocated to state $\mathbf{n}$. A detailed interpretation about this is given in Appendix B.2.

4.3.2 TD

Now we consider the time-division case where, in each fading state $\mathbf{n}$, the information for the M users will be divided and sent in time-slots which are functions of $\mathbf{n}$. For a given power and time allocation policy $\mathcal{P}$, let $P_j(\mathbf{n})$ and $\tau_j(\mathbf{n})$ $(0 \leq \tau_j(\mathbf{n}) \leq 1)$ be the transmit power and fraction of transmission time allocated to User j $(j = 1, 2, \ldots, M)$, respectively, for fading state $\mathbf{n}$, and let $\mathcal{F}$ be the set of all such possible power and time allocation policies satisfying

$$\begin{cases} E_{\mathbf{n}}\left[\sum_{j=1}^{M} \tau_j(\mathbf{n}) P_j(\mathbf{n})\right] \leq \bar{P}, & \text{and} \\ \sum_{j=1}^{M} \tau_j(\mathbf{n}) = 1, & \forall \mathbf{n} \in \mathcal{N}. \end{cases} \tag{4.17}$$

Theorem 4.3 *The achievable rate region for the variable power and variable transmission time scheme is:*

$$\mathcal{C}(\bar{P}) = \bigcup_{\mathcal{P} \in \mathcal{F}} \mathcal{C}_{TD}(\mathcal{P}), \tag{4.18}$$

where

$$\mathcal{C}_{TD}(\mathcal{P}) = \left\{ R_j \leq E_{\mathbf{n}} \left[\tau_j(\mathbf{n}) B \log \left(1 + \frac{P_j(\mathbf{n})}{n_j B} \right) \right], j = 1, 2, \cdots, M \right\}. \tag{4.19}$$

Moreover, the rate region $\mathcal{C}(\bar{P})$ is convex.

Proof: See Appendix B.3. □

Note that in this chapter we will refer to this achievable rate region as the capacity region for TD, though we do not have a converse proof due to the fact that the converse only holds for the optimal transmission strategy for this channel, which, according to *Theorem 4.1*, is CD with successive decoding.

Due to the convexity of this capacity region, $\forall \boldsymbol{\mu} \in \Re_+^M$ with $\sum_{j=1}^M \mu_j = 1$, if a rate vector $\mathbf{R}$ is a solution to

$$\max_{\mathbf{R} \in \mathcal{C}(\bar{P})} \boldsymbol{\mu R}, \tag{4.20}$$

it will be on the boundary surface of $\mathcal{C}(\bar{P})$. Based on the expression for $\mathbf{R}$ in (4.19) and the average total power constraint in (4.17), we can decompose the maximization problem (4.20) into the following two problems:

(1) Assuming that $\forall \mathbf{n} \in \mathcal{N}$, $P(\mathbf{n})$ is the total power assigned to the M users, i.e., $P(\mathbf{n}) = \sum_{j=1}^M \tau_j(\mathbf{n}) P_j(\mathbf{n})$, we must determine how to distribute $P(\mathbf{n})$ among the M users so that the total weighted rate in state $\mathbf{n}$ is maximized. That is, we must find

$$\begin{cases} J(P(\mathbf{n})) \stackrel{\triangle}{=} \max_{\boldsymbol{\tau}(\mathbf{n}), \mathbf{P}(\mathbf{n})} \sum_{j=1}^M \tau_j(\mathbf{n}) f_j(P_j(\mathbf{n})) \\ \text{subject to:} \quad \sum_{j=1}^M \tau_j(\mathbf{n}) P_j(\mathbf{n}) = P(\mathbf{n}), \end{cases} \tag{4.21}$$

where $\boldsymbol{\tau}(\mathbf{n}) = [\tau_1(\mathbf{n}), \tau_2(\mathbf{n}), \cdots, \tau_M(\mathbf{n})]$ with $\sum_{j=1}^M \tau_j(\mathbf{n}) = 1$, $\mathbf{P}(\mathbf{n}) = [P_1(\mathbf{n}), P_2(\mathbf{n}), \cdots, P_M(\mathbf{n})]$, and

$$f_j(x) \stackrel{\triangle}{=} \mu_j \ln \left[1 + \frac{x}{n_j B} \right]; \tag{4.22}$$

(2) After we obtain the expression for $J(\cdot)$ by solving (4.21), the remaining problem is how to assign the total power $P(\mathbf{n})$ of the M users for each state $\mathbf{n}$ so that the total weighted rate averaged over all fading states as expressed in (4.20) is maximized. That is,

$$\begin{cases} \max_{P(\mathbf{n})} \{ E_{\mathbf{n}}[J(P(\mathbf{n}))] - \lambda E_{\mathbf{n}}[P(\mathbf{n})] \} \\ \text{subject to:} \quad E_{\mathbf{n}}[P(\mathbf{n})] = \bar{P}, \end{cases} \tag{4.23}$$

where λ is the Lagrangian multiplier.

We solve the maximization problems (4.21) and (4.23) for the two-user case first and then generalize the result to the M-user case.

Two-User Case

Lemma 4.1 *When $M = 2$, assuming that $\mu_1 < \mu_2$, the solution to (4.21) is:*

1. *if $\frac{\mu_1}{n_1} \leq \frac{\mu_2}{n_2}$, then $J(P(\mathbf{n})) = f_2(P(\mathbf{n}))$, which is achieved when $\tau_1(\mathbf{n}) = 0$, $\tau_2(\mathbf{n}) = 1$, $P_1(\mathbf{n}) = 0$, $P_2(\mathbf{n}) = P(\mathbf{n})$;*

2. *if $\frac{\mu_1}{n_1} > \frac{\mu_2}{n_2}$, let*

$$h(x, \mathbf{n}) \triangleq (n_2 - n_1)Bx + \left(\mu_2 \ln \frac{\mu_2}{xn_2B} - \mu_1 \ln \frac{\mu_1}{xn_1B}\right) - (\mu_2 - \mu_1). \tag{4.24}$$

Then

$$J(P(\mathbf{n})) = \begin{cases} f_1(P(\mathbf{n})) & \text{if } 0 < P(\mathbf{n}) \leq P_a(\mathbf{n}), \\ f_1(P_a(\mathbf{n})) + \lambda_0(\mathbf{n})[P(\mathbf{n}) - P_a(\mathbf{n})] & \text{if } P_a(\mathbf{n}) < P(\mathbf{n}) < P_b(\mathbf{n}), \\ f_2(P(\mathbf{n})) & \text{if } P(\mathbf{n}) \geq P_b(\mathbf{n}), \end{cases} \tag{4.25}$$

where $P_a(\mathbf{n}) = \frac{\mu_1}{\lambda_0(\mathbf{n})} - n_1B$, $P_b(\mathbf{n}) = \frac{\mu_1}{\lambda_0(\mathbf{n})} - n_2B$, and $\lambda_0(\mathbf{n})$ satisfies

$$h(\lambda_0(\mathbf{n}), \mathbf{n}) = 0. \tag{4.26}$$

$J(P(\mathbf{n}))$ in (4.25) is achieved by letting

$$\left.\begin{aligned} \tau_1(\mathbf{n}) &= 1 \\ \tau_2(\mathbf{n}) &= 0 \\ P_1(\mathbf{n}) &= P(\mathbf{n}) \\ P_2(\mathbf{n}) &= 0 \end{aligned}\right\} \text{if } 0 < P(\mathbf{n}) \leq P_a(\mathbf{n}), \tag{4.27}$$

$$\left.\begin{aligned} \tau_1(\mathbf{n}) &= \tfrac{P_b(\mathbf{n})-P(\mathbf{n})}{P_b(\mathbf{n})-P_a(\mathbf{n})} \\ \tau_2(\mathbf{n}) &= \tfrac{P(\mathbf{n})-P_a(\mathbf{n})}{P_b(\mathbf{n})-P_a(\mathbf{n})} \\ P_1(\mathbf{n}) &= P_a(\mathbf{n}) \\ P_2(\mathbf{n}) &= P_b(\mathbf{n}) \end{aligned}\right\} \textit{if } P_a(\mathbf{n}) < P(\mathbf{n}) < P_b(\mathbf{n}), \tag{4.28}$$

$$\left.\begin{aligned} \tau_1(\mathbf{n}) &= 0 \\ \tau_2(\mathbf{n}) &= 1 \\ P_1(\mathbf{n}) &= 0 \\ P_2(\mathbf{n}) &= P(\mathbf{n}) \end{aligned}\right\} \textit{if } P(\mathbf{n}) \geq P_b(\mathbf{n}). \tag{4.29}$$

Proof. See Appendix B.4. □

From (4.21) we see that when $M = 2$, $J(P(\mathbf{n}))$ is a linear combination of $f_j(\cdot)$, $j = 1, 2$. The proof of *Lemma 4.1* in Appendix B.4 shows that if $\frac{\mu_1}{n_1} \leq \frac{\mu_2}{n_2}$, then $\forall P(\mathbf{n}) > 0$, $f_1(P(\mathbf{n})) < f_2(P(\mathbf{n}))$ and $J(P(\mathbf{n}))$ is simply $f_2(P(\mathbf{n}))$. If $\frac{\mu_1}{n_1} > \frac{\mu_2}{n_2}$, it is proved that for $P(\mathbf{n}) > 0$, $f_1(P(\mathbf{n}))$ and $f_2(P(\mathbf{n}))$ will cross each other once at some positive value $P_0(\mathbf{n})$, as shown in Figure 4.5[5]. In this figure, the slope of the tangent line between the curves $f_1(P(\mathbf{n}))$ and $f_2(P(\mathbf{n}))$ is $\lambda_0(\mathbf{n})$ and it satisfies

$$f_1'(P_a(\mathbf{n})) = f_2'(P_b(\mathbf{n})) = \lambda_0(\mathbf{n}), \tag{4.30}$$

i.e.,

$$h(\lambda_0(\mathbf{n}), \mathbf{n}) = 0.$$

Thus, in this case, $J(P(\mathbf{n}))$ is the continuous contour in Figure 4.5 which consists of part of the curve $f_1(P(\mathbf{n}))$, the tangent line, and part of the curve $f_2(P(\mathbf{n}))$, as indicated with the dash-dotted line which is offset slightly for clarity. The expression of $J(P(\mathbf{n}))$ is therefore as given in (4.25). Note that the slope of the tangent of the curve $J(P(\mathbf{n}))$ is continuous and it decreases with the increase of $P(\mathbf{n})$.

For a given fading state $\mathbf{n}$, from (4.23) we know that the optimal power $P^*(\mathbf{n})$ satisfies

$$J'(P^*(\mathbf{n})) - \lambda = 0. \tag{4.31}$$

[5]In this figure, P, P_a, P_0 and P_b are all functions of $\mathbf{n}$, which is not shown for simplicity.

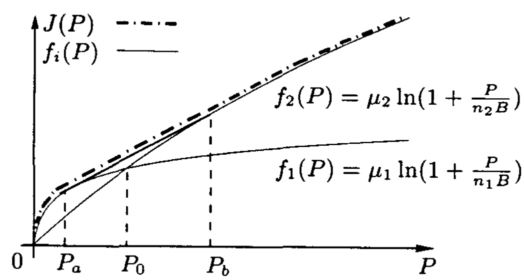


Figure 4.5: The functions $f_1(P)$, $f_2(P)$ and $J(P)$ when $\frac{\mu_1}{n_1} > \frac{\mu_2}{n_2}$.

Therefore, for any given $\lambda > 0$, $P^*(\mathbf{n})$ is determined by the point(s) on the curve $J(P(\mathbf{n}))$ whose tangent has a slope λ. In the case where $\frac{\mu_1}{n_1} > \frac{\mu_2}{n_2}$ and $\lambda = \lambda_0(\mathbf{n})$, since all the points on the tangent line between $f_1(P(\mathbf{n}))$ and $f_2(P(\mathbf{n}))$ in Figure 4.5 have the same tangent slope λ, $P^*(\mathbf{n})$ can be any value between $P_a(\mathbf{n})$ and $P_b(\mathbf{n})$: if $P_a(\mathbf{n}) < P^*(\mathbf{n}) < P_b(\mathbf{n})$, from *Lemma 4.1* we know that $P^*(\mathbf{n})$ will be time-shared by the two users; if $P^*(\mathbf{n})$ is simply chosen as $P_a(\mathbf{n})$ or $P_b(\mathbf{n})$, then it is only assigned to User 1 or to User 2, respectively. In all other cases, the point that has a tangent with slope λ is unique and hence so is $P^*(\mathbf{n})$. The unique choice of $P^*(\mathbf{n})$ is then allocated to a single user based on its relative value compared to $P_a(\mathbf{n})$ and $P_b(\mathbf{n})$ as discussed in *Lemma 4.1*. Consequently, we have the following theorem:

Theorem 4.4 *When $M = 2$, assuming that $\mu_1 < \mu_2$, the optimal power and time allocation policy that achieves the TD capacity region boundary for each fading state $\mathbf{n} = (n_1, n_2)$ is:*

1. if $\frac{\mu_1}{n_1} \leq \frac{\mu_2}{n_2}$, then

$$\begin{cases} \tau_1^*(\mathbf{n}) = 0, \\ \tau_2^*(\mathbf{n}) = 1, \end{cases} \quad \textit{and} \quad \begin{cases} P_1^*(\mathbf{n}) = 0, \\ P_2^*(\mathbf{n}) = \left[\frac{\mu_2}{\lambda} - n_2 B\right]_+; \end{cases}$$

2. if $\frac{\mu_1}{n_1} > \frac{\mu_2}{n_2}$, then

(a) if $\lambda \geq \frac{\mu_2 - \mu_1}{n_2 B - n_1 B}$ or if $\lambda < \frac{\mu_2 - \mu_1}{n_2 B - n_1 B}$ and $h(\lambda, \mathbf{n}) < 0$,

$$\begin{cases} \tau_1^*(\mathbf{n}) = 1, \\ \tau_2^*(\mathbf{n}) = 0, \end{cases} \quad \textit{and} \quad \begin{cases} P_1^*(\mathbf{n}) = \left[\frac{\mu_1}{\lambda} - n_1 B\right]_+, \\ P_2^*(\mathbf{n}) = 0; \end{cases}$$

(b) if $\lambda < \frac{\mu_2 - \mu_1}{n_2 B - n_1 B}$ *and* $h(\lambda, \mathbf{n}) > 0$,

$$\begin{cases} \tau_1^*(\mathbf{n}) = 0, \\ \tau_2^*(\mathbf{n}) = 1, \end{cases} \quad and \quad \begin{cases} P_1^*(\mathbf{n}) = 0, \\ P_2^*(\mathbf{n}) = \left[\frac{\mu_2}{\lambda} - n_2 B\right]_+ ; \end{cases}$$

(c) if $\lambda < \frac{\mu_2 - \mu_1}{n_2 B - n_1 B}$ *and* $h(\lambda, \mathbf{n}) = 0$,

$$\begin{cases} \tau_1^*(\mathbf{n}) = \tau_0^*, \\ \tau_2^*(\mathbf{n}) = 1 - \tau_0^*, \end{cases} \quad and \quad \begin{cases} P_1^*(\mathbf{n}) = \left[\frac{\mu_1}{\lambda} - n_1 B\right]_+ , \\ P_2^*(\mathbf{n}) = \left[\frac{\mu_2}{\lambda} - n_2 B\right]_+ . \end{cases}$$

In the above expressions, $h(x, \mathbf{n})$ *is given in (4.24) and* $s^* \triangleq \frac{1}{\lambda}$ *is the water-filling power level.* λ *and* τ_0^* *satisfy the total average power constraint:*

$$E_{\mathbf{n}} \left[\sum_{i=1}^{2} \tau_i^*(\mathbf{n}) P_i^*(\mathbf{n}) \right] = \bar{P}, \tag{4.32}$$

and they may not be unique.

Proof. See Appendix B.5. □

When $\mu_1 > \mu_2$, the optimal power policy for User 1 and User 2 is similarly derived using appropriate substitutions for all subscripts. When $\mu_1 = \mu_2$ the optimal power policy is simplified as follows: if $n_1 \leq n_2$ then $P_1^*(\mathbf{n}) = \left[\frac{\mu_1}{\lambda} - n_1 B\right]_+$, $P_2^*(\mathbf{n}) = 0$; otherwise $P_1^*(\mathbf{n}) = 0$, $P_2^*(\mathbf{n}) = \left[\frac{\mu_2}{\lambda} - n_2 B\right]_+$. In fact this is the same power allocation as for CD discussed in Section 4.3.1.

Note that when the c.d.f. $F(\mathbf{n})$ is continuous, $\forall \lambda > 0$, the probability measure of the set

$$L \triangleq \{\mathbf{n} : h(\lambda, \mathbf{n}) = 0\} \tag{4.33}$$

is zero. Since according to *Theorem 4.4*, τ_0^* is defined on a subset of L and the probability measure of any subset of L must also be zero, the value of τ_0^* will not affect the power constraint (4.32) and λ is therefore uniquely determined by (4.32). Moreover, in this case, with probability 1, at most a single user transmits in each fading state $\mathbf{n}$. If $F(\mathbf{n})$ is not continuous, the set L may have positive probability measure and for any fading state in this set, the broadcast channel can be either time-shared by the two users, occupied by a single user, or not used by any user if the fading is too severe.

<u>*M*-User Case</u>

The optimal power and time allocation policy that achieves the capacity region boundary for the two-user case can be generalized to the M-user case ($M > 2$). In the M-user case, the optimal power allocation is again obtained based on the values of functions $f_j(\cdot)$, $j = 1, 2, \cdots, M$. Specifically, for a given $f_j(\cdot)$ in (4.22), if $\exists i \neq j$ such that $\forall x > 0$, $f_j(x) \leq f_i(x)$, then we can show that $f_j(\cdot)$ will not appear in the expression of $J(P(\mathbf{n}))$, $\forall P(\mathbf{n}) > 0$. Thus, no resources should be assigned to User j in the state $\mathbf{n}$. That is, the optimal $\tau_j^*(\mathbf{n})$ and $P_j^*(\mathbf{n})$ are $\tau_j^*(\mathbf{n}) = 0$, $P_j^*(\mathbf{n}) = 0$. For any $i, j = 1, 2, \ldots, M$, assuming $\mu_j < \mu_i$, we know that, as shown in the proof of *Lemma 4.1*, if $\frac{\mu_j}{n_j} \leq \frac{\mu_i}{n_i}$ then $\forall x > 0$, $f_j(x) < f_i(x)$; if $\frac{\mu_j}{n_j} > \frac{\mu_i}{n_i}$ then $f_i(x)$ and $f_j(x)$ will cross each other once at some positive x. In both cases, since $\mu_j < \mu_i$, for x large enough ($x > 0$), $f_j(x) < f_i(x)$.

Thus, assuming without loss of generality (WLOG) that $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_M$, we first remove any user j (i.e., let $\tau_j^*(\mathbf{n}) = 0$, $P_j^*(\mathbf{n}) = 0$) for which $\exists i$, $j < i \leq M$ with $\frac{\mu_j}{n_j} \leq \frac{\mu_i}{n_i}$ or which satisfies $\mu_{j-1} = \mu_j$ and $\frac{\mu_j}{n_j} < \frac{\mu_{j-1}}{n_{j-1}}$. For the remaining users, there are still some users j whose corresponding $f_j(\cdot)$ may not appear in the expression of $J(P(\mathbf{n}))$. For example, assume WLOG that the remaining users are User 1-User 4. Due to the removal criterion, we know that

$$\frac{\mu_1}{n_1} > \frac{\mu_2}{n_2} > \frac{\mu_3}{n_3} > \frac{\mu_4}{n_4}, \tag{4.34}$$

and

$$\mu_1 < \mu_2 < \mu_3 < \mu_4, \tag{4.35}$$

which means that their corresponding $f_j(P(\mathbf{n}))$ ($1 \leq j \leq 4$) will cross one another once at some positive $P(\mathbf{n})$, and

$$f_1(P(\mathbf{n})) < f_2(P(\mathbf{n})) < f_3(P(\mathbf{n})) < f_4(P(\mathbf{n}))$$

for $P(\mathbf{n})$ large enough. Thus, if $\{f_j(P(\mathbf{n}))\}_{j=1}^4$ are as shown in Figure 4.6[6], it is clear that $f_2(\cdot)$ will not appear in the expression of $J(P(\mathbf{n}))$ since, $\forall P(\mathbf{n}) > 0$, the curve $f_2(P(\mathbf{n}))$ is always under the dash-dotted contour of $J(P(\mathbf{n}))$ formed by part of the curves $f_1(P(\mathbf{n}))$, $f_3(P(\mathbf{n}))$, $f_4(P(\mathbf{n}))$ and the straight tangent lines between them. Note that the dash-dotted curve in this figure is offset slightly for clarity. In the following, we use an iterative procedure

[6]In this figure, P, P_{a_j} and P_{b_j} ($j = 1, 2$) are all functions of $\mathbf{n}$, which is not shown for simplicity.

to find all the users j among the remaining users whose corresponding $f_j(\cdot)$ will not appear in the expression of $J(P(\mathbf{n}))$ and identify all other users to whom the resources will be allocated later. An interpretation of this procedure based on Figure 4.6 will then be given.

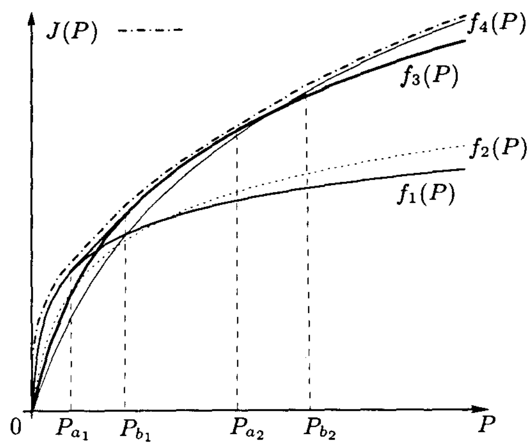


Figure 4.6: The functions $J(P)$ and $f_j(P) = \mu_j \ln\left[1 + \frac{P}{n_j B}\right]$ $(1 \leq j \leq 4)$ which satisfy (4.34) and (4.35).

Initialization: Let $m = 1$.

Step 1: Denote the number of remaining users as K and define the permutation $\psi(\cdot)$ such that

$$\mu_{\psi(1)} < \mu_{\psi(2)} < \cdots < \mu_{\psi(K)}.$$

Then, due to the removal criterion, we have

$$\frac{\mu_{\psi(1)}}{n_{\psi(1)}} > \frac{\mu_{\psi(2)}}{n_{\psi(2)}} > \cdots > \frac{\mu_{\psi(K)}}{n_{\psi(K)}}. \tag{4.36}$$

Step 2: Let $\pi_0(m) = \psi(1)$. If $K < 2$, all the users j whose corresponding $f_j(\cdot)$ will not appear in the expression of $J(P(\mathbf{n}))$ have been removed and stop; if $K \geq 2$, go to Step 3.

Step 3: For $1 \leq i \leq K-1$, define

$$h_i(x, \mathbf{n}) \triangleq \left[\mu_{\psi(i+1)} \ln \frac{\mu_{\psi(i+1)}}{x n_{\psi(i+1)} B} - \mu_{\psi(1)} \ln \frac{\mu_{\psi(1)}}{x n_{\psi(1)} B}\right]$$

$$+ (n_{\psi(i+1)} - n_{\psi(1)})Bx - (\mu_{\psi(i+1)} - \mu_{\psi(1)}),$$

and let t_i satisfy $h_i(t_i, \mathbf{n}) = 0$. Let $\lambda_0(m) = max_i\{t_i\}$, $i^* = arg\ max_i\{t_i\}$. If $i^* \geq 2$, remove those users $\psi(j)$ for which $2 \leq j \leq i^*$ (i.e., let $\tau^*_{\psi(j)}(\mathbf{n}) = 0$, $P^*_{\psi(j)}(\mathbf{n}) = 0$, and remove them from further consideration). Also remove User $\psi(1)$. Increase m by 1 and return to Step 1.

In this procedure we observe that in the first iteration, $f_{\psi(1)}(P(\mathbf{n}))$ of User $\psi(1)$ must be the first part of the curve $J(P(\mathbf{n}))$ where $P(\mathbf{n})$ is close to 0, since $\forall 1 \leq j \leq K$,

$$\left.\frac{df_{\psi(j)}(x)}{dx}\right|_{x=0} = \frac{\mu_{\psi(j)}}{n_{\psi(j)}},$$

and according to (4.36), for x close to 0 ($x > 0$), it must be true that

$$f_{\psi(1)}(x) > f_{\psi(2)}(x) > \cdots > f_{\psi(K)}(x).$$

For $1 \leq i \leq K-1$, t_i satisfying $h_i(t_i, \mathbf{n}) = 0$ corresponds to the slope of the common tangent between the two curves $f_{\psi(1)}(P(\mathbf{n}))$ and $f_{\psi(i+1)}(P(\mathbf{n}))$. Since $t_{i^*} = max_i\{t_i\}$, if $i^* > 1$, all the curves $\left\{f_{\psi(j)}(P(\mathbf{n}))\right\}_{j=2}^{i^*}$ will always be under the contour formed by part of the curve $f_{\psi(1)}(P(\mathbf{n}))$, part of the curve $f_{\psi(i^*+1)}(P(\mathbf{n}))$, and the common tangent between them. Thus, no power should be assigned to users $\psi(j)$, $2 \leq j \leq i^*$. In this case, we know that part of the curve $f_{\psi(i^*+1)}(P(\mathbf{n}))$ of User $\psi(i^*+1)$ as well as the common tangent between the curves $f_{\psi(1)}(P(\mathbf{n}))$ and $f_{\psi(i^*+1)}(P(\mathbf{n}))$ must be part of $J(P(\mathbf{n}))$. For example, in Figure 4.6, since the number of remaining users is $K = 4$ and $\psi(j) = j$, $j = 1, 2, 3, 4$, if we draw the common tangents between curves $f_1(P(\mathbf{n}))$ and $f_2(P(\mathbf{n}))$, $f_1(P(\mathbf{n}))$ and $f_3(P(\mathbf{n}))$, and also $f_1(P(\mathbf{n}))$ and $f_4(P(\mathbf{n}))$, the slopes of which are t_1, t_2, and t_3, respectively, then it is clear that $t_2 = \max\{t_1, t_2, t_3\}$ and $i^* = 2$, i.e., the slope of the tangent between $f_1(P(\mathbf{n}))$ and $f_3(P(\mathbf{n}))$ is the largest among the slopes of the three tangents. Thus, $f_2(P(\mathbf{n}))$ will not be part of $J(P(\mathbf{n}))$ but $f_3(P(\mathbf{n}))$ and the common tangent between $f_1(P(\mathbf{n}))$ and $f_3(P(\mathbf{n}))$ will.

After removing those users $\psi(j)$, $2 \leq j \leq i^*$ and User $\psi(1)$, the number of remaining users is reduced from K to $K - i^*$ and User $\psi(i^*+1)$ in the first iteration becomes User $\psi(1)$ in the second iteration. Similarly, in the second iteration, by comparing the slopes of the common tangents between curves $f_{\psi(1)}(P(\mathbf{n}))$ and $f_{\psi(i+1)}(P(\mathbf{n}))$, $1 \leq i \leq K-1$, we may remove more users and find a new User $\psi(i^*+1)$ in this iteration whose corresponding $f_{\psi(i^*+1)}(P(\mathbf{n}))$ as well as the common tangent between curves $f_{\psi(1)}(P(\mathbf{n}))$ and $f_{\psi(i^*+1)}(P(\mathbf{n}))$ must be part of $J(P(\mathbf{n}))$. This User $\psi(i^*+1)$ becomes User $\psi(1)$ in the third iteration and the iterative procedure goes on and on until all the users j whose corresponding $f_j(x)$ will be part of $J(P(\mathbf{n}))$ have been identified and all other users have been removed.

Note that in each iteration, the value of m is different. Assume that by the time the iteration stops, $m = M_0$. If $M_0 = 1$, then $J(P(\mathbf{n}))$ is simply $f_{\pi_0(1)}$; if $M_0 > 1$, then $J(P(\mathbf{n}))$ is composed of $\left\{f_{\pi_0(m)}(\cdot)\right\}_{m=1}^{M_0}$ as well as the common tangents between curves $f_{\pi_0(m)}(\cdot)$ and $f_{\pi_0(m+1)}(\cdot)$, $1 \leq m \leq M_0 - 1$, the slopes of which are $\{\lambda_0(m)\}_{m=1}^{M_0-1}$. That is, by denoting $P_{a_0}(\mathbf{n}) = P_{b_0}(\mathbf{n}) = 0$, $P_{a_{M_0}}(\mathbf{n}) = P_{b_{M_0}}(\mathbf{n}) = \infty$, $\lambda_0(M_0) = 0$, and letting $P_{a_m}(\mathbf{n})$ and $P_{b_m}(\mathbf{n})$ $(1 \leq m \leq M_0 - 1)$ be the points satisfying

$$f'_{\pi_0(m)}(P_{a_m}(\mathbf{n})) = f'_{\pi_0(m+1)}(P_{b_m}(\mathbf{n})) = \lambda_0(m), \tag{4.37}$$

i.e.,

$$P_{a_m}(\mathbf{n}) \triangleq \frac{\mu_{\pi_0(m)}}{\lambda_0(m)} - n_{\pi_0(m)}B,$$

$$P_{b_m}(\mathbf{n}) \triangleq \frac{\mu_{\pi_0(m+1)}}{\lambda_0(m)} - n_{\pi_0(m+1)}B,$$

we can express $J(P(\mathbf{n}))$ as

$$J(P(\mathbf{n})) = \begin{cases} f_{\pi_0(m)}(P(\mathbf{n})) & \text{if } P_{b_{m-1}}(\mathbf{n}) \leq P(\mathbf{n}) < P_{a_m}(\mathbf{n}), \\ f_{\pi_0(m)}(P_{a_m}(\mathbf{n})) + \lambda_0(m)[P(\mathbf{n}) - P_{a_m}(\mathbf{n})] & \text{if } P_{a_m}(\mathbf{n}) \leq P(\mathbf{n}) < P_{b_m}(\mathbf{n}), \end{cases} \tag{4.38}$$

where $1 \leq m \leq M_0$. For example, in the case shown in Figure 4.6, since only User 2 will be removed from further consideration, by the time the iterative procedure stops, we have $M_0 = 3$ and

$$\begin{cases} \pi_0(1) = 1, \\ \pi_0(2) = 3, \\ \pi_0(3) = 4. \end{cases}$$

The $P_{a_m}(\mathbf{n})$ and $P_{b_m}(\mathbf{n})$ $(m = 1, 2)$ satisfying (4.37) are as shown in Figure 4.6 and $\lambda_0(1)$

and $\lambda_0(2)$ denote the slopes of the tangent lines for which $P(\mathbf{n})$ ranges from $P_{a_1}(\mathbf{n})$ to $P_{b_1}(\mathbf{n})$, and from $P_{a_2}(\mathbf{n})$ to $P_{b_2}(\mathbf{n})$, respectively. Thus, $J(P(\mathbf{n}))$ can be expressed as

$$J(P(\mathbf{n})) = \begin{cases} f_{\pi_0(1)}(P(\mathbf{n})) & \text{if } 0 \leq P(\mathbf{n}) < P_{a_1}(\mathbf{n}), \\ f_{\pi_0(1)}(P(\mathbf{n})) + \lambda_0(1)[P(\mathbf{n}) - P_{a_1}(\mathbf{n})] & \text{if } P_{a_1}(\mathbf{n}) \leq P(\mathbf{n}) < P_{b_1}(\mathbf{n}), \\ f_{\pi_0(2)}(P(\mathbf{n})) & \text{if } P_{b_1}(\mathbf{n}) \leq P(\mathbf{n}) < P_{a_2}(\mathbf{n}), \\ f_{\pi_0(2)}(P(\mathbf{n})) + \lambda_0(2)[P(\mathbf{n}) - P_{a_2}(\mathbf{n})] & \text{if } P_{a_2}(\mathbf{n}) \leq P(\mathbf{n}) < P_{b_2}(\mathbf{n}), \\ f_{\pi_0(3)}(P(\mathbf{n})) & \text{if } P(\mathbf{n}) \geq P_{b_2}(\mathbf{n}). \end{cases}$$

Once we obtain the curve $J(P(\mathbf{n}))$, similar to the two-user case, $\forall \lambda > 0$ fixed, since the optimal power $P^*(\mathbf{n})$ satisfies the condition (4.31), i.e., $J'(P^*(\mathbf{n})) = \lambda$, $P^*(\mathbf{n})$ is determined by the point(s) on the curve $J(P(\mathbf{n}))$ whose tangent has a slope λ. If λ equals any $\lambda_0(j)$ $(1 \leq j \leq M_0 - 1)$, since all the points on the common tangent between curves $f_{\pi_0(j)}$ and $f_{\pi_0(j+1)}$ share the same tangent, $P^*(\mathbf{n})$ can be any value between $P_{a_j}(\mathbf{n})$ and $P_{b_j}(\mathbf{n})$: if $P_{a_j}(\mathbf{n}) < P^*(\mathbf{n}) < P_{b_j}(\mathbf{n})$, $P^*(\mathbf{n})$ will be time-shared by the two users $\pi_0(j)$ and $\pi_0(j+1)$; if $P^*(\mathbf{n})$ is simply chosen as $P_{a_j}(\mathbf{n})$ or $P_{b_j}(\mathbf{n})$, then it is only assigned to User $\pi_0(j)$ or User $\pi_0(j+1)$, respectively. If $\lambda \neq \lambda_0(j)$ for all $1 \leq j \leq M_0 - 1$, then $P^*(\mathbf{n})$ is uniquely determined by the single point on $J(P(\mathbf{n}))$ that has a tangent with slope λ. Therefore, based on the expression of $J(P(\mathbf{n}))$ in (4.38) and the condition (4.31), for $\lambda > 0$ fixed, the optimal power and time allocation policy for the remaining M_0 users $\{\pi_0(m)\}_{m=1}^{M_0}$ is:

(a) if $M_0 = 1$, then

$$\begin{cases} \tau^*_{\pi_0(1)}(\mathbf{n}) = 1, \\ P^*_{\pi_0(1)}(\mathbf{n}) = \left[\frac{\mu_{\pi_0(1)}}{\lambda} - n_{\pi_0(1)}B\right]_+ ; \end{cases}$$

(b) if $M_0 > 1$, by denoting $\lambda_0(0) \triangleq \infty$, we know that for the given λ, there exists a $j \in \{1, 2, \ldots, M_0\}$ such that $\lambda_0(j-1) > \lambda > \lambda_0(j)$ or $\lambda = \lambda_0(j)$, since

$$\infty = \lambda_0(0) > \lambda_0(1) > \lambda_0(2) > \cdots > \lambda_0(M_0 - 1) > \lambda_0(M_0) = 0,$$

which results from the iterative procedure generating $\{\lambda_0(m)\}_{m=1}^{M_0}$ and from the fact that $J(P(\mathbf{n}))$ is an increasing, concave function. If $\exists j$ such that $\lambda_0(j) < \lambda < \lambda_0(j-1)$, from (4.38) we have $J(P(\mathbf{n})) = f_{\pi_0(j)}(P(\mathbf{n}))$, since only when $P_{b_{j-1}}(\mathbf{n}) \leq P(\mathbf{n}) < P_{a_j}(\mathbf{n})$ does the tangent slope of $J(P(\mathbf{n}))$ decrease from $\lambda_0(j-1)$ to $\lambda_0(j)$. In this case, we set

$$\begin{cases} \tau^*_{\pi_0(j)}(\mathbf{n}) = 1, \\ P^*_{\pi_0(j)}(\mathbf{n}) = \left[\frac{\mu_{\pi_0(j)}}{\lambda} - n_{\pi_0(j)}B\right]_+, \end{cases}$$

which corresponds to transmitting the information of User $\pi_0(j)$ only. If $\exists j$ such that $\lambda = \lambda(j)$, then $J(P(\mathbf{n})) = f_{\pi_0(j)}(P(\mathbf{n})) + \lambda(j)[P(\mathbf{n}) - P_{a_j}(\mathbf{n})]$, since only when $P_{a_j}(\mathbf{n}) \leq P(\mathbf{n}) < P_{b_j}(\mathbf{n})$ does the tangent slope of $J(P(\mathbf{n}))$ equal $\lambda(j)$. Therefore, as in the two-user case, we can set

$$\begin{cases} \tau^*_{\pi_0(j)}(\mathbf{n}) = \tau^*_0, \\ \tau^*_{\pi_0(j+1)}(\mathbf{n}) = 1 - \tau^*_0, \end{cases} \tag{4.39}$$

and

$$\begin{cases} P^*_{\pi_0(j)}(\mathbf{n}) = \left[\frac{\mu_{\pi_0(j)}}{\lambda} - n_{\pi_0(j)}B\right]_+, \\ P^*_{\pi_0(j+1)}(\mathbf{n}) = \left[\frac{\mu_{\pi_0(j+1)}}{\lambda} - n_{\pi_0(j+1)}B\right]_+, \end{cases}$$

which indicates that the channel is time-shared by User $\pi_0(j)$ and User $\pi_0(j+1)$ if choosing $0 < \tau^*_0 < 1$, and is occupied by User $\pi_0(j)$ or User $\pi_0(j+1)$ along if choosing $\tau^*_0 = 1$ or $\tau^*_0 = 0$, respectively.

In the above policy, λ and τ^*_0 satisfy the average power constraint

$$E_{\mathbf{n}}\left[\sum_{i=1}^{M} \tau^*_i(\mathbf{n}) P^*_i(\mathbf{n})\right] = \bar{P}, \tag{4.40}$$

and they may not be unique, since τ^*_0 can be any value between 0 and 1. Notice that as in the two-user case, if the c.d.f. $F(\mathbf{n})$ is continuous, $\forall \lambda > 0$, the probability measure of the set[7]

$$L \triangleq \{\mathbf{n} : \ \exists j \text{ such that } 1 \leq j \leq M_0 \text{ and } \lambda = \lambda_0(j)\} \tag{4.41}$$

is zero and λ is uniquely determined by (4.40). Moreover, in this case, the above optimal power and time allocation policy for the M-user broadcast channel implies that with probability 1, the information of at most a single user is transmitted in each fading state. If $F(\mathbf{n})$ is not continuous, then the set L may have non-zero probability measure and for $\mathbf{n} \in L$, as discussed before, the channel capacity region is achieved by time-sharing between two

[7]Note that in (4.41), M_0 is a function of $\mathbf{n}$, which results from the iterative procedure described earlier in this section.

users or dedicated transmission to just one user; for any other channel state $\mathbf{n}$ ($\mathbf{n} \notin L$), the information of at most one user is transmitted. Thus, in all cases, the capacity region boundary of TD can be achieved by sending information to just one user in every fading state. This motivates the sub-optimal TD policy we propose in the next section.

4.3.3 Sub-Optimal TD Policy

The optimal power and time allocation policy in Section 4.3.2 indicates that the capacity region $\mathcal{C}(\bar{P})$ in (4.18) is achieved by transmitting the information of at most a single user in each fading state $\mathbf{n}$, although it can also be achieved by a strategy that transmits the information of two users in some states by time-sharing and assigns the channel to a single user or no user in the other states. Based on this observation, we now propose a sub-optimal method for resource allocation. This method selects a single user in each channel state and allocates appropriate power to him according to the fading state. We now describe our method to choose the single user and his corresponding power.

In each state $\mathbf{n}$, $\forall i = 1, 2, \ldots, M$, define

$$P_i(\mathbf{n}) = \left[\frac{\mu_i}{\lambda} - n_i B\right]_+,$$

and let

$$G(\mathbf{n}) = \max_{1 \le i \le M} f_i(P_i(\mathbf{n})),$$

$$S = \{i : f_i(P_i(\mathbf{n})) = G(\mathbf{n})\},$$

$$i^* \triangleq \arg\min_{i \in S} P_i(\mathbf{n}),$$

where $f_i(x)$ is given in (4.22). Then, in the state $\mathbf{n}$, only the information for User i^* is sent with transmit power $P_{i^*}(\mathbf{n})$, where λ satisfies the M-user average power constraint (4.40).

For example, in the two-user case, by denoting

$$g_i(x, \mathbf{n}) \triangleq \mu_i \ln \frac{\mu_i}{x n_i B}, \qquad i = 1, 2, \tag{4.42}$$

we can express the sub-optimal power and time allocation policy as follows:

(a) if $g_1(\lambda, \mathbf{n}) > g_2(\lambda, \mathbf{n})$, or if $g_1(\lambda, \mathbf{n}) = g_2(\lambda, \mathbf{n})$ and $\lambda \ge \frac{\mu_2 - \mu_1}{n_2 B - n_1 B}$ (i.e., $\frac{\mu_1}{\lambda} - n_1 B \le$

$\frac{\mu_2}{\lambda} - n_2 B$), then

$$\begin{cases} \tau_1^*(\mathbf{n}) = 1, \\ \tau_2^*(\mathbf{n}) = 0, \end{cases} \text{and} \quad \begin{cases} P_1^*(\mathbf{n}) = \left[\frac{\mu_1}{\lambda} - n_1 B\right]_+, \\ P_2^*(\mathbf{n}) = 0, \end{cases}$$

(b) if $g_1(\lambda, \mathbf{n}) < g_2(\lambda, \mathbf{n})$, or if $g_1(\lambda, \mathbf{n}) = g_2(\lambda, \mathbf{n})$ and $\lambda < \frac{\mu_2 - \mu_1}{n_2 B - n_1 B}$ (i.e., $\frac{\mu_1}{\lambda} - n_1 B > \frac{\mu_2}{\lambda} - n_2 B$), then

$$\begin{cases} \tau_1^*(\mathbf{n}) = 0, \\ \tau_2^*(\mathbf{n}) = 1, \end{cases} \text{and} \quad \begin{cases} P_1^*(\mathbf{n}) = 0, \\ P_2^*(\mathbf{n}) = \left[\frac{\mu_2}{\lambda} - n_2 B\right]_+, \end{cases}$$

where λ satisfies the two-user average power constraint (4.32).

Compared to the optimal power and time allocation policy, the advantage of this sub-optimal scheme is that it is much easier to compute the water-filling power level $s^* \triangleq \frac{1}{\lambda}$ using (4.40). As will be shown in Section 4.4, the resulting rate region comes very close to that of the optimal TD policy. This is due to the fact that the two policies are identical except over a small set of fading states. Specifically, in the case where the c.d.f. $F(\mathbf{n})$ is continuous, the detailed comparison of the optimal and sub-optimal decision regions for the two-user fading broadcast channel in Appendix B.6 shows that for a given $\lambda > 0$, the two policies transmit the information of different users only in the rare occasions when $\mathbf{n} \in L_D$, where

$$L_D \triangleq \left\{ \mathbf{n} : \frac{\mu_1}{n_1} > \frac{\mu_2}{n_2} \text{ and } \lambda_0 < \lambda < \lambda_0^{sub} \text{ with } h(\lambda_0, \mathbf{n}) = 0,\ g_2(\lambda_0^{sub}, \mathbf{n}) - g_1(\lambda_0^{sub}, \mathbf{n}) = 0 \right\}. \tag{4.43}$$

4.3.4 CD without Successive Decoding

For CD without successive decoding each receiver treats the signals for other users as interference noise. For a given power allocation policy $\mathcal{P}$, by denoting $P_j(\mathbf{n})$ as the transmit power allocated to User j and $\mathcal{F}$ as the set of all possible power policies satisfying the average power constraint $E_{\mathbf{n}}\left[\sum_{j=1}^{M} P_j(\mathbf{n})\right] \leq \bar{P}$, we have the following theorem:

Theorem 4.5 *The achievable rate region for CD without successive decoding is given by:*

$$\mathcal{C}(\bar{P}) = \bigcup_{\mathcal{P}\in\mathcal{F}} \mathcal{C}_{CDWO}(\mathcal{P}), \tag{4.44}$$

where

$$\mathcal{C}_{CDWO}(\mathcal{P}) = \left\{ R_j \leq E_{\mathbf{n}} \left[B\log\left(1 + \frac{P_j(\mathbf{n})}{n_jB + \sum_{i=1,i\neq j}^{M} P_i(\mathbf{n})}\right)\right], \ j = 1, 2, \cdots, M \right\}.$$

The proof of the achievability follows along the same lines as that for the capacity region of CD given in Appendix B.1 and is therefore omitted. Note that in this chapter, as in the case of TD, we refer to this achievable rate region as the capacity region for CD without successive decoding, though we do not have a converse proof since the converse only applies to the optimal transmission strategy, which is CD with successive decoding.

In order to show that $\mathcal{C}(\bar{P})$ in (4.44) cannot be larger than the capacity region of TD in (4.18), we give the following lemma:

Lemma 4.2 $\forall P_i \geq 0,\ \mu_i \geq 0,\ n_iB > 0,\ i = 1, 2, \ldots, M,$

$$\sum_{i=1}^{M} \mu_i \log\left(1 + \frac{P_i}{n_iB + \sum_{j=1,j\neq i}^{M} P_j}\right) \leq \max_{i\in\{1,2,\ldots,M\}} \left[\mu_i \log\left(1 + \frac{\sum_{j=1}^{M} P_j}{n_iB}\right)\right]. \tag{4.45}$$

Proof: See Appendix B.7. □

Theorem 4.6

$$\bigcup_{\mathcal{P}\in\mathcal{F}} \mathcal{C}_{CDWO}(\mathcal{P}) \subseteq \bigcup_{\mathcal{P}\in\mathcal{F}} \mathcal{C}_{TD}(\mathcal{P}), \tag{4.46}$$

where equality is achieved using the optimal TD policy with power allocated to at most one user in each fading state.

Proof. Recall that the optimal power and time allocation policy discussed in Section 4.3.2 indicates that if $F(\mathbf{n})$ is continuous, then with probability 1, no more than one user is using the broadcast channel in each fading state $\mathbf{n}$; if $F(\mathbf{n})$ is discontinuous, in some fading states with non-zero probability, we can either choose two users and transmit their information by time-sharing the channel, or just select one of them and transmit his information alone. Therefore, in any case, the boundary of the capacity region $\mathcal{C}(\bar{P})$ in (4.18) can be achieved by an optimal TD policy which transmits the information of at most one user through the fading broadcast channel in each channel state $\mathbf{n}$. Obviously this optimal policy can be used as a power allocation policy for CD without successive decoding to eliminate interference from all other users and therefore to achieve the same capacity region boundary as TD. Thus, we need only to show that the capacity region of CD without successive decoding in (4.44) cannot be larger than the capacity region of TD in (4.18). We use *Lemma 4.2* to prove this as follows.

For CD without successive decoding, $\forall \mathbf{n} \in \mathcal{N}$, denote $P(\mathbf{n}) = \sum_{i=1}^{M} P_j(\mathbf{n})$. For $1 \leq i \leq M$, let

$$C_i(\mathbf{n}) = B \log \left[1 + \frac{P(\mathbf{n})}{n_i B}\right],$$

$$R_i(\mathbf{n}) = B \log \left[1 + \frac{P_i(\mathbf{n})}{n_i B + \sum_{j=1, j \neq i}^{M} P_j(\mathbf{n})}\right],$$

and let $\mu_i = \frac{B}{C_i(\mathbf{n})}$. Then according to *Lemma 4.2*, we have:

$$\sum_{i=1}^{M} \frac{R_i(\mathbf{n})}{C_i(\mathbf{n})} \leq 1. \tag{4.47}$$

Now consider the equal power TD strategy which assigns the power $P(\mathbf{n})$ to each user i for a fraction $\tau_i(\mathbf{n})$ of the total transmission time in the state $\mathbf{n}$. By denoting

$$R_i'(\mathbf{n}) = \tau_i(\mathbf{n}) B \log \left[1 + \frac{P(\mathbf{n})}{n_i B}\right],$$

we have

$$\sum_{i=1}^{M} \frac{R_i'(\mathbf{n})}{C_i(\mathbf{n})} = 1, \tag{4.48}$$

since $\sum_{i=1}^{M} \tau_i(\mathbf{n}) = 1$. Therefore, from (4.47) and (4.48), it is clear that given a fading state $\mathbf{n}$, $\forall P(\mathbf{n}) > 0$, the capacity region of the equivalent AWGN broadcast channel for CD without successive decoding is within that for equal power TD and is therefore within that for optimal TD. Consequently, the capacity region of CD without successive decoding in (4.44) cannot be larger than the capacity region of TD in (4.18). □

Note that since the sub-optimal TD scheme proposed at the end of Section 4.3.2 indicates that the broadcast channel is used by no more than one user, this suboptimal policy can also be applied to CD without successive decoding.

4.4 Numerical Results

In this section we present numerical results for the two-user ergodic capacity regions of the Rician and Rayleigh fading channels under different spectrum-sharing techniques. In the figures, as in [75], the equal power TD scheme refers to the strategy that assigns the constant transmission power $\bar{P}$ and total bandwidth B to User 1 for a fraction τ of the total transmission time, and then to User 2 for the remainder of the transmission. The optimal TD scheme for both the AWGN channel and the fading channels is obtained by allocating different power to the two users. We refer to CD without successive decoding as CDWO. Since TD and FD are equivalent in the sense that they have the same capacity region, all results for TD in the figures also apply for FD.

In Figure 4.7, the ergodic capacity regions of the Rician and Rayleigh fading broadcast channels are compared to that of the Gaussian broadcast channel using the CD, TD, equal power TD, and CDWO techniques. The SNR difference between the two users is 3 dB ($\bar{n}_1$ and $\bar{n}_2$ denote the average noise densities of User 1's channel and User 2's channel, respectively). The total transmission power is $\bar{P} = 10$ dB and the signal bandwidth $B = 100$ KHz. The ratio of the direct-path power to the scattered-path power in the Rician fading sub-channels is $k = 6$ dB.

In this figure we see that while the single-user ergodic rate (the R_1-axis or R_2-axis intercept) in fading is smaller than the rate in AWGN, the two-user capacity regions of both the Rician fading and the Rayleigh fading broadcast channels using either optimal CD or sub-optimal TD techniques in some places dominate that of the AWGN broadcast channel using optimal CD. That is, for the fading broadcast channel, ergodic rate pairs beyond the capacity region of the non-fading broadcast channel can be achieved by applying optimal resource allocation over the joint fading channel states. For simplicity, we calculate the rate region of the fading broadcast channel for TD by applying the simple sub-optimal TD power allocation policy. The resulting rate region turns out to be very close to the capacity region for optimal CD. Therefore, the capacity region using the optimal TD power policy will also come very close to that of optimal CD. This observation implies that, due to the small SNR difference between the two users, superposition encoding with successive decoding is not necessary, since time-sharing is near-optimal. For the AWGN broadcast channel, the capacity region boundary of the optimal TD scheme is indistinguishable from

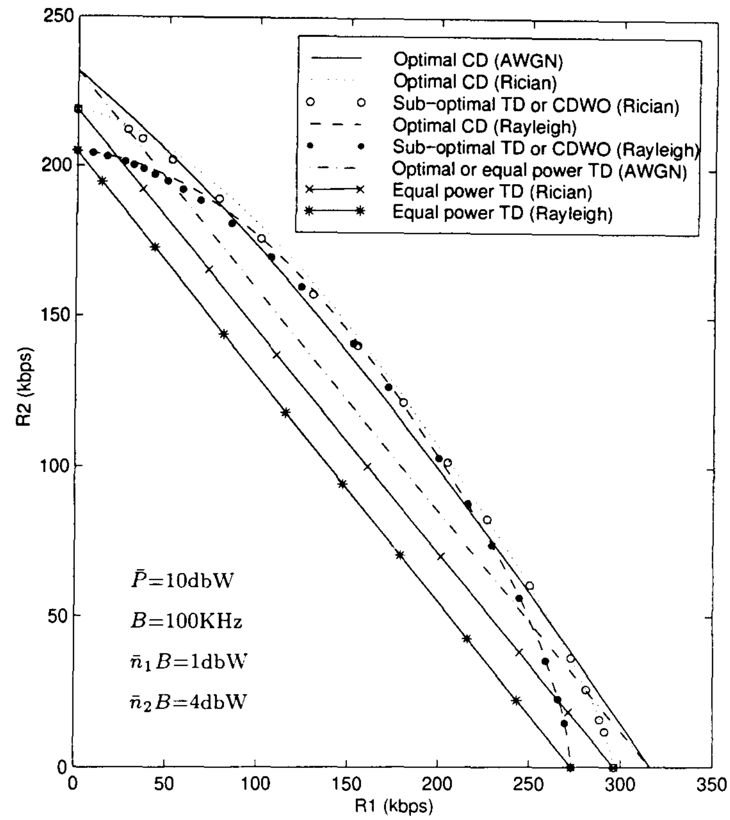


Figure 4.7: Two-user ergodic capacity region comparison: 3 dB SNR difference.

the equal power TD straight line, which means that when the two users have a similar channel noise power, constant power allocation is good enough for TD. The CDWO capacity region boundary (omitted from Figure 4.7 but shown in figures in [75]) includes the two end points of the equal power TD line but is below this straight line due to its convexity. However, for the fading channels, optimal or sub-optimal TD has a much larger capacity region than equal power TD and the capacity region for CDWO is the same as that for optimal TD.

We show in Figure 4.8 that when the SNR difference between the two users is 20 dB and the total average power is 25 dB, the ergodic capacity region for CD in Rayleigh fading is now completely within the region for CD in AWGN. However, optimal TD in fading can achieve some rate pairs far beyond the capacity region of the AWGN broadcast channel using optimal TD, and the sub-optimal TD power policy for Rayleigh fading results in a rate region almost as large as the capacity region with the optimal TD power policy. For both AWGN and fading channels, due to the large SNR difference between the two users, the capacity region for optimal CD is noticeably larger than that for optimal TD, and the capacity region for optimal TD is noticeably larger than that for equal power TD.

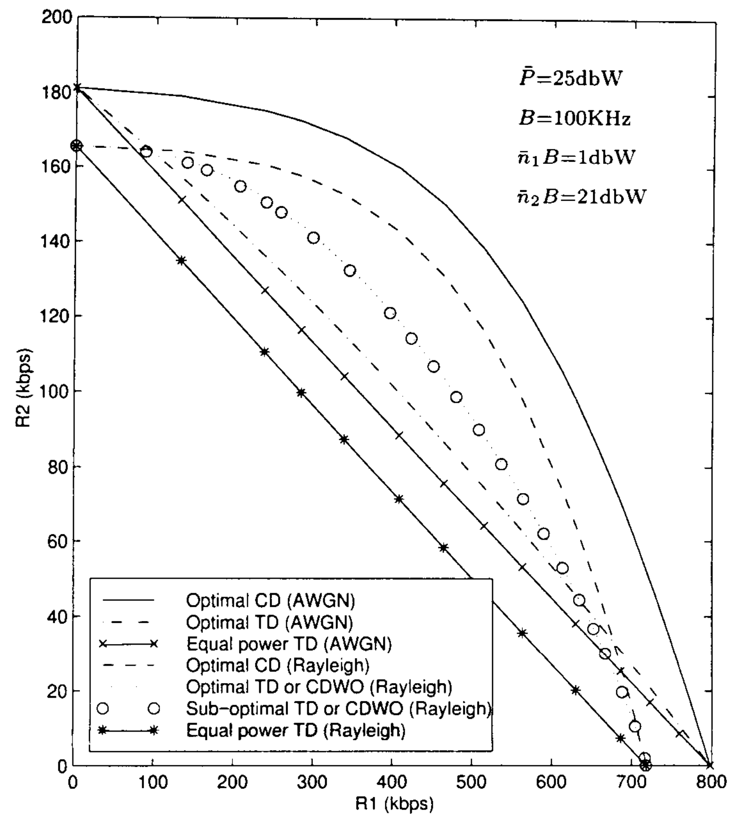


Figure 4.8: Two-user throughput capacity region comparison: 20 dB SNR difference with strong signal power.

Figure 4.9 shows the case where the SNR difference between the two users is 20 dB and the total average power is only 10 dB. Unlike the previous cases, we see here that the ergodic single-user capacity of User 2 for the Rayleigh fading channel is larger than that for the AWGN channel due to its very low average SNR. Thus, the optimal CD or sub-optimal TD scheme for fading yields a large rate region that is not achievable for the AWGN channel. However, as in Figure 4.8, due to the great SNR difference between the two users, optimal CD results in a capacity region much larger than that for sub-optimal TD and the capacity region for optimal TD is significantly larger than that for equal power TD. This observation holds for both fading and AWGN channels.

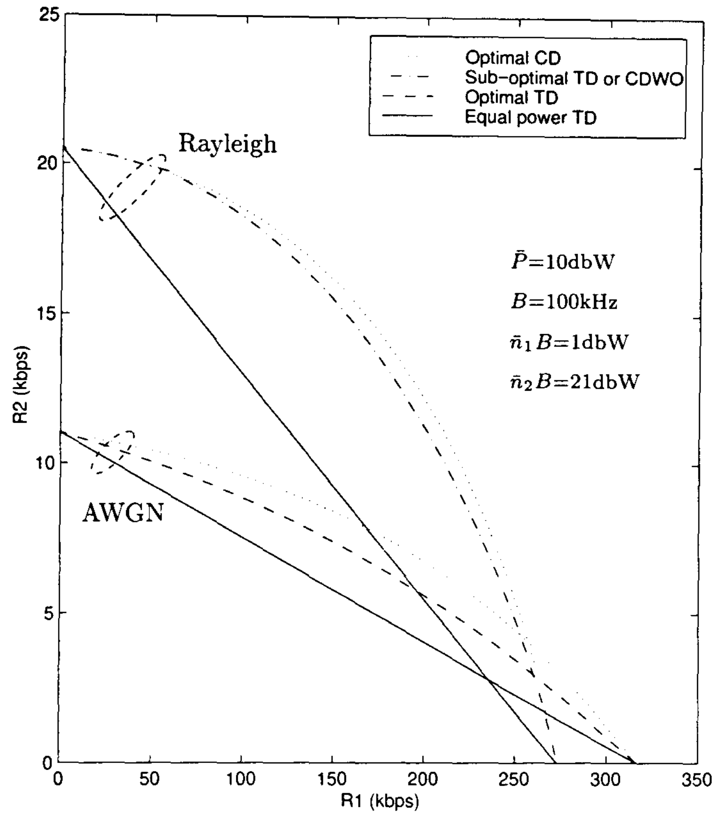


Figure 4.9: Two-user throughput capacity region comparison: 20 dB SNR difference with weak signal power.

4.5 Conclusions

We have obtained the ergodic capacity region and the optimal dynamic resource allocation strategy for fading broadcast channels with perfect CSI at both the transmitter and the receiver. These results are obtained for CD with and without successive decoding, TD and FD. Comparisons of the capacity regions show that CD with successive decoding has the largest capacity region, while TD and FD are equivalent and they have the same capacity region as CD without successive decoding. For CD without successive decoding, the optimal power policy is to transmit the information of at most one user in each joint fading state. This policy is also optimal for TD, though other strategies which allow at most two users to time-share the channel may also be optimal. When the average channel fading condition for each user is similar, the capacity regions for optimal CD and TD are quite close to each other. However, when each user has an average channel condition quite different from that of the others, optimal CD can achieve a much larger ergodic capacity region than the other techniques. In the next chapter, we will derive the zero-outage capacity region and the capacity region with non-zero outage for fading broadcast channels. Since each rate vector in the outage capacity region must be achievable in every fading state $\mathbf{n}$ unless an outage is declared, we cannot average over different fading conditions. However, the outage and ergodic capacity regions exhibit similar relative performance between the various channel-sharing techniques.

Chapter 5 Outage Capacities and Optimal Resource Allocation for Fading Broadcast Channels

5.1 Introduction

In the previous chapter, by applying optimal dynamic power and rate allocation strategies, we have obtained the ergodic (Shannon) capacities of fading broadcast channels under different spectrum-sharing techniques. This kind of capacity is a measure of the long-term achievable rate averaged over the time-varying channel. For real-time applications that cannot tolerate the variable delays exhibited by the coding strategy that achieves the ergodic capacity, we have to consider the information rate that can be maintained in all fading conditions through optimal power control. In order to maintain a constant rate during severe fading, much power is needed. Therefore, given an average power constraint, the channel fading may be so severe that no constant rate greater than zero is possible. For example, the maximum instantaneous mutual information rate that can be supported continuously on the single-user Rayleigh fading channel with a finite average transmit power constraint is zero [3]. However, if we allow some transmission outage under severe fading conditions, the maximum instantaneous mutual information rate that can be maintained during non-outage will increase. Finding the optimal resource allocation strategy that achieves the outage capacity with a given outage probability is tantamount to deriving the strategy that minimizes the outage probability for a given rate vector. In [76], the minimum outage probability problem is solved for the single-user fading channel. In addition, it is shown that under a long-term average power constraint, the optimal power allocation depends on the fading statistics through a threshold-decision rule: no transmission is allowed in a fading state where the required power is above a threshold value.

For an M-user broadcast fading channel and a given rate vector $\mathbf{R}$, we consider a similar minimum common outage probability problem under the assumption that the broadcast channel is either not used at all when fading is severe or is used simultaneously for all users when fading is tolerable. Under the more complex assumption that an outage can be declared for each user individually, we obtain an optimal power allocation policy that achieves boundaries of outage probability regions for time-division (TD), frequency-division (FD) and code-division (CD) with and without successive decoding. This optimal power allocation strategy is a multi-user generalization of the single-user threshold-decision rule.

As a special case, if no outage is allowed during the transmission, the outage capacity with a given outage probability becomes the zero-outage capacity. In [3], with an average power constraint for each user, under the assumption that CSI is available at both the transmitters and the receiver, the zero-outage capacity region[1] and the optimal power allocation scheme are derived for the fading MAC by exploiting the special polymatroidal structure of the region. It is shown that the boundary of this capacity region can be achieved through successive decoding and applying a greedy optimal power allocation scheme. The successive decoding order depends on both the current fading state and the power price for each user.

In this chapter, for convenience we first obtain directly zero-outage capacity regions and the associate optimal resource allocation strategies of an M-user broadcast fading channel for TD, FD, and CD with and without successive decoding. For CD with successive decoding, we will show that the superposition coding and successive decoding order depends only on the current fading state. For the Nakagami-m fading model [77] we prove that the limiting zero-outage capacity region converges to that of the Gaussian broadcast channel for CD with and without successive decoding when $m \to \infty$.

The remainder of this chapter is organized as follows: the broadcast fading channel model is briefly described in Section 5.2. In Section 5.3, the zero-outage capacity regions are derived for each of the different spectrum-sharing techniques. We derive strategies to minimize the common outage probability and achieve the boundary of the outage probability region for TD, FD, and CD with or without successive decoding in Section 5.4. Section 5.5 shows numerical results, followed by our conclusions in the last section.

[1]The zero-outage capacity is called "delay-limited capacity" in [3], since the coding strategy that achieves the zero-outage capacity has a delay that is independent of the channel variation.

5.2 The Fading Broadcast Channel

We consider the same discrete-time M-user broadcast fading channel model as in Chapter 4, where the signal source $X[i]$ is composed of M independent information sources and the broadcast channel consists of M independent fading sub-channels. The time-varying sub-channel gains are denoted as $\sqrt{g_1[i]}, \sqrt{g_2[i]}, \ldots, \sqrt{g_M[i]}$ and the Gaussian noises of these sub-channels are denoted as $z_1[i], z_2[i], \ldots, z_M[i]$. Let $\bar{P}$ be the total average transmit power, B the received signal bandwidth, and ν_j the noise density of $z_j[i]$, $j = 1, 2, \ldots, M$. Since the time-varying received SNR $\gamma_j[i] = \bar{P}g_j[i]/(\nu_j B)$, $j = 1, 2, \ldots, M$, by denoting[2] $n_j[i] = \nu_j/g_j[i]$, we have $\gamma_j[i] = \bar{P}/(n_j[i]B)$.

For a slowly time-varying broadcast channel, we assume that the $n_j[i]$, $j = 1, 2, \ldots, M$, are known to the transmitter and all M receivers at time i. Thus, the transmitter can vary the transmit power $P_j[i]$ for each user relative to the noise density vector $\mathbf{n}[i] = (n_1[i], n_2[i], \ldots, n_M[i])$, subject only to the average power constraint $\bar{P}$. For TD or FD, it can also vary the fraction of transmission time or bandwidth $\tau_j[i]$ assigned to each user, subject to the constraint $\sum_{j=1}^{M} \tau_j[i] = 1$ for all i. For CD, the superposition code can be varied at each transmission. We call $\mathbf{n}[i]$ the joint fading process and denote $\mathcal{N}$ as the set of all possible joint fading states. $F(\mathbf{n})$ denotes a given cumulative distribution function (c.d.f.) on $\mathcal{N}$.

5.3 Zero-Outage Capacity Region

For an M-user fading broadcast channel with stationary distribution $\mathcal{Q}$ and a total average power constraint $\bar{P}$, we give the following definition for the zero-outage capacity region $C_{zero}(\bar{P})$, which is similar to that of the delay-limited capacity region for the MAC in [3]:

Definition 5.1 *For a given rate vector* $\mathbf{R} = (R_1, R_2, \cdots, R_M)$, *if* $\forall\epsilon > 0$, *there exists a coding delay* T *such that for every fading process with stationary distribution* $\mathcal{Q}$, *there exist codebooks and a decoding scheme with probability of error* $P_e^{(T)} < \epsilon$, *then* $\mathbf{R} \in C_{zero}(\bar{P})$. *Moreover, the codewords can be chosen as a function of the realization of the fading processes.*

[2]See Section 4.2 for a discussion of the case $g_j[i] = 0$.

In this section, the zero-outage capacity region of an M-user broadcast fading channel is obtained for CD with and without successive decoding and for TD. For FD, using the same argument as in [75], it can be easily shown that the zero-outage capacity region is the same as for TD and the optimal power and bandwidth allocation policy for FD can be derived directly from that of TD.

5.3.1 CD

For an M-user broadcast system, we first consider superposition coding and successive decoding where, in each joint fading state, the channel can be viewed as a degraded Gaussian broadcast channel with noise densities $n_1[i], n_2[i], \ldots, n_M[i]$ and the multiresolution signal constellation is optimized relative to these instantaneous noises. In this case, the users with smaller noise densities will subtract the interference from the users with larger noise densities. Given a power allocation policy $\mathcal{P}$, let $P_j(\mathbf{n})$ be the transmit power allocated to User j for the joint fading state $\mathbf{n} = (n_1, n_2, \ldots, n_M)$ and denote $\mathcal{F}$ as the set of all possible power policies satisfying the average power constraint $E_{\mathbf{n}}\left[\sum_{j=1}^{M} P_j(\mathbf{n})\right] \leq \bar{P}$, where $E[\cdot]$ denotes the expectation function. For simplicity, assume that the stationary distributions of the fading processes have continuous densities[3], i.e., $Pr\{n_i = n_j\} = 0, \forall i \neq j$.

Theorem 5.1 *The zero-outage capacity region for the fading broadcast channel when the transmitter and all the receivers know the current channel state is given by:*

$$C_{zero}(\bar{P}) = \bigcup_{\mathcal{P} \in \mathcal{F}} \bigcap_{\mathbf{n} \in \mathcal{N}} C_{CD}(\mathbf{n}, \mathcal{P}), \tag{5.1}$$

where $C_{CD}(\mathbf{n}, \mathcal{P})$ is the capacity region of the time-invariant Gaussian broadcast channel. That is,

$$C_{CD}(\mathbf{n}, \mathcal{P}) = \left\{ R_j \leq B \log\left(1 + \frac{P_j(\mathbf{n})}{n_j B + \sum_{i=1}^{M} P_i(\mathbf{n}) \mathbf{1}[n_j > n_i]}\right), \quad j = 1, 2, \cdots, M \right\}, \tag{5.2}$$

where $\mathbf{1}[\cdot]$ denotes the indicator function ($\mathbf{1}[x] = 1$ if x is true and zero otherwise).

Proof: See Appendix C.1. □

[3]If $Pr\{n_i = n_j\} \neq 0$ for some i, j then, in state $\mathbf{n}$, User i and User j can be viewed as a single user and superposition coding and successive decoding are applied to $M - 1$ users. The information for User i and User j are then transmitted by time-sharing the channel.

For a given rate vector $\mathbf{R}$ and a fading state $\mathbf{n}$, from (5.2) we can calculate the minimum required power $P_j^{min}(\mathbf{n})$ ($j = 1, 2, \ldots, M$) that can support the rate vector $\mathbf{R}$. Specifically, let $\pi(\cdot)$ be the permutation such that

$$n_{\pi(1)} < n_{\pi(2)} < \cdots < n_{\pi(M)}.$$

Then according to (5.2), we have

$$\begin{cases} R_{\pi(1)} \le B \log\left(1 + \frac{P_{\pi(1)}(\mathbf{n})}{n_{\pi(1)}B}\right), \\ R_{\pi(i)} \le B \log\left(1 + \frac{P_{\pi(i)}(\mathbf{n})}{n_{\pi(i)}B + \sum_{j=1}^{i-1} P_{\pi(j)}(\mathbf{n})}\right), \quad i = 2, 3, \cdots, M. \end{cases}$$

Thus, to support rate vector $\mathbf{R}$, we require

$$\begin{cases} P_{\pi(1)}(\mathbf{n}) \ge n_{\pi(1)}B\left(2^{R_{\pi(1)}/B} - 1\right), \\ P_{\pi(i)}(\mathbf{n}) \ge \left(n_{\pi(i)}B + \sum_{j=1}^{i-1} P_{\pi(j)}(\mathbf{n})\right)\left(2^{R_{\pi(i)}/B} - 1\right), \quad i = 2, 3, \cdots, M. \end{cases}$$

The minimum power required to support $\mathbf{R}$ for each user is

$$\begin{cases} P_{\pi(1)}^{min}(\mathbf{n}) = n_{\pi(1)}B\left(2^{R_{\pi(1)}/B} - 1\right), \\ P_{\pi(i)}^{min}(\mathbf{n}) = \left(n_{\pi(i)}B + \sum_{j=1}^{i-1} P_{\pi(j)}^{min}(\mathbf{n})\right)\left(2^{R_{\pi(i)}/B} - 1\right), \quad i = 2, 3, \cdots, M. \end{cases}$$

Consequently, the minimum required total power $P^{min}(\mathbf{R}, \mathbf{n})$ that can support $\mathbf{R}$ in fading state $\mathbf{n}$ is:

$$\begin{aligned} P^{min}(\mathbf{R}, \mathbf{n}) &= \sum_{i=1}^{M} P_{\pi(i)}^{min}(\mathbf{n}) \\ &= \sum_{i=1}^{M-1}\left[2^{\sum_{j=i+1}^{M} R_{\pi(j)}/B}\left(2^{R_{\pi(i)}/B} - 1\right) n_{\pi(i)}B\right] \\ &\quad + \left(2^{R_{\pi(M)}/B} - 1\right) n_{\pi(M)}B. \end{aligned} \tag{5.3}$$

For a given $\mathbf{R}$, if $\mathbf{R} \in \mathcal{C}_{zero}(\bar{P})$, then by (5.1), the minimum required average power $E_{\mathbf{n}}[P^{min}(\mathbf{R}, \mathbf{n})]$ satisfies the total average power constraint

$$E_{\mathbf{n}}[P^{min}(\mathbf{R}, \mathbf{n})] \le \bar{P}, \tag{5.4}$$

where $P^{min}(\mathbf{R}, \mathbf{n})$ is given by (5.3). If $\mathbf{R}$ is on the boundary surface of $\mathcal{C}_{zero}(\bar{P})$, then the equality in (5.4) is achieved. Note that for the single user case ($M = 1$), if R_1 is on the boundary of $\mathcal{C}_{zero}(\bar{P})$, from (5.3) and (5.4) we have

$$E_{n_1}\left[\left(2^{R_1/B} - 1\right) n_1 B\right] = \bar{P}.$$

Thus,

$$R_1 = B \log\left(1 + \frac{\bar{P}}{E_{n_1}[n_1 B]}\right),$$

which is the same as derived in [2].

5.3.2 CD without Successive Decoding

In CD without successive decoding, each receiver treats the signals for other users as interfering noise. For a given power allocation policy $\mathcal{P}$, let $P_j(\mathbf{n})$ denote the transmit power allocated to User j in the state $\mathbf{n}$ and let $\mathcal{F}$ denote the set of all possible power policies satisfying the average power constraint $E_{\mathbf{n}}\left[\sum_{j=1}^{M} P_j(\mathbf{n})\right] \leq \bar{P}$. Then the achievable zero-outage rate region for CD without successive decoding is given by:

$$\mathcal{C}_{zero}(\bar{P}) = \bigcup_{\mathcal{P}\in\mathcal{F}} \bigcap_{\mathbf{n}\in\mathcal{N}} \mathcal{C}_{CDWO}(\mathbf{n}, \mathcal{P}), \tag{5.5}$$

where $\mathcal{C}_{CDWO}(\mathbf{n}, \mathcal{P})$ is the rate region of the time-invariant Gaussian broadcast channel using CD without successive decoding:

$$\mathcal{C}_{CDWO}(\mathbf{n}, \mathcal{P}) = \left\{R_j \leq B \log\left(1 + \frac{P_j(\mathbf{n})}{n_j B + \sum_{i=1,i\neq j}^{M} P_i(\mathbf{n})}\right), \quad j = 1, 2, \cdots, M\right\}. \tag{5.6}$$

The proof of the achievability follows along the same lines as that for the capacity region of CD given in Appendix C.1 and is therefore omitted. Note that in this chapter, we refer to this achievable rate region as the zero-outage capacity region for CD without successive decoding, though we do not have a converse proof since the converse only applies to the optimal transmission strategy, which is CD with successive decoding.

For a given rate vector $\mathbf{R}$ and a fading state $\mathbf{n}$, we know from (5.6) that

$$P_i(\mathbf{n}) \geq \left(n_i B + \textstyle\sum_{j=1,j\neq i}^{M} P_j(\mathbf{n})\right)\left(2^{R_i/B} - 1\right), \quad i = 1, 2, \cdots, M. \tag{5.7}$$

Denoting $P_i^{min}(\mathbf{n})$ $(i = 1, 2, \cdots, M)$ as the minimum power required for User i in order to support rate vector $\mathbf{R}$, by (5.7) we have

$$P_i^{min}(\mathbf{n}) \geq \left(n_i B + \sum_{j=1, j\neq i}^{M} P_j^{min}(\mathbf{n}) \right) \left(2^{R_i/B} - 1 \right), \quad i = 1, 2, \cdots, M.$$

Therefore, $P_i^{min}(\mathbf{n})$ must satisfy

$$P_i^{min}(\mathbf{n}) = \left(n_i B + \sum_{j=1, j\neq i}^{M} P_j^{min}(\mathbf{n}) \right) \left(2^{R_i/B} - 1 \right), \quad i = 1, 2, \cdots, M. \tag{5.8}$$

By defining matrix $A = (a_{ij})$, $i, j = 1, 2, \cdots, M$, where

$$a_{ij} = \begin{cases} \frac{1}{2^{R_i/B} - 1}, & \text{if } i = j, \\ -1, & \text{if } i \neq j, \end{cases} \tag{5.9}$$

we prove in Appendix C.2 that the M linear equations in (5.8) have positive solutions for all $P_i^{min}(\mathbf{n})$ $(1 \leq i \leq M)$ in every fading state $\mathbf{n}$ if and only if $\det A > 0$. Assuming that $\det A > 0$, it is clear that the explicit solution to $\mathbf{P}^{min}(\mathbf{n}) = (P_1^{min}(\mathbf{n}), P_2^{min}(\mathbf{n}), \cdots, P_M^{min}(\mathbf{n}))$ is

$$\mathbf{P}^{min}(\mathbf{n}) = A^{-1} \cdot B\mathbf{n}^T, \tag{5.10}$$

where A^{-1} denotes the inverse of matrix A and $\mathbf{n}^T$ denotes the transpose of vector $\mathbf{n}$. Thus, the minimum required total power $P^{min}(\mathbf{R}, \mathbf{n})$ is

$$P^{min}(\mathbf{R}, \mathbf{n}) = \sum_{i=1}^{M} P_i^{min}(\mathbf{n}). \tag{5.11}$$

For example, in the two-user case $(M = 2)$, if $\det A > 0$, i.e., if $2^{R_1/B} + 2^{R_2/B} > 2^{(R_1+R_2)/B}$, the solution for $P_1^{min}(\mathbf{n})$ and $P_2^{min}(\mathbf{n})$ will be

$$\begin{cases} P_1^{min}(\mathbf{n}) = \frac{(2^{R_1/B}-1)n_1 B + (2^{R_1/B}-1)(2^{R_2/B}-1)n_2 B}{2^{R_1/B} + 2^{R_2/B} - 2^{(R_1+R_2)/B}}, \\ P_2^{min}(\mathbf{n}) = \frac{(2^{R_2/B}-1)n_2 B + (2^{R_2/B}-1)(2^{R_1/B}-1)n_1 B}{2^{R_1/B} + 2^{R_2/B} - 2^{(R_1+R_2)/B}}. \end{cases}$$

Thus,

$$\begin{aligned} P^{min}(\mathbf{R},\mathbf{n}) &= P_1^{min}(\mathbf{n}) + P_2^{min}(\mathbf{n}) \\ &= \frac{\left(2^{(R_1+R_2)/B} - 2^{R_2/B}\right) n_1 B + \left(2^{(R_1+R_2)/B} - 2^{R_1/B}\right) n_2 B}{2^{R_1/B} + 2^{R_2/B} - 2^{(R_1+R_2)/B}}. \end{aligned} \tag{5.12}$$

Therefore, a given $\mathbf{R} \in \mathcal{C}_{zero}(\bar{P})$ if the average power constraint in (5.4) is satisfied with $P^{min}(\mathbf{R},\mathbf{n})$ given by (5.11). If $\mathbf{R}$ is on the boundary surface of $\mathcal{C}_{zero}(\bar{P})$, then the equality in (5.4) is achieved.

5.3.3 TD

Now we consider the time-division case where, in each fading state $\mathbf{n}$, the information for the M users will be divided and sent in time-slots which are functions of $\mathbf{n}$. For a given power and time allocation policy $\mathcal{P}$, let $P_j(\mathbf{n})$ and $\tau_j(\mathbf{n})$ $(0 \leq \tau_j(\mathbf{n}) \leq 1)$ be the transmit power and fraction of transmission time allocated to User j $(j = 1, 2, \ldots, M)$, respectively, for fading state $\mathbf{n}$, and let $\mathcal{F}$ be the set of all such possible power and time allocation policies satisfying

$$\begin{cases} E_{\mathbf{n}}\left[\sum_{j=1}^{M} \tau_j(\mathbf{n}) P_j(\mathbf{n})\right] \leq \bar{P}, & \text{and} \\ \sum_{j=1}^{M} \tau_j(\mathbf{n}) = 1, \quad \forall \mathbf{n} \in \mathcal{N}. \end{cases}$$

Then the achievable zero-outage capacity region for the variable power and transmission time scheme is:

$$\mathcal{C}_{zero}(\bar{P}) = \bigcup_{\mathcal{P} \in \mathcal{F}} \bigcap_{\mathbf{n} \in \mathcal{N}} \mathcal{C}_{TD}(\mathbf{n}, \mathcal{P}), \tag{5.13}$$

where $\mathcal{C}_{TD}(\mathbf{n}, \mathcal{P})$ is the rate region of the time-invariant Gaussian broadcast channel using the TD technique:

$$\mathcal{C}_{TD}(\mathbf{n}, \mathcal{P}) = \left\{ R_j \leq \tau_j(\mathbf{n}) B \log\left(1 + \frac{P_j(\mathbf{n})}{n_j B}\right), \quad j = 1, 2, \cdots, M \right\}. \tag{5.14}$$

The proof of the achievability follows along the same lines as that for the capacity region of CD given in Appendix C.1 and is therefore omitted. Note that as in the case of CD without successive, we refer to this achievable rate region as the zero-outage capacity region for TD, though we do not have a converse proof due to the fact that the converse only holds for the optimal transmission strategy for this channel, which, according to *Theorem 5.1*, is CD with successive decoding.

For a given rate vector $\mathbf{R}$ and a fading state $\mathbf{n}$, from (5.14) we have

$$P_i(\mathbf{n}) \geq n_i B \left(2^{\frac{R_i}{B\tau_i(\mathbf{n})}} - 1\right), \quad i = 1, 2, \ldots, M.$$

Therefore, the required total power $P(\mathbf{R}, \mathbf{n})$ of the M users satisfies

$$\begin{aligned} P(\mathbf{R}, \mathbf{n}) &= \sum_{i=1}^{M} \tau_i(\mathbf{n}) P_i(\mathbf{n}) \\ &\geq \sum_{i=1}^{M} \tau_i(\mathbf{n}) n_i B \left(2^{\frac{R_i}{B\tau_i(\mathbf{n})}} - 1\right). \end{aligned}$$

Let $\boldsymbol{\tau}(\mathbf{n}) \triangleq [\tau_1(\mathbf{n}), \tau_2(\mathbf{n}), \cdots, \tau_M(\mathbf{n})]$ and let $P^{min}(\mathbf{R}, \mathbf{n})$ be the minimum required total power of the M users for fading state $\mathbf{n}$, then

$$\begin{cases} P^{min}(\mathbf{R}, \mathbf{n}) = \min_{\boldsymbol{\tau}(\mathbf{n})} \left\{ \sum_{i=1}^{M} \tau_i(\mathbf{n}) n_i B \left(2^{\frac{R_i}{B\tau_i(\mathbf{n})}} - 1\right) \right\} \\ \\ \text{subject to: } \sum_{i=1}^{M} \tau_i(\mathbf{n}) = 1, \quad \forall \mathbf{n} \in \mathcal{N}. \end{cases} \tag{5.15}$$

By applying the Lagrangian technique, we can find the optimal $\boldsymbol{\tau}(\mathbf{n})$ which achieves $P^{min}(\mathbf{R}, \mathbf{n})$ in (5.15). For example, in a two-user system ($M = 2$), let

$$P(\tau_1(\mathbf{n})) \triangleq \tau_1(\mathbf{n}) n_1 B \left(2^{\frac{R_1}{B\tau_1(\mathbf{n})}} - 1\right) + (1 - \tau_1(\mathbf{n})) n_2 B \left(2^{\frac{R_2}{B(1-\tau_1(\mathbf{n}))}} - 1\right)$$

and let $\tau_1^*(\mathbf{n})$ be the solution to the non-linear equation

$$\frac{dP(\tau_1(\mathbf{n}))}{d\tau_1(\mathbf{n})} = 0.$$

Since it is easy to verify that for $\tau_1(\mathbf{n}) > 0$, $\frac{d^2 P(\tau_1(\mathbf{n}))}{d\tau_1^2(\mathbf{n})} > 0$, we have

$$\begin{aligned} P^{min}(\mathbf{R}, \mathbf{n}) &= \min_{\tau_1(\mathbf{n})} P(\tau_1(\mathbf{n})) \\ &= P(\tau_1^*(\mathbf{n})). \end{aligned}$$

Therefore, $\mathbf{R} \in \mathcal{C}_{zero}(\bar{P})$ if $P^{min}(\mathbf{R}, \mathbf{n})$ in (5.15) satisfies (5.4). If $\mathbf{R}$ is on the boundary surface of $\mathcal{C}_{zero}(\bar{P})$ then the equality in (5.4) is achieved.

5.3.4 The Limiting Zero-Outage Capacity Region for Nakagami Fading

It is well known that the Nakagami-m fading model [15] can be used to describe different fading conditions ranging from Rayleigh ($m = 1$) to Rician channels with strong line-of-sight components. As the fading parameter m ($m \geq 1/2$) goes to infinity, the Nakagami-m fading channel converges to an AWGN channel. Therefore, it is expected that the limiting zero-outage capacity region of the Nakagami-m fading broadcast channel converges to the capacity region of an AWGN broadcast channel as $m \to \infty$. In this subsection, we prove this to be true for CD with and without successive decoding in a two-user system. These results can be easily extended to more users.

CD with Successive Decoding

For a two-user broadcast channel with fading, given rate vector $\mathbf{R} = (R_1, R_2)$, we know by (5.3) that the minimum required total power to support $\mathbf{R}$ in fading state $\mathbf{n} = (n_1, n_2)$ is:

$$P^{min}(\mathbf{R}, \mathbf{n}) = \begin{cases} 2^{R_2/B}(2^{R_1/B} - 1)n_1 B + (2^{R_2/B} - 1)n_2 B, & \text{if } n_1 < n_2, \\ 2^{R_1/B}(2^{R_2/B} - 1)n_2 B + (2^{R_1/B} - 1)n_1 B, & \text{if } n_1 > n_2. \end{cases} \tag{5.16}$$

If $\mathbf{R}$ is on the boundary surface of $\mathcal{C}_{zero}(\bar{P})$ in (5.1), by substituting (5.16) into (5.4) with equality, we obtain

$$\begin{aligned} \bar{P} &= E_{\mathbf{n}}[P^{min}(\mathbf{R}, \mathbf{n})] \\ &= 2^{R_2/B}(2^{R_1/B} - 1)E_{n_1<n_2}[n_1 B] + (2^{R_2/B} - 1)E_{n_1<n_2}[n_2 B] \\ &\quad + 2^{R_1/B}(2^{R_2/B} - 1)E_{n_1>n_2}[n_2 B] + (2^{R_1/B} - 1)E_{n_1>n_2}[n_1 B]. \end{aligned} \tag{5.17}$$

Let $\bar{n}_1 B$ and $\bar{n}_2 B$ be the average noise variances of the channels for User 1 and User 2, respectively. Assuming that the signal power is normalized to 1, the signal-to-noise ratio (SNR) γ_i of User i ($i = 1, 2$) for a given channel state $\mathbf{n}$ is:

$$\gamma_i = \frac{1}{n_i B}. \tag{5.18}$$

We use the following lemma to show that for Nakagami-m fading, as $m \to \infty$, (5.17) converges to the boundary equation for the capacity region of an AWGN broadcast channel with sub-channel noise variances $\bar{n}_1 B$ and $\bar{n}_2 B$.

Lemma 5.1 *Given $0 < p < 1/2$ and a fixed integer r,*

$$\lim_{k\to\infty} \sum_{i=0}^{k+r} \binom{i+k-1}{i} p^i (1-p)^k = 1.$$

Proof: See Appendix C.3. □

For Nakagami-m fading, the probability density function (p.d.f.) of γ_i in (5.18) is:

$$p_i(\gamma_i) = \frac{(m\bar{n}_i B)^m}{(m-1)!}\gamma_i^{m-1} e^{-m\bar{n}_i B\gamma_i}, \qquad i = 1, 2. \tag{5.19}$$

Thus, from (5.17) and (5.18) we know that

$$\begin{aligned} \bar{P} &= 2^{R_2/B}(2^{R_1/B}-1)D_a(m) + (2^{R_2/B}-1)D_b(m) \\ &\quad + 2^{R_1/B}(2^{R_2/B}-1)D_c(m) + (2^{R_1/B}-1)D_d(m), \end{aligned} \tag{5.20}$$

where

$$D_a(m) = E_{n_1<n_2}[n_1 B] = \int_{\gamma_1>\gamma_2} \frac{1}{\gamma_1} p_1(\gamma_1)p_2(\gamma_2)d\boldsymbol{\gamma}, \tag{5.21}$$

$$D_b(m) = E_{n_1<n_2}[n_2 B] = \int_{\gamma_1>\gamma_2} \frac{1}{\gamma_2} p_1(\gamma_1)p_2(\gamma_2)d\boldsymbol{\gamma}, \tag{5.22}$$

$$D_c(m) = E_{n_1>n_2}[n_2 B] = \int_{\gamma_1<\gamma_2} \frac{1}{\gamma_2} p_1(\gamma_1)p_2(\gamma_2)d\boldsymbol{\gamma}, \tag{5.23}$$

$$D_d(m) = E_{n_1>n_2}[n_1 B] = \int_{\gamma_1<\gamma_2} \frac{1}{\gamma_1} p_1(\gamma_1)p_2(\gamma_2)d\boldsymbol{\gamma}. \tag{5.24}$$

By applying *Lemma 5.1*, we obtain the following lemma:

Lemma 5.2 *For Nakagami-m fading, assuming that $\bar{n}_1 B < \bar{n}_2 B$,*

$$\lim_{m\to\infty} D_a(m) = \bar{n}_1 B,$$

$$\lim_{m\to\infty} D_b(m) = \bar{n}_2 B,$$

$$\lim_{m\to\infty} D_c(m) = 0,$$

$$\lim_{m\to\infty} D_d(m) = 0.$$

Proof: See Appendix C.4. □

Theorem 5.2 *As $m \to \infty$, assuming that $\bar{n}_1 B < \bar{n}_2 B$, the boundary of the capacity region for Nakagami-m fading broadcast channel (5.20) becomes*

$$\bar{P} = 2^{R_2/B}(2^{R_1/B} - 1)\bar{n}_1 B + (2^{R_2/B} - 1)\bar{n}_2 B, \tag{5.25}$$

which is the same as the boundary of the capacity region for the AWGN broadcast channel using CD with successive decoding.

Proof: Applying *Lemma 5.2* to (5.20) directly yields (5.25). For the two-user degraded AWGN broadcast channel with noise variances $\bar{n}_1 B$ and $\bar{n}_2 B$ ($\bar{n}_1 B < \bar{n}_2 B$), the capacity region for CD with successive decoding is [78, 79]:

$$\mathcal{C}_{CD} = \left\{ R_1 \leq B \log\left(1 + \frac{P_1}{\bar{n}_1 B}\right), R_2 \leq B \log\left(1 + \frac{P_2}{\bar{n}_2 B + P_1}\right); \forall P_1 + P_2 \leq \bar{P} \right\}. \tag{5.26}$$

Therefore, if $\mathbf{R}$ is on the boundary of the capacity region (5.26), i.e., all the equalities in (5.26) are achieved, then

$$\begin{aligned} \bar{P} &= P_1 + P_2 \\ &= (2^{R_1/B} - 1)\bar{n}_1 B + (2^{R_2/B} - 1)[\bar{n}_2 B + (2^{R_1/B} - 1)\bar{n}_1 B] \\ &= 2^{R_2/B}(2^{R_1/B} - 1)\bar{n}_1 B + (2^{R_2/B} - 1)\bar{n}_2 B, \end{aligned}$$

which means that $\mathbf{R}$ also satisfies (5.25). □

CD without Successive Decoding

For a two-user broadcast channel with fading, given rate vector $\mathbf{R} = (R_1, R_2)$, we know by (5.12) that the minimum required total power to support $\mathbf{R}$ in a fading state $\mathbf{n} = (n_1, n_2)$ is

$$P^{min}(\mathbf{R}, \mathbf{n}) = \frac{\left(2^{(R_1+R_2)/B} - 2^{R_2/B}\right) n_1 B + \left(2^{(R_1+R_2)/B} - 2^{R_1/B}\right) n_2 B}{2^{R_1/B} + 2^{R_2/B} - 2^{(R_1+R_2)/B}}. \tag{5.27}$$

If $\mathbf{R}$ is on the boundary surface of $\mathcal{C}_{zero}(\bar{P})$ in (5.5), substituting (5.27) into (5.4) with equality we obtain

$$\begin{aligned}\bar{P} &= \frac{\left(2^{(R_1+R_2)/B} - 2^{R_2/B}\right) E_{n_1}[n_1 B] + \left(2^{(R_1+R_2)/B} - 2^{R_1/B}\right) E_{n_2}[n_2 B]}{2^{R_1/B} + 2^{R_2/B} - 2^{(R_1+R_2)/B}} \\ &= \frac{\left(2^{(R_1+R_2)/B} - 2^{R_2/B}\right) [D_a(m) + D_d(m)]}{2^{R_1/B} + 2^{R_2/B} - 2^{(R_1+R_2)/B}} \\ &+ \frac{\left(2^{(R_1+R_2)/B} - 2^{R_1/B}\right) [D_b(m) + D_c(m)]}{2^{R_1/B} + 2^{R_2/B} - 2^{(R_1+R_2)/B}}, \end{aligned} \tag{5.28}$$

where $D_a(m)$, $D_b(m)$, $D_c(m)$ and $D_d(m)$ are as defined in (5.21)-(5.24) for the Nakagami-m fading channel. From *Lemma 5.2* we know that

$$\lim_{m\to\infty} [D_a(m) + D_d(m)] = \bar{n}_1 B,$$

$$\lim_{m\to\infty} [D_b(m) + D_c(m)] = \bar{n}_2 B.$$

Thus, as $m \to \infty$, (5.28) becomes

$$\bar{P} = \frac{\left(2^{(R_1+R_2)/B} - 2^{R_2/B}\right) \bar{n}_1 B + \left(2^{(R_1+R_2)/B} - 2^{R_1/B}\right) \bar{n}_2 B}{2^{R_1/B} + 2^{R_2/B} - 2^{(R_1+R_2)/B}}, \tag{5.29}$$

which is the same as the boundary of the capacity region for the AWGN broadcast channel using CD without successive decoding, since for the two-user degraded AWGN broadcast channel with noise variances $\bar{n}_1 B$ and $\bar{n}_2 B$ ($\bar{n}_1 B < \bar{n}_2 B$), the capacity region for CD without successive decoding is [75]:

$$\mathcal{C}_{CDWO} = \left\{ R_1 \le B \log\left(1 + \frac{P_1}{\bar{n}_1 B + P_2}\right), R_2 \le B \log\left(1 + \frac{P_2}{\bar{n}_2 B + P_1}\right); \forall P_1 + P_2 \le \bar{P} \right\}.$$

5.4 Outage Capacities and Minimum Outage Probability

In the previous section we obtained the zero-outage capacity region of an M-user broadcast channel, where the transmitter was required to maintain a constant rate for each user no matter how severe its fading. We now consider the outage capacity region, where the transmitter may suspend transmission over a subset of fading states with a given probability. Specifically, for a given average power constraint $\bar{P}$, the outage capacity regions $C_{out}(\bar{P}, Pr)$

and $C_{out}(\bar{P}, \mathbf{Pr})$ are defined as follows:

Definition 5.2 *Assuming that the transmission to all users is turned on or off simultaneously so that the outage probability for each user is the same (common outage probability), for a given* $0 \leq Pr \leq 1$*, the outage capacity region* $C_{out}(\bar{P}, Pr)$ *consists of all rate vectors* $\mathbf{R} = (R_1, R_2, \cdots, R_M)$ *which can be maintained with a common outage probability no larger than* Pr *under the the power constraint* $\bar{P}$*.*

Definition 5.3 *Assuming that the transmission to each user is turned on or off independently so that the outage probability for each user may be different, for a given probability vector* $\mathbf{Pr} = (Pr_1, Pr_2, \cdots, Pr_M)$*, the outage capacity region* $C_{out}(\bar{P}, \mathbf{Pr})$ *consists of all rate vectors* $\mathbf{R} = (R_1, R_2, \cdots, R_M)$ *which can be maintained with the outage probability for User* j *no larger than* Pr_j $(\forall 1 \leq j \leq M)$ *under the the given power constraint* $\bar{P}$*.*

With these definitions, we wish to find: a) the optimal resource allocation strategy that achieves the boundary of the outage capacity region $C_{out}(\bar{P}, Pr)$; b) the optimal resource allocation strategy that achieves the boundary of $C_{out}(\bar{P}, \mathbf{Pr})$. The first optimization problem is equivalent to deriving the resource allocation policy that minimizes the common outage probability for a given rate vector $\mathbf{R}$ and we have the following definition for the corresponding minimum common outage probability $Pr_{min}(\bar{P}, \mathbf{R})$:

Definition 5.4 *Assuming that the transmission to all users is turned on or off simultaneously, the minimum common outage probability* $Pr_{min}(\bar{P}, \mathbf{R})$ *is the smallest common outage probability with which the rate vector* $\mathbf{R}$ *can be maintained under the given power constraint* $\bar{P}$*.*

The second optimization problem is equivalent to obtaining the resource allocation policy that achieves the boundary of the outage probability region $\mathcal{O}(\bar{P}, \mathbf{R})$ or the usage probability region $\bar{\mathcal{O}}(\bar{P}, \mathbf{R})$ defined as follows:

Definition 5.5 *Assuming that the transmission to each user is turned on or off independently, for a given rate vector* $\mathbf{R}$*, the outage probability region* $\mathcal{O}(\bar{P}, \mathbf{R})$ *consists of all outage probability vectors* $\mathbf{Pr}$ *for which* $\mathbf{R}$ *can be maintained for the* M *users under the given power constraint* $\bar{P}$*.*

Definition 5.6 *The usage probability region $\bar{\mathcal{O}}(\bar{P}, \mathbf{R})$ is the complementary region of the outage probability region $\mathcal{O}(\bar{P}, \mathbf{R})$, i.e., if a probability vector $\mathbf{Pr} = (Pr_1, Pr_2, \cdots, Pr_M) \in \mathcal{O}(\bar{P}, \mathbf{R})$, then the probability vector $\mathbf{Pr}^{on} = (Pr_1^{on}, Pr_2^{on}, \cdots, Pr_M^{on}) \in \bar{\mathcal{O}}(\bar{P}, \mathbf{R})$, where*

$$Pr_j^{on} = 1 - Pr_j, \quad \forall 1 \le j \le M.$$

With the above definitions, it is easily seen that given $0 \le Pr \le 1$, the outage capacity region $C_{out}(\bar{P}, Pr)$ is implicitly obtained once the minimum common outage probability $Pr_{min}(\bar{P}, \mathbf{R})$ for a given rate vector is calculated under the optimal resource allocation, since $\forall \mathbf{R}$, we can determine that $\mathbf{R} \in C_{out}(\bar{P}, Pr)$ if $Pr_{min}(\bar{P}, \mathbf{R}) \le Pr$, and $\mathbf{R} \notin C_{out}(\bar{P}, Pr)$ otherwise. Similarly, given a probability vector $\mathbf{Pr}$, the outage capacity region $C_{out}(\bar{P}, \mathbf{Pr})$ is implicitly obtained once the boundary of the outage probability region $\mathcal{O}(\bar{P}, \mathbf{R})$ (and so the whole region $\mathcal{O}(\bar{P}, \mathbf{R})$) for a given rate vector $\mathbf{R}$ is derived through the optimal resource allocation, since $\forall \mathbf{R}$, we can determine that $\mathbf{R} \in C_{out}(\bar{P}, \mathbf{Pr})$ if $\mathbf{Pr} \in \mathcal{O}(\bar{P}, \mathbf{R})$, and $\mathbf{R} \notin C_{out}(\bar{P}, \mathbf{Pr})$ otherwise. We now derive the minimum common outage probability $Pr_{min}(\bar{P}, \mathbf{R})$ and the corresponding optimal resource allocation strategy in Section 5.4.1. We obtain the outage probability region boundary of $\mathcal{O}(\bar{P}, \mathbf{R})$ as well as the optimal resource allocation strategy in Section 5.4.2 for the case of independent outage problems.

5.4.1 Minimum Common Outage Probability

Under the assumption that an outage is declared for all users simultaneously, the minimum common outage probability problem for the M-user broadcast channel is similar to that of the single user case [76]. For each joint fading state $\mathbf{n}$ and a given rate vector $\mathbf{R}$, the minimum required total power $P^{min}(\mathbf{R}, \mathbf{n})$ for the M users using CD with or without successive decoding or using TD can be calculated as in (5.3), (5.11) or (5.15), respectively. Thus, $\forall s > 0$, we define the sets of fading states $\mathcal{R}(s)$ and $\tilde{\mathcal{R}}(s)$ as:

$$\mathcal{R}(s) = \{\mathbf{n} : P^{min}(\mathbf{R}, \mathbf{n}) < s\}, \tag{5.30}$$

$$\tilde{\mathcal{R}}(s) = \{\mathbf{n} : P^{min}(\mathbf{R}, \mathbf{n}) \le s\}. \tag{5.31}$$

The corresponding average power over the two sets are:

$$P(s) = E_{\mathbf{n}\in\mathcal{R}(s)} P^{min}(\mathbf{R}, \mathbf{n}), \tag{5.32}$$

$$\tilde{P}(s) = E_{\mathbf{n}\in\tilde{\mathcal{R}}(s)} P^{min}(\mathbf{R}, \mathbf{n}). \tag{5.33}$$

For a given total power $\bar{P} > 0$, let

$$s^* \triangleq \sup\left\{s : P(s) < \bar{P}\right\},$$

$$w^* \triangleq \frac{\bar{P} - P(s^*)}{\tilde{P}(s^*) - P(s^*)}. \tag{5.34}$$

By using *Lemma 3* in [76], for each fading state $\mathbf{n}$, the optimal power policy that minimizes the common outage probability is: if $\mathbf{n} \notin \tilde{\mathcal{R}}(s^*)$, no power is assigned to any user; if $\mathbf{n} \in \mathcal{R}(s^*)$, a total power of $P^{min}(\mathbf{R}, \mathbf{n})$ is assigned to the M users and the power to each user is allocated as described in Section 5.3; if $\mathbf{n} \notin \mathcal{R}(s^*)$ but $\mathbf{n} \in \tilde{\mathcal{R}}(s^*)$, then with probability w^*, $P^{min}(\mathbf{R}, \mathbf{n})$ is assigned to the M users and with probability $1-w^*$, no power is assigned to any user. The minimum common outage probability $Pr_{min}(\bar{P}, \mathbf{R})$ is:

$$Pr_{min}(\bar{P}, \mathbf{R}) = 1 - Pr\left\{\mathbf{n} \in \mathcal{R}(s^*)\right\} - w^* Pr\left\{\mathbf{n} \in \tilde{\mathcal{R}}(s^*) \text{ and } \mathbf{n} \notin \mathcal{R}(s^*)\right\}, \tag{5.35}$$

where $Pr\{\cdot\}$ denotes the probability function.

5.4.2 Outage Probability Region

We now consider the case where an outage can be declared independently for each user. From *Definition 5.5* and *Definition 5.6*, it is clear that for a given rate vector $\mathbf{R}$ and an average power constraint $\bar{P}$, deriving the boundary of the outage probability region $\mathcal{O}(\bar{P}, \mathbf{R})$ is equivalent to deriving the boundary of the usage probability region $\bar{\mathcal{O}}(\bar{P}, \mathbf{R})$. We will require the following definition and lemma to derive the boundary of $\bar{\mathcal{O}}(\bar{P}, \mathbf{R})$ and the corresponding optimal power allocation that achieves this boundary:

Definition 5.7 *For a given rate vector* $\mathbf{R} = (R_1, R_2, \cdots, R_M)$*, assume that rate* R_i *is maintained with probability* $Pr_i^{on}(\mathbf{R})$*,* $1 \le i \le M$*. Denoting* $\mathbf{Pr}^{on}(\mathbf{R}) = [Pr_1^{on}(\mathbf{R}), Pr_2^{on}(\mathbf{R}), \cdots,$

$Pr_M^{on}(\mathbf{R})]$, the total usage reward $W(\mathbf{R})$ is

$$W(\mathbf{R}) = \boldsymbol{\mu}\mathbf{Pr}^{on}(\mathbf{R}) = \sum_{i=1}^{M} \mu_i Pr_i^{on}(\mathbf{R}),$$

where $\boldsymbol{\mu} \in \Re_+^M$ with $\sum_{j=1}^{M} \mu_j = 1$, and μ_i is the relative reward if the information for User i is transmitted[4].

Lemma 5.3 *The usage probability region $\bar{O}(\bar{P}, \mathbf{R})$ of the fading broadcast channel is convex.*

Proof: See Appendix C.5. □

Since $\bar{O}(\bar{P}, \mathbf{R})$ is convex, $\forall \boldsymbol{\mu} \in \Re_+^M$ with $\sum_{j=1}^{M} \mu_j = 1$, a usage probability vector $\mathbf{Pr}^{on}(\mathbf{R})$ will be on the boundary surface of $\bar{O}(\bar{P}, \mathbf{R})$ if it is a solution to

$$\max_{\mathbf{Pr}^{on}(\mathbf{R}) \in \bar{O}(\bar{P}, \mathbf{R})} W(\mathbf{R}), \tag{5.36}$$

where the total usage reward $W(\mathbf{R})$ is defined in *Definition 5.7.*

For any given fading state $\mathbf{n}$, $\sum_{i=1}^{M} \binom{M}{i} = 2^M - 1$ different combinations of the M users may be transmitting over the channel. We will represent each of these $2^M - 1$ possible combinations of users as a vector $[\psi(k,1), \psi(k,2), \cdots, \psi(k,M)]$ equal to the binary expansion of k, $1 \leq k \leq 2^M - 1$. For each vector $[\psi(k,1), \psi(k,2), \cdots, \psi(k,M)]$, if $\psi(k,i) = 1$, then User i is transmitting; otherwise User i is not. For $0 \leq k \leq 2^M - 1$, we define the set of active users $\mathcal{U}_k$ relative to k as $\mathcal{U}_k = \{j : \psi(k,j) = 1, 1 \leq j \leq M\}$. $\mathcal{U}_0$ denotes the empty set (no active users). For any fading state $\mathbf{n}$ suppose that the broadcast channel only transmits information to users in the non-empty set $\mathcal{U}_k$. Then, as discussed in Section 5.3, we can calculate the minimum total power $P_k^{min}(\mathbf{R}, \mathbf{n})$ $(1 \leq k \leq 2^M - 1)$ required to support a sub-vector of $\mathbf{R}$ composed of the required rates of the users in $\mathcal{U}_k$ under those different spectrum-sharing techniques. For the fading state $\mathbf{n}$ let $w_k(\mathbf{R}, \mathbf{n})$ denote the probability that the broadcast channel transmits information to the subset of users in $\mathcal{U}_k$. Then obviously

$$\sum_{k=1}^{2^M-1} w_k(\mathbf{R}, \mathbf{n}) = 1 - w_0(\mathbf{R}, \mathbf{n}) \leq 1.$$

[4] μ_i can also be viewed as the relative penalty if an outage is declared for User i.

For a given rate vector $\mathbf{R}$ and a fading state $\mathbf{n}$, let $Pr_i^{on}(\mathbf{R}, \mathbf{n})$ be the probability that information is sent to User i:

$$Pr_i^{on}(\mathbf{R}, \mathbf{n}) = \sum_{k=1}^{2^M-1} w_k(\mathbf{R}, \mathbf{n}) \mathbf{1}[i \in \mathcal{U}_k], \tag{5.37}$$

where $\mathbf{1}[\cdot]$ denotes the indicator function. Then the average outage probability $Pr_i(\mathbf{R})$ of User i $(1 \leq i \leq M)$ is:

$$\begin{aligned} Pr_i(\mathbf{R}) &= 1 - Pr_i^{on}(\mathbf{R}) \\ &= 1 - E_{\mathbf{n}} \left[Pr_i^{on}(\mathbf{R}, \mathbf{n}) \right]. \end{aligned}$$

For a given fading state $\mathbf{n}$, according to (5.37), the total usage reward $W(\mathbf{R}, \mathbf{n})$ is

$$\begin{aligned} W(\mathbf{R}, \mathbf{n}) &= \sum_{i=1}^{M} \mu_i Pr_i^{on}(\mathbf{R}, \mathbf{n}) \\ &= \sum_{i=1}^{M} \mu_i \left(\sum_{k=1}^{2^M-1} w_k(\mathbf{R}, \mathbf{n}) \mathbf{1}[i \in \mathcal{U}_k] \right) \\ &= \sum_{k=1}^{2^M-1} w_k(\mathbf{R}, \mathbf{n}) \left(\sum_{i=1}^{M} \mu_i \mathbf{1}[i \in \mathcal{U}_k] \right) \\ &= \sum_{k=1}^{2^M-1} w_k(\mathbf{R}, \mathbf{n}) \eta_k, \end{aligned} \tag{5.38}$$

where the reward for transmitting information to the users in set $\mathcal{U}_k$ is

$$\eta_k \triangleq \sum_{i=1}^{M} \mu_i \mathbf{1}[i \in \mathcal{U}_k]. \tag{5.39}$$

Thus, the total usage reward averaged over the time-varying channel is

$$W(\mathbf{R}) = E_{\mathbf{n}} \left[W(\mathbf{R}, \mathbf{n}) \right]. \tag{5.40}$$

Since in fading state $\mathbf{n}$, the total required minimum power to support $\mathbf{R}$ with usage probability $Pr_i^{on}(\mathbf{R}, \mathbf{n})$ for each user i $(\forall 1 \leq i \leq M)$ is

$$P^{min}(\mathbf{R}, \mathbf{n}) = \sum_{k=1}^{2^M-1} P_k^{min}(\mathbf{R}, \mathbf{n}) w_k(\mathbf{R}, \mathbf{n}), \tag{5.41}$$

the total required minimum average power to achieve $W(\mathbf{R})$ will be $E_{\mathbf{n}}\left[P^{min}(\mathbf{R},\mathbf{n})\right]$.

For a given rate vector $\mathbf{R}$, we wish to solve the maximization problem (5.36), which is equivalent to finding the optimal $w_k^*(\mathbf{R},\mathbf{n})$ ($1 \le k \le 2^M - 1$, $\forall \mathbf{n} \in \mathcal{N}$) that maximizes $W(\mathbf{R})$ in (5.40) under the total power constraint. That is, we can re-write the maximization problem (5.36) as:

$$\begin{cases} \max_{\mathbf{w}(\mathbf{R},\mathbf{n})} E_{\mathbf{n}}\left[W(\mathbf{R},\mathbf{n})\right] \quad \text{subject to:} \\ \\ E_{\mathbf{n}}\left[P^{min}(\mathbf{R},\mathbf{n})\right] \le \bar{P}, \quad \sum_{k=1}^{2^M-1} w_k(\mathbf{R},\mathbf{n}) \le 1 \quad \text{and} \quad 0 \le w_k(\mathbf{R},\mathbf{n}) \le 1, \end{cases} \tag{5.42}$$

where $\mathbf{w}(\mathbf{R},\mathbf{n}) = [w_1(\mathbf{R},\mathbf{n}), w_2(\mathbf{R},\mathbf{n}), \cdots, w_N(\mathbf{R},\mathbf{n})]$ with $N \triangleq 2^M - 1$, $W(\mathbf{R},\mathbf{n})$ and $P^{min}(\mathbf{R},\mathbf{n})$ are as given in (5.38) and (5.41), respectively and $\bar{P}$ is the total average transmit power. The maximization problem (5.42) can be decomposed into the following two problems:

1. Assuming that $\forall \mathbf{n} \in \mathcal{N}$, $P(\mathbf{n})$ is the total average power assigned to the N sets of users in state $\mathbf{n}$, i.e., $P(\mathbf{n}) = \sum_{k=1}^{N} w_k(\mathbf{R},\mathbf{n}) P_k^{min}(\mathbf{R},\mathbf{n})$, we must choose $\mathbf{w}(\mathbf{R},\mathbf{n})$ so that the total usage reward in state $\mathbf{n}$ is maximized. That is, we must find

$$\begin{cases} J(P(\mathbf{n})) \triangleq \max\limits_{\mathbf{w}(\mathbf{R},\mathbf{n})} \sum_{k=1}^{N} w_k(\mathbf{R},\mathbf{n}) \eta_k \quad \text{subject to:} \\ \\ \sum_{k=1}^{N} w_k(\mathbf{R},\mathbf{n}) P_k^{min}(\mathbf{R},\mathbf{n}) \le P(\mathbf{n}), \ \sum_{k=1}^{N} w_k(\mathbf{R},\mathbf{n}) \le 1, \ \text{and} \ 0 \le w_k(\mathbf{R},\mathbf{n}) \le 1, \end{cases} \tag{5.43}$$

   where η_k is given in (5.39).

2. After we obtain the expression $J(\cdot)$ by solving (5.43), the remaining problem is how to assign the total power $P(\mathbf{n})$ of the N sets of users for each state $\mathbf{n}$ so that the total usage reward averaged over all fading states as expressed in (5.40) is maximized. That is,

$$\begin{cases} \max\limits_{P(\mathbf{n})} E_{\mathbf{n}}[J(P(\mathbf{n}))] - \frac{1}{s} E_{\mathbf{n}}[P(\mathbf{n})] \\ \text{subject to } E_{\mathbf{n}}[P(\mathbf{n})] \le \bar{P} \end{cases}, \tag{5.44}$$

   where $\frac{1}{s}$ is the Lagrangian multiplier.

We solve the maximization problem (5.43) by first defining the permutation $\pi(\cdot)$ such that

$$0 < \eta_{\pi(1)} \leq \eta_{\pi(2)} \leq \cdots \leq \eta_{\pi(N)}.$$

For simplicity we denote the reward and power needed for transmitting information to the users in set $\mathcal{U}_{\pi(i)}$ as λ_i and v_i, respectively, where

$$\begin{cases} \lambda_i \triangleq \eta_{\pi(i)}, \\ v_i \triangleq P_{\pi(i)}^{min}(\mathbf{R}, \mathbf{n}), \end{cases} \quad \forall 1 \leq i \leq N. \tag{5.45}$$

Note that the power v_i's are all functions of rate vector $\mathbf{R}$ and fading state $\mathbf{n}$. For a given state $\mathbf{n}$, $\forall 1 \leq k \leq N$, if $\exists j$ that satisfies

$$\frac{\lambda_k}{v_k} \leq \frac{\lambda_j}{v_j}, \quad k < j \leq N, \tag{5.46}$$

or satisfies

$$\begin{cases} \lambda_j = \lambda_k \\ \frac{\lambda_j}{v_j} > \frac{\lambda_k}{v_k} \end{cases}, \quad j = k - 1, \tag{5.47}$$

then we will get a larger reward by assigning the same average power $P(\mathbf{n})$ to set $\mathcal{U}_{\pi(j)}$ instead of set $\mathcal{U}_{\pi(k)}$, $\forall P(\mathbf{n}) > 0$. Specifically, if (5.46) is true, since $\lambda_k \leq \lambda_j$ when $k < j$, if $v_k \geq v_j$, then obviously by transmitting information to users in set $\mathcal{U}_{\pi(k)}$, we need more power and get less reward than transmitting information to users in set $\mathcal{U}_{\pi(j)}$. If (5.46) is true and $v_k < v_j$ then, assuming that $w^*_{\pi(k)}(\mathbf{R}, \mathbf{n}) \neq 0$ ($0 < w^*_{\pi(k)}(\mathbf{R}, \mathbf{n}) \leq 1$), the reward we get from assigning power v_k to set $\mathcal{U}_{\pi(k)}$ with a fraction $w^*_{\pi(k)}(\mathbf{R}, \mathbf{n})$ of the transmission time in state $\mathbf{n}$ is $\lambda_k w^*_{\pi(k)}(\mathbf{R}, \mathbf{n})$ and the average power needed is $P(\mathbf{n}) = v_k w^*_{\pi(k)}(\mathbf{R}, \mathbf{n})$, while the reward we get from the same average power $P(\mathbf{n})$ by assigning power v_j to set $\mathcal{U}_{\pi(j)}$ with a fraction $\frac{P(\mathbf{n})}{v_j} = \frac{v_k w^*_{\pi(k)}(\mathbf{R},\mathbf{n})}{v_j}$ of the transmission time will be $\lambda_j \cdot \frac{v_k w^*_{\pi(k)}(\mathbf{R},\mathbf{n})}{v_j}$, which is larger than $\lambda_k w^*_{\pi(k)}(\mathbf{R}, \mathbf{n})$ by (5.46). Moreover, the fraction of transmission time needed for set $\mathcal{U}_{\pi(j)}$ is less than that for set $\mathcal{U}_{\pi(k)}$, since when $v_k < v_j$,

$$\frac{v_k w^*_{\pi(k)}(\mathbf{R}, \mathbf{n})}{v_j} < w^*_{\pi(k)}(\mathbf{R}, \mathbf{n}).$$

If (5.47) is true, it is obvious that $v_j < v_k$. Thus, to obtain the same reward $\lambda_j = \lambda_k$, set $\mathcal{U}_{\pi(j)}$ requires less power than set $\mathcal{U}_{\pi(k)}$. Therefore, in order to get the largest average reward under the given average power constraint, we do not consider assigning any power to those sets $\mathcal{U}_{\pi(k)}$ for which $\exists j$ satisfying either (5.46) or (5.47). That is, we remove them from further consideration and set $w^*_{\pi(k)}(\mathbf{R}, \mathbf{n}) = 0$.

For example, when $N = 3$ and the relative values of v_1, v_2, v_3 and λ_1, λ_2, λ_3 are as shown in Figure 5.1, where λ_i and v_i correspond to the reward and power needed for set $\mathcal{U}_{\pi(i)}$ ($i = 1, 2, 3$), respectively, it is obvious that $\frac{\lambda_1}{v_1} < \frac{\lambda_2}{v_2}$. Thus, if the available power $P(\mathbf{n}) \leq v_1$ and we assign it to set $\mathcal{U}_{\pi(1)}$, the reward we can get is the straight line OA; if $v_1 < P(\mathbf{n}) < v_2$ and we assign it to sets $\mathcal{U}_{\pi(1)}$ and $\mathcal{U}_{\pi(2)}$ by time-sharing, the reward we can get is the straight line AB. However, in both cases, if we assign the power $P(\mathbf{n})$ to set $\mathcal{U}_{\pi(2)}$, we get a larger reward which is indicated by the straight line OB. Thus, no power should be assigned to set $\mathcal{U}_{\pi(1)}$ and $J(P(\mathbf{n}))$ defined in (5.43) is as shown by the solid curve in Figure 5.1.

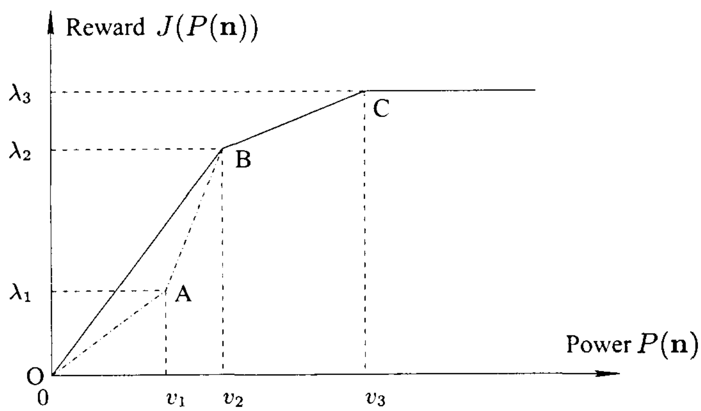


Figure 5.1: Power $P(\mathbf{n})$ vs. Reward $J(P(\mathbf{n}))$ for $N = 3$.

Generally, for the remaining sets of users, it is possible that there are still some sets $\mathcal{U}_{\pi(k)}$ to which no power should be assigned in order to get the largest average reward. For example, suppose that the remaining sets are $\mathcal{U}_{\pi(1)}$, $\mathcal{U}_{\pi(2)}$ and $\mathcal{U}_{\pi(3)}$, and the relative values of v_1, v_2, v_3 and λ_1, λ_2, λ_3 are as shown in Figure 5.2. From this figure we see that neither (5.46) nor (5.47) is satisfied for any $k = 1, 2, 3$, since $\lambda_1 < \lambda_2 < \lambda_3$ and

$$\frac{\lambda_1}{v_1} > \frac{\lambda_2}{v_2} > \frac{\lambda_3}{v_3}.$$

However, it is obvious that no power should be assigned to set $\mathcal{U}_{\pi(2)}$, because a larger reward can be obtained when the same average power is time-shared by sets $\mathcal{U}_{\pi(1)}$ and $\mathcal{U}_{\pi(3)}$ instead of by sets $\mathcal{U}_{\pi(1)}$ and $\mathcal{U}_{\pi(2)}$, or by sets $\mathcal{U}_{\pi(2)}$ and $\mathcal{U}_{\pi(3)}$.

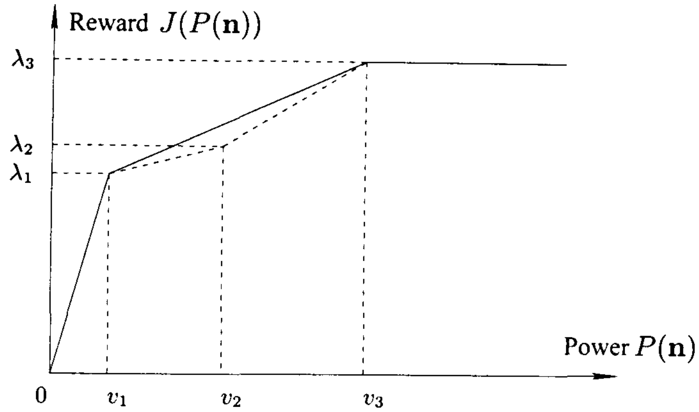


Figure 5.2: Power $P(\mathbf{n})$ vs. Reward $J(P(\mathbf{n}))$ for three remaining sets of users.

In the following, we use an iterative procedure to find all the sets $\mathcal{U}_{\pi(k)}$ in the remaining sets that should be assigned no power and remove them from further consideration (i.e., let $w^*_{\pi(k)}(\mathbf{R}, \mathbf{n}) = 0$). An interpretation of this procedure based on Figure 5.2 will be given shortly.

Initialization: Let $m = 1$.

Step 1: Denote the number of remaining sets as G_u and let the permutation $\rho(\cdot)$ be defined such that for the remaining G_u sets,

$$\lambda_{\rho(1)} < \lambda_{\rho(2)} < \cdots < \lambda_{\rho(G_u)}. \tag{5.48}$$

Due to the removal criterion, it must be true that

$$\frac{\lambda_{\rho(1)}}{v_{\rho(1)}} > \frac{\lambda_{\rho(2)}}{v_{\rho(2)}} > \cdots > \frac{\lambda_{\rho(G_u)}}{v_{\rho(G_u)}}. \tag{5.49}$$

Step 2: Let $\rho_0(m) = \rho(1)$ and $z_{\rho_0(m)} = \frac{\lambda_{\rho(1)}}{v_{\rho(1)}}$. If $G_u < 2$, all the sets that should be assigned no power have been removed and the procedure terminates; if $G_u \geq 2$, go to Step 3.

Step 3: For $2 \leq k \leq G_u$, decrease $\lambda_{\rho(k)}$ and $v_{\rho(k)}$ by $\lambda_{\rho(1)}$ and $v_{\rho(1)}$, respectively. Do not assign any power to those sets of users $\mathcal{U}_{\pi[\rho(k)]}$ for which $\exists j$ that satisfies

$$\frac{\lambda_{\rho(k)}}{v_\rho(k)} \leq \frac{\lambda_{\rho(j)}}{v_\rho(j)}, \quad k < j \leq G_u,$$

and remove them from further consideration (i.e., let $w^*_{\pi[\rho(k)]}(\mathbf{R}, \mathbf{n}) = 0$). Also remove set $\mathcal{U}_{\pi[\rho(1)]}$. Increase m by 1 and return to Step 1.

In this procedure we observe that in the first iteration, from (5.48) and (5.49) it is clear that

$$v_{\rho(1)} < v_{\rho(2)} < \cdots < v_{\rho(G_u)}.$$

Since $\frac{\lambda_{\rho(1)}}{v_{\rho(1)}} > \frac{\lambda_{\rho(i)}}{v_{\rho(i)}}$ for all $1 < j \leq G_u$, when the total average power $P(\mathbf{n}) \leq v_{\rho(1)}$, we get the largest reward $\frac{\lambda_{\rho(1)}}{v_{\rho(1)}} \cdot P(\mathbf{n})$ by assigning it to set $\mathcal{U}_{\pi[\rho(1)]}$. That is, in (5.43), as shown in the example in Figure 5.2[5],

$$J(P(\mathbf{n})) = \frac{\lambda_{\rho(1)}}{v_{\rho(1)}} \cdot P(\mathbf{n}), \quad \text{if } P(\mathbf{n}) \leq v_{\rho(1)},$$

and we store the index $\rho(1)$ of set $\mathcal{U}_{\pi[\rho(1)]}$ and the tangent $\frac{\lambda_{\rho(1)}}{v_{\rho(1)}}$ in $\rho_0(1)$ and $z_{\rho_0(1)}$, respectively. Next we wish to identify those sets $\mathcal{U}_{\pi[\rho(k)]}$ $(1 < k \leq G_u)$ for which $\exists j$ that satisfies

$$\frac{\lambda_{\rho(k)} - \lambda_{\rho(1)}}{v_{\rho(k)} - v_{\rho(1)}} < \frac{\lambda_{\rho(j)} - \lambda_{\rho(1)}}{v_{\rho(j)} - v_{\rho(1)}}, \quad k < j \leq G_u, \tag{5.50}$$

and we do not assign any power to them since, for $k < j$, $v_{\rho(k)} < v_{\rho(j)}$, $\lambda_{\rho(k)} < \lambda_{\rho(j)}$, and if (5.50) is true, on the Power-Reward plane as shown in Figure 5.2, point $(v_{\rho(k)}, \lambda_{\rho(k)})$ will be under the straight line formed by connecting point $(v_{\rho(1)}, \lambda_{\rho(1)})$ and point $(v_{\rho(j)}, \lambda_{\rho(j)})$[6]. Therefore, in (5.43),

$$J(P(\mathbf{n})) = \frac{\lambda_{\rho(i^*)} - \lambda_{\rho(1)}}{v_{\rho(i^*)} - v_{\rho(1)}} \cdot [P(\mathbf{n}) - v_{\rho(1)}], \quad \text{if } v_{\rho(1)} < P(\mathbf{n}) \leq v_{\rho(i^*)},$$

[5]Note that in this example, $G_u = 3$ and $\rho(i) = i$, $i = 1, 2, 3$.

[6]In Figure 5.2, point (v_2, λ_2) is under the straight line formed by connecting point (v_1, λ_1) and point (v_3, λ_3).

where the index i^* is given by:

$$i^* = \arg\max_{1<i\leq G_u} \left\{ \frac{\lambda_{\rho(i)} - \lambda_{\rho(1)}}{v_{\rho(i)} - v_{\rho(1)}} \right\}. \tag{5.51}$$

After removing those sets $\mathcal{U}_{\pi[\rho(k)]}$ (i.e., let $w^*_{\pi[\rho(k)]}(\mathbf{R},\mathbf{n}) = 0$) for which (5.50) holds and also removing set $\mathcal{U}_{\pi[\rho(1)]}$, the index $\rho(i^*)$ in the first iteration becomes $\rho(1)$ of a new permutation $\rho(\cdot)$ in the second iteration (otherwise (5.51) cannot be true) and is stored in $\rho_0(2)$, and the corresponding tangent $\frac{\lambda_{\rho(i^*)}-\lambda_{\rho(1)}}{v_{\rho(i^*)}-v_{\rho(1)}}$ is stored in $z_{\rho_0(2)}$. In the second iteration, similarly, the new index $\rho(i^*)$ that satisfies (5.51) will be identified, and new set(s) $\mathcal{U}_{\pi[\rho(k)]}$ for which (5.50) holds will be removed, since the point(s) $(v_{\rho(k)}, \lambda_{\rho(k)})$ will be under the straight line formed by connecting point $(v_{\rho(1)}, \lambda_{\rho(1)})$ and $(v_{\rho(j)}), \lambda_{\rho(j)})$, where j satisfies (5.50). The new index $\rho(i^*)$ will become $\rho(1)$ in the third iteration and be stored in $\rho_0(3)$, and the corresponding tangent $\frac{\lambda_{\rho(i^*)}-\lambda_{\rho(1)}}{v_{\rho(i^*)}-v_{\rho(1)}}$ will be stored in $z_{\rho_0(3)}$. The iterative procedure continues until all the sets that should be assigned no power have been removed and the curve of $J(P(\mathbf{n}))$ in (5.43) is obtained by connecting the origin (0,0) and points $(v_{\rho_0(1)}, \lambda_{\rho_0(1)})$-$(v_{\rho_0(m_0)}, \lambda_{\rho_0(m_0)})$, where m_0 is the value of m when the iteration stops $(1 \leq m_0 \leq N)$. That is,

$$J(P(\mathbf{n})) = \begin{cases} z_{\rho_0(1)} \cdot P(\mathbf{n}), & 0 < P(\mathbf{n}) \leq v_{\rho_0(1)}, \\ z_{\rho_0(j)} \cdot [P(\mathbf{n}) - v_{\rho_0(j-1)}], & v_{\rho_0(j-1)} < P(\mathbf{n}) \leq v_{\rho_0(j)} \quad \text{for some } j, 1 < j \leq m_0, \\ \lambda_{\rho_0(m_0)}, & P(\mathbf{n}) > v_{\rho_0(m_0)}. \end{cases} \tag{5.52}$$

Note that from the iterative procedure, it is clear that

$$\lambda_{\rho_0(1)} < \lambda_{\rho_0(2)} < \cdots < \lambda_{\rho_0(m_0)}, \tag{5.53}$$

$$v_{\rho_0(1)} < v_{\rho_0(2)} < \cdots < v_{\rho_0(m_0)}, \tag{5.54}$$

$$z_{\rho_0(1)} > z_{\rho_0(2)} > \cdots > z_{\rho_0(m_0)}, \tag{5.55}$$

where

$$z_{\rho_0(j)} = \begin{cases} \frac{\lambda_{\rho_0(1)}}{v_{\rho_0(1)}}, & j = 1, \\ \\ \frac{\lambda_{\rho_0(j)} - \lambda_{\rho_0(j-1)}}{v_{\rho_0(j)} - v_{\rho_0(j-1)}}, & 1 < j \le m_0. \end{cases} \tag{5.56}$$

For example, in Figure 5.2, if we execute the above three-step procedure, we will have $m_0 = 2$, $\rho_0(1) = 1$, $\rho_0(2) = 3$, $z_{\rho_0(1)} = \frac{\lambda_1}{v_1}$, and $z_{\rho_0(2)} = \frac{\lambda_3 - \lambda_1}{v_3 - v_1}$. Therefore, from (5.52) we obtain

$$J(P(\mathbf{n})) = \begin{cases} \frac{\lambda_1}{v_1} \cdot P(\mathbf{n}), & 0 < P(\mathbf{n}) \le v_1, \\ \frac{\lambda_3 - \lambda_1}{v_3 - v_1} \cdot [P(\mathbf{n}) - v_1], & v_1 < P(\mathbf{n}) \le v_3, \\ \lambda_3, & P(\mathbf{n}) > v_3, \end{cases}$$

which is exactly as shown by the solid curve in Figure 5.2.

Once the curve $J(P(\mathbf{n}))$ is obtained, from (5.44) we know that $\forall \frac{1}{s} > 0$ fixed, the optimal power $P^*(\mathbf{n})$ satisfies $J'(P^*(\mathbf{n})) = \frac{1}{s}$ if the tangent of $J(P(\mathbf{n}))$ is continuous. However, in our case, the tangent of $J(P(\mathbf{n}))$ is discrete and $P^*(\mathbf{n})$ cannot be determined directly. Therefore, we will use the following theorem to find the optimal $w^*_{\pi[\rho_0(j)]}(\mathbf{R}, \mathbf{n})$ $(1 \le j \le m_0)$ for the remaining m_0 sets $\left\{\mathcal{U}_{\pi[\rho_0(j)]}\right\}_{j=1}^{m_0}$.

Before stating the theorem, we first define some additional notations and parameters. In (5.53)-(5.56), the indices $\{\rho_0(i)\}_{i=1}^{m_0}$ are all functions of $\mathbf{n}$. Therefore, we will refer to them as $\{\rho_0(i, \mathbf{n})\}_{i=1}^{m_0}$ and for simplicity, $\forall 1 \le i \le m_0$, we denote[7]

$$\begin{cases} \lambda_i(\mathbf{n}) \triangleq \lambda_{\rho_0(i)}, \\ v_i(\mathbf{n}) \triangleq v_{\rho_0(i)}, \\ z_i(\mathbf{n}) \triangleq z_{\rho_0(i)}. \end{cases} \tag{5.57}$$

Thus, equations (5.53)-(5.56) become

$$\lambda_1(\mathbf{n}) < \lambda_2(\mathbf{n}) < \cdots < \lambda_{m_0}(\mathbf{n}), \tag{5.58}$$

$$v_1(\mathbf{n}) < v_2(\mathbf{n}) < \cdots < v_{m_0}(\mathbf{n}), \tag{5.59}$$

$$z_1(\mathbf{n}) > z_2(\mathbf{n}) > \cdots > z_{m_0}(\mathbf{n}), \tag{5.60}$$

[7]Note that $\{\rho_0(i)\}_{i=1}^{m_0}$ are also functions of the given rate vector $\mathbf{R}$; however, we only write out $\mathbf{n}$ explicity for its related variables for simplicity.

and

$$z_j(\mathbf{n}) = \begin{cases} \frac{\lambda_1(\mathbf{n})}{v_1\mathbf{n})}, & j = 1, \\ \\ \frac{\lambda_j(\mathbf{n})-\lambda_{j-1}(\mathbf{n})}{v_j(\mathbf{n})-v_{j-1}(\mathbf{n})}, & 1 < j \leq m_0. \end{cases} \tag{5.61}$$

Moreover, we define Ω_{m_0} as the set of all fading states $\mathbf{n}$ for which the final value of the loop parameter m is m_0 $(1 \leq m_0 \leq N)$ when the three-step iteration procedure terminates. By denoting $z_{m_0+1}(\mathbf{n}) \triangleq 0$, $\forall s > 0$, for $j = 1, 2, \cdots, m_0$, we define sets $L_j(m_0, s)$ and $\tilde{L_j}(m_0, s)$ as

$$L_j(m_0, s) = \left\{\mathbf{n} : \mathbf{n} \in \Omega_{m_0}, z_j(\mathbf{n}) > \frac{1}{s} > z_{j+1}(\mathbf{n})\right\},$$

$$\tilde{L_j}(m_0, s) = \left\{\mathbf{n} : \mathbf{n} \in \Omega_{m_0}, \frac{1}{s} = z_j(\mathbf{n})\right\}.$$

Thus, $\forall s > 0$,

$$\begin{aligned} \mathcal{N} &= \bigcup_{1 \leq m_0 \leq N} \Omega_{m_0} \\ &= \bigcup_{1 \leq m_0 \leq N} \bigcup_{1 \leq j \leq m_0} \left(L_j(m_0, s) \cup \tilde{L_j}(m_0, s)\right), \end{aligned}$$

where $\mathcal{N}$ is the set of all possible fading states. For a given total average power constraint $\bar{P} > 0$, define s^* as

$$s^* \triangleq sup\left\{s : P(s) < \bar{P}\right\},$$

where

$$P(s) \triangleq \sum_{m_0=1}^{N} \sum_{j=1}^{m_0} \int_{\mathbf{n} \in L_j(m_0,s)} v_j(\mathbf{n}) dF(\mathbf{n}).$$

Therefore, $\forall \mathbf{n} \in \Omega_{m_0}$, $1 \leq m_0 \leq N$, finding the optimal $w^*_{\pi[\rho_0(j)]}(\mathbf{R}, \mathbf{n})$ $(1 \leq j \leq m_0)$ in (5.42) for the remaining m_0 sets $\left\{\mathcal{U}_{\pi[\rho_0(j)]}\right\}_{j=1}^{m_0}$ after the iterative procedure is equivalent to solving the maximization problem

$$\begin{cases} \max_{\mathbf{u}(\mathbf{n})} \sum_{m_0=1}^{N} E_{\mathbf{n} \in \Omega_{m_0}} \left[\sum_{j=1}^{m_0} \lambda_j(\mathbf{n}) u_j(\mathbf{n})\right] \quad \text{subject to:} \\ \sum_{m_0=1}^{N} E_{\mathbf{n} \in \Omega_{m_0}} \left[\sum_{j=1}^{m_0} v_j(\mathbf{n}) u_j(\mathbf{n})\right] \leq \bar{P}, \quad 0 \leq \sum_{j=1}^{m_0} u_j(\mathbf{n}) \leq 1, \quad \text{and} \quad 0 \leq u_j(\mathbf{n}) \leq 1, \end{cases} \tag{5.62}$$

where $\mathbf{u}(\mathbf{n}) = [u_1(\mathbf{n}), u_2(\mathbf{n}), \cdots, u_{m_0}(\mathbf{n})]$.

Theorem 5.3 *$\forall \mathbf{n} \in \Omega_{m_0}$, $1 \le m_0 \le N$, by denoting $u_j^*(\mathbf{n})$ ($1 \le j \le m_0$) as the solution to the maximization problem (5.62), we have*

(a) if $\frac{1}{s^*} > z_1(\mathbf{n})$, then $u_j^*(\mathbf{n}) = 0$, $\forall 1 \le j \le m_0$;

(b) if $\exists j \in \{1, 2, \cdots, m_0\}$, $z_j(\mathbf{n}) > \frac{1}{s^*} > z_{j+1}(\mathbf{n})$, then $u_j^*(\mathbf{n}) = 1$, $u_i^*(\mathbf{n}) = 0$, $\forall i \ne j, 1 \le i \le m_0$;

(c) if $\exists j \in \{1, 2, \cdots, m_0\}$, $\frac{1}{s^*} = z_j(\mathbf{n})$, then $u_j^*(\mathbf{n}) = \tau^*$, $u_{j-1}^*(\mathbf{n}) = 1 - \tau^*$, $u_i^*(\mathbf{n}) = 0$, $\forall i \ne j, j-1, 1 \le i \le m_0$, where τ^* satisfies

$$P(s^*) + \sum_{m_0=1}^{N} \sum_{i=1}^{m_0} [v_j(\mathbf{n})\tau^* + v_{j-1}(\mathbf{n})(1-\tau^*)] Pr\left\{\tilde{L}_j(m_0, s^*)\right\} = \bar{P}.$$

Proof: See Appendix C.6. □

Note that this theorem is a generalization of *Lemma 3* in [76], which corresponds to $N = 1$. Therefore, $\forall \mathbf{n} \in \Omega_{m_0}$ ($1 \le m_0 \le N$), given the remaining m_0 sets $\left\{\mathcal{U}_{\pi[\rho_0(j)]}\right\}_{j=1}^{m_0}$ after the iterative procedure, *Theorem 5.3* determines which set(s) of users should be chosen for transmission by solving (5.62), since after removing those users to which no power should be assigned, the maximization problems (5.42) and (5.62) are equivalent and

$$w_{\pi[\rho_0(j)]}^*(\mathbf{R}, \mathbf{n}) = u_j^*(\mathbf{n}), \quad 1 \le j \le m_0.$$

In particular, the theorem indicates that, based on the total power constraint, there is a threshold power lever s^* which is important in determining the optimal set(s) of users. Moreover, in each fading state in set $\tilde{L}_j(m_0, s)$ ($\forall 1 \le m_0 \le N$, $\forall 1 \le j \le m_0$), at most two sets of users are chosen and the information for the selected two sets are sent by time-sharing the channel. In each of the other fading states, at most one set of users is chosen. Therefore, if the c.d.f. $F(\mathbf{n})$ is continuous, with probability 1, at most one set of users is chosen in each state, since $\forall 1 \le m_0 \le N$, $\forall 1 \le j \le m_0$, $Pr[\tilde{L}_j(m_0, s)] = 0$. If $F(\mathbf{n})$ is discontinuous, $Pr[\tilde{L}_j(m_0, s)]$ may be larger than zero for some j and m_0, and the probability that two sets of users are chosen in some fading states may be larger than zero.

5.4.3 Multimedia Outage Probability Region

In an M-user broadcast system, some users may require constant-rate transmission without any outage (e.g., voice users), while other users allow certain outages in the transmission of their information (e.g., data users). Let M_0 be the number of those users allowing no outage. Then the M-user outage probability region contracts to an $(M-M_0)$-user outage probability region. Since in each fading state, the channel can be used for $\sum_{i=1}^{M-M_0} \binom{M-M_0}{i} = 2^{M-M_0}-1$ different sets of users, by applying the same optimal strategy discussed in Section 5.4.2, we can obtain the boundary of the outage probability region for the $M - M_0$ users.

For example, in a two-user system where one user (say, User 1) allows some outage and the other user (say, User 2) requires no outage ($M = 2$, $N = 1$), the minimum outage probability problem for User 1 is a modified threshold-decision rule similar to that of the single-user case:

For each joint fading state $\mathbf{n} = (n_1, n_2)$ and a given rate vector $\mathbf{R} = (R_1, R_2)$, if the information for User 1 is not transmitted, we denote the minimum required total power as $P_{off}(\mathbf{R}, \mathbf{n})$ and it is just the power needed to support rate R_2 for User 2; if the information for User 1 is transmitted, we denote the minimum required total power as $P_{on}(\mathbf{R}, \mathbf{n})$ and it is $P^{min}(\mathbf{R}, \mathbf{n})$ given in (5.3), (5.11) and (5.15) for CD with or without successive decoding and for TD, respectively. Let $\bar{P}$ be the total average power and assume that in fading state $\mathbf{n}$, the channel transmits the information for User 1 with probability $w(\mathbf{n}, \mathbf{R})$ and for User 2 with probability 1 (no outage). The maximization problem

$$\begin{cases} \max E_{\mathbf{n}}\left[w(\mathbf{R}, \mathbf{n})\right] \quad \text{subject to:} \\ E_{\mathbf{n}}\left\{P_{on}(\mathbf{R}, \mathbf{n})w(\mathbf{R}, \mathbf{n}) + P_{off}(\mathbf{R}, \mathbf{n})[1 - w(\mathbf{R}, \mathbf{n})]\right\} = \bar{P} \end{cases} \tag{5.63}$$

is equivalent to

$$\begin{cases} \max E_{\mathbf{n}}\left[w(\mathbf{R}, \mathbf{n})\right] \quad \text{subject to:} \\ E_{\mathbf{n}}\left\{[P_{on}(\mathbf{R}, \mathbf{n}) - P_{off}(\mathbf{R}, \mathbf{n})]w(\mathbf{R}, \mathbf{n})\right\} = \bar{P} - E_{\mathbf{n}}\left\{P_{off}(\mathbf{R}, \mathbf{n})\right\}. \end{cases}$$

Therefore, by substituting $\bar{P}$ with $\bar{P} - E_{\mathbf{n}}\left[P_{off}(\mathbf{R}, \mathbf{n})\right]$ and $P^{min}(\mathbf{R}, \mathbf{n})$ with $P_{on}(\mathbf{R}, \mathbf{n}) - P_{off}(\mathbf{R}, \mathbf{n})$ into the definitions of $\mathcal{R}(s)$, $\tilde{\mathcal{R}}(s)$, $P(s)$, $\tilde{P}(s)$, s^* and w^* in (5.30)-(5.34), we obtain the solution to (5.63): if $\mathbf{n} \notin \tilde{\mathcal{R}}(s^*)$, $w(\mathbf{R}, \mathbf{n}) = 0$, i.e., only the information of User 2 is transmitted; if $\mathbf{n} \in \mathcal{R}(s^*)$, $w(\mathbf{R}, \mathbf{n}) = 1$, i.e., the information of User 1 is transmitted together with that of User 2; if $\mathbf{n} \notin \mathcal{R}(s^*)$ but $\mathbf{n} \in \tilde{\mathcal{R}}(s^*)$, then $w(\mathbf{R}, \mathbf{n}) = w^*$, i.e., with probability w^*, the information of User 1 is transmitted with that of User 2. The minimum outage probability $Pr_{min}(\bar{P}, \mathbf{R})$ for User 1 is as given in (5.35).

5.5 Numerical Results

In this section, we present numerical results for zero-outage capacity regions, outage capacity regions and outage probability regions under different spectrum-sharing techniques. In all figures, the Nakagami-m fading model is used, the total average transmit power is denoted as $\bar{P}$, and the average noise density of the ith sub-channel is denoted as $\bar{n}_i$, $i = 1, 2$. We refer to the CD without successive decoding technique as CDWO. Since TD and FD are equivalent in the sense that they have the same capacity region of any kind, all results for TD in the figures also apply for FD.

In Figure 5.3, the two-user zero-outage capacity region for the Nakagami-m fading broadcast channel is shown for $m = 2, 3, 4$ and ∞. The SNR difference between the two users is

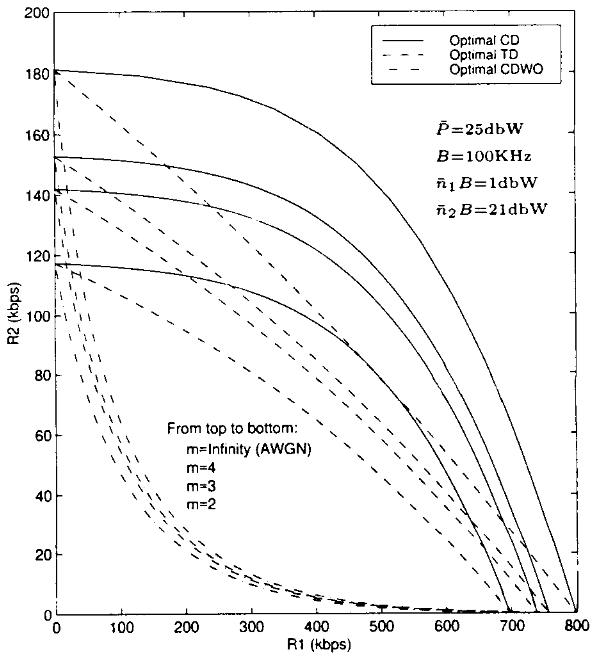


Figure 5.3: Two-user zero-outage capacity region in Nakagami fading: 20 dB SNR difference.

20 dB and the total average power $\bar{P} = 25$ dB. Similar to the ergodic (Shannon) capacity

region comparison in Chapter 4, optimal CD results in a much larger zero-outage capacity region than optimal TD. But the zero-outage capacity region of optimal TD is now much larger than that of the optimal CDWO[8], the boundary of which is convex. Note that the zero-outage capacity region increases as m increases for all of the three types of spectrum-sharing techniques, since smaller m corresponds to more severe fading. However, unlike using optimal CD or TD, the capacity region using CDWO does not increase much with the increase of m. Also note that for the Rayleigh fading channel ($m = 1$), the zero-outage capacity region is zero, which is why it is not shown. When $m \to \infty$, the Nakagami-m fading channel approaches the Gaussian channel and as proved in Section 5.3.4, the limiting zero-outage capacity region of the Nakagami-m fading channel is the same as that of the AWGN channel for CD or CDWO.

Figure 5.4 shows the case where the SNR difference between the two users is 3 dB and the total average power is 10 dB. Since the SNR difference between the users is relatively

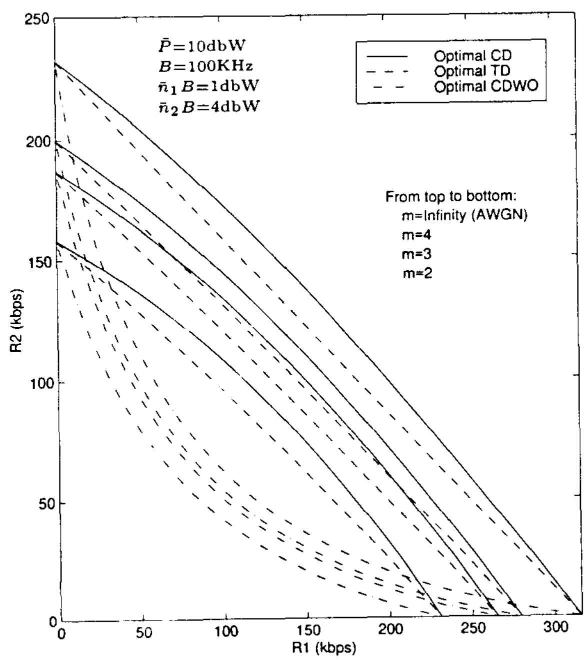


Figure 5.4: Two-user zero-outage capacity region in Nakagami fading: 3 dB SNR difference.

small, the differences of the zero-outage capacity region between using CD, TD, or CDWO is not so dramatic as in the previous case. When m increases, the capacity region of CDWO

[8] As shown in Chapter 4, TD and CDWO have the same ergodic capacity region.

now increases faster than it does in Figure 5.3.

In Figures 5.5-5.11, as discussed in Section 5.4.1, for any fading state the broadcast fading channel is either used for all users or not used for any user. The resulting common outage probability is denoted as Pr. Given a common outage probability of $Pr = 0.1$, the two-user capacity region for the Nakagami-m fading channel ($m = 2, 3, 4$) using optimal CD is shown for 3 dB and 20 dB SNR differences between the users in Figure 5.5 and Figure 5.6, respectively. In both cases, notice that the increase in capacity obtained by

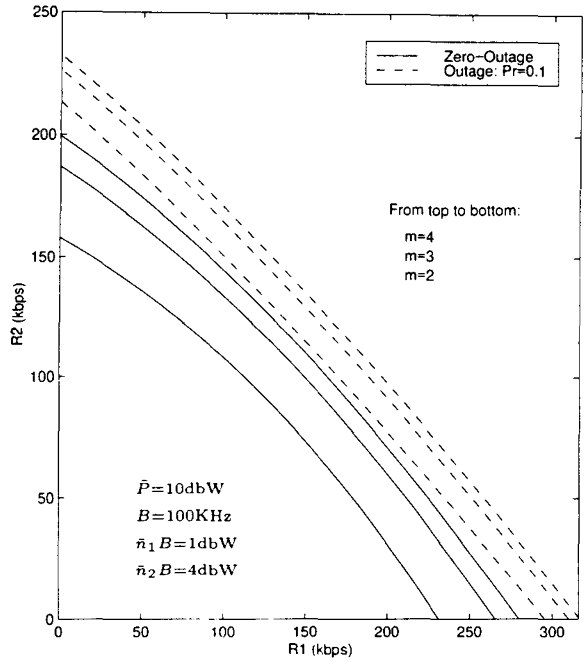


Figure 5.5: Two-user capacity region for a given common outage probability in Nakagami fading using CD: 3 dB SNR difference.

allowing a non-zero outage probability is larger for smaller m. This is because a smaller m corresponds to more severe fading, which is difficult to compensate for in the zero-outage case. In Figure 5.5, since the SNR difference between the users is small, the increase of R_1 obtained by allowing outage is pretty much independent of R_2. However, when the SNR difference is large, we notice in Figure 5.6 that in the region where R_2 is large, we obtain a large increase in User 1's rate R_1 by allowing some outage. This increase is much smaller in the region where R_2 is small. Since User 2 has much more noise on the average, for R_2 large and no outage, most of the total transmit power is used to send the information to

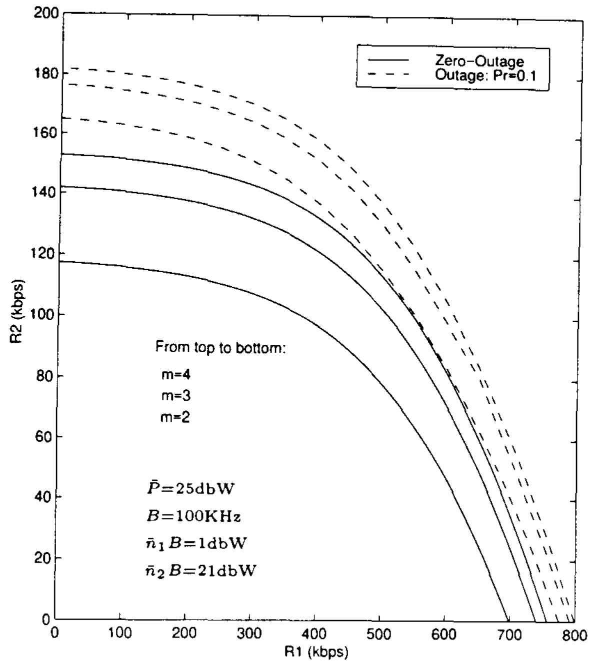


Figure 5.6: Two-user capacity region for a given common outage probability in Nakagami fading using CD: 20 dB SNR difference.

User 2 and allowing some common outage probability will then save relatively more power for User 1 than in the case where R_2 is small.

In Figure 5.7, the capacity regions using CD in Nakagami fading ($m = 2$) with common outage probability $Pr = 0.02$, $Pr = 0.05$ and $Pr = 0.1$ are compared when the SNR difference between the two users is 3 dB and the total average power is 10 dB. We see that by allowing even a small outage probability, we obtain a significant capacity increase relative to the zero-outage case. Figure 5.8 shows the minimum common outage probability Pr as a function of the total average power $\bar{P}$ at a given rate pair $(R_1, R_2) = (100, 130)$ kbps using CD under the same channel conditions as in Figure 5.7. According to (5.3) and (5.4), this rate vector is on the boundary of the zero-outage capacity region for a total average power $S \approx 10.9$ dB, as is shown in the figure.

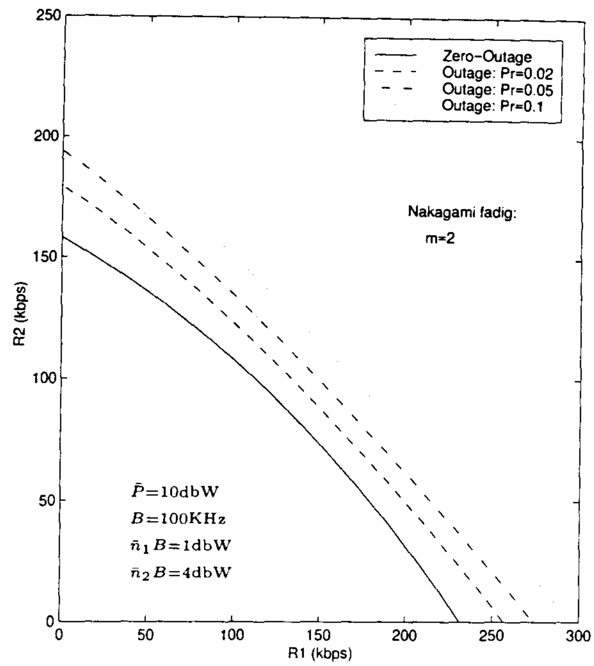


Figure 5.7: Two-user capacity region comparison of different common outage probabilities in Nakagami fading using CD: 3 dB SNR difference.

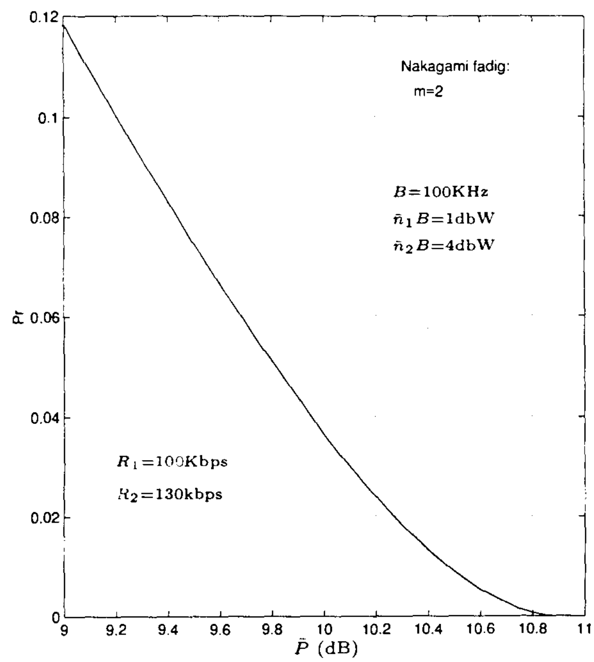


Figure 5.8: Minimum common outage probability for a given rate vector vs. average transmit power in Nakagami fading using CD.

Figure 5.9 and Figure 5.10 show the two-user capacity region of CDWO in Nakagami-m fading for a common outage probability of $Pr = 0.1$. The SNR differences between the two users in these figures are 3 dB and 20 dB, respectively. Similar to the zero-capacity region, when the SNR difference between the users is small, the capacity regions with a given common outage probability increase faster with the increase of the Nakagami channel parameter m than when the SNR difference is large. However, in both cases, the increase of the capacity region from zero-outage to an outage of 0.1 for each m is not that much and the differences between the outage capacity regions with different m are even smaller than that of the zero-outage capacity regions. This means that the optimal power policy that allows a certain common outage probability does not help much in increasing the capacity region of CDWO, especially when there is a great difference between the average channel conditions of the two users.

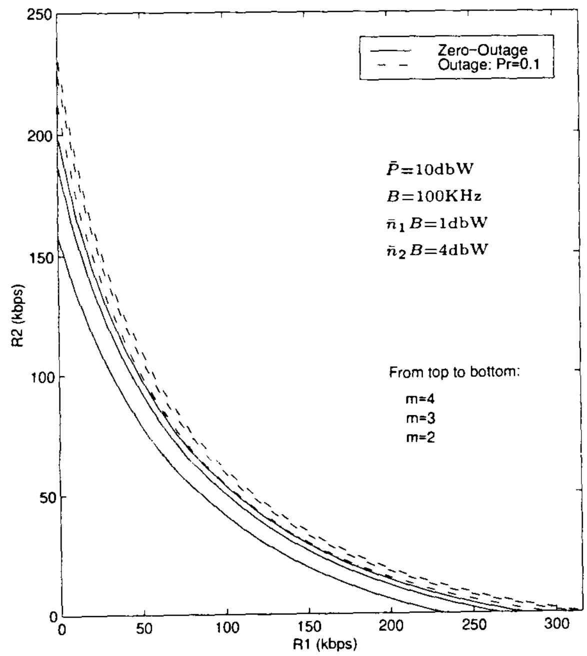


Figure 5.9: Two-user capacity region for a given common outage probability in Nakagami fading using CDWO: 3 dB SNR difference.

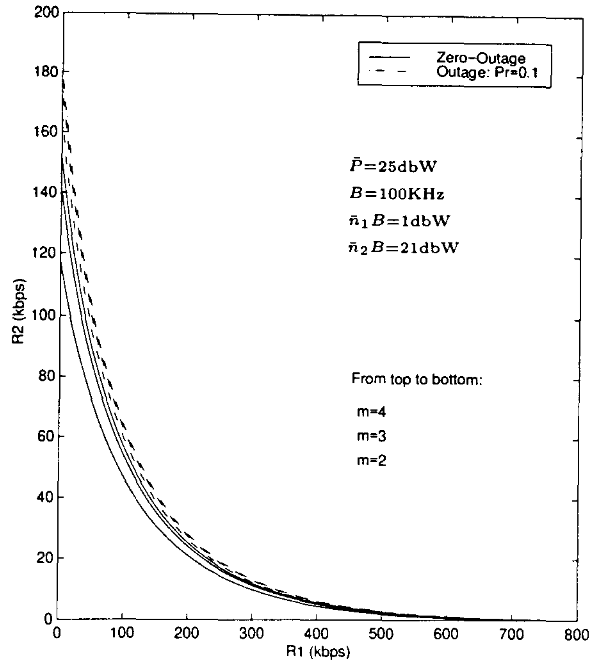


Figure 5.10: Two-user capacity region for a given common outage probability in Nakagami fading using CDWO: 20 dB SNR difference.

In Figure 5.11, the capacity region with a common outage probability $Pr = 0.1$ using optimal TD for the Nakagami fading channel ($m = 2$) is shown and compared to that of the CD and CDWO techniques. The SNR difference between the two users is 3 dB and the total power $S = 10$ dB. As in the case where there is no outage, the capacity region with a common outage probability using TD is smaller than that of CD but is much larger than that of CDWO. Note that by allowing some common outage probability, there is a large increase of the capacity region for both CD and TD, but the increase is relatively small for CDWO.

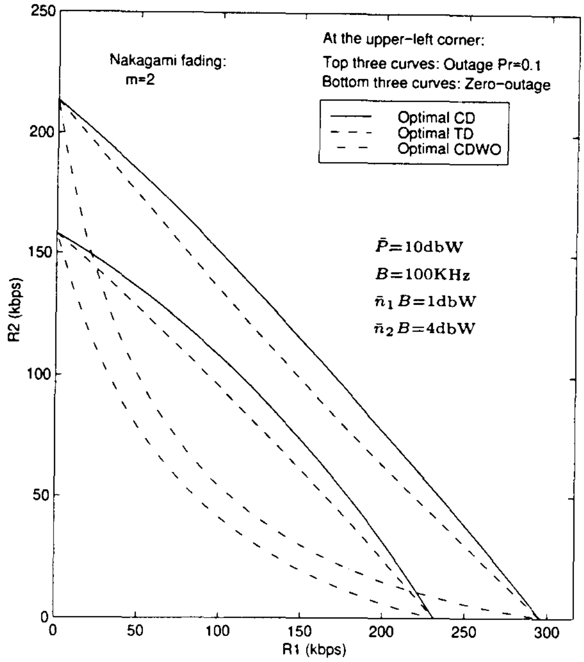


Figure 5.11: Two-user capacity region comparison for a given common outage probability in Nakagami fading using CD, TD and CDWO: 3 dB SNR difference.

In Figures 5.12-5.16, we assume that a different outage probability can be declared for each user. The corresponding optimal power policy is obtained by applying the three-step procedure described in Section 5.4.2 and *Theorem 5.3*, which is then used to calculate either the capacity region for a given outage probability vector or the outage probability region for a given rate vector. We obtain the capacity regions with outage or the outage probability regions for CD only, since the relative behavior of TD or CDWO is similar to that of CD.

In a two-user system, let Pr_1 and Pr_2 denote the outage probabilities for User 1 and User 2, respectively. Given $(Pr_1, Pr_2) = (0.02, 0.03)$, the two-user capacity regions with this outage for the Nakagami fading channels ($m = 2, 3, 4$) are shown in Figure 5.12 and Figure 5.13 for SNR difference between the users of 3 dB and 20 dB, respectively. In both cases, as was true for the capacity region with a common outage probability, allowing some outage probability for each user results in a capacity increase that is larger for smaller m. Thus, the optimal power policy is more effective in increasing the capacity region when the overall broadcast channel fading is more severe. For different m, the differences between the capacity regions with outage are smaller than those between the capacity regions with no outage.

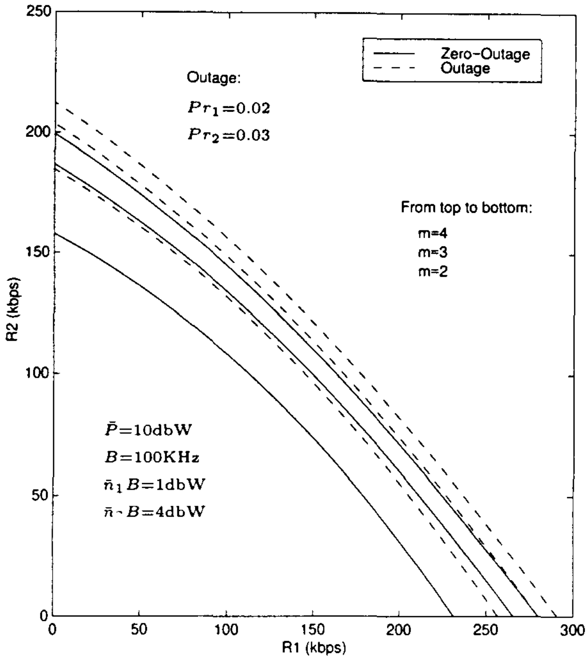


Figure 5.12: Two-user capacity region for a given outage probability vector in Nakagami fading using CD: 3 dB SNR difference.

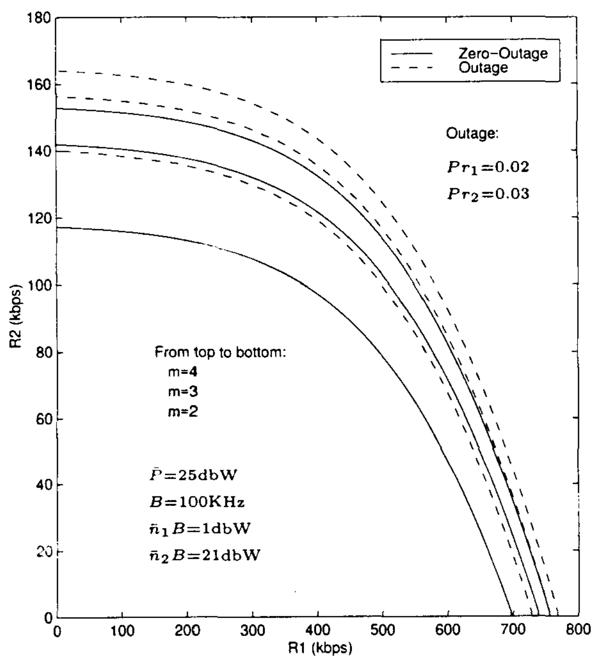


Figure 5.13: Two-user capacity region for a given outage probability vector in Nakagami fading using CD: 20 dB SNR difference.

Figure 5.14 shows the two-user outage probability regions with different total transmit power $\bar{P}$ for a given rate vector $(R_1, R_2) = (100, 130)$ kbps. Note that the region below

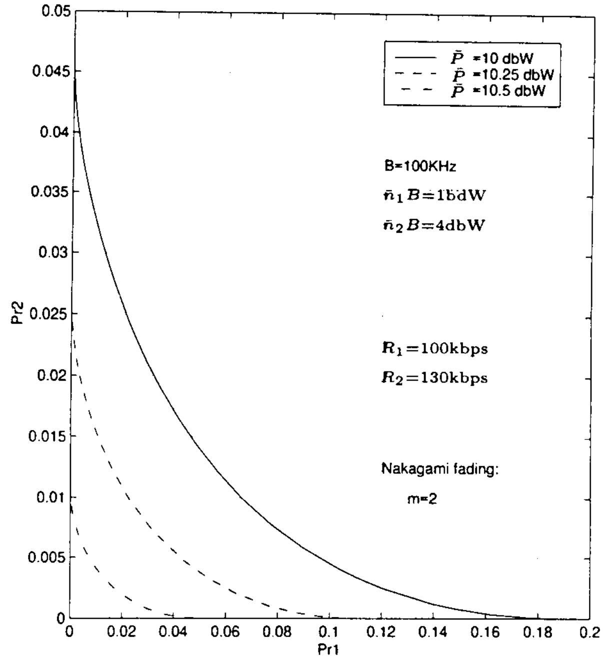


Figure 5.14: Two-user outage probability region comparison for different average transmit power in Nakagami fading using CD: 3 dB SNR difference.

each curve is the outage probability region not achievable with the corresponding transmit power $\bar{P}$. This non-achievable region shrinks quickly with the increase of transmit power $\bar{P}$ and disappears when $\bar{P} > 10.9$ dB since, when $\bar{P} \approx 10.9$ dB, according to (5.3) and (5.4), the rate vector $(R_1, R_2) = (100, 130)$ kbps is on the boundary of the zero-outage capacity region. For a given transmit power $\bar{P}$, when the outage probability Pr_2 of User 2 decreases, there is a fast increase in the outage probability Pr_1 of User 1, since the average channel condition of User 2 is worse than that of User 1 and thus the total power required to support R_2 increases fast with the decrease of its outage probability. The intersections of the curves with the two axes in this figure denote the minimum outage probabilities for one user when there is no outage in the transmission for the other user.

Figure 5.15 shows the two-user capacity regions with several different outage probability vectors and a total transmit power $\bar{P} = 10$ dB. In this figure, the points A_0, A_1 and A_2 are the single-user capacities of User 1 when the allowed outage probabilities are $Pr_1 = 0$,

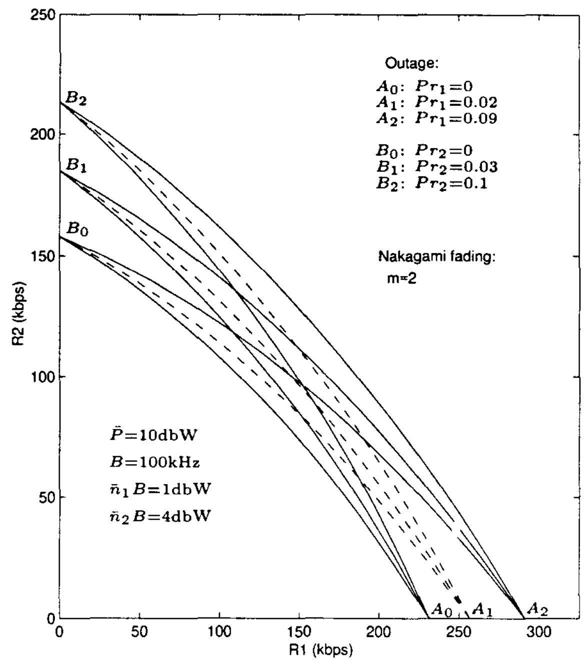


Figure 5.15: Two-user capacity region comparison for different outage probability vectors in Nakagami fading using CD: 3 dB SNR difference.

$Pr_1 = 0.02$ and $Pr_1 = 0.09$, respectively; the points B_0, B_1 and B_2 are the single-user capacities of User 2 when the allowed outage probabilities are $Pr_2 = 0$, $Pr_2 = 0.03$ and $Pr_2 = 0.1$, respectively. Let $Pa_0 = 0$, $Pa_1 = 0.02$, $Pa_2 = 0.09$, $Pb_0 = 0$, $Pb_1 = 0.03$ and $Pb_2 = 0.1$. The curves between points A_i and B_j are the boundaries of the capacity regions when the allowed outage probability vectors are $(Pr_1, Pr_2) = (Pa_i, Pb_j)$, $i, j = 0, 1, 2$. Note that when one of the outage probabilities Pr_1 and Pr_2 is zero, regardless of the time-varying channel state, the information of the corresponding user is always transmitted and the optimal power policy discussed in Section 5.4.3 will be used for the other user to achieve the demonstrated capacity region in this figure under the constraint of its given outage probability.

Finally, in Figure 5.16 and Figure 5.17 where the SNR differences between the two users are 3 dB and 20 dB, respectively, the capacity regions using the optimal CD power policy with a common outage probability $Pr = 0.1$ as discussed in Section 5.4.1 and the optimal CD power policy with an outage probability vector $(Pr_1, Pr_2) = (0.1, 0.1)$ as discussed in Section 5.4.2 are compared in Nakagami fading ($m = 2$). Since the outage probability for

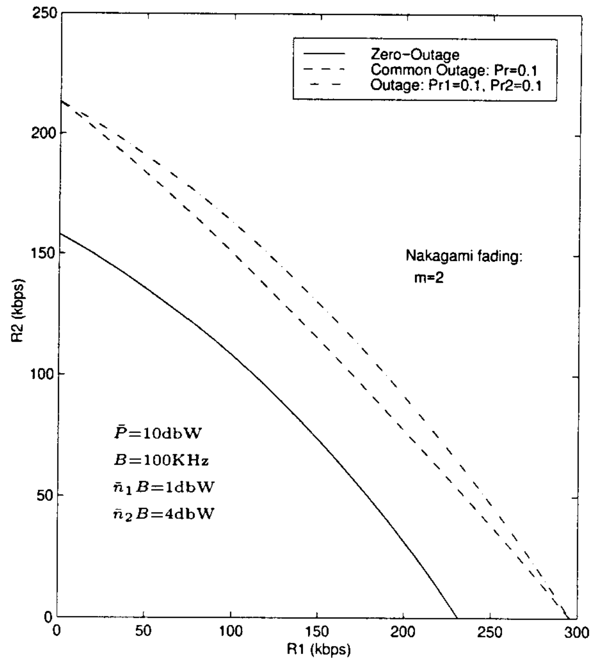


Figure 5.16: Two-user capacity regions with a common outage probability and with an outage probability vector in Nakagami fading using CD: 3 dB SNR difference.

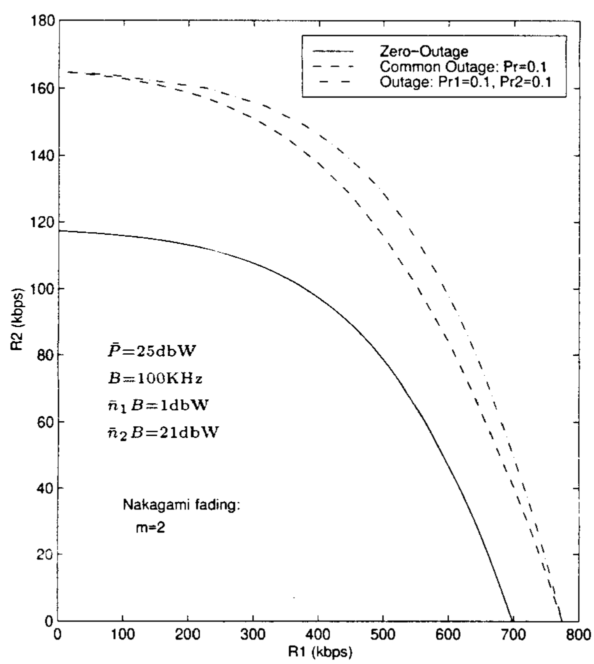


Figure 5.17: Two-user capacity regions with a common outage probability and with an outage probability vector in Nakagami fading using CD: 20 dB SNR difference.

each user is 0.1 using either of the two power policies, from the figures it is clear that by allowing a separate outage declaration for each user and using the corresponding optimal CD power policy, a larger capacity region can be achieved than by simply turning on or off the transmission for both users simultaneously based on the optimal power policy discussed in Section 5.4.1. However, the system applying the optimal power policy with a common outage probability is less complex.

5.6 Conclusions

We have obtained both the zero-outage capacity region and the minimum-outage capacity region of the broadcast fading channels for TD, FD and CD with and without successive decoding, assuming that perfect CSI is available both at the transmitter and at the receiver. It is shown that optimal CD has the largest zero-outage capacity region, as expected. Moreover, we show that the capacity region can be greatly expanded by allowing some outage probability for each user. For a given rate vector, we have derived the optimal power policy that minimizes the common outage probability when transmission to all users is turned off simultaneously. When an outage can be declared for each user individually, we have also derived a general power allocation strategy to achieve boundaries of the outage probability regions under different spectrum-sharing techniques. We observe that these regions can increase dramatically with an increase in the total transmit power. Therefore, by applying the optimal dynamic power allocation strategies derived herein, tradeoffs between the maximum transmission rate, the outage probability for each user, and the total transmit power may be evaluated for the design of a broadcast communication system in a fading environment.

Chapter 6 Outage Capacities and Optimal Power Allocation for Fading Multiple-Access Channels

6.1 Introduction

Wireless communication channels vary over time due to user mobility. As discussed in Chapter 5, the outage capacity of a fading channel is defined as the maximum instantaneous information rate that can be maintained under any fading condition during non-outage such that the allowed average transmission outage probability is satisfied. Finding the optimal power allocation that achieves the outage capacity with a given outage probability is tantamount to deriving the optimal power allocation that minimizes the outage probability of a given rate. In [76], the minimum outage probability problem is solved for the single-user fading channel. For an M-user fading broadcast channel, under different assumptions about whether the transmission to all users is turned off simultaneously or independently, we have derived in Chapter 5 the optimal power allocation strategy that minimizes the common outage probability or bounds the outage probability region of the M users under a total average power constraint of all users.

In this chapter we derive the outage capacity region and the optimal power allocation policy for an M-user fading MAC under similar assumptions about whether the outage declaration from each user is simultaneous or independent. This problem is solved in [3] for the case of zero-outage. Specifically, it is shown in [3] that the zero-outage capacity region is implicitly obtained by determining, for each given rate vector $\mathbf{R} = (R_1, R_2, \cdots, R_M)$, whether there exists an optimal power allocation such that the average transmit power required of each user to support rate R_i under any fading condition satisfies his given average power constraint. For the general case where the allowed outage probability of each user is larger than zero, we will show that the outage capacity region is implicitly obtained by determining, for each given rate vector $\mathbf{R}$, whether there exists an optimal power allocation such that the average outage probability of each user is no larger than his allowed outage probability and the average transmit power required of each user to support

R_i under any non-outage fading condition satisfies his given average power constraint. In particular, given the average power constraint of each user and rate vector $\mathbf{R}$, assuming that the outage declaration from each user is simultaneous, we derive the optimal power allocation that minimizes the common outage probability of the M users and obtain the minimum common outage probability. Under the alternative assumption that each user can declare an outage during transmission independently, we derive the outage probability region of the M users and the optimal power allocation that achieves the outage probability region boundary. Under both assumptions, given the allowed outage probability of each user and a rate vector $\mathbf{R}$, we solve the dual problem of finding the average power region of the M users required to support $\mathbf{R}$ for the given outage probability vector. Note that the optimal power allocation is conceptually simpler and easier to analyze when the outage declaration is simultaneous for all users. Common outage may be desirable in scenarios when users are trying to avoid detection.

For a given rate vector $\mathbf{R}$, in order to solve the optimization problem of minimizing the common outage probability or bounding the outage probability region for a given average power constraint on each of the M users, we use the Lagrangian method with multiple constraints. Since there is an independent average power constraint for each of the M users, M Lagrangian multipliers are needed. For each given Lagrangian multiplier vector $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_M)$, the optimization problem under a single constraint of the total weighted average power of the M users (weighted by their corresponding Lagrangian multipliers) is readily solved by applying the techniques developed in [3] and Chapter 5 of this thesis. Iterative algorithms are then proposed to provide the appropriate Lagrangian multiplier vector for the M users such that the M average power constraints are satisfied simultaneously.

The remainder of this chapter is organized as follows. In Section 6.2 we present the fading MAC model. In Section 6.3 we give the definitions and notations that will be used in the rest of the paper. In Section 6.4, the minimum common outage probability for a given rate vector $\mathbf{R}$ and the corresponding optimal power allocation strategy are derived and the average power region for supporting $\mathbf{R}$ with a given common outage probability is obtained for the case of simultaneous outage declaration. As for independent outage declaration, we derive the outage probability region boundary for a given rate vector $\mathbf{R}$, the corresponding optimal power allocation strategy, and the required average power region for supporting $\mathbf{R}$ with the given outage probability of each user in Section 6.5. The iterative algorithms that provide the appropriate Lagrangian multiplier vector for the M users are presented in Section 6.6. In Section 6.7 we present the main difference in the solutions to the above problems when additional peak power constraints are imposed on the M users. Our conclusions are given in Section 6.8.

6.2 The Fading Multiple-Access Channel

We consider a discrete-time M-user fading MAC model as discussed in [3]:

$$Y(n) = \sum_{i=1}^{M} \sqrt{H_i(n)} X_i(n) + Z(n), \tag{6.1}$$

where $X_i(n)$ and $H_i(n)$ are the transmitted waveform and the fading process of the ith user, respectively, and $Z(n)$ is Gaussian noise with variance σ^2. Let $\mathbf{H}(n) = [H_1(n), H_2(n), \cdots, H_M(n)]$ denote the joint fading process, and let $\bar{P}_i$ be the average power constraint of User i. We assume that the joint fading process of the M users is stationary and ergodic, and the stationary distribution has continuous density and is bounded[1]. For a slowly time-varying MAC, let $\mathbf{h} = (h_1, h_2, \cdots, h_M)$ be the joint fading state at a particular time n, i.e., $\mathbf{H}(n) = \mathbf{h}$, and let $\mathcal{H}_{all}$ denote the set of all possible joint fading states. We assume that all the M transmitters and the receiver know the current joint fading state $\mathbf{h}$. Therefore, each transmitter can vary its transmit power and codewords relative to the joint fading condition of the M sub-channels, and the receiver can vary its decoding order of the M users. We will show that successive decoding is optimal in order to achieve the outage capacity region boundary. By using successive decoding at the receiver, the signals from users that are decoded earlier can be subtracted from the total received signal $Y(n)$, and the users that are decoded later will have less interference and thus require less transmit power to support their given data rates.

[1]As in the single-user case [76] and the broadcast system discussed in Chapter 5, our analysis can be easily extended to discrete distributions.

6.3 Definitions and Notations

In this section we define the outage capacity region of an M-user MAC, where each transmitter may suspend transmission over a subset of fading states under a given average power constraint and an average outage probability constraint. Specifically, for a given average power constraint vector $\bar{\mathbf{P}} = (\bar{P}_1, \bar{P}_2, \cdots, \bar{P}_M)$ of the M users, the outage capacity regions $C_{out}(\bar{\mathbf{P}}, Pr)$ and $C_{out}(\bar{\mathbf{P}}, \mathbf{Pr})$ are defined as follows:

Definition 6.1 *Assume that the transmission from all users is turned on (non-outage) or off (outage) simultaneously, and thus the average outage probability for all users is the same (common outage probability). Then for a given outage probability Pr, $\forall \mathbf{R} = (R_1, R_2, \cdots, R_M)$, if user i can transmit at rate R_i with arbitrarily small error probability under any fading condition during non-outage such that the average common outage probability is no larger than Pr and the average power constraint $\bar{P}_i$ is satisfied for all i $(1 \leq i \leq M)$ then $\mathbf{R} \in C_{out}(\bar{\mathbf{P}}, Pr)$.*

Definition 6.2 *Assume that the transmission from each user is turned on (non-outage) or off (outage) independently, and thus the average outage probability for each user may be different. Then for a given probability vector $\mathbf{Pr} = (Pr_1, Pr_2, \cdots, Pr_M)$, $\forall \mathbf{R} = (R_1, R_2, \cdots, R_M)$, if user i can transmit at rate R_i with arbitrarily small error probability under any fading condition during non-outage such that his average outage probability is no larger than Pr_i and the power constraint $\bar{P}_i$ is satisfied for all i $(1 \leq i \leq M)$, then $\mathbf{R} \in C_{out}(\bar{\mathbf{P}}, \mathbf{Pr})$.*

Definition 6.3 *The boundary surface of $C_{out}(\bar{\mathbf{P}}, Pr)$ (or $C_{out}(\bar{\mathbf{P}}, \mathbf{Pr})$) is the set of those rate vectors for which we cannot increase one component and remain in $C_{out}(\bar{\mathbf{P}}, Pr)$ (or $C_{out}(\bar{\mathbf{P}}, \mathbf{Pr})$) without decreasing another component.*

With these definitions, we wish to find: a) the optimal power allocation strategy that achieves the boundary of the outage capacity region $C_{out}(\bar{\mathbf{P}}, Pr)$; b) the optimal power allocation strategy that achieves the boundary of $C_{out}(\bar{\mathbf{P}}, \mathbf{Pr})$. The regions $C_{out}(\bar{\mathbf{P}}, Pr)$ and $C_{out}(\bar{\mathbf{P}}, \mathbf{Pr})$ are easily determined given these optimal power allocation strategies. We will show that the first optimization problem is equivalent to deriving the power allocation policy that minimizes the common outage probability for a given rate vector $\mathbf{R}$ and we have the following definitions for the common outage probability set $\mathcal{O}_C(\bar{\mathbf{P}}, \mathbf{R})$, the common transmission (usage) probability set $\overline{\mathcal{O}}_C(\bar{\mathbf{P}}, \mathbf{R})$, and the minimum common outage probability $Pr_{min}(\bar{\mathbf{P}}, \mathbf{R})$:

Definition 6.4 *Assuming that the transmission from all users is turned on or off simultaneously, the common outage probability set $\mathcal{O}_C(\bar{\mathbf{P}}, \mathbf{R})$ is defined as the set of all average common outage probabilities for which under any non-outage fading condition, each user i can transmit at rate R_i with arbitrarily small error probability under the given power constraint $\bar{P}_i$ $(1 \le i \le M)$.*

Definition 6.5 *The common usage probability set $\overline{\mathcal{O}}_C(\bar{\mathbf{P}}, \mathbf{R})$ is the complementary probability set of $\mathcal{O}_C(\bar{\mathbf{P}}, \mathbf{R})$, i.e., if a probability $Pr \in \mathcal{O}_C(\bar{\mathbf{P}}, \mathbf{R})$, then the probability $Pr^{on} \in \overline{\mathcal{O}}_C(\bar{\mathbf{P}}, \mathbf{R})$, where $Pr^{on} = 1 - Pr$.*

Definition 6.6 *The minimum common outage probability $Pr_{min}(\bar{\mathbf{P}}, \mathbf{R})$ is the smallest probability in set $\mathcal{O}_C(\bar{\mathbf{P}}, \mathbf{R})$.*

We will also show that the second optimization problem is equivalent to obtaining the power allocation policy that achieves the boundary of the outage probability region $\mathcal{O}_I(\bar{\mathbf{P}}, \mathbf{R})$ or the usage probability region $\overline{\mathcal{O}}_I(\bar{\mathbf{P}}, \mathbf{R})$, defined as follows:

Definition 6.7 *Assuming that the transmission to each user is turned on or off independently, for a given rate vector $\mathbf{R}$, $\forall \mathbf{Pr} = (Pr_1, Pr_2, \ldots, Pr_M)$, if user i can transmit at rate R_i with arbitrarily small error probability under any non-outage fading condition such that his average outage probability is no larger than Pr_i and the power constraint $\bar{P}_i$ $(1 \le i \le M)$ is satisfied, then $\mathbf{Pr} \in \mathcal{O}_I(\bar{\mathbf{P}}, \mathbf{R})$.*

Definition 6.8 *The usage probability region $\overline{\mathcal{O}}_I(\bar{\mathbf{P}}, \mathbf{R})$ is the complementary region of the outage probability region $\mathcal{O}_I(\bar{\mathbf{P}}, \mathbf{R})$, i.e., if a probability vector $\mathbf{Pr} = (Pr_1, Pr_2, \cdots, Pr_M) \in \mathcal{O}_I(\bar{\mathbf{P}}, \mathbf{R})$, then the probability vector $\mathbf{Pr}^{on} = (Pr_1^{on}, Pr_2^{on}, \cdots, Pr_M^{on}) \in \overline{\mathcal{O}}_I(\bar{\mathbf{P}}, \mathbf{R})$, where*

$$Pr_i^{on} = 1 - Pr_i, \quad \forall 1 \le i \le M.$$

With the above definitions[2], it is easily seen that given $0 \le Pr \le 1$, the outage capacity region $C_{out}(\bar{\mathbf{P}}, Pr)$ is implicitly obtained once the minimum common outage probability

[2]Note that the definitions of $C_{out}(\bar{\mathbf{P}}, Pr)$, $C_{out}(\bar{\mathbf{P}}, \mathbf{Pr})$, $Pr_{min}(\bar{\mathbf{P}}, \mathbf{R})$, $\mathcal{O}_I(\bar{\mathbf{P}}, \mathbf{R})$, and $\overline{\mathcal{O}}_I(\bar{\mathbf{P}}, \mathbf{R})$ are similar to those in Chapter 5, where the power constraint is a total average power $\bar{P}$ instead of a vector $\bar{\mathbf{P}}$ for the M users.

$Pr_{min}(\bar{\mathbf{P}}, \mathbf{R})$ for a given rate vector $\mathbf{R}$ is calculated under the optimal power allocation. That is, for any rate vector $\mathbf{R}$, $\mathbf{R} \in C_{out}(\bar{\mathbf{P}}, Pr)$ if $Pr_{min}(\bar{\mathbf{P}}, \mathbf{R}) \leq Pr$, and $\mathbf{R} \notin C_{out}(\bar{\mathbf{P}}, Pr)$ otherwise. Similarly, given a probability vector $\mathbf{Pr}$, the outage capacity region $C_{out}(\bar{\mathbf{P}}, \mathbf{Pr})$ is implicitly obtained once the boundary of the outage probability region $\mathcal{O}_I(\bar{\mathbf{P}}, \mathbf{R})$ (and so the whole region $\mathcal{O}_I(\bar{\mathbf{P}}, \mathbf{R})$) for a given rate vector $\mathbf{R}$ is derived through the optimal power allocation since, for any rate vector $\mathbf{R}$, $\mathbf{R} \in C_{out}(\bar{\mathbf{P}}, \mathbf{Pr})$ if $\mathbf{Pr} \in \mathcal{O}_I(\bar{\mathbf{P}}, \mathbf{R})$ and $\mathbf{R} \notin C_{out}(\bar{\mathbf{P}}, \mathbf{Pr})$ otherwise.

We treat common outage in the next section. Specifically, we derive the minimum common outage probability $Pr_{min}(\bar{\mathbf{P}}, \mathbf{R})$ and the corresponding optimal power allocation strategy in Sections 6.4.1-6.4.3. We solve the dual problem of finding the average power region of the M users required to support $\mathbf{R}$ with a given common outage probability Pr in Section 6.4.4.

6.4 Simultaneous Outage Declaration

Under the assumption that an outage is declared from all users simultaneously, for any given fading state $\mathbf{h}$, the channel is either used by all users or not used by any user.

6.4.1 Power Allocation Policy

For a given rate vector $\mathbf{R}$, a "power allocation policy $\mathcal{P}$" is defined by the following two items:

(1) A probability $w(\mathbf{R}, \mathbf{h})$ ($\forall \mathbf{h} \in \mathcal{H}_{all}$) that the channel is used by the M users simultaneously [3] in state $\mathbf{h}$ with each user i transmitting at rate R_i, $1 \leq i \leq M$.

(2) A power $P_i(\mathbf{R}, \mathbf{h})$ ($\forall \mathbf{h} \in \mathcal{H}_{all}$) denoting the transmit power of user i, $1 \leq i \leq M$, when the M users transmit information simultaneously in state $\mathbf{h}$. Let $\mathbf{P}(\mathbf{R}, \mathbf{h}) \triangleq [P_1(\mathbf{R}, \mathbf{h}), P_2(\mathbf{R}, \mathbf{h}), \cdots, P_M(\mathbf{R}, \mathbf{h})]$.

Under the power allocation policy $\mathcal{P}$, the average common transmission (usage) probability $Pr^{on}(\mathbf{R})$ is:

$$Pr^{on}(\mathbf{R}) = E_{\mathbf{h}}\left[w(\mathbf{R}, \mathbf{h})\right], \tag{6.2}$$

[3] As will be shown later in our derivation of the optimal power allocation policy, $w(\mathbf{R}, \mathbf{h})$ is either 1 or 0 under the assumption that the stationary distribution of the joint fading process of the M users has continuous density, since the probability measure of each fading state $\mathbf{h}$ is zero. However, as in the single-user case [76], this is not true if the c.d.f. (cumulative density function) of the joint fading process is discontinuous.

where $E[\cdot]$ denotes the expectation function, and the average common outage probability $Pr(\mathbf{R})$ is:

$$
\begin{aligned}
Pr(\mathbf{R}) &= 1 - Pr^{on}(\mathbf{R}) \\
&= 1 - E_{\mathbf{h}}\left[w(\mathbf{R},\mathbf{h})\right].
\end{aligned}
$$

The average transmit power $\bar{P}_i(\mathbf{R})$ of each user i, $1 \le i \le M$, for rate vector $\mathbf{R}$ is:

$$
\bar{P}_i(\mathbf{R}) = E_{\mathbf{h}}\left[w(\mathbf{R},\mathbf{h})P_i(\mathbf{R},\mathbf{h})\right]. \tag{6.3}
$$

6.4.2 Minimum Common Outage Probability

For a given average power constraint vector $\bar{\mathbf{P}}^*$ and rate vector $\mathbf{R}$, from *Definition 6.4-6.6* we know that deriving the minimum common outage probability $Pr_{min}(\bar{\mathbf{P}}^*, \mathbf{R})$ is equivalent to deriving the maximum common usage probability in the set $\overline{\mathcal{O}}_C(\bar{\mathbf{P}}^*, \mathbf{R})$. That is, we need to solve the maximization problem

$$
\max_{\mathcal{P}} Pr^{on}(\mathbf{R}) \quad \text{subject to: } Pr^{on}(\mathbf{R}) \in \overline{\mathcal{O}}_C(\bar{\mathbf{P}}^*, \mathbf{R}), \tag{6.4}
$$

where, for a given power allocation policy $\mathcal{P}$, $Pr^{on}(\mathbf{R})$ can be calculated from (6.2). Thus, (6.4) is equivalent to

$$
\max_{\mathcal{P}} E_{\mathbf{h}}[w(\mathbf{R},\mathbf{h})] \tag{6.5}
$$

subject to:

$$
\begin{cases}
\mathbf{R} \in C_g(\mathbf{h}, \mathbf{P}(\mathbf{R},\mathbf{h})), \quad \forall \mathbf{h} \in \mathcal{H}_{tran}, \\
\\
E_{\mathbf{h}}\left[w(\mathbf{R},\mathbf{h})P_i(\mathbf{R},\mathbf{h})\right] \le \bar{P}_i^*, \quad \forall 1 \le i \le M,
\end{cases} \tag{6.6}
$$

where

$$
\mathcal{H}_{tran} = \{\mathbf{h} : w(\mathbf{R},\mathbf{h}) > 0\}, \tag{6.7}
$$

$C_g(\mathbf{h}, \mathbf{P}(\mathbf{R}, \mathbf{h}))$ is the capacity region of the M-user time-invariant Gaussian MAC and is given by

$$C_g(\mathbf{h}, \mathbf{P}(\mathbf{R}, \mathbf{h})) = \left\{ \mathbf{R} : \sum_{i \in F} R_i \leq \frac{1}{2} \log\left(1 + \frac{\sum_{i \in F} h_i P_i(\mathbf{R}, \mathbf{h})}{\sigma^2}\right), \ \ \forall F \subseteq \{1, 2, \cdots, M\} \right\}. \tag{6.8}$$

For a given rate vector $\mathbf{R}$, define

$$\mathcal{Q}_C = \{(Pr^{on}, \bar{\mathbf{P}}) : Pr^{on} \in \overline{\mathcal{O}}_C(\bar{\mathbf{P}}, \mathbf{R})\}. \tag{6.9}$$

We will require the following lemma to find the solution to (6.4), or equivalently, to (6.5) under the constraints in (6.6).

Lemma 6.1 *The set $\mathcal{Q}_C$ is convex.*

Proof: See Appendix D.1. □

Due to the convexity of the set $\mathcal{Q}_C$, Pr^{on*} solves (6.4) for a given rate vector $\mathbf{R}$ if and only if there exists a Lagrangian multiplier vector $\boldsymbol{\lambda} \in \Re_+^M$ such that $(Pr^{on*}, \bar{\mathbf{P}}^*)$ is a solution to the problem

$$\max_{(Pr^{on}(\mathbf{R}), \bar{\mathbf{P}}(\mathbf{R})) \in \mathcal{Q}_C} \left[Pr^{on}(\mathbf{R}) - \boldsymbol{\lambda} \cdot \bar{\mathbf{P}}(\mathbf{R})\right], \tag{6.10}$$

where $\bar{\mathbf{P}}(\mathbf{R}) = [\bar{P}_1(\mathbf{R}), \bar{P}_2(\mathbf{R}), \cdots, \bar{P}_M(\mathbf{R})]$ and, for a given power allocation policy $\mathcal{P}$, $\bar{P}_i(\mathbf{R})$ $(1 \leq i \leq M)$ can be calculated from (6.3). Note that we transform the maximization problem (6.5) under the constraint (6.6) into the problem (6.10) because the scalar to be maximized in (6.10) includes the power constraints in (6.6) and is therefore an easier maximization problem to solve. In (6.10),

$$\begin{aligned}
\boldsymbol{\lambda} \cdot \bar{\mathbf{P}}(\mathbf{R}) &= \sum_{i=1}^{M} \lambda_i \bar{P}_i(\mathbf{R}) \\
&= \sum_{i=1}^{M} \lambda_i E_{\mathbf{h}} \left[w(\mathbf{R}, \mathbf{h}) P_i(\mathbf{R}, \mathbf{h}\right] \\
&= E_{\mathbf{h}} \left[w(\mathbf{R}, \mathbf{h}) \left(\sum_{i=1}^{M} \lambda_i P_i(\mathbf{R}, \mathbf{h})\right)\right] \\
&= E_{\mathbf{h}} \left[w(\mathbf{R}, \mathbf{h}) \left(\boldsymbol{\lambda} \cdot \mathbf{P}(\mathbf{R}, \mathbf{h})\right)\right].
\end{aligned} \tag{6.11}$$

From (6.10) we see that Pr^{on*} will be the maximum common usage probability in set $\overline{\mathcal{O}}_C(\bar{\mathbf{P}}^*, \mathbf{R})$ if and only if $\bar{\mathbf{P}}^*$ is a solution to

$$\min_{\bar{\mathbf{P}}(\mathbf{R})} \boldsymbol{\lambda} \cdot \bar{\mathbf{P}}(\mathbf{R}) \quad \text{subject to} \quad Pr^{on*} \in \overline{\mathcal{O}}_C(\bar{\mathbf{P}}(\mathbf{R}), \mathbf{R}),$$

i.e., if and only if there exists a power allocation policy $\mathcal{P}^*$ such that for all $\mathbf{h} \in \mathcal{H}_{all}$, $\mathbf{P}^*(\mathbf{R}, \mathbf{h})$ and $w^*(\mathbf{R}, \mathbf{h})$ solve the problem

$$\min_{\mathcal{P}} E_{\mathbf{h}} \left[w(\mathbf{R}, \mathbf{h}) \boldsymbol{\lambda} \cdot \mathbf{P}(\mathbf{R}, \mathbf{h}) \right] \quad \text{subject to} \quad \begin{cases} \mathbf{R} \in C_g(\mathbf{h}, \mathbf{P}(\mathbf{R}, \mathbf{h})), \quad \forall \mathbf{h} \in \mathcal{H}_{tran}, \\ E_{\mathbf{h}} \left[w(\mathbf{R}, \mathbf{h}) \right] \geq Pr^{on*}, \end{cases} \tag{6.12}$$

and

$$E_{\mathbf{h}} \left[w^*(\mathbf{R}, \mathbf{h}) \mathbf{P}^*(\mathbf{R}, \mathbf{h}) \right] = \bar{\mathbf{P}}^*. \tag{6.13}$$

Assuming that the Lagrangian multiplier vector $\boldsymbol{\lambda}$ is known, to achieve the minimum in (6.12), the optimal transmit power vector $\mathbf{P}^*(\mathbf{R}, \mathbf{h})$ of the M users must be a solution to

$$\min_{\mathbf{P}(\mathbf{R},\mathbf{h})} \boldsymbol{\lambda} \cdot \mathbf{P}(\mathbf{R}, \mathbf{h}) \quad \text{subject to:} \quad \mathbf{R} \in C_g(\mathbf{h}, \mathbf{P}(\mathbf{R}, \mathbf{h})), \quad \forall \mathbf{h} \in \mathcal{H}_{tran}. \tag{6.14}$$

Since $\forall \mathbf{h} \in \mathcal{H}_{tran}$, the set of *received powers* that can support the given rate vector $\mathbf{R}$ of the M users is

$$\mathcal{G}(\mathbf{R}, \mathbf{h}) \triangleq \{ \mathbf{Q} : Q_i = h_i P_i(\mathbf{R}, \mathbf{h}), \forall 1 \leq i \leq M, \quad \mathbf{R} \in C_g(\mathbf{h}, \mathbf{P}(\mathbf{R}, \mathbf{h})) \}, \tag{6.15}$$

which is shown to be a contra-polymatroid with rank function [3]

$$f(F) = \sigma^2 \left[\exp\left(2 \sum_{i \in F} R_i \right) - 1 \right], \quad \forall F \subseteq \{1, 2, \cdots, M\}, \tag{6.16}$$

for any given $\boldsymbol{\lambda} \in \Re_+^M$, $\forall \mathbf{h} \in \mathcal{H}_{tran}$, the solution $\mathbf{P}_{\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}) = [P_{1,\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}),\ P_{2,\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}), \cdots,$ $P_{M,\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})]$ to the optimization problem (6.14) is readily obtained by applying *Lemma 3.4* in [3], i.e.,

$$P_{\pi(i),\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}) = \begin{cases} \frac{\sigma^2}{h_{\pi(1)}} \left[\exp\left(2 R_{\pi(1)} \right) - 1 \right], & \text{if } i = 1, \\ \frac{\sigma^2}{h_{\pi(i)}} \left[\exp\left(2 \sum_{k=1}^{i} R_{\pi(k)} \right) - \exp\left(2 \sum_{k=1}^{i-1} R_{\pi(k)} \right) \right], & \forall 2 \leq i \leq M, \end{cases} \tag{6.17}$$

where the permutation $\pi(\cdot)$ satisfies:

$$\frac{\lambda_{\pi(1)}}{h_{\pi(1)}} \geq \frac{\lambda_{\pi(2)}}{h_{\pi(2)}} \geq \cdots \geq \frac{\lambda_{\pi(M)}}{h_{\pi(M)}}.$$

The solution in (6.17) indicates that $\pi(\cdot)$ determines the optimal decoding order at the receiver. That is, the signals from User $\pi(M)$ are decoded first, with the signals from all other users being treated as interference. Then the signals from User $\pi(M-1)$ are decoded, with the signals from User $\pi(M)$ being subtracted and those from the other $M-2$ users being treated as interference. The signals from User $\pi(1)$ are decoded last, with the signals from all other users being known and thus being subtracted from the total received signal. Since $\pi(\cdot)$ is determined by vector $\boldsymbol{\lambda}$ in each fading state, the optimal decoding order in each state is actually determined by vector $\boldsymbol{\lambda}$. Therefore, $\boldsymbol{\lambda}$ can be viewed as the *power price vector* for the M users, which we will use to refer to $\boldsymbol{\lambda}$ hereafter.

Now denote the minimum in (6.14) as $P_{\boldsymbol{\lambda}}^{min}(\mathbf{R}, \mathbf{h})$, i.e.,

$$P_{\boldsymbol{\lambda}}^{min}(\mathbf{R}, \mathbf{h}) = \boldsymbol{\lambda} \cdot \mathbf{P}_{\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}), \tag{6.18}$$

where the M components of vector $\mathbf{P}_{\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})$ are given in (6.17). Then $P_{\boldsymbol{\lambda}}^{min}(\mathbf{R}, \mathbf{h})$ can be viewed as the required minimum total weighted power of the M users for transmitting their information at rate vector $\mathbf{R}$ in state $\mathbf{h}$. Since the M average power constraints in (6.6) imply that there is a total average weighted power constraint $\boldsymbol{\lambda} \cdot \bar{\mathbf{P}}^*$ for a given power price vector $\boldsymbol{\lambda}$, we solve the optimization problem (6.5) subject to (6.6) by first finding the solution $w_{\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})$ to the following single-constraint maximization problem for the given power price vector $\boldsymbol{\lambda}$:

$$\begin{cases} \max\limits_{w(\mathbf{R},\mathbf{h})} E_{\mathbf{h}} \left[w(\mathbf{R}, \mathbf{h}) \right] \\ \\ \text{subject to: } E_{\mathbf{h}} \left[w(\mathbf{R}, \mathbf{h}) P_{\boldsymbol{\lambda}}^{min}(\mathbf{R}, \mathbf{h}) \right] \leq \boldsymbol{\lambda} \cdot \bar{\mathbf{P}}^*, \end{cases}$$

which is equivalent to

$$\max_{w(\mathbf{R},\mathbf{h})} \left\{ E_{\mathbf{h}} \left[w(\mathbf{R}, \mathbf{h}) \right] - \frac{1}{s^*} E_{\mathbf{h}} \left[w(\mathbf{R}, \mathbf{h}) P_{\boldsymbol{\lambda}}^{min}(\mathbf{R}, \mathbf{h}) \right] \right\} \tag{6.19}$$

subject to:

$$E_{\mathbf{h}}\left[w(\mathbf{R},\mathbf{h})P_{\boldsymbol{\lambda}}^{min}(\mathbf{R},\mathbf{h})\right] = \boldsymbol{\lambda}\cdot\bar{\mathbf{P}}^{*}, \tag{6.20}$$

where $\frac{1}{s^*}$ is the Lagrangian multiplier to be chosen such that (6.20) is satisfied. In Section 6.6 an iterative algorithm is proposed to find the optimal power price vector $\boldsymbol{\lambda}^*$ such that the solutions $\mathbf{P}_{\boldsymbol{\lambda}^*}(\mathbf{R},\mathbf{h})$ and $w_{\boldsymbol{\lambda}^*}(\mathbf{R},\mathbf{h})$ satisfy the M average power constraints in (6.13), i.e.,

$$E_{\mathbf{h}}\left[w_{\boldsymbol{\lambda}^*}(\mathbf{R},\mathbf{h})P_{i,\boldsymbol{\lambda}^*}(\mathbf{R},\mathbf{h})\right] = \bar{P}_i^*, \quad \forall 1\le i\le M. \tag{6.21}$$

Note that the maximization problem (6.19) subject to (6.20) is similar to the minimum outage probability problem of the single-user fading channel [76] and so is the solution given below.

$\forall s>0$, define the set of fading states $\mathcal{R}(s,\boldsymbol{\lambda})$ as:

$$\mathcal{R}(s,\boldsymbol{\lambda}) = \{\mathbf{h} : P_{\boldsymbol{\lambda}}^{min}(\mathbf{R},\mathbf{h}) \le s\}. \tag{6.22}$$

The corresponding total average weighted power over this set is:

$$\bar{P}(s,\boldsymbol{\lambda}) = E_{\mathbf{h}\in\mathcal{R}(s,\boldsymbol{\lambda})}\left[P_{\boldsymbol{\lambda}}^{min}(\mathbf{R},\mathbf{h})\right].$$

Obviously $\bar{P}(s,\boldsymbol{\lambda})$ is a monotonically increasing function of s. Given the total average weighted power constraint (6.20), we choose s^* to be the point that satisfies

$$\bar{P}(s^*,\boldsymbol{\lambda}) = \boldsymbol{\lambda}\cdot\bar{\mathbf{P}}^*. \tag{6.23}$$

Since we assume that the joint fading process has a continuous stationary distribution, $\forall\boldsymbol{\lambda}\in\Re_+^M$, with probability one the decoding order as defined by $\pi(\cdot)$ in each state $\mathbf{h}$ is unique and so is the power allocation $\mathbf{P}_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h})$ in (6.17). Therefore, by using *Lemma 3* in [76], the solution to the maximization problem (6.19) subject to (6.20) is the *Common Outage Transmission Policy* given in Section 6.4.3 below.

6.4.3 Common Outage Transmission Policy

For each fading state $\mathbf{h}\in\mathcal{H}_{all}$, the optimal transmission policy that solves (6.19) subject to (6.20) is:

1. if $\mathbf{h} \notin \mathcal{R}(s^*, \boldsymbol{\lambda})$, then $w_{\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}) = 0$, i.e., an outage is declared for all users and no power is assigned to any user;

2. if $\mathbf{h} \in \mathcal{R}(s^*, \boldsymbol{\lambda})$, then $w_{\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}) = 1$, i.e., a total weighted power of $P_{\boldsymbol{\lambda}}^{min}(\mathbf{R}, \mathbf{h})$ is assigned to the M users and each user i will transmit at rate R_i with power $P_{i,\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})$ given in (6.17).

In the above transmission policy, we see that s^* can be viewed as a threshold power, since simultaneous transmission for the M users is allowed if and only if the required minimum total weighted power $P_{\boldsymbol{\lambda}}^{min}(\mathbf{R}, \mathbf{h})$ for the M users to transmit their information at rate vector $\mathbf{R}$ in state $\mathbf{h}$ is less than s^*.

Under this transmission policy, the resulting average common outage probability $Pr_{\boldsymbol{\lambda}}(\mathbf{R})$ is:

$$\begin{aligned} Pr_{\boldsymbol{\lambda}}(\mathbf{R}) &= 1 - E_{\mathbf{h}}[w_{\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})] \\ &= 1 - Prob\{\mathbf{h} \in \mathcal{R}(s^*, \boldsymbol{\lambda})\}, \end{aligned} \tag{6.24}$$

where $Prob\{\cdot\}$ denotes the probability function. If the power price vector $\boldsymbol{\lambda}$ is the optimal $\boldsymbol{\lambda}^*$ satisfying (6.21), then the minimum average common outage probability $Pr_{min}(\bar{\mathbf{P}}^*, \mathbf{R})$ is:

$$Pr_{min}(\bar{\mathbf{P}}^*, \mathbf{R}) = Pr_{\boldsymbol{\lambda}^*}(\mathbf{R}).$$

6.4.4 Average Power Region

In Sections 6.4.2-6.4.3, given the power constraint vector $\bar{\mathbf{P}}^*$ and rate vector $\mathbf{R}$ of the M users, we derived the minimum common outage probability $Pr_{min}(\bar{\mathbf{P}}^*, \mathbf{R})$ and the corresponding optimal power allocation policy. In this subsection we find for a given rate vector $\mathbf{R}$ and common outage probability Pr^*, the required *average power region* $APV_{out}(Pr^*, \mathbf{R})$, defined as the set of all possible average power vectors that can support rate vector $\mathbf{R}$ with a common outage probability no larger than Pr^*. That is,

$$APV_{out}(Pr^*, \mathbf{R}) \triangleq \{\bar{\mathbf{P}}(\mathbf{R}) : Pr^* \in \mathcal{O}_C(\bar{\mathbf{P}}(\mathbf{R}), \mathbf{R})\}. \tag{6.25}$$

By convexity of the set $\mathcal{Q}_C$ defined in (6.9), it is clear from (6.10) that an average power vector will be on the boundary surface of $APV_{out}(Pr^*, \mathbf{R})$ if it is a solution to the minimization problem

$$\min_{\bar{\mathbf{P}}(\mathbf{R})} \boldsymbol{\lambda} \cdot \bar{\mathbf{P}}(\mathbf{R}) \quad \text{subject to: } \; Pr^* \in \mathcal{O}_C(\bar{\mathbf{P}}(\mathbf{R}), \mathbf{R}) \tag{6.26}$$

for some power price vector $\boldsymbol{\lambda} \in \Re_+^M$. Therefore, for each given power price vector $\boldsymbol{\lambda} \in \Re_+^M$, an average power vector $\bar{\mathbf{P}}_{\boldsymbol{\lambda}}^M(\mathbf{R})$ solves (6.26) if and only if there exists a Lagrangian multiplier s^* such that $\left(1 - Pr^*, \bar{\mathbf{P}}_{\boldsymbol{\lambda}}^M(\mathbf{R})\right)$ is a solution to the problem

$$\min_{(Pr^{on}(\mathbf{R}), \bar{\mathbf{P}}(\mathbf{R})) \in \mathcal{Q}_C} \left\{ \boldsymbol{\lambda} \cdot \bar{\mathbf{P}}(\mathbf{R}) - s^* \cdot Pr^{on}(\mathbf{R}) \right\}. \tag{6.27}$$

Substituting (6.2) and (6.11) into (6.27) and by definition of the set $\mathcal{Q}_C$, it is clear that $\bar{\mathbf{P}}_{\boldsymbol{\lambda}}(\mathbf{R})$ solves (6.26) if and only if there exist a Lagrange multiplier s^* and a power allocation policy $\mathcal{P}^*$ such that for any $\mathbf{h} \in \mathcal{H}_{all}$, $\mathbf{P}^*(\mathbf{R}, \mathbf{h})$ and $w^*(\mathbf{R}, \mathbf{h})$ solve the optimization problem

$$\min_{\mathcal{P}} \left\{ E_{\mathbf{h}} \left[w(\mathbf{R}, \mathbf{h}) \boldsymbol{\lambda} \cdot \mathbf{P}(\mathbf{R}, \mathbf{h}) \right] - s^* \cdot E_{\mathbf{h}} \left[w(\mathbf{R}, \mathbf{h}) \right] \right\} \tag{6.28}$$

subject to:

$$\begin{cases} \mathbf{R} \in C_g(\mathbf{h}, \mathbf{P}(\mathbf{R}, \mathbf{h})), \quad \forall \mathbf{h} \in \mathcal{H}_{tran}, \\ \\ E_{\mathbf{h}} \left[w(\mathbf{R}, \mathbf{h}) \right] = 1 - Pr^*. \end{cases} \tag{6.29}$$

It is easily shown, as in Section 6.4.2, that $\forall \mathbf{h} \in \mathcal{H}_{tran}$, the optimal transmit power vector $\mathbf{P}^*(\mathbf{R}, \mathbf{h})$ of the M users is $\mathbf{P}_{\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})$, the components of which are given in (6.17). Therefore, the optimization problem (6.28) subject to (6.29) is equivalent to the one in (6.19) subject to:

$$E_{\mathbf{h}} \left[w(\mathbf{R}, \mathbf{h}) \right] = 1 - Pr^*. \tag{6.30}$$

Note that as described in Section 6.4.3, $\forall \mathbf{h} \in \mathcal{H}_{all}$, the solution $w_{\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})$ to (6.19) given by the *Common Outage Transmission Policy* is readily obtained once the threshold power s^* is known. In Section 6.4.3 s^* is determined by the total average weighted power constraint in (6.20). Now we choose s^* such that the average common outage probability constraint (6.30) is satisfied. That is, s^* satisfies

$$Prob\{\mathbf{h} \in \mathcal{R}(s^*, \boldsymbol{\lambda})\} = 1 - Pr^*. \tag{6.31}$$

We then obtain the solution $w_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h})$ to (6.19) ($\forall \mathbf{h} \in \mathcal{H}_{all}$) by applying the *Common Outage Transmission Policy* in Section 6.4.3. Given the derived $\mathbf{P}_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h})$ and $w_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h})$ for every state $\mathbf{h} \in \mathcal{H}_{all}$, the complete power allocation policy is known and the corresponding average power vector $\bar{\mathbf{P}}_{\boldsymbol{\lambda}}(\mathbf{R})$ can be easily obtained as shown in Section 6.4.1, i.e.,

$$\begin{aligned} \bar{\mathbf{P}}_{\boldsymbol{\lambda}}(\mathbf{R}) &= E_{\mathbf{h}}[w_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h})\mathbf{P}_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h})] \\ &= E_{\mathbf{h}\in\mathcal{R}(s^*,\boldsymbol{\lambda})}[\mathbf{P}_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h})]. \end{aligned} \tag{6.32}$$

Therefore, by varying the power price vector $\boldsymbol{\lambda} \in \Re_+^M$, we can obtain different average power vectors $\bar{\mathbf{P}}_{\boldsymbol{\lambda}}(\mathbf{R})$ that lie on the boundary surface of $APV_{out}(Pr^*,\mathbf{R})$.

However, as discussed in [3] for the zero-outage capacity case, there are other average power vectors on the boundary surface of $APV_{out}(Pr^*,\mathbf{R})$ that cannot be parameterized by any $\boldsymbol{\lambda} \in \Re_+^M$ since, as will be shown shortly, they correspond to decoding rules that cannot be represented simply by any $\boldsymbol{\lambda} \in \Re_+^M$. Therefore, it is necessary to use a more general method [3] for describing all possible decoding orders. Specifically, let $\mathcal{L}$ denote a set of subsets of $G \triangleq \{1,2,\cdots,M\}$ with all subsets in $\mathcal{L}$ nested. That is, if G_i and G_j are two subsets of G in $\mathcal{L}$, then $G_i \subseteq G_j$ or $G_j \subseteq G_i$. Consider successive decoding with the ordering determined by $\boldsymbol{\lambda}$ as before, except that now the users in a set $G_i \in \mathcal{L}$ have priority over the users in set G_i^c for every fading state $\mathbf{h}$, i.e., they are always decoded later. The decoding order of the users within set G_i is still determined by their power prices $(\lambda_k)_{k\in G_i}$. For example, if $G_1 \subseteq G_2 \subseteq \cdots \subseteq G$, then all the users in G_i are decoded after users in G_i^c and therefore provide interference to each user in G_i^c, $i = 1,2,\cdots$. Therefore, $(\lambda_k)_{k\in G_1}$ is used to determine the decoding order of users in G_1, and all these users provide interference to users in $G_2\backslash G_1$, the decoding order of which is determined by $(\lambda_k)_{k\in G_2\backslash G_1}$ completely. Inductively, $(\lambda_k)_{k\in G_i\backslash G_{i-1}}$ is used to determine the decoding order of users in $G_i\backslash G_{i-1}$, and these users provide interference to each user in $G_{i+1}\backslash G_i$. If $\mathcal{L} = \{G\}$, then all the M users are of the same priority and their decoding order is then determined by the power price vector $\boldsymbol{\lambda}$ completely. Therefore, a decoding order determined by $(\boldsymbol{\lambda},\mathcal{L})$ with $\mathcal{L} \neq \{G\}$ is different from any of the orders determined by $(\boldsymbol{\lambda},\mathcal{L})$ with ($\mathcal{L} = \{G\}$), and $(\boldsymbol{\lambda},\mathcal{L})$ can be viewed as the *power allocation parameter pair* that describes all possible decoding orders.

For each given power allocation parameter pair $(\boldsymbol{\lambda},\mathcal{L})$, the decoding order in every fading state $\mathbf{h}$ is determined and we have shown for $\mathcal{L} = \{G\}$ that the boundary average power vector $\bar{\mathbf{P}}_{(\boldsymbol{\lambda},\{G\})}(\mathbf{R}) \triangleq \bar{\mathbf{P}}_{\boldsymbol{\lambda}}(\mathbf{R})$ of region $APV_{out}(Pr^*, \mathbf{R})$ can be obtained from (6.32). For each power price vector $\boldsymbol{\lambda} \in \Re_+^M$, the average power vector $\bar{\mathbf{P}}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R})$ with $\mathcal{L} = \{G\}$ is called a *regular point* on the boundary surface of $APV_{out}(Pr^*, \mathbf{R})$, and the average power vector $\bar{\mathbf{P}}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R})$ with $\mathcal{L} \neq \{G\}$ is called a *nonregular point*. We now show how to calculate the nonregular point $\bar{\mathbf{P}}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R})$ on the boundary surface of $APV_{out}(Pr^*, \mathbf{R})$.

For a given set $\mathcal{L}$ of nested subsets of G, let the subsets be denoted as $G_1, G_2, \cdots, G_I$, where $G_1 \subseteq G_2 \subseteq \cdots \subseteq G_I \equiv G$, $I \geq 2$. Then $(\lambda_k)_{k \in G_1}$ determines the decoding order of the users in G_1 and they are decoded after the signals from the users in G_1^c are decoded and subtracted out for every fading state $\mathbf{h}$. Therefore, there is no interference from any user in G_1^c to the users in G_1, and the required average power for each user in G_1 is independent of the power prices and channel conditions of the users in G_1^c. Since the users in G_1 have absolute priority over all other users, we determine the fading states where no power should be assigned to any user by considering users in G_1 only. That is, $\forall \mathbf{h} \in \mathcal{H}_{all}$, first define the total weighted power $P^{min}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}, \mathbf{h})$ of all the users in G_1 as:

$$P^{min}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}, \mathbf{h}) = \boldsymbol{\lambda}^{(G_1)} \cdot \mathbf{P}^{(G_1)}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}, \mathbf{h}), \tag{6.33}$$

where the subscript G_1 in vectors $\boldsymbol{\lambda}^{(G_1)}$ and $\mathbf{P}^{(G_1)}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}, \mathbf{h})$ implies that they are subvectors of $\boldsymbol{\lambda}$ and $\mathbf{P}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}, \mathbf{h})$, consisting of components corresponding to the users in set G_1. Here $\mathbf{P}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}, \mathbf{h}) = [P_{1,(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}, \mathbf{h}), P_{2,(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}, \mathbf{h}), \cdots, P_{M,(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}, \mathbf{h})]$ denotes the required transmit power vector of the M users when the decoding order is determined by the power allocation parameter pair $(\boldsymbol{\lambda}, \mathcal{L})$. That is, $\forall \mathbf{h} \in \mathcal{H}_{all}$, the required transmit power of each user in set G_1 is:

$$P_{\pi_{G_1}(i),(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}, \mathbf{h}) = \begin{cases} \frac{\sigma^2}{h_{\pi_{G_1}(1)}} \left[\exp(2R_{\pi_{G_1}(1)}) - 1\right], & \text{if } i = 1, \\ \frac{\sigma^2}{h_{\pi_{G_1}(i)}} \left[\exp\left(2\sum_{k=1}^{i} R_{\pi_{G_1}(k)}\right) - \exp\left(2\sum_{k=1}^{i-1} R_{\pi_{G_1}(k)}\right)\right], & \\ & \forall 2 \leq i \leq |G_1|, \end{cases} \tag{6.34}$$

where $|G_1|$ denotes the number of users in set G_1 and the permutation $\pi_{G_1}(\cdot)$ of the $|G_1|$ users satisfies:

$$\frac{\lambda_{\pi_{G_1}(1)}}{h_{\pi_{G_1}(1)}} \geq \frac{\lambda_{\pi_{G_1}(2)}}{h_{\pi_{G_1}(2)}} \geq \cdots \geq \frac{\lambda_{\pi_{G_1}(|G_1|)}}{h_{\pi_{G_1}(|G_1|)}}.$$

The required transmit power of each user in set $G_j \backslash G_{j-1}$ $(2 \leq j \leq I)$ for every state

$\mathbf{h} \in \mathcal{H}_{all}$ is:

$$P_{\pi_{G_j}(i),(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R},\mathbf{h}) = \begin{cases} \frac{\sigma^2}{h_{\pi_{G_j}(1)}} \cdot \exp\left(2 \sum_{k \in G_{j-1}} R_k\right) \cdot \left[\exp(2R_{\pi_{G_j}(1)}) - 1\right], & \text{if } i = 1, \\ \frac{\sigma^2}{h_{\pi_{G_j}(i)}} \cdot \exp\left(2 \sum_{k \in G_{j-1}} R_k\right) \cdot \left[\exp\left(2\sum_{k=1}^{i} R_{\pi_{G_j}(k)}\right)\right. \\ \qquad \left. - \exp\left(2\sum_{k=1}^{i-1} R_{\pi_{G_j}(k)}\right)\right], & \forall 2 \leq i \leq |G_j \backslash G_{j-1}|, \end{cases} \tag{6.35}$$

where $|G_j \backslash G_{j-1}|$ denotes the number of users in set $G_j \backslash G_{j-1}$ and the permutation $\pi_{G_j}(\cdot)$ of the $|G_j \backslash G_{j-1}|$ users satisfies:

$$\frac{\lambda_{\pi_{G_j}(1)}}{h_{\pi_{G_j}(1)}} \geq \frac{\lambda_{\pi_{G_j}(2)}}{h_{\pi_{G_j}(2)}} \geq \cdots \geq \frac{\lambda_{\pi_{G_j}(|G_j \backslash G_{j-1}|)}}{h_{\pi_{G_j}(|G_j \backslash G_{j-1}|)}}.$$

Now $\forall s > 0$, define the set of fading states $\mathcal{R}(s, \boldsymbol{\lambda}, \mathcal{L})$ as:

$$\mathcal{R}(s, \boldsymbol{\lambda}, \mathcal{L}) = \{\mathbf{h} : P^{min}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}, \mathbf{h}) \leq s\}, \tag{6.36}$$

where $P^{min}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}, \mathbf{h})$ is given in (6.33), and let the threshold power s^* be the point that satisfies:

$$Prob\{\mathbf{h} \in \mathcal{R}(s^*, \boldsymbol{\lambda}, \mathcal{L})\} = 1 - Pr^*. \tag{6.37}$$

Given s^*, the set $\mathcal{R}(s^*, \boldsymbol{\lambda}, \mathcal{L})$ is then determined, which corresponds to the set of all fading states where the M users transmit information at rate vector $\mathbf{R}$. For any fading state $\mathbf{h} \notin \mathcal{R}(s^*, \boldsymbol{\lambda}, \mathcal{L})$, an outage will be declared for all the M users. Therefore, since in every state $\mathbf{h} \in \mathcal{R}(s^*, \boldsymbol{\lambda}, \mathcal{L})$, the decoding order of the M users is determined by $(\boldsymbol{\lambda}, \mathcal{L})$ and so is the required transmit power of the M users given in (6.34) and (6.35), the required average power vector $\bar{\mathbf{P}}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R})$ for supporting rate vector $\mathbf{R}$ with common outage probability Pr^* is given by:

$$\bar{\mathbf{P}}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}) = E_{\mathbf{h} \in \mathcal{R}(s^*, \boldsymbol{\lambda}, \mathcal{L})}[\mathbf{P}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}, \mathbf{h})], \tag{6.38}$$

which will lie on the boundary surface of $APV_{out}(Pr^*, \mathbf{R})$.

6.5 Independent Outage Declaration

We now consider the case where each user can declare an outage independently. We derive the boundary of the outage probability region $\mathcal{O}_I(\bar{\mathbf{P}}, \mathbf{R})$ as well as the optimal power allocation strategy in Section 6.5.1- 6.5.3, and obtain the average power region of the M users required to support $\mathbf{R}$ with a given outage probability vector $\mathbf{Pr}$ in Section 6.5.4.

For any given fading state $\mathbf{h}$, since some users may declare an outage, the channel may be only used by a subset of the M users. Obviously there are $\sum_{i=0}^{M}\binom{M}{i} = 2^M$ different combinations (subsets) of the M users, including the empty set. We will represent each of these 2^M possible combinations of users as an M-dimensional vector $[\psi(k,1), \psi(k,2), \cdots, \psi(k,M)]$ equal to the binary expansion of k, $0 \leq k \leq 2^M - 1$. For each vector $[\psi(k,1), \psi(k,2), \cdots, \psi(k,M)]$, if $\psi(k,i) = 1$ then User i is transmitting; otherwise User i is not. For each k, $0 \leq k \leq 2^M - 1$, we define the set of active users S_k relative to k as $S_k = \{i : \psi(k,i) = 1, 1 \leq i \leq M\}$. S_0 denotes the empty set (no active users).

6.5.1 Power Allocation Policy

For a given rate vector $\mathbf{R}$, in each fading state $\mathbf{h}$, by "power allocation policy $\mathcal{P}$," we mean the following:

(1) Let $w(\mathbf{R}, \mathbf{h}, S_k)$ denote the probability that S_k defines the set of active users (i.e., in state $\mathbf{h}$ only users in set S_k are active and each user $i \in S_k$ transmits at rate R_i)[4].

(2) Let $P_i(\mathbf{R}, \mathbf{h}, S_k)$ denote the transmit power of each user $i \in S_k$ when S_k defines the set of active users in state $\mathbf{h}$. Let $\mathbf{P}(\mathbf{R}, \mathbf{h}, S_k) \triangleq [P_1(\mathbf{R}, \mathbf{h}, S_k),\ P_2(\mathbf{R}, \mathbf{h}, S_k),\ \cdots,\ P_M(\mathbf{R}, \mathbf{h}, S_k)]$, $\forall 0 \leq k \leq 2^M - 1$.

Then obviously,

$$P_i(\mathbf{R}, \mathbf{h}, S_k) = 0, \quad \forall i \notin S_k, \tag{6.39}$$

and

$$\sum_{k=1}^{2^M-1} w(\mathbf{R}, \mathbf{h}, S_k) = 1 - w(\mathbf{R}, \mathbf{h}, S_0) \leq 1.$$

[4]As will be shown later in the optimal power allocation policy, $w(\mathbf{R}, \mathbf{h}, S_k)$ is either 1 or 0 under the assumption that the stationary distribution of the joint fading process of the M users has continuous density, since the probability measure of each fading state $\mathbf{h}$ is zero. However, as in the broadcast communication case discussed in Chapter 5, this is not true if the c.d.f. (cumulative density function) of the joint fading process is discontinuous.

Moreover, in each fading state $\mathbf{h}$, by denoting $Pr_i^{on}(\mathbf{R},\mathbf{h})$ and $P_i(\mathbf{R},\mathbf{h})$ as, respectively, the probability and transmit power with which each user i transmits information to the base station at rate R_i, we have

$$Pr_i^{on}(\mathbf{R},\mathbf{h}) = \sum_{k=1}^{2^M-1} w(\mathbf{R},\mathbf{h},S_k)\mathbf{1}[i \in S_k], \tag{6.40}$$

$$P_i(\mathbf{R},\mathbf{h}) = \sum_{k=1}^{2^M-1} w(\mathbf{R},\mathbf{h},S_k)P_i(\mathbf{R},\mathbf{h},S_k),$$

where $\mathbf{1}[\cdot]$ denotes the indicator function ($\mathbf{1}[x] = 1$ if x is true and zero otherwise). Therefore, the average usage probability $Pr_i^{on}(\mathbf{R})$ and average transmit power $\bar{P}_i(\mathbf{R})$ of each user i ($1 \le i \le M$) for rate vector $\mathbf{R}$ are:

$$\begin{aligned} Pr_i^{on}(\mathbf{R}) &= E_{\mathbf{h}}\left[Pr_i^{on}(\mathbf{R},\mathbf{h})\right] \\ &= E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w(\mathbf{R},\mathbf{h},S_k)\mathbf{1}[i \in S_k]\right], \end{aligned} \tag{6.41}$$

$$\begin{aligned} \bar{P}_i(\mathbf{R}) &= E_{\mathbf{h}}\left[P_i(\mathbf{R},\mathbf{h})\right] \\ &= E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w(\mathbf{R},\mathbf{h},S_k)P_i(\mathbf{R},\mathbf{h},S_k)\right]. \end{aligned} \tag{6.42}$$

The average outage probability $Pr_i(\mathbf{R})$ of each user i ($1 \le i \le M$) is

$$Pr_i(\mathbf{R}) = 1 - Pr_i^{on}(\mathbf{R}), \tag{6.43}$$

where $Pr_i^{on}(\mathbf{R})$ is given in (6.41).

6.5.2 Outage Probability Region

In this subsection, for any given vector $\mathbf{X} = (X_1, X_2, \cdots, X_M)$ and any set $S \in \{S_k\}_{k=1}^{2^M-1}$, let $|S|$ denote the number of users in set S, and let $\mathbf{X}^{(S)}$ denote the subvector of $\mathbf{X}$ consisting of components corresponding to the $|S|$ users in set S.

From *Definition 6.7* and *Definition 6.8* it is clear that for a given average power constraint vector $\bar{\mathbf{P}}^*$ and rate vector $\mathbf{R}$, deriving the boundary of the outage probability region $\mathcal{O}_I(\bar{\mathbf{P}}^*, \mathbf{R})$ is equivalent to deriving the boundary of the usage probability region $\overline{\mathcal{O}}_I(\bar{\mathbf{P}}^*, \mathbf{R})$.

Define

$$\mathcal{Q}_I = \{(\mathbf{Pr}^{on}, \bar{\mathbf{P}}) : \mathbf{Pr}^{on} \in \overline{\mathcal{O}}_I(\bar{\mathbf{P}}, \mathbf{R})\}. \tag{6.44}$$

We will require the following lemma and definition to derive the boundary of $\overline{\mathcal{O}}_I(\bar{\mathbf{P}}^*, \mathbf{R})$ and the corresponding optimal power allocation that achieves this boundary:

Lemma 6.2 *Both the usage probability region $\overline{\mathcal{O}}_I(\bar{\mathbf{P}}^*, \mathbf{R})$ and the set $\mathcal{Q}_I$ are convex.*

Proof: See Appendix D.2. □

Definition 6.9 *Assume that the M users transmit at a given rate vector $\mathbf{R}$ and the average outage probability for User i is $Pr_i(\mathbf{R})$, $1 \le i \le M$. The average usage probability for User i is denoted as $Pr_i^{on}(\mathbf{R}) = 1 - Pr_i(\mathbf{R})$. Let $\mathbf{Pr}^{on}(\mathbf{R}) \triangleq [Pr_1^{on}(\mathbf{R}), Pr_2^{on}(\mathbf{R}), \cdots, Pr_M^{on}(\mathbf{R})]$. Then for any given nonnegative vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_M) \in \Re^M$, the total average usage reward $W(\mathbf{R})$ is defined as:*

$$\begin{aligned} W(\mathbf{R}) &= \boldsymbol{\mu} \cdot \mathbf{Pr}^{on}(\mathbf{R}) \\ &= \sum_{i=1}^{M} \mu_i Pr_i^{on}(\mathbf{R}). \end{aligned} \tag{6.45}$$

In (6.45), μ_i can be viewed as the channel usage reward if the information from User i is transmitted[5].

Due to the convexity of $\overline{\mathcal{O}}_I(\bar{\mathbf{P}}^*, \mathbf{R})$, an average usage probability vector will be on the boundary surface of $\overline{\mathcal{O}}_I(\bar{\mathbf{P}}^*, \mathbf{R})$ if it is a solution to

$$\max_{\mathbf{Pr}^{on}(\mathbf{R}) \in \overline{\mathcal{O}}_I(\bar{\mathbf{P}}^*, \mathbf{R})} W(\mathbf{R}) \tag{6.46}$$

for some nonnegative vector $\boldsymbol{\mu} \in \Re^M$, where $W(\mathbf{R})$ is defined in (6.45). In this subsection we focus on the strictly positive vector $\boldsymbol{\mu}$ for a parameterization of the boundary surface of $\overline{\mathcal{O}}_I(\bar{\mathbf{P}}^*, \mathbf{R})$, and we call the corresponding outage probability vector a *regular point* of the boundary surface. The *nonregular points* correspond to the case where some components of vector $\boldsymbol{\mu}$ equal zero. Since we can get arbitrarily close to a nonregular point by letting some components of vector $\boldsymbol{\mu}$ go to zero, it suffices to focus on deriving the regular points of the boundary surface. Moreover, we will show how to obtain the nonregular points explicitly based on the the method for deriving a regular point at the end of Section 6.5.3.

[5] μ_i can also be viewed as the channel outage penalty if an outage is declared from User i.

Since the set $\mathcal{Q}_I$ is convex, for a given channel usage reward vector $\boldsymbol{\mu} \in \Re_+^M$, vector $\mathbf{Pr}_{\boldsymbol{\mu}}^{on}(\mathbf{R}) = \left[Pr_{1,\boldsymbol{\mu}}^{on}(\mathbf{R}),\ Pr_{2,\boldsymbol{\mu}}^{on}(\mathbf{R}),\ \cdots,\ Pr_{M,\boldsymbol{\mu}}^{on}(\mathbf{R})\right]$ solves (6.46) if and only if there exists a Lagrangian multiplier vector $\boldsymbol{\lambda} \in \Re_+^M$ such that $\left(\mathbf{Pr}_{\boldsymbol{\mu}}^{on}(\mathbf{R}), \bar{\mathbf{P}}^*\right)$ is a solution to the problem

$$\max_{(\mathbf{Pr}^{on}(\mathbf{R}), \bar{\mathbf{P}}(\mathbf{R})) \in \mathcal{Q}_I} \left[W(\mathbf{R}) - \boldsymbol{\lambda} \cdot \bar{\mathbf{P}}(\mathbf{R})\right], \tag{6.47}$$

where $\bar{\mathbf{P}}(\mathbf{R}) = [\bar{P}_1(\mathbf{R}), \bar{P}_2(\mathbf{R}), \cdots, \bar{P}_M(\mathbf{R})]$ and, for a given power allocation policy $\mathcal{P}$, $\bar{P}_i(\mathbf{R})$ $(1 \le i \le M)$ can be calculated from (6.42). Note that we transform the maximization problem (6.46) into the problem (6.47) because the scalar to be maximized in (6.47) includes the average power constraints, and is therefore an easier maximization problem to solve. In (6.47),

$$\begin{aligned}
\boldsymbol{\lambda} \cdot \bar{\mathbf{P}}(\mathbf{R}) &= \sum_{i=1}^{M} \lambda_i \bar{P}_i(\mathbf{R}) \\
&= \sum_{i=1}^{M} \lambda_i E_{\mathbf{h}} \left[\sum_{k=1}^{2^M-1} w(\mathbf{R}, \mathbf{h}, S_k) P_i(\mathbf{R}, \mathbf{h}, S_k)\right] \\
&= E_{\mathbf{h}} \left[\sum_{k=1}^{2^M-1} w(\mathbf{R}, \mathbf{h}, S_k) \left(\sum_{i=1}^{M} \lambda_i P_i(\mathbf{R}, \mathbf{h}, S_k)\right)\right] \\
&= E_{\mathbf{h}} \left[\sum_{k=1}^{2^M-1} w(\mathbf{R}, \mathbf{h}, S_k) \left(\boldsymbol{\lambda}^{(S_k)} \cdot \mathbf{P}^{(S_k)}(\mathbf{R}, \mathbf{h}, S_k)\right)\right],
\end{aligned} \tag{6.48}$$

where the last equation results from (6.39).

For a given rate vector $\mathbf{R}$ and fading state $\mathbf{h}$, let $W(\mathbf{R}, \mathbf{h})$ denote the total usage reward of the M users in state $\mathbf{h}$, i.e.,

$$\begin{aligned}
W(\mathbf{R}, \mathbf{h}) &= \boldsymbol{\mu} \cdot \mathbf{Pr}^{on}(\mathbf{R}, \mathbf{h}) \\
&= \sum_{i=1}^{M} \mu_i Pr_i^{on}(\mathbf{R}, \mathbf{h}).
\end{aligned} \tag{6.49}$$

Substituting (6.40) into (6.49), we have

$$\begin{aligned}
W(\mathbf{R}, \mathbf{h}) &= \sum_{i=1}^{M} \mu_i \left(\sum_{k=1}^{2^M-1} w(\mathbf{R}, \mathbf{h}, S_k) \mathbf{1}[i \in S_k]\right) \\
&= \sum_{k=1}^{2^M-1} w(\mathbf{R}, \mathbf{h}, S_k) \left(\sum_{i=1}^{M} \mu_i \mathbf{1}[i \in S_k]\right)
\end{aligned}$$

$$= \sum_{k=1}^{2^M-1} w(\mathbf{R},\mathbf{h},S_k)\eta_k, \tag{6.50}$$

where the total reward for all users in set S_k to transmit information is

$$\eta_k \triangleq \sum_{i=1}^{M} \mu_i \mathbf{1}[i \in S_k], \quad 1 \le k \le 2^M - 1. \tag{6.51}$$

Thus, the total usage reward averaged over the time-varying channel is

$$\begin{aligned} W(\mathbf{R}) &= E_{\mathbf{h}}\left[W(\mathbf{R},\mathbf{h})\right] \\ &= E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w(\mathbf{R},\mathbf{h},S_k)\eta_k\right]. \end{aligned} \tag{6.52}$$

For a given $\boldsymbol{\mu} \in \Re_+^M$, we know that vector $\mathbf{Pr}_{\boldsymbol{\mu}}^{on}(\mathbf{R})$ solves (6.46) (i.e., $\mathbf{Pr}_{\boldsymbol{\mu}}^{on}(\mathbf{R})$ is the boundary average usage probability vector of $\overline{\mathcal{O}}_I(\bar{\mathbf{P}}^*,\mathbf{R})$ for the given $\boldsymbol{\mu}$) if and only if there exists a Lagrangian multiplier vector $\boldsymbol{\lambda} \in \Re_+^M$ such that $\left(\mathbf{Pr}_{\boldsymbol{\mu}}^{on}(\mathbf{R}), \bar{\mathbf{P}}^*\right)$ is a solution to (6.47). Substituting (6.48) and (6.52) into (6.47), it is clear that $\mathbf{Pr}_{\boldsymbol{\mu}}^{on}(\mathbf{R})$ is the boundary vector of $\overline{\mathcal{O}}_I(\bar{\mathbf{P}}^*,\mathbf{R})$ for the given $\boldsymbol{\mu}$ if and only if there exist a Lagrangian multiplier vector $\boldsymbol{\lambda} \in \Re_+^M$ and a power allocation policy $\mathcal{P}^*$ such that for all $\mathbf{h} \in \mathcal{H}_{all}$, $\{\mathbf{P}^*(\mathbf{R},\mathbf{h},S_k)\}_{k=1}^{2^M-1}$ and $\{w^*(\mathbf{R},\mathbf{h},S_k)\}_{k=1}^{2^M-1}$ solve the maximization problem:

$$\max_{\mathcal{P}} \left\{ E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w(\mathbf{R},\mathbf{h},S_k)\eta_k\right] - E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w(\mathbf{R},\mathbf{h},S_k)\boldsymbol{\lambda}^{(S_k)} \cdot \mathbf{P}^{(S_k)}(\mathbf{R},\mathbf{h},S_k)\right]\right\} \tag{6.53}$$

subject to:

$$\begin{cases} \mathbf{R}^{(S)} \in C_g\left(\mathbf{h}, \mathbf{P}^{(S)}(\mathbf{R},\mathbf{h},S)\right), \quad \forall \mathbf{h} \in \mathcal{H}_S, \quad \forall S \in \{S_k\}_{k=1}^{2^M-1}, \\ E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w(\mathbf{R},\mathbf{h},S_k)P_i(\mathbf{R},\mathbf{h},S_k)\right] = \bar{P}_i^*, \quad \forall 1 \le i \le M, \end{cases} \tag{6.54}$$

where

$$\mathcal{H}_S = \{\mathbf{h} : w(\mathbf{R},\mathbf{h},S) > 0\}, \tag{6.55}$$

$C_g\left(\mathbf{h}, \mathbf{P}^{(S)}(\mathbf{R},\mathbf{h},S)\right)$ is the capacity region of the $|S|$-user time-invariant Gaussian MAC and is given by

$$C_g\left(\mathbf{h},\mathbf{P}^{(S)}(\mathbf{R},\mathbf{h},S)\right)=\left\{\mathbf{R}^{(S)}:\sum_{i\in F}R_i\le\frac{1}{2}\log\left(1+\frac{\sum_{i\in F}h_iP_i(\mathbf{R},\mathbf{h},S)}{\sigma^2}\right),\quad\forall F\subseteq S\right\},\tag{6.56}$$

and the average usage probability $Pr^{on}_{i,\boldsymbol{\mu}}(\mathbf{R})$ of each user i is:

$$Pr^{on}_{i,\boldsymbol{\mu}}(\mathbf{R})=E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1}w^*(\mathbf{R},\mathbf{h},S_k)\mathbf{1}[i\in S_k]\right],\quad\forall 1\le i\le M.\tag{6.57}$$

Therefore, for a given rate vector $\mathbf{R}$ and channel usage reward vector $\boldsymbol{\mu}$, in order to derive the average usage probability boundary vector $\mathbf{Pr}^{on}_{\boldsymbol{\mu}}(\mathbf{R})$ of $\overline{\mathcal{O}}_I(\bar{\mathbf{P}}^*,\mathbf{R})$, we wish to find the appropriate Lagrangian multiplier vector $\boldsymbol{\lambda}$, the optimal transmit power vector $\mathbf{P}^*(\mathbf{R},\mathbf{h},S_k)$, and the optimal transmission probability $w^*(\mathbf{R},\mathbf{h},S_k)$ for each set S_k ($1\le k\le 2^M-1$, $\forall\mathbf{h}\in\mathcal{H}_{all}$) that achieve the maximum in (6.53) under the constraints in (6.54). Assuming that the Lagrangian multiplier vector $\boldsymbol{\lambda}$ is known, to achieve the maximum in (6.53), $\forall S\in\{S_k\}_{k=1}^{2^M-1}$, the optimal transmit power vector $\mathbf{P}^{*(S)}(\mathbf{R},\mathbf{h},S)$ of the $|S|$ users in set S must be a solution to

$$\min_{\mathbf{P}^{(S)}(\mathbf{R},\mathbf{h},S)}\ \boldsymbol{\lambda}^{(S)}\cdot\mathbf{P}^{(S)}(\mathbf{R},\mathbf{h},S)\quad\text{subject to:}\quad\mathbf{R}^{(S)}\in C_g\left(\mathbf{h},\mathbf{P}^{(S)}(\mathbf{R},\mathbf{h},S)\right),\quad\forall\mathbf{h}\in\mathcal{H}_S.\tag{6.58}$$

Note that according to (6.39), $\forall i\notin S$, $P_i^*(\mathbf{R},\mathbf{h},S)=0$.

Since $\forall S\in\{S_k\}_{k=1}^{2^M-1}$, $\forall\mathbf{h}\in\mathcal{H}_S$, the set of *received powers* that can support the given rate vector $\mathbf{R}^{(S)}$ of the users in set S is

$$\mathcal{G}(\mathbf{R}^{(S)},\mathbf{h})\triangleq\left\{\mathbf{Q}^{(S)}:Q_i=h_iP_i(\mathbf{R},\mathbf{h},S),\forall i\in S,\ \ \mathbf{R}^{(S)}\in C_g\left(\mathbf{h},\mathbf{P}^{(S)}(\mathbf{R},\mathbf{h},S)\right)\right\},$$

which is shown to be a contra-polymatroid with rank function [3]

$$f(F)=\sigma^2\left[\exp\left(2\sum_{i\in F}R_i\right)-1\right],\quad\forall F\subseteq S,$$

for any given $\boldsymbol{\lambda}\in\Re_+^M$, $\forall\mathbf{h}\in\mathcal{H}_S$, applying *Lemma 3.4* in [3], the solution $\mathbf{P}^{(S)}_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S)$ to the optimization problem (6.58) is readily obtained. That is, the required transmit power for each user in set S is:

$$P_{\pi_s(i),\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S) = \begin{cases} \frac{\sigma^2}{h_{\pi_s(1)}}\left[\exp(2R_{\pi_s(1)}) - 1\right], & \text{if } i = 1, \\ \frac{\sigma^2}{h_{\pi_s(i)}}\left[\exp(2\sum_{k=1}^{i} R_{\pi_s(k)}) - \exp(2\sum_{k=1}^{i-1} R_{\pi_s(k)})\right], & \forall 2 \leq i \leq |S|, \end{cases} \tag{6.59}$$

where the permutation $\pi_s(\cdot)$ of the $|S|$ users satisfies:

$$\frac{\lambda_{\pi_s(1)}}{h_{\pi_s(1)}} \geq \frac{\lambda_{\pi_s(2)}}{h_{\pi_s(2)}} \geq \cdots \geq \frac{\lambda_{\pi_s(|S|)}}{h_{\pi_s(|S|)}}.$$

As for the common outage case, the solution in (6.59) indicates that $\boldsymbol{\lambda}^{(S)}$ determines the decoding order of the users in set S in each fading state $\mathbf{h} \in \mathcal{H}_S$, $\forall S \in \{S_k\}_{k=1}^{2^M-1}$. Therefore, $\boldsymbol{\lambda}$ can be viewed as the *power price vector* of the M users, which will be used to refer to the vector $\boldsymbol{\lambda}$ hereafter.

Now $\forall S \in \{S_k\}_{k=1}^{2^M-1}$, denote the minimum in (6.58) as $P_{\boldsymbol{\lambda}}^{min}(\mathbf{R},\mathbf{h},S)$, i.e.,

$$P_{\boldsymbol{\lambda}}^{min}(\mathbf{R},\mathbf{h},S) = \boldsymbol{\lambda}^{(S)} \cdot \mathbf{P}_{\boldsymbol{\lambda}}^{(S)}(\mathbf{R},\mathbf{h},S), \tag{6.60}$$

where the components of vector $\mathbf{P}_{\boldsymbol{\lambda}}^{(S)}(\mathbf{R},\mathbf{h},S)$ are given in (6.59). Then $\forall 1 \leq k \leq 2^M - 1$, $P_{\boldsymbol{\lambda}}^{min}(\mathbf{R},\mathbf{h},S_k)$ can be viewed as the minimum total weighted power of the $|S_k|$ users required to transmit their information at rate $\mathbf{R}^{(S_k)}$ and achieve the usage reward η_k in state $\mathbf{h}$. Since the M average power constraints in (6.54) imply that there is a total weighted average power constraint $\boldsymbol{\lambda} \cdot \bar{\mathbf{P}}^*$ for a given power price vector $\boldsymbol{\lambda}$, in order to solve (6.53) for the given channel usage reward vector $\boldsymbol{\mu} \in \Re_+^M$ (η_k in (6.53) is a function of $\boldsymbol{\mu}$ as shown in (6.51)), we first find the solution $\left\{w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_k)\right\}_{k=1}^{2^M-1}$ to the following single-constraint maximization problem:

$$\begin{cases} \max_{\{w(\mathbf{R},\mathbf{h},S_k)\}_{k=1}^{2^M-1}} E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w(\mathbf{R},\mathbf{h},S_k)\eta_k\right] \\ \\ \text{subject to: } E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w(\mathbf{R},\mathbf{h},S_k) P_{\boldsymbol{\lambda}}^{min}(\mathbf{R},\mathbf{h},S_k)\right] \leq \boldsymbol{\lambda} \cdot \bar{\mathbf{P}}^*, \end{cases}$$

which is equivalent to

$$\max_{\{w(\mathbf{R},\mathbf{h},S_k)\}_{k=1}^{2^M-1}} \left\{ E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w(\mathbf{R},\mathbf{h},S_k)\eta_k\right] - \frac{1}{s^*} E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w(\mathbf{R},\mathbf{h},S_k) P_{\boldsymbol{\lambda}}^{min}(\mathbf{R},\mathbf{h},S_k)\right] \right\} \tag{6.61}$$

subject to:

$$E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w(\mathbf{R},\mathbf{h},S_k) P_{\boldsymbol{\lambda}}^{min}(\mathbf{R},\mathbf{h},S_k)\right] = \boldsymbol{\lambda}\cdot\bar{\mathbf{P}}^*, \tag{6.62}$$

where $\frac{1}{s^*}$ is the Lagrangian multiplier to be chosen such that the average total weighted power constraint $\boldsymbol{\lambda}\cdot\bar{\mathbf{P}}^*$ is satisfied. We propose an iterative algorithm in Section 6.6 to find the optimal power price vector $\boldsymbol{\lambda}^*$ such that the solutions $\left\{\mathbf{P}_{\boldsymbol{\lambda}^*}(\mathbf{R},\mathbf{h},S_k)\right\}_{k=1}^{2^M-1}$ and $\left\{w_{\boldsymbol{\mu},\boldsymbol{\lambda}^*}(\mathbf{R},\mathbf{h},S_k)\right\}_{k=1}^{2^M-1}$ satisfy the M average power constraints in (6.54), i.e.,

$$E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w_{\boldsymbol{\mu},\boldsymbol{\lambda}^*}(\mathbf{R},\mathbf{h},S_k)\mathbf{P}_{\boldsymbol{\lambda}^*}(\mathbf{R},\mathbf{h},S_k)\right] = \bar{\mathbf{P}}^*. \tag{6.63}$$

Therefore, given the power price vector $\boldsymbol{\lambda}^*$, the power allocation policy defined by the solutions $\left\{\mathbf{P}_{\boldsymbol{\lambda}^*}(\mathbf{R},\mathbf{h},S_k)\right\}_{k=1}^{2^M-1}$ and $\left\{w_{\boldsymbol{\mu},\boldsymbol{\lambda}^*}(\mathbf{R},\mathbf{h},S_k)\right\}_{k=1}^{2^M-1}$ solves (6.53) subject to (6.54).

Note that the maximization problem (6.61) subject to (6.62) for $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ fixed is the same as (5.44) in Chapter 5 and so is the solution. In the following subsection we rewrite this solution for completeness.

6.5.3 Independent Outage Transmission Policy

Before describing the optimal *Independent Outage Transmission Policy* for the M users in each fading state $\mathbf{h}$, we define the permutation $\hat{\pi}(\cdot)$ such that $0 < \eta_{\hat{\pi}(1)} < \eta_{\hat{\pi}(2)} < \cdots < \eta_{\hat{\pi}(N)}$, where $N \triangleq 2^M - 1$ and η_i, $1 \le i \le 2^M - 1$, is given in (6.51). For simplicity denote $\xi_i \triangleq \eta_{\hat{\pi}(i)}$ and $v_i \triangleq P_{\boldsymbol{\lambda}}^{min}(\mathbf{R},\mathbf{h},S_{\hat{\pi}(i)})$, where $P_{\boldsymbol{\lambda}}^{min}(\mathbf{R},\mathbf{h},S)$ is given in (6.60). In order to get the largest usage reward in a given state $\mathbf{h}$, first we use the following iterative procedure to remove those sets of users $S_{\hat{\pi}(k)}$ to which no power should be assigned:

Initialization: Do not assign any power to those sets $S_{\hat{\pi}(k)}$ if $\exists j$ that satisfies $k < j \le 2^M - 1$ and $\xi_k/v_k \le \xi_j/v_j$. Remove them from further consideration and set $w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_{\hat{\pi}(k)}) = 0$. Let $m = 1$.

Step 1: Denote the number of remaining sets as G_u and let the permutation $\rho(\cdot)$ be defined such that for the remaining G_u sets, $\xi_{\rho(1)} < \xi_{\rho(2)} < \cdots < \xi_{\rho(G_u)}$.

Step 2: Let $\rho_0(m) = \rho(1)$ and $z_{\rho_0(m)} = \frac{\xi_{\rho(1)}}{v_{\rho(1)}}$. If $G_u < 2$, all the sets that should be assigned no power have been removed and the procedure terminates; otherwise go to Step 3.

Step 3: For $2 \leq k \leq G_u$, decrease $\xi_{\rho(k)}$ and $v_{\rho(k)}$ by $\xi_{\rho(1)}$ and $v_{\rho(1)}$, respectively. Do not assign any power to those sets of users $S_{\hat{\pi}[\rho(k)]}$ for which $\exists j$ that satisfies $k < j \leq G_u$ and $\xi_{\rho(k)}/v_{\rho(k)} \leq \xi_{\rho(j)}/v_{\rho(j)}$, and remove them from further consideration (i.e., let $w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_{\hat{\pi}[\rho(k)]}) = 0$). Also remove set $S_{\hat{\pi}[\rho(1)]}$. Increase m by 1 and return to Step 1.

Assume that when the above procedure terminates, the value of the loop parameter m is m_0 ($1 \leq m_0 \leq 2^M - 1$). By denoting $z_{\rho_0(m_0+1)} \triangleq 0$, the optimal $\left\{w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_{\hat{\pi}[\rho_0(j)]})\right\}_{j=1}^{m_0}$ for the remaining m_0 sets $\left\{S_{\hat{\pi}[\rho_0(j)]}\right\}_{j=1}^{m_0}$ are given by:

(a) if $\frac{1}{s^*} > z_{\rho_0(1)}$, then $w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_{\hat{\pi}[\rho_0(j)]}) = 0$, $\forall 1 \leq j \leq m_0$;

(b) if $\exists j \in \{1, 2, \cdots, m_0\}$, $z_{\rho_0(j)} > \frac{1}{s^*} > z_{\rho_0(j+1)}$, then $w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_{\hat{\pi}[\rho_0(j)]}) = 1$, and $w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_{\hat{\pi}[\rho_0(i)]}) = 0$, $\forall i \neq j, 1 \leq i \leq m_0$. □

This optimal transmission policy is a multi-user generalization of the single-user threshold-decision rule in [76], where s^* can be viewed as the threshold power determined by the total average weighted power $\boldsymbol{\lambda} \cdot \bar{\mathbf{P}}^*$. For the given channel usage reward vector $\boldsymbol{\mu}$ and power price vector $\boldsymbol{\lambda}$, we see that once s^* is determined, $\left\{w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_k)\right\}_{k=1}^{2^M-1}$ will be determined for all $\mathbf{h} \in \mathcal{H}_{all}$, and the corresponding average outage probability $Pr_{i,\boldsymbol{\mu},^M\boldsymbol{\lambda}}(\mathbf{R})$ and average transmit power $\bar{P}_{i,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R})$ of each user i ($1 \leq i \leq M$) can be easily calculated. That is,

$$Pr_{i,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}) = 1 - E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_k)\mathbf{1}[i \in S_k]\right], \quad 1 \leq i \leq M, \tag{6.64}$$

$$\bar{P}_{i,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}) = E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_k)P_{i,\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_k)\right], \quad 1 \leq i \leq M, \tag{6.65}$$

where, $\forall S \in \{S_k\}_{k=1}^{2^M-1}$, $P_{i,\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S)$ is given by (6.59) if $i \in S$ and equals zero otherwise. For the given channel usage reward vector $\boldsymbol{\mu} \in \Re_+^M$, if the power price vector $\boldsymbol{\lambda}$ is the optimal $\boldsymbol{\lambda}^*$ such that $\bar{P}_{i,\boldsymbol{\mu},\boldsymbol{\lambda}^*}(\mathbf{R}) = \bar{P}_i^*$, $\forall 1 \leq i \leq M$, i.e., (6.63) holds, then the resulting average outage probability vector $\mathbf{Pr}_{\boldsymbol{\mu},\boldsymbol{\lambda}^*}(\mathbf{R})$ with the M components given in (6.64) will be on the boundary surface of the outage probability region $\mathcal{O}_I(\bar{\mathbf{P}}^*, \mathbf{R})$ and $\mathbf{Pr}_{\boldsymbol{\mu}}^{on}(\mathbf{R}) = \mathbf{1} - \mathbf{Pr}_{M\boldsymbol{\mu},\boldsymbol{\lambda}^*}(\mathbf{R})$ will be on the boundary surface of the usage probability region $\overline{\mathcal{O}}_I(\bar{\mathbf{P}}^*, \mathbf{R})$. Therefore, by varying the channel usage reward vector $\boldsymbol{\mu} \in \Re_+^M$, we can obtain all regular points on the boundary surface of the outage probability region $\mathcal{O}_I(\bar{\mathbf{P}}^*, \mathbf{R})$. However, it is not easy to obtain the optimal power price vector $\boldsymbol{\lambda}^*$ and the corresponding boundary outage probability vector directly for a given channel usage reward vector $\boldsymbol{\mu} \in \Re_+^M$. An iterative algorithm is proposed in Section 6.6 for obtaining all regular points on the boundary surface of region $\mathcal{O}_I(\bar{\mathbf{P}}^*, \mathbf{R})$ and the optimal power price vector for each regular point.

A regular point on the boundary surface of region $\mathcal{O}_I(\bar{\mathbf{P}}^*, \mathbf{R})$ corresponds to the case where all the M users are of the same priority and in each fading state $\mathbf{h}$, the optimal decoding order is determined by their power price vector $\boldsymbol{\lambda}$ and their sub-channel fading gains. Now consider the case where some users are given absolute priority over some other users in any fading condition, i.e., the absolute decoding rules can be represented by a set $\mathcal{L}$ of nested subsets $G_1, G_2, \cdots, G_I$ $(I \geq 2)$ of $G = \{1, 2, \cdots, M\}$ as described in Section 6.4.4, where $G_1 \subseteq G_2 \subseteq \cdots \subseteq G_I \equiv G$. Let $U_1 = G_1$ and $U_j = G_j \backslash G_{j-1}$, $\forall 2 \leq j \leq I$. Since users in U_i are given absolute priority over users in U_j for $i < j$ and users in U_1 are given the highest priority, for each given vector $\boldsymbol{\mu} \in \Re_+^M$, first we can derive the outage probability subvector $\mathbf{Pr}_{\boldsymbol{\mu}}^{(U_1)}(\mathbf{R})$ for users in set U_1 as if users in other sets were nonexistent. For $j \geq 2$, once the optimal power allocation for users in $\cup_{1 \leq i \leq j-1} U_i$ is determined, the outage probability subvector $\mathbf{Pr}_{\boldsymbol{\mu}}^{(U_j)}(\mathbf{R})$ for users in set U_j can be derived as if users in $\cup_{j+1 \leq i \leq I} U_i$ were nonexistent and the signals from users in $\cup_{1 \leq i \leq j-1} U_i$ were background noise. Note that subvector $\mathbf{Pr}_{\boldsymbol{\mu}}^{(U_j)}(\mathbf{R})$ $(1 \leq j \leq I)$ is a regular point on the boundary surface of the outage probability region $\mathcal{O}_I\left(\bar{\mathbf{P}}^{*(U_j)}, \mathbf{R}^{(U_j)}\right)$ for the users in set U_j. Since we have already shown how to obtain a regular point on the boundary surface of region $\mathcal{O}_I(\bar{\mathbf{P}}^*, \mathbf{R})$ for an M-user system, the regular point $\mathbf{Pr}_{\boldsymbol{\mu}}^{(U_j)}(\mathbf{R})$ of region $\mathcal{O}_I\left(\bar{\mathbf{P}}^{*(U_j)}, \mathbf{R}^{(U_j)}\right)$ for a $|U_j|$-user system can be similarly derived.

Therefore for each given pair $(\boldsymbol{\mu}, \mathcal{L})$ with $\mathcal{L} \neq \{G\}$, we can obtain the corresponding boundary outage probability vector of region $\mathcal{O}_I(\bar{\mathbf{P}}^*, \mathbf{R})$. Such boundary vectors are the nonregular points on the boundary surface of region $\mathcal{O}_I(\bar{\mathbf{P}}^*, \mathbf{R})$ and we can obtain all nonregular points explicitly by varying $(\boldsymbol{\mu}, \mathcal{L})$ with $\boldsymbol{\mu} \in \Re_+^M$ and $\mathcal{L} \neq \{G\}$. When $\mathcal{L} = \{G\}$, the corresponding boundary outage probability vectors are the regular points discussed before.

6.5.4 Average Power Region

In Section 6.5.2-6.5.3, given the power constraint vector $\bar{\mathbf{P}}^*$ and rate vector $\mathbf{R}$ of the M users, we derived the outage probability region $\mathcal{O}_I(\bar{\mathbf{P}}^*, \mathbf{R})$ and the corresponding optimal power allocation policy. Now for a given rate vector $\mathbf{R}$ and average outage probability vector $\mathbf{Pr}^*$, we consider the *average power region* $APV_{out}(\mathbf{Pr}^*, \mathbf{R})$, defined as the set of all possible average power vectors that can support rate vector $\mathbf{R}$ with the average outage probability of each user i no larger than Pr_i^*, $\forall 1 \le i \le M$. That is,

$$APV_{out}(\mathbf{Pr}^*, \mathbf{R}) \triangleq \{\bar{\mathbf{P}} : \mathbf{Pr}^* \in \mathcal{O}_I(\bar{\mathbf{P}}, \mathbf{R})\}. \tag{6.66}$$

By convexity of the set $\mathcal{Q}_I$ defined in (6.44), it is clear from (6.47) that an average power vector will be on the boundary surface of $APV_{out}(\mathbf{Pr}^*, \mathbf{R})$ if it is a solution to the minimization problem

$$\min_{\bar{\mathbf{P}}(\mathbf{R})} \boldsymbol{\lambda} \cdot \bar{\mathbf{P}}(\mathbf{R}) \quad \text{subject to:} \quad \mathbf{1} - \mathbf{Pr}^* \in \overline{\mathcal{O}}_I \left(\bar{\mathbf{P}}(\mathbf{R}), \mathbf{R}\right) \tag{6.67}$$

for some vector $\boldsymbol{\lambda} \in \Re_+^M$. Therefore, for each given power price vector $\boldsymbol{\lambda} \in \Re_+^M$, an average power vector $\bar{\mathbf{P}}_{\boldsymbol{\lambda}}(\mathbf{R})$ solves (6.67) if and only if there exists a Lagrangian multiplier vector $\boldsymbol{\mu}$ such that $(\mathbf{1} - \mathbf{Pr}^*, \bar{\mathbf{P}}_{\boldsymbol{\lambda}}(\mathbf{R}))$ is a solution to the problem

$$\min_{(\mathbf{Pr}^{on}(\mathbf{R}), \bar{\mathbf{P}}(\mathbf{R})) \in \mathcal{Q}} \left\{ \boldsymbol{\lambda} \cdot \bar{\mathbf{P}}(\mathbf{R}) - \boldsymbol{\mu} \cdot \mathbf{Pr}^{on}(\mathbf{R}) \right\}. \tag{6.68}$$

This minimization problem is similar to the maximization problem of (6.47). However, in (6.47), the average transmit power vector $\bar{\mathbf{P}}^*$ and the channel usage reward vector $\boldsymbol{\mu}$ are given, but the appropriate power price vector $\boldsymbol{\lambda}$ must be found such that the average power constraint $\bar{P}_i^*$ of each user i $(1 \le i \le M)$ is satisfied. Then the resulting average outage probability vector will lie on the boundary surface of $\mathcal{O}(\bar{\mathbf{P}}^*, \mathbf{R})$ for the given $\boldsymbol{\mu}$. In (6.68) the power price vector $\boldsymbol{\lambda}$ and the average outage probability vector $\mathbf{Pr}^*$ are given, but the usage reward vector $\boldsymbol{\mu}$ must be found such that the average outage probability constraint Pr_i^* of each user i $(1 \le i \le M)$ is satisfied. Then the resulting average power vector will lie on the boundary surface of $APV_{out}(\mathbf{Pr}^*, \mathbf{R})$ for the given $\boldsymbol{\lambda}$.

From (6.68) we see that for a given power price vector $\boldsymbol{\lambda} \in \Re_+^M$, power vector $\bar{\mathbf{P}}_{\boldsymbol{\lambda}}(\mathbf{R})$ will be the corresponding boundary vector of $APV_{out}(\mathbf{Pr}^*, \mathbf{R})$ if and only if there exist a Lagrangian multiplier vector $\boldsymbol{\mu}$ and a power allocation policy $\mathcal{P}^*$ such that for any $\mathbf{h} \in \mathcal{H}_{all}$, $\{\mathbf{P}^*(\mathbf{R}, \mathbf{h}, S_k)\}_{k=1}^{2^M-1}$ and $\{w^*(\mathbf{R}, \mathbf{h}, S_k)\}_{k=1}^{2^M-1}$ solve the maximization problem (6.53) subject to:

$$\begin{cases} \mathbf{R}^{(S)} \in C_g(\mathbf{h}, \mathbf{P}^{(S)}(\mathbf{R}, \mathbf{h}, S)), \quad \forall \mathbf{h} \in \mathcal{H}_S, \quad \forall S \in \{S_k\}_{k=1}^{2^M-1}, \\ E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w(\mathbf{R}, \mathbf{h}, S_k)\mathbf{1}[i \in S_k]\right] = 1 - Pr_i^*, \quad \forall 1 \le i \le M. \end{cases} \tag{6.69}$$

The boundary average power vector $\bar{\mathbf{P}}_{\boldsymbol{\lambda}}(\mathbf{R})$ is then given by:

$$\bar{\mathbf{P}}_{\boldsymbol{\lambda}}(\mathbf{R}) = E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w^*(\mathbf{R}, \mathbf{h}, S_k)\mathbf{P}^*(\mathbf{R}, \mathbf{h}, S_k)\right].$$

It is easily shown, as in Section 6.5.2, that for the given power price vector $\boldsymbol{\lambda}$, $\forall S \in \{S_k\}_{k=1}^{2^M-1}$, the optimal transmit power vector $\mathbf{P}^{*(S)}(\mathbf{R}, \mathbf{h}, S)$ of the $|S|$ users in set S is $\mathbf{P}_{\boldsymbol{\lambda}}^{(S)}(\mathbf{R}, \mathbf{h}, S)$, the components of which are given in (6.59), and $\forall i \notin S$, $P_i^*(\mathbf{R}, \mathbf{h}, S) = 0$. Therefore, for the given power price vector $\boldsymbol{\lambda}$, finding the appropriate usage reward vector $\boldsymbol{\mu}$ and solving the corresponding maximization problem (6.53) subject to (6.69) is now equivalent to determining the appropriate usage reward vector $\boldsymbol{\mu}$ and the Lagrangian multiplier $\frac{1}{s^*}$ in (6.61) and finding the solution $\left\{w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_k)\right\}_{k=1}^{2^M-1}$ to (6.61) subject to:

$$E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w(\mathbf{R}, \mathbf{h}, S_k)\mathbf{1}[i \in S_k]\right] = 1 - Pr_i^*. \ \forall 1 \le i \le M. \tag{6.70}$$

Note that, as described in Section 6.5.3, given the channel usage reward vector $\boldsymbol{\mu} \in \Re_+^M$ fixed and a power price vector $\boldsymbol{\lambda} \in \Re_+^M$ fixed, $\forall \mathbf{h} \in \mathcal{H}_{all}$, the solution $\left\{w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_k)\right\}_{k=1}^{2^M-1}$ to (6.61) is readily obtained using the *Independent Outage Transmission Policy* once the threshold power s^* in (6.61) is known. In Section 6.5.3, given the M average power constraints in (6.54), s^* is determined by the total average weighted power constraint $\boldsymbol{\lambda} \cdot \bar{\mathbf{P}}^*$ in (6.62) for each $\boldsymbol{\lambda} \in \Re_+^M$, and changing the power price vector $\boldsymbol{\lambda}$ will then result in different average transmit power vectors. Now for each power price vector $\boldsymbol{\lambda}$ fixed, given the M average outage probability constraints in (6.70), the appropriate s^* and channel usage reward vector $\boldsymbol{\mu}$ are to be chosen such that all constraints in (6.70) are satisfied. However, it is obvious that in (6.61), adjusting both $\boldsymbol{\mu}$ and $\frac{1}{s^*}$ such that (6.70) holds is equivalent to fixing s^* (e.g., setting $s^* = 1$) and changing $\boldsymbol{\mu}$ alone. Since by setting $s^* = 1$, given $\boldsymbol{\lambda}$ fixed, the solution $\left\{w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_k)\right\}_{k=1}^{2^M-1}$ ($\forall \mathbf{h} \in \mathcal{H}_{all}$) to (6.61) for each different usage reward vector $\boldsymbol{\mu}$ is then obtained using the *Independent Outage Transmission Policy*, we will obtain different average outage probability vectors $\mathbf{Pr}_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R})$ by varying $\boldsymbol{\mu}$. In the following we propose a simple iterative algorithm to find the optimal channel usage reward vector $\boldsymbol{\mu}^*$ that satisfies $\mathbf{Pr}_{\boldsymbol{\mu}^*,\boldsymbol{\lambda}}(\mathbf{R}) = \mathbf{Pr}^*$. This algorithm is similar to those in [2] and [74][6].

Algorithm 6.1 For a given power price vector $\boldsymbol{\lambda}$, let s^* in the *Independent Outage Transmission Policy* be fixed, e.g., $s^* = 1$, and denote the value of $\boldsymbol{\mu}$ in the nth iteration as $\boldsymbol{\mu}(n)$. Let $\boldsymbol{\mu}(0) = 0$. Given $\boldsymbol{\mu}(n)$, $\boldsymbol{\mu}(n+1)$ is obtained as follows: $\forall 1 \le i \le M$, adjust the usage reward $\mu_i(n+1)$ for User i while the usage rewards for all other users remain fixed at their values in the nth iteration such that the resulting average outage probability for User i is Pr_i^*. Specifically, given $\boldsymbol{\mu}(n)$, $\forall 1 \le i \le M$, let $\boldsymbol{\mu}^i$ denote the usage reward vector of the M users with the kth $(1 \le k \le M)$ component being $\mu_k(n)$ if $k \ne i$ and $\mu_i(n+1)$ if $k = i$, then $\mu_i(n+1)$ is chosen such that

$$Pr_{i,\boldsymbol{\mu}^i,\boldsymbol{\lambda}}(\mathbf{R}) = Pr_i^*,$$

where

$$Pr_{i,\boldsymbol{\mu}^i,\boldsymbol{\lambda}}(\mathbf{R}) = 1 - E_{\mathbf{h}}\left[\sum_{k}^{2^M-1} w_{\boldsymbol{\mu}^i,\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_k)\mathbf{1}[i \in S_k]\right],$$

with $\left\{w_{\boldsymbol{\mu}^i,\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_k)\right\}_{k=1}^{2^M-1}$ obtained from the *Independent Outage Transmission Policy* by setting $s^* = 1$ in it. $\square$

In order to prove that the iterative procedure converges, i.e., the sequence $\{\boldsymbol{\mu}(n)\}$ converges to $\boldsymbol{\mu}^*$ that satisfies $\mathbf{Pr}_{\boldsymbol{\mu}^*,\boldsymbol{\lambda}}(\mathbf{R}) = \mathbf{Pr}^*$, we need the following monotonicity lemma:

Lemma 6.3 *For a given power price vector $\boldsymbol{\lambda} \in Re_+^M$, let $s^* = 1$ be fixed in the Independent Outage Transmission Policy. Then $\forall 1 \le i \le M$, if the usage reward μ_i for User i increases while the usage rewards for all other users remain fixed, the resulting average usage prob-*

[6] It is easily seen that this algorithm can also be applied to the fading broadcast channel to obtain the minimum required total average power of the M users for supporting rate vector $\mathbf{R}$ with a given outage probability vector.

ability $Pr^{on}_{i,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}) = 1 - Pr_{i,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R})$ *of User* i *increases or remains unchanged, and the resulting average usage probabilities for all other users decrease or remain unchanged.*

Proof: See Appendix D.3. □

In each iteration of *Algorithm 6.1*, if the usage reward for all other users remain fixed while the usage reward μ_i for User i increases from 0 to ∞, it is obvious that the average usage probability $Pr^{on}_{i,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R})$ will increase monotonically from 0 to 1. Therefore, in each iteration there exists a μ_i such that $Pr_{i,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}) = Pr^*_i$. Since each iteration in *Algorithm 6.1* can be represented as a map

$$\begin{aligned} T : \Re^M_+ &\rightarrow \Re^M_+ \\ \boldsymbol{\mu}(n) &\mapsto \boldsymbol{\mu}(n+1), \end{aligned}$$

any fixed point of T will be the solution $\boldsymbol{\mu}^*$ satisfying $\mathbf{Pr}_{\boldsymbol{\mu}^*,\boldsymbol{\lambda}}(\mathbf{R}) = \mathbf{Pr}^*$. Applying *Lemma 6.3* we see that T is order preserving. That is,

$$\boldsymbol{\mu}^{(a)} \leq \boldsymbol{\mu}^{(b)} \Rightarrow T(\boldsymbol{\mu}^{(a)}) \leq T(\boldsymbol{\mu}^{(b)}),$$

where the notation $\mathbf{x} \leq \mathbf{y}$ means that $x_i \leq y_i$, $\forall 1 \leq i \leq M$. Therefore, starting with $\boldsymbol{\mu}(0) = \mathbf{0}$, $\boldsymbol{\mu}(1) = T(\boldsymbol{\mu}(0)) \geq \boldsymbol{\mu}(0)$, and by induction, $\boldsymbol{\mu}(n+1) = T(\boldsymbol{\mu}(n)) \geq \boldsymbol{\mu}(n)$. If $\boldsymbol{\mu}^*$ is a fixed point of T, then since $\boldsymbol{\mu}(0) \leq \boldsymbol{\mu}^*$. $\boldsymbol{\mu}(n) = T^n(\boldsymbol{\mu}(0)) \leq T^n(\boldsymbol{\mu}^*) = \boldsymbol{\mu}^*$. Thus, $\{\boldsymbol{\mu}(n)\}$ is a monotonically increasing sequence bounded from above and must converge to a limit. The limit has to be a fixed point of T and hence a solution that satisfies $\mathbf{Pr}_{\boldsymbol{\mu}^*,\boldsymbol{\lambda}}(\mathbf{R}) = \mathbf{Pr}^*$.

For the given power price vector $\boldsymbol{\lambda} \in \Re^M_+$, once we find the optimal channel usage reward vector $\boldsymbol{\mu}^*$ satisfying $\mathbf{Pr}_{\boldsymbol{\mu}^*,\boldsymbol{\lambda}}(\mathbf{R}) = \mathbf{Pr}^*$, the corresponding average transmit power vector $\bar{\mathbf{P}}_{\boldsymbol{\mu}^*,\boldsymbol{\lambda}}(\mathbf{R})$ will be the boundary vector $\bar{\mathbf{P}}_{\boldsymbol{\lambda}}(\mathbf{R})$ of region $APV_{out}(\mathbf{Pr}^*, \mathbf{R})$. That is,

$$\begin{aligned} \bar{\mathbf{P}}_{\boldsymbol{\lambda}}(\mathbf{R}) &\triangleq \bar{\mathbf{P}}_{\boldsymbol{\mu}^*,\boldsymbol{\lambda}}(\mathbf{R}) \\ &= E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w_{\boldsymbol{\mu}^*,\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_k)\mathbf{P}_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_k)\right]. \end{aligned} \tag{6.71}$$

However, as in the case with a given common outage probability or with zero-outage, there are other points on the boundary surface of $APV_{out}(\mathbf{Pr}^*, \mathbf{R})$ that cannot be parameterized by any $\boldsymbol{\lambda} \in \Re^M_+$, since they correspond to decoding rules that cannot be represented simply by any $\boldsymbol{\lambda} \in \Re_+^M$. Instead, those decoding rules can be represented by a set $\mathcal{L}$ of nested subsets of $G = \{1, 2, \cdots, M\}$ as described in Section 6.4.4. Therefore, we use the power allocation parameter pair $(\boldsymbol{\lambda}, \mathcal{L})$ to describe all possible decoding orders. When $\mathcal{L} = \{G\}$, the boundary average power vector $\bar{\mathbf{P}}_{(\boldsymbol{\lambda},\{G\})}(\mathbf{R}) \triangleq \bar{\mathbf{P}}_{\boldsymbol{\lambda}}(\mathbf{R})$ of region $APV_{out}(\mathbf{Pr}^*, \mathbf{R})$ can be obtained from (6.71), and it is called the *regular point* of the boundary surface of $APV_{out}(\mathbf{Pr}^*, \mathbf{R})$. When $\mathcal{L} \neq \{G\}$, the corresponding boundary average power vector $\bar{\mathbf{P}}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R})$ is called the *nonregular point* of the boundary surface of $APV_{out}(\mathbf{Pr}^*, \mathbf{R})$. We now show how to calculate the nonregular point $\bar{\mathbf{P}}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R})$ for a given power allocation parameter pair $(\boldsymbol{\lambda}, \mathcal{L})$ with an average outage probability vector $\mathbf{Pr}^*$.

For a given set $\mathcal{L}$ of nested subsets of G, suppose $G_1 \subsetneq G_2 \subsetneq \cdots \subsetneq G_I \equiv G$, $I \geq 2$. Let $U_1 = G_1$ and $U_j = G_j \backslash G_{j-1}$, $\forall 2 \leq j \leq I$. Then users in U_i are given absolute priority over users in U_j for $i < j$ and users in U_1 are given the highest priority. That is, for $j = 1$, $(\lambda_k)_{k \in U_1}$ determines the decoding order of the users in U_1 and they are decoded after the signals from the users in $U_2 \cup \cdots \cup U_I$ are decoded and subtracted out in every fading state $\mathbf{h}$. Since there will be no interference from any user in set $U_1^c = U_2 \cup \cdots \cup U_I$ to users in U_1 in any fading state $\mathbf{h}$, by applying the *Independent Outage Transmission Policy* to the $|U_1|$ users instead of to the M users, we can determine the users in U_1 who will transmit information in each state $\mathbf{h}$ and calculate the required average power vector $\bar{\mathbf{P}}^{(U_1)}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R})$ for supporting rate vector $\mathbf{R}^{(U_1)}$ with given average outage probability vector $\mathbf{Pr}^{*(U_1)}$ as if users in U_1^c did not exist. Similarly, for $j \geq 2$, $(\lambda_k)_{k \in U_j}$ determines the decoding order of the users in U_j and they are decoded after the signals from the users in $U_{j+1} \cup \cdots \cup U_I$ are decoded and subtracted out in every fading state $\mathbf{h}$. Therefore, there will be no interference from any user in $U_{j+1} \cup \cdots \cup U_I$ to users in U_j. Since the users in $U_1 \cup \cdots \cup U_{j-1}$ who will transmit information in each state $\mathbf{h}$ and therefore cause interference to each user in U_j are already determined, the total received power of the users in $U_1 \cup \cdots \cup U_{j-1}$ can be viewed as background noise to users in U_j. Thus, for $j \geq 2$, by applying the *Independent Outage Transmission Policy* to the $|U_j|$ users instead of to the M users, we can determine the users in U_j who will transmit information in each state $\mathbf{h}$ and calculate the required average power vector $\bar{\mathbf{P}}^{(U_j)}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R})$ for supporting rate vector $\mathbf{R}^{(U_j)}$ with given average outage probability vector $\mathbf{Pr}^{*(U_j)}$.

We now show how to calculate the average power vector $\bar{\mathbf{P}}^{(U_j)}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R})$, $1 \leq j \leq I$. For the $|U_j|$ users in set U_j, given the power price vector $\boldsymbol{\lambda}^{(U_j)}$ and the allowed average outage probability vector $\mathbf{Pr}^{*(U_j)}$, the length-$|U_j|$ required average power vector $\bar{\mathbf{P}}^{(U_j)}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R})$ corresponds to a regular point on the boundary surface of region $APV_{out}\left(\mathbf{Pr}^{*(U_j)}, \mathbf{R}^{(U_j)}\right)$, the set of all possible average power vectors of the $|U_j|$ users for which rate vector $\mathbf{R}^{(U_j)}$ can be supported with the given average outage probability vector $\mathbf{Pr}^{*(U_j)}$. Since we have already shown how to calculate a regular point of the average power region $APV_{out}(\mathbf{Pr}^*, \mathbf{R})$ for an M-user system, the regular point $\bar{\mathbf{P}}^{(U_j)}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R})$ of the average power region $APV_{out}\left(\mathbf{Pr}^{*(U_j)}, \mathbf{R}^{(U_j)}\right)$ for an $|U_j|$-user system can be similarly obtained. Specifically, for the given power price vector $\boldsymbol{\lambda}^{(U_j)} \in \Re_+^{|U_j|}$, first we use *Algorithm 6.1* for the $|U_j|$ users instead of for the M users to find the optimal channel usage reward vector $\boldsymbol{\mu}^{*(U_j)}$ for which the average outage probability of each user i equals Pr_i^*, $\forall i \in U_j$. Then the required average power vector $\bar{\mathbf{P}}^{(U_j)}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R})$ $(1 \leq j \leq I)$ is given by:

$$\bar{\mathbf{P}}^{(U_j)}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}) = E_{\mathbf{h}}\left[\sum_{k=1}^{2^{|U_j|}-1} w_{\boldsymbol{\mu}^{*(U_j)},\boldsymbol{\lambda}^{(U_j)}}(\mathbf{R}, \mathbf{h}, S_k) \cdot \mathbf{P}^{(U_j)}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}, \mathbf{h}, S_k)\right], \qquad (6.72)$$

where $\left\{w_{\boldsymbol{\mu}^{*(U_j)},\boldsymbol{\lambda}^{(U_j)}}(\mathbf{R}, \mathbf{h}, S_k)\right\}_{k=1}^{2^{|U_j|}-1}$ $(\forall \mathbf{h} \in \mathcal{H}_{all})$ is obtained by applying the *Independent Outage Transmission Policy* to the $|U_j|$ users with $s^* = 1$ in it, and $\forall S \in \{S_k\}_{k=1}^{2^{|U_j|}-1}$, the $|U_j|$ components of vector $\mathbf{P}^{(U_j)}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}, \mathbf{h}, S)$ are: $P_{i,(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}, \mathbf{h}, S) = 0$ if $i \notin S$ and, for each user in set S,

$$P_{\hat{\rho}_s(i),(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}, \mathbf{h}, S) = \begin{cases} \frac{\sigma^2}{h_{\hat{\rho}_s(1)}} \cdot \exp\left(2\sum_{k\in S_{j-1}} R_k\right) \cdot \left[\exp\left(2R_{\hat{\rho}_s(1)}\right) - 1\right], & \text{if } i = 1, \\ \frac{\sigma^2}{h_{\hat{\rho}_s(i)}} \cdot \exp\left(2\sum_{k\in S_{j-1}} R_k\right) \cdot \left[\exp\left(2\sum_{k=1}^{i} R_{\hat{\rho}_s(k)}\right) \right. & \\ \qquad \left. - \exp\left(2\sum_{k=1}^{i-1} R_{\hat{\rho}_s(k)}\right)\right], & \forall 2 \leq i \leq |S|, \end{cases} \qquad (6.73)$$

where S_{j-1} denotes the empty set $(j = 1)$ or the subset of users in $U_1 \cup \cdots \cup U_{j-1}$ who are transmitting information in state $\mathbf{h}$ $(j \geq 2)$, and the permutation $\hat{\rho}_s(\cdot)$ of the $|S|$ users satisfies:

$$\frac{\lambda_{\hat{\rho}_s(1)}}{h_{\hat{\rho}_s(1)}} \geq \frac{\lambda_{\hat{\rho}_s(2)}}{h_{\hat{\rho}_s(2)}} \geq \cdots \geq \frac{\lambda_{\hat{\rho}_s(|S|)}}{h_{\hat{\rho}_s(|S|)}}.$$

Therefore, all the nonregular points $\bar{\mathbf{P}}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R})$ on the boundary surface of the average power region $APV_{out}(\mathbf{Pr}^*, \mathbf{R})$ can be obtained by varying the power allocation parameter pair $(\boldsymbol{\lambda}, \mathcal{L})$ with $\mathcal{L} \neq \{G\}$, and all the regular points $\bar{\mathbf{P}}_{(\boldsymbol{\lambda},\{G\})}(\mathbf{R})$ on the boundary surface of region $APV_{out}(\mathbf{Pr}^*, \mathbf{R})$ can be obtained by varying the power price vector $\boldsymbol{\lambda} \in \Re_+^M$ of the M users.

6.6 Iterative Algorithms for Power Allocation

In Section 6.4.1-6.4.4, by assuming that all users turn on and off transmission simultaneously, we derived the power allocation strategy that minimizes the common outage probability. In Section 6.5.1-6.5.4, by assuming that each user declares an outage independently, we derived the power allocation strategy that achieves the boundary of the outage probability region. The outage capacity regions $C_{out}(\bar{\mathbf{P}}, Pr)$ and $C_{out}(\bar{\mathbf{P}}, \mathbf{Pr})$ are then characterized implicitly. These strategies assumed the existence of an optimal power price vector $\boldsymbol{\lambda}^*$ or, more generally, an optimal power allocation parameter pair $(\boldsymbol{\lambda}^*, \mathcal{L}^*)$ to determine user decoding priorities such that the average power constraints for all users are satisfied and the minimum common outage probability or the boundary outage probability vector is achieved. It is shown in Section 6.4 and Section 6.5 that once $(\boldsymbol{\lambda}^*, \mathcal{L}^*)$ is given, the optimal successive decoding order of the M users in each fading state is determined and so are the optimal power allocation and the average power required from each user. Therefore, for a given power constraint vector $\bar{\mathbf{P}}^*$ and rate vector $\mathbf{R}$, an important question is how to obtain the optimal power allocation parameter pair $(\boldsymbol{\lambda}^*, \mathcal{L}^*)$. It is this question that we address in this section.

In the zero-outage capacity case, for a given power allocation parameter pair $(\boldsymbol{\lambda}, \mathcal{L})$, let $\bar{\mathbf{P}}(\mathbf{R}, \boldsymbol{\lambda}, \mathcal{L}) \triangleq [\bar{P}_1(\mathbf{R}, \boldsymbol{\lambda}, \mathcal{L})\ \bar{P}_2(\mathbf{R}, \boldsymbol{\lambda}, \mathcal{L}), \cdots, \bar{P}_M(\mathbf{R}, \boldsymbol{\lambda}, \mathcal{L})]$ denote the corresponding boundary average power vector of region $APV_{zero}(\mathbf{R})$, defined as the set of all possible average power vectors that can support rate vector $\mathbf{R}$ in all fading conditions without any outage, i.e.,

$$\bar{\mathbf{P}}(\mathbf{R}, \boldsymbol{\lambda}, \mathcal{L}) = \begin{cases} E_{\mathbf{h}}\left[\mathbf{P}_{\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})\right], & \text{if } \mathcal{L} = \{G\}, \\ E_{\mathbf{h}}\left[\mathbf{P}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}, \mathbf{h})\right], & \text{if } \mathcal{L} \neq \{G\}, \end{cases} \tag{6.74}$$

where components of $\mathbf{P}_{\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})$ are given in (6.17) and components of $\mathbf{P}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R}, \mathbf{h})$ are given in (6.34) and (6.35). Given the average power constraint vector $\bar{\mathbf{P}}^*$, an iterative algorithm (we will refer to it as the *Hanly-Tse (HT) Algorithm*) is proposed in [3] for obtaining the optimal power allocation parameter pair $(\boldsymbol{\lambda}^*, \mathcal{L}^*)$ that achieves the infimum in the following optimization problem:

$$\inf_{(\boldsymbol{\lambda},\mathcal{L})} \max_{1\le i\le M} \frac{\bar{P}_i(\mathbf{R},\boldsymbol{\lambda},\mathcal{L})}{\bar{P}_i^*}. \tag{6.75}$$

Therefore, rate vector $\mathbf{R}$ lies in the zero-outage capacity region if and only if the infimum in (6.75) is no greater than 1. Thus, the solution to (6.75) implicitly defines the zero-outage capacity region. Moreover, if the corresponding average power vector $\bar{\mathbf{P}}(\mathbf{R},\boldsymbol{\lambda}^*,\mathcal{L}^*)$ is a regular point on the boundary surface of $APV_{zero}(\mathbf{R})$, then the optimal $\mathcal{L}^*$ must be $\{G\}$ and the *HT Algorithm* will converge to the optimal $\boldsymbol{\lambda}^* \in \Re_+^M$. If $\bar{\mathbf{P}}(\mathbf{R},\boldsymbol{\lambda}^*,\mathcal{L}^*)$ is a nonregular point on the boundary surface of $APV_{zero}(\mathbf{R})$, the *HT Algorithm* will provide the optimal power allocation parameter pair $(\boldsymbol{\lambda}^*,\mathcal{L}^*)$. The basic idea of the *HT Algorithm* is that at each iteration the required normalized average powers of the M users (normalized by their given average power constraints) are balanced as much as possible by increasing the power prices of the users with larger normalized average powers in the last iteration. A brief description of the *HT Algorithm* is given in Appendix D.4 for convenience.

Now if the infimum in (6.75) is larger than 1, the given rate vector $\mathbf{R}$ can only be maintained with certain outage probability for each of the M users. In this case, under the assumption that the transmission from all users is turned on or off simultaneously, we wish to obtain the minimum common outage probability $Pr^* \triangleq Pr_{min}(\bar{\mathbf{P}}^*,\mathbf{R})$. Under the alternative assumption that the transmission from each user is turned on or off independently, we wish to obtain the outage probability region $\mathcal{O}_I(\bar{\mathbf{P}}^*,\mathbf{R})$. In this section, for notational convenience, given a power allocation parameter pair $(\boldsymbol{\lambda},\mathcal{L})$, let $\bar{\mathbf{P}}(Pr,\mathbf{R},\boldsymbol{\lambda},\mathcal{L}) \triangleq [\bar{P}_1(Pr,\mathbf{R},\boldsymbol{\lambda},\mathcal{L}), \bar{P}_2(Pr,\mathbf{R},\boldsymbol{\lambda},\mathcal{L}), \cdots, \bar{P}_M(Pr,\mathbf{R},\boldsymbol{\lambda},\mathcal{L})]$ denote the boundary average power vector of region $APV_{out}(Pr,\mathbf{R})$ for a given common outage probability $Pr > 0$ under the first assumption, and let $\bar{\mathbf{P}}(\mathbf{Pr},\mathbf{R},\boldsymbol{\lambda},\mathcal{L}) \triangleq [\bar{P}_1(\mathbf{Pr},\mathbf{R},\boldsymbol{\lambda},\mathcal{L}), \bar{P}_2(\mathbf{Pr},\mathbf{R},\boldsymbol{\lambda},\mathcal{L}), \cdots, \bar{P}_M(\mathbf{Pr}, \mathbf{R}, \boldsymbol{\lambda}, \mathcal{L})]$ denote the boundary average power vector of region $APV_{out}(\mathbf{Pr},\mathbf{R})$ for a given outage probability vector $\mathbf{Pr} \geq \mathbf{0}$[7] under the second assumption. Note that we have already shown how to obtain these boundary vectors of region $APV_{out}(Pr,\mathbf{R})$ and $APV_{out}(\mathbf{Pr},\mathbf{R})$ in Sections 6.4.4 and 6.5.4, respectively (for simplicity, in Section 6.4.4 we used $\mathbf{P}_{\boldsymbol{\lambda}}(\mathbf{R})$ ($\mathcal{L} = \{G\}$) or $\mathbf{P}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R})$ ($\mathcal{L} \neq \{G\}$) to denote $\bar{\mathbf{P}}(Pr,\mathbf{R},\boldsymbol{\lambda},\mathcal{L})$ and in Section 6.5.4 we used $\mathbf{P}_{\boldsymbol{\lambda}}(\mathbf{R})$ ($\mathcal{L} = \{G\}$) or $\mathbf{P}_{(\boldsymbol{\lambda},\mathcal{L})}(\mathbf{R})$ ($\mathcal{L} \neq \{G\}$) to denote $\bar{\mathbf{P}}(\mathbf{Pr},\mathbf{R},\boldsymbol{\lambda},\mathcal{L})$).

[7]In this chapter, $\mathbf{a} \geq \mathbf{b} \Leftrightarrow a_i \geq b_i, \forall i$.

Under the first assumption, for each given common outage probability $Pr > 0$, if we can find the power allocation parameter pair $(\boldsymbol{\lambda}, \mathcal{L})$ that solves

$$\inf_{(\boldsymbol{\lambda},\mathcal{L})} \max_{1\le i\le M} \frac{\bar{P}_i(Pr,\mathbf{R},\boldsymbol{\lambda},\mathcal{L})}{\bar{P}_i^*}, \tag{6.76}$$

then by denoting the infimum in (6.76) as $Inf(Pr)$, the following lemma holds.

Lemma 6.4 *$Inf(Pr)$ is a strictly decreasing function of the common outage probability Pr.*

Proof: See Appendix D.5. □

Therefore, it is clear that

$$\begin{cases} Inf(Pr) > 1, & \text{if } Pr < Pr^*, \\ Inf(Pr) < 1, & \text{if } Pr > Pr^*, \\ Inf(Pr) = 1, & \text{if } Pr = Pr^*. \end{cases}$$

That is, if the infimum $Inf(Pr)$ in (6.76) equals 1, then the corresponding common outage probability Pr of the M users is the target minimum common outage probability Pr^* and the pair $(\boldsymbol{\lambda}, \mathcal{L})$ that achieves this infimum is the target optimal power allocation parameter pair $(\boldsymbol{\lambda}^*, \mathcal{L}^*)$.

Similarly, under the second assumption, for each given outage probability vector $\mathbf{Pr} > \mathbf{0}$, if we can find the power allocation parameter pair $(\boldsymbol{\lambda}, \mathcal{L})$ that solves

$$\inf_{(\boldsymbol{\lambda},\mathcal{L})} \max_{1\le i\le M} \frac{\bar{P}_i(\mathbf{Pr},\mathbf{R},\boldsymbol{\lambda},\mathcal{L})}{\bar{P}_i^*}, \tag{6.77}$$

then by denoting the infimum in (6.77) as $Inf(\mathbf{Pr})$, the following lemma holds.

Lemma 6.5 *For two given average outage probability vectors $\mathbf{Pr}^{(1)}$ and $\mathbf{Pr}^{(2)}$, if $\mathbf{Pr}^{(1)} \le \mathbf{Pr}^{(2)}$ and $\mathbf{Pr}^{(1)} \ne \mathbf{Pr}^{(2)}$, then $Inf(\mathbf{Pr}^{(1)}) > Inf(\mathbf{Pr}^{(2)})$.*

Proof: See Appendix D.6. □

Therefore, it is obvious that

$$\begin{cases} Inf(\mathbf{Pr}) > 1, & \text{if } \mathbf{Pr} \notin \mathcal{O}_I(\bar{\mathbf{P}}^*, \mathbf{R}), \\ Inf(\mathbf{Pr}) \le 1, & \text{if } \mathbf{Pr} \in \mathcal{O}_I(\bar{\mathbf{P}}^*, \mathbf{R}), \end{cases}$$

and $Inf(\mathbf{Pr}) = 1$ if $\mathbf{Pr}$ is on the boundary surface of the outage probability region $\mathcal{O}_I(\bar{\mathbf{P}}^*, \mathbf{R})$. This is because if $\mathbf{Pr} \notin \mathcal{O}_I(\bar{\mathbf{P}}^*, \mathbf{R})$ then there must exist a boundary average outage probability vector $\mathbf{Pr}^{(0)}$ of $\mathcal{O}_I(\bar{\mathbf{P}}^*, \mathbf{R})$ satisfying $\mathbf{Pr}^{(0)} \geq \mathbf{Pr}$, and $\mathbf{Pr}^{(0)} \neq \mathbf{Pr}$, which means $Inf(\mathbf{Pr}) > Inf(\mathbf{Pr}^{(0)}) = 1$ by *Lemma 6.5.* Moreover, if $\mathbf{Pr}$ is a regular point on the boundary surface, then the power allocation parameter pair $(\boldsymbol{\lambda}^*, \mathcal{L}^*)$ that achieves the infimum in (6.77) must satisfy $\mathcal{L}^* = \{G\}$ and

$$\frac{\bar{P}_i(\mathbf{Pr}, \mathbf{R}, \boldsymbol{\lambda}^*, \{G\})}{\bar{P}_i^*} = Inf(\mathbf{Pr}) = 1, \quad \forall 1 \leq i \leq M. \tag{6.78}$$

In the following we will propose an iterative algorithm (*Algorithm 6.2* below) that finds, for the given rate vector $\mathbf{R}$, the target minimum common outage probability Pr^* and the corresponding power allocation parameter pair $(\boldsymbol{\lambda}^*, \mathcal{L}^*)$ that achieves the infimum in (6.76). Another iterative algorithm (*Algorithm 6.3* below) is proposed to obtain all regular points on the boundary surface of the target outage probability region $\mathcal{O}_I(\bar{\mathbf{P}}^*, \mathbf{R})$ and the corresponding power allocation parameter pair $(\boldsymbol{\lambda}^*, \{G\})$ that satisfies (6.78) for each regular point $\mathbf{Pr}$. The nonregular points on the boundary surface of $\mathcal{O}_I(\bar{\mathbf{P}}^*, \mathbf{R})$ are obtained as described at the end of Section 6.5.3.

Before stating our algorithms we need to extend the *HT Algorithm* to the case of non-zero outage. That is, when only simultaneous outage declaration from the M users is allowed, given a common outage probability Pr, we use a similar algorithm (we will refer to it as the *HT^* Algorithm*) as for the zero-outage case ($Pr = 0$) to find the solution to (6.76). Note that rate vector $\mathbf{R}$ lies in the outage capacity region $\mathcal{C}_{out}(\bar{\mathbf{P}}^*, Pr)$ if and only if the infimum $Inf(Pr)$ in (6.76) satisfies $Inf(Pr) \leq 1$. Therefore, the solution to (6.76) implicitly defines the outage capacity region $\mathcal{C}_{out}(\bar{\mathbf{P}}^*, Pr)$. When independent outage declaration from each user is allowed, given an outage probability vector $\mathbf{Pr}$, we also use a similar algorithm (we will refer to it as the *HT^{**} Algorithm*) as for the zero-outage case ($\mathbf{Pr} = \mathbf{0}$) to solve (6.77). Consequently, rate vector $\mathbf{R}$ lies in the outage capacity region $\mathcal{C}_{out}(\bar{\mathbf{P}}^*, \mathbf{Pr})$ if and only if the infimum $Inf(\mathbf{Pr})$ in (6.77) satisfies $Inf(\mathbf{Pr}) \leq 1$. Therefore, the solution to (6.77) implicitly defines the outage capacity region $\mathcal{C}_{out}(\bar{\mathbf{P}}^*, \mathbf{Pr})$. Details of the *$HT^*$ Algorithm* and the *HT^{**} Algorithm* are given in Appendix D.7 and Appendix D.8, respectively, and these algorithms will be used, respectively, in *Algorithm 6.2* and *Algorithm 6.3* given below.

Algorithm 6.2 For a given rate vector $\mathbf{R}$ and average power constraint vector $\bar{\mathbf{P}}^*$, denote the common outage probability at the nth iteration as $Pr(n)$.

Initialization: Let $Pr(0) = 0$. Use the *HT Algorithm* for the zero-outage case to solve (6.75) and denote the solution $(\boldsymbol{\lambda}, \mathcal{L})$ as $(\boldsymbol{\lambda}(0), \mathcal{L}(0))$. If the infimum in (6.75) is no greater than 1, then the minimum common outage probability $Pr^* = 0$ and we terminate the algorithm. Otherwise, set $n = 1$ and go to Step n.

Step n:

(1) Given power allocation parameter pair $(\boldsymbol{\lambda}(n-1), \mathcal{L}(n-1))$, calculate the minimum common outage probability $Pr(n)$ for which the average power vector $\bar{\mathbf{P}}\,(Pr(n), \mathbf{R}, \boldsymbol{\lambda}(n-1), \mathcal{L}(n-1))$ of the M users satisfies

$$\max_{1\leq i\leq M} \frac{\bar{P}_i\,(Pr(n), \mathbf{R}, \boldsymbol{\lambda}(n-1), \mathcal{L}(n-1))}{\bar{P}_i^*} = 1. \tag{6.79}$$

Specifically, if $\mathcal{L}(n-1) = \{G\}$, $Pr(n)$ is obtained from (6.24) with $\boldsymbol{\lambda}$ in (6.24) replaced by $\boldsymbol{\lambda}(n-1)$. The threshold power s^* in (6.24) is now implicitly determined by (6.79), since all components of $\bar{\mathbf{P}}\,(Pr(n), \mathbf{R}, \boldsymbol{\lambda}(n-1), \{G\})$ in (6.79) are implicit functions of s^* as shown in (6.32) with $\boldsymbol{\lambda}$ replaced by $\boldsymbol{\lambda}(n-1)$.

In the case where $\mathcal{L}(n-1) \neq \{G\}$, $Pr(n)$ is given by

$$Pr(n) = 1 - Prob\{\mathbf{h} \in \mathcal{R}(s^*, \boldsymbol{\lambda}(n-1), \mathcal{L}(n-1))\}, \tag{6.80}$$

where the function $\mathcal{R}(s, \boldsymbol{\lambda}, \mathcal{L})$ is defined in (6.36) and s^* is implicitly determined by (6.79), since all components of $\bar{\mathbf{P}}(Pr(n), \mathbf{R}, \boldsymbol{\lambda}(n-1), \mathcal{L}(n-1))$ are implicit functions of s^* as shown in (6.38) with $(\boldsymbol{\lambda}, \mathcal{L})$ replaced by $(\boldsymbol{\lambda}(n-1), \mathcal{L}(n-1))$.

(2) Given the common outage probability $Pr(n)$, use the HT^* *Algorithm* to solve (6.76). Denote the infimum in (6.76) as $Inf(Pr(n))$ and the solution $(\boldsymbol{\lambda}, \mathcal{L})$ as $(\boldsymbol{\lambda}(n), \mathcal{L}(n))$. Go to Step $n+1$. □

The iterative procedure of this algorithm is shown in Figure 6.1. For the given average power constraint vector $\bar{\mathbf{P}}^*$ and rate vector $\mathbf{R}$, suppose that the infimum in (6.75) is larger than 1, i.e.,

$$\max_{1\leq i\leq M} \frac{\bar{P}_i(\mathbf{R}, \boldsymbol{\lambda}(0), \mathcal{L}(0))}{\bar{P}_i^*} > 1, \tag{6.81}$$

then the target minimum common outage probability $Pr^* > 0$, and there exists a power allocation parameter pair $(\boldsymbol{\lambda}^*, \mathcal{L}^*)$ that achieves the infimum in (6.76) with the common outage probability Pr equal to Pr^* and satisfies $Inf(Pr^*) = 1$. We now show that *Algorithm 6.2* converges to this target Pr^* and the corresponding power allocation parameter pair $(\boldsymbol{\lambda}^*, \mathcal{L}^*)$.

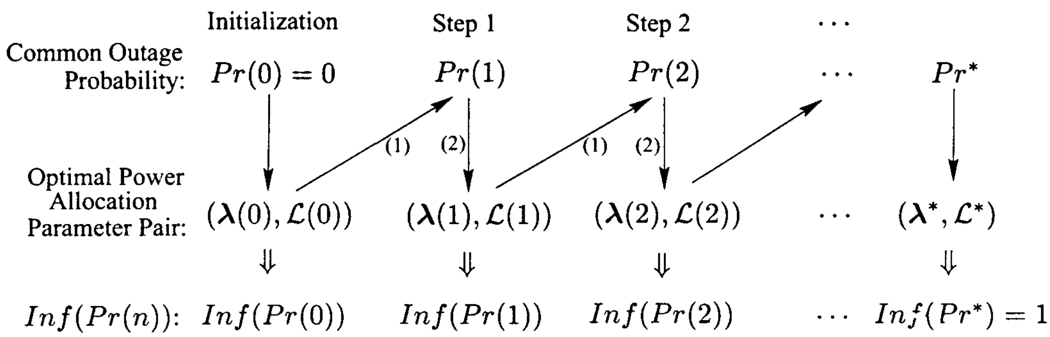


Figure 6.1: The iterative procedure of *Algorithm 6.2.*

Theorem 6.1 *In Algorithm 6.2,* $\{Pr(n)\}$ $(n \geq 1)$ *is a monotonically decreasing sequence and converges to the target minimum common outage probability* Pr^*. $\{Inf(Pr(n))\}$ $(n \geq 1)$ *is a monotonically increasing sequence and converges to* $Inf(Pr^*)$ *that satisfies* $Inf(Pr^*)$ $= 1$. *The sequence* $\{(\boldsymbol{\lambda}(n), \mathcal{L}(n))\}$ *converges to the power allocation parameter pair* $(\boldsymbol{\lambda}^*, \mathcal{L}^*)$ *that solves (6.76) with* $Pr = Pr^*$ *in it.*

Proof: $\forall n \geq 1$, we know from *Step n (1)* that (6.79) holds when the common outage probability is $Pr(n)$. Given common outage probability $Pr(n)$, it is clear from *Step n (2)* that $(\boldsymbol{\lambda}(n), \mathcal{L}(n))$ satisfies

$$\begin{aligned}
\max_{1 \leq i \leq M} \frac{\bar{P}_i\left(Pr(n), \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}(n)\right)}{\bar{P}_i^*} &= Inf(Pr(n)) \\
&= \inf_{(\boldsymbol{\lambda}, \mathcal{L})} \max_{1 \leq i \leq M} \frac{\bar{P}_i\left(Pr(n), \mathbf{R}, \boldsymbol{\lambda}, \mathcal{L}\right)}{\bar{P}_i^*} \\
&\leq \frac{\bar{P}_i\left(Pr(n), \mathbf{R}, \boldsymbol{\lambda}(n-1), \mathcal{L}(n-1)\right)}{\bar{P}_i^*} \\
&= 1, \qquad (6.82)
\end{aligned}$$

where the last equality is due to (6.79). Note that in *Step n (2)* , $\bar{\mathbf{P}}(Pr(n), \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}(n))$ is a function of the threshold power s^* as given in (6.32) ($\mathcal{L}(n) = \{G\}$) or (6.38) ($\mathcal{L}(n) \neq \{G\}$). Here the threshold power s^* is chosen such that the common outage probability equals

$Pr(n)$, and we denote it as $s^*(n)$. Now in *Step* $n+1$ *(1)*, a common outage probability $Pr(n+1)$ is to be chosen such that (6.79) holds with n replaced by $n+1$, i.e.,

$$\max_{1\le i\le M} \frac{\bar{P}_i\left(Pr(n+1), \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}(n)\right)}{\bar{P}_i^*} = 1. \tag{6.83}$$

Therefore, for the given rate vector $\mathbf{R}$ and the same power allocation parameter pair $(\boldsymbol{\lambda}(n), \mathcal{L}(n))$, we have to re-determine the threshold power s^* (we will refer to it as $s^*(n+1)$) such that (6.83) is satisfied. Comparison of (6.82) and (6.83) shows that the threshold power $s^*(n+1)$ must be no less than $s^*(n)$ in order to satisfy (6.83), i.e., $s^*(n+1) \ge s^*(n)$. since the *Common Outage Transmission Policy* implies that a larger threshold s^* means that the M users will transmit simultaneously in more fading states and thus require larger average transmit power. Thus, the corresponding common outage probability $Pr(n+1)$ must satisfy $Pr(n+1) \le Pr(n)$. Consequently, $\{Pr(n)\}$ $(n \ge 1)$ is a monotonically decreasing sequence and, by *Lemma 6.4*, $\{Inf(Pr(n)\}$ $(n \ge 1)$ is a monotonically increasing sequence.

Since each iteration in *Algorithm 6.2* can be represented as a map

$$\begin{aligned} T : \Re_+ &\rightarrow \Re_+ \\ Inf(Pr(n)) &\mapsto Inf(Pr(n+1)), \end{aligned}$$

any fixed point of T will be the solution $Inf(Pr^*)$ satisfying $Inf(Pr^*) = 1$. From (6.82) it is clear that $\forall n \ge 1$, $Inf(Pr(n)) \le 1$, i.e., it is bounded from above. Therefore, since $\{Inf(Pr(n)\}$ $(n \ge 1)$ is a monotonically increasing sequence bounded from above, it must converge to a limit, which is a fixed point of T and hence the solution $Inf(Pr^*)$ that satisfies $Inf(Pr^*) = 1$. This implies that sequence $\{Pr(n)\}$ $(n \ge 1)$ converges to the target minimum common outage probability Pr^*, and the sequence $\{(\boldsymbol{\lambda}(n), \mathcal{L}(n))\}$ converges to $(\boldsymbol{\lambda}^*, \mathcal{L}^*)$ that solves (6.76) with the common outage probability Pr equal to Pr^*. □

Algorithm 6.3 For a given rate vector $\mathbf{R}$ and average power constraint vector $\bar{\mathbf{P}}^*$, denote the average outage probability vector at the nth iteration as $\mathbf{Pr}(n) = [Pr_1(n), Pr_2(n), \cdots, Pr_M(n)]$.

Initialization: Let $\mathbf{Pr}(0) = \mathbf{0}$. Use the *HT Algorithm* for the zero-outage case to solve (6.75) and denote the optimal $(\boldsymbol{\lambda}, \mathcal{L})$ as $(\boldsymbol{\lambda}(0), \mathcal{L}(0))$. If the infimum in (6.75) is no greater than 1, then rate vector $\mathbf{R}$ can be supported with the given average power vector $\bar{\mathbf{P}}^*$ without any outage and we terminate the algorithm. Otherwise, assuming that none of the M users can support his given rate without any outage under his average power constraint[8], for a given channel usage reward vector $\boldsymbol{\mu}(0) \in \Re_+^M$, set $n = 1$ and go to Step n.

Step n:

(1) Given the power allocation parameter pair $(\boldsymbol{\lambda}(n-1), \mathcal{L}(n-1))$ and usage reward vector $\boldsymbol{\mu}(n-1)$, calculate the average outage probability vector $\mathbf{Pr}(n)$ for which the required average power vector $\bar{\mathbf{P}}\left(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}(n-1), \mathcal{L}(n-1)\right)$ of the M users satisfies

$$\max_{1 \le i \le M} \frac{\bar{P}_i(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}(n-1), \mathcal{L}(n-1))}{\bar{P}_i^*} = 1. \tag{6.84}$$

The lengthy details about how to obtain $\mathbf{Pr}(n)$ are given in Appendix D.10.

(2) Given the average outage probability vector $\mathbf{Pr}(n)$, use the *HT*** *Algorithm* to solve (6.77). Denote the infimum in (6.77) as $Inf(\mathbf{Pr}(n))$ and the solution $(\boldsymbol{\lambda}, \mathcal{L})$ as $(\boldsymbol{\lambda}(n), \mathcal{L}(n))$. For the given power allocation parameter pair $(\boldsymbol{\lambda}(n), \mathcal{L}(n))$, let $\boldsymbol{\mu}(n)$ denote the corresponding channel usage reward vector that results in the average outage probability vector $\mathbf{Pr}(n)$ and go to Step $n+1$. □

The iterative procedure of this algorithm is shown in Figure 6.2. For the given average power constraint vector $\bar{\mathbf{P}}^*$ and rate vector $\mathbf{R}$, suppose that the infimum in (6.75) is larger than 1 and none of the M users can support his given rate without any outage under his average power constraint. Then by varying the initial channel usage reward vector $\boldsymbol{\mu}(0) \in \Re_+^M$ with $\sum_{i=1}^M \mu_i(0) = 1$ in *Algorithm 6.3*, we will obtain all regular points[9] on the boundary surface of the outage probability region $\mathcal{O}_I(\bar{\mathbf{P}}^*, \mathbf{R})$ based on the following theorem.

Theorem 6.2 *In Algorithm 6.3, for each given channel usage reward vector* $\boldsymbol{\mu}(0) \in \Re_+^M$ *with* $\sum_{i=1}^M \mu_i(0) = 1$*, the sequence* $\{\mathbf{Pr}(n)\}$ $(n \ge 1)$ *converges to a boundary outage probability vector* $\mathbf{Pr}^*$ *of the target outage probability region* $\mathcal{O}_I(\bar{\mathbf{P}}^*, \mathbf{R})$*.* $\{Inf(\mathbf{Pr}(n))\}$ $(n \ge 1)$ *is a monotonically increasing sequence and converges to* $Inf(\mathbf{Pr}^*) = 1$*. The sequence*

[8] This assumption is not always true even if the infimum in (6.75) is larger than 1. If there are users who can support their given rates with zero-outage, we will apply our algorithm to the remaining users. Details about how to find those users who do not have to declare any outage are given in Appendix D.9.

[9] In *Algorithm 6.3*, it is easily seen that each different vector $\boldsymbol{\mu}(0) \in \Re_+^M$ with $\sum_{i=1}^M \mu_i(0) = 1$ will result in a distinct sequence $\{\mathbf{Pr}(n)\}$.

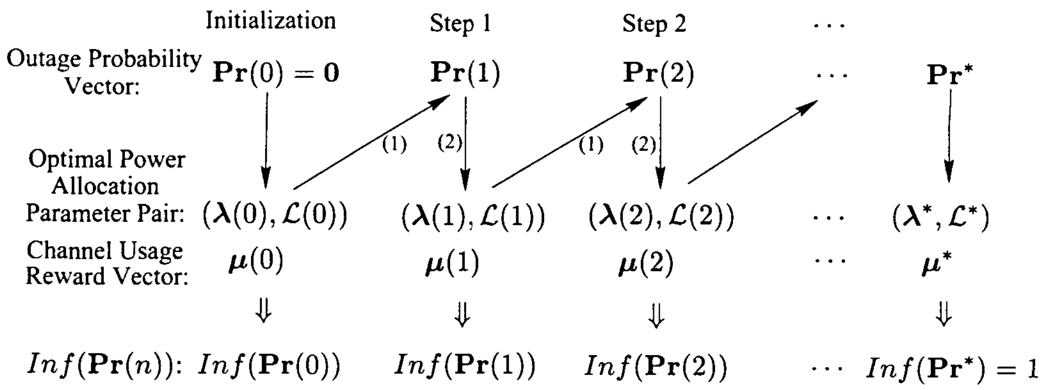


Figure 6.2: The iterative procedure of *Algorithm 6.3.*

$\{(\boldsymbol{\lambda}(n), \mathcal{L}(n))\}$ *converges to* $(\boldsymbol{\lambda}^*, \mathcal{L}^*)$ *that satisfies* $\mathcal{L}^* = \{G\}$ *and*

$$\frac{\bar{P}_i(\mathbf{Pr}^*, \mathbf{R}, \boldsymbol{\lambda}^*, \{G\})}{\bar{P}_i^*} = Inf(\mathbf{Pr}^*) = 1. \quad \forall 1 \leq i \leq M. \tag{6.85}$$

The proof of this theorem follows similar steps as that of *Theorem 6.1* and is given in Appendix D.11.

6.7 Auxiliary Constraints on Transmit Power

In Section 6.4 and Section 6.6 the power constraints we considered are average transmit power constraints of the M users. In practice, sometimes we have to consider the peak transmit power constraint of each user as well. That is, in addition to the average transmit power constraint vector $\bar{\mathbf{P}}^*$ of the M users, for each fading state $\mathbf{h}$, the transmit power vector of the M users must be no larger than $\hat{\mathbf{P}} = (\hat{P}_1, \hat{P}_2, \cdots, \hat{P}_M)$. Under these auxiliary constraints, given a rate vector $\mathbf{R}$, the problem of deriving the minimum average common outage probability of all users or deriving the average outage probability region of the M users can be similarly solved as shown in Section 6.4, except that now we have to solve the minimization problems (6.14) and (6.58) subject to the additional peak power constraint vector $\hat{\mathbf{P}}$, i.e., the additional constraint for problem (6.14) is:

$$\mathbf{P}(\mathbf{R}, \mathbf{h}) \leq \hat{\mathbf{P}}, \quad \forall \mathbf{h} \in \mathcal{H}_{tran},$$

and the additional constraint for problem (6.58) is:

$$\mathbf{P}^{(S)}(\mathbf{R}, \mathbf{h}, S) \leq \hat{\mathbf{P}}, \quad \forall \mathbf{h} \in \mathcal{H}_S,$$

where $\mathcal{H}_S$ is given in (6.55).

The solution to problems (6.14) and (6.58) under the additional peak power constraint vector $\hat{\mathbf{P}}$ is given in [3]. That is, for a given power price vector $\boldsymbol{\lambda}$, assuming that $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_M$, $\forall \mathbf{h} \in \mathcal{H}_{tran}$ or $\forall \mathbf{h} \in \mathcal{H}_S$, the transmit power $P_{i,\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})$ $(1 \leq i \leq M)$ or $P_{i,\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S)$ $(i \in S)$ of each user i is now obtained through a greedy algorithm. Specifically, $\forall \mathbf{h} \in \mathcal{H}_{tran}$, by denoting $x_i^{(k)}$ as the value of $P_{i,\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})$ in the kth step of the following algorithm, the solution $P_{i,\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})$ for each user i $(1 \leq i \leq M)$ is obtained after M steps:

- *Initialization*: Set $x_i^{(0)} = \hat{P}_i$ for all i $(1 \leq i \leq M)$. If $\mathbf{h} \cdot \mathbf{x}^{(0)} \notin \mathcal{G}(\mathbf{R}, \mathbf{h})$ then stop, where $\mathcal{G}(\mathbf{R}, \mathbf{h})$ is defined in (6.15). Otherwise set $k = 1$.

- *Step k*: Let $\pi^{(k)}(\cdot)$ be a permutation on $\{1, \cdots, k-1, k+1, \cdots, M\}$ (note that $\pi^{(k)}(k)$ does not exist) such that

$$\begin{aligned} \frac{x_{\pi^{(k)}(1)} h_{\pi^{(k)}(1)}}{R_{\pi^{(k)}(1)}} \leq \cdots &\leq \frac{x_{\pi^{(k)}(k-1)} h_{\pi^{(k)}(k-1)}}{R_{\pi^{(k)}(k-1)}} \\ &\leq \frac{x_{\pi^{(k)}(k+1)} h_{\pi^{(k)}(k+1)}}{R_{\pi^{(k)}(k+1)}} \leq \cdots \leq \frac{x_{\pi^{(k)}(M)} h_{\pi^{(k)}(M)}}{R_{\pi^{(k)}(M)}}. \end{aligned}$$

  Then set

$$x_i^{(k)} = \begin{cases} x_i^{(k-1)}, & \text{if } i \neq k, \\ \frac{1}{h_i} \cdot \max_{j \neq k} [f(S_j \cup \{k\}) - \sum_{l \in S_j} x_l], & \text{if } i = k, \end{cases}$$

  where $f(\cdot)$ is defined in (6.16) and $S_j \equiv \{\pi^{(k)}(1), \cdots, \pi^{(k)}(j)\}$ with $\pi^{(k)}(k)$ nonexistent. Go to Step $k+1$.

- Stop after M steps. □

Similarly, $\forall \mathbf{h} \in \mathcal{H}_S$, $P_{i,\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S)$ $(i \in S)$ can be obtained by applying the above algorithm to the $|S|$ users in set S instead of to the M users. Note that the solution $P_{i,\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})$ $(1 \leq i \leq M)$ in (6.17) and the solution $P_{i,\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S)$ $(i \in S)$ in (6.59) that can be achieved by successive decoding are obtained through an algorithm that is actually a special case of the above greedy algorithm [3]. However, when peak power constraints are imposed on the

M users, the solution $P_{i,\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h})$ or $P_{i,\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S)$ obtained through this greedy algorithm cannot be achieved by successive decoding in general.

6.8 Conclusions

In this paper, under the assumption that perfect CSI is available both at the transmitters and at the receiver, we have obtained the outage capacity regions of fading MACs implicitly by deriving the minimum common outage probability or the outage probability region for a given rate vector. Given the average power constraint of each user, we have derived the power allocation policy that minimizes the common outage probability for a given rate vector when transmission to all users is turned off simultaneously. When an outage can be declared for each user individually, we have derived a power allocation strategy to achieve the outage probability region boundary for the given rate vector. In both cases, similar to the non-outage scenario, successive decoding is optimal and in each fading state, the decoding order is determined by the power prices of the users and their fading gains. Iterative algorithms are proposed for obtaining the appropriate power allocation parameters, which determine the optimal decoding order and power allocation for each user. By applying these optimal power allocation strategies, the average power regions that can support a rate vector with a given common outage probability or an outage probability vector for the M users are also obtained. When there are additional peak power constraints, the optimal power allocation for the M users in each fading state can be obtained through a greedy algorithm, but in general cannot be achieved by successive decoding.

Chapter 7 Summary of Contributions

The contributions of this thesis are as follows. In Chapter 3 we have proposed two feedback decoders, a decision-feedback decoder and an output-feedback decoder, for coded signals transmitted over channels with slow, flat fading. While the decision-feedback decoder was originally proposed in [42] for finite-state Markov channels, its bit-error-rate (BER) performance for coded signals on fading channels was not investigated. The output-feedback decoder has a novel channel state estimator and has similar complexity and BER performance as the decision-feedback decoder. However, unlike the decision-feedback decoder, it does not suffer from error propagation and therefore is robust. Based on the extremely poor performance of the conventional decoder, we have also proposed a simple improvement to conventional decoding which uses a weighted metric. In order to analyze and compare the performance of the various decoding algorithms, we extend the sliding window decoding method, originally proposed in [54] for very noisy AWGN channels, to channels with fading, which provides a good approximation to the BER performance of decoders with complex metrics. Note that it is very difficult to analyze the performance of decoders with complex metrics in fading directly. Simulation results for our decoding techniques are also presented for several different fading models and modulation types.

In Chapter 4, we have obtained the ergodic capacity region and the optimal dynamic resource allocation strategy for an M-user fading broadcast channel with perfect CSI at both the transmitter and the receiver. These results are obtained for code-division (CD) with and without successive decoding, time-division (TD), and frequency-division (FD). We have shown that CD with successive decoding has the largest capacity region, while TD and FD are equivalent and they have the same capacity region as CD without successive decoding. For CD with successive decoding, it is demonstrated that the water-filling optimal power allocation procedure can be obtained through a greedy algorithm. For CD without successive decoding, we have found that the optimal power policy is to transmit the information of at most one user in each joint fading state. This policy is also optimal for TD, though other strategies which allow at most two users to time-share the channel may also be optimal. A simple sub-optimal power policy is also proposed for TD and CD without successive decoding that results in a rate region quite close to the ergodic capacity region.

In Chapter 5, we have obtained both the zero-outage capacity region and the outage capacity region of an M-user fading broadcast channel for CD with and without successive decoding, TD, and FD, assuming that perfect CSI is available at both the transmitter and all the receivers. The zero-outage capacity region is obtained implicitly by determining whether the total average required minimum power of the M users for supporting a given rate vector in all fading conditions satisfies the total average power constraint. For CD with and without successive decoding, we have proved that the zero-outage capacity region of a broadcast channel with Nakagami-m fading converges to the capacity region of an AWGN broadcast channel when m goes to infinity. The outage capacity region of a fading broadcast channel is implicitly obtained by deriving the minimum common outage probability or the outage probability region for a given rate vector. Given the required rate of each user, we have derived the optimal power policy that minimizes the common outage probability when transmission to all users is turned on or off simultaneously. When an outage can be declared for each user individually, we have also derived a general power allocation strategy to achieve boundaries of the outage probability regions under different spectrum-sharing techniques. The corresponding optimal power allocation scheme is a multi-user generalization of the single-user threshold-decision rule.

For an M-user fading multiple-access channel, under the assumption that perfect CSI is available both at the transmitters and at the receiver, we have obtained in Chapter 6 its outage capacity region implicitly using similar methods as for the fading broadcast channels. Given the average power constraint of each user, we have derived the minimum common outage probability and the corresponding optimal power allocation policy for a given rate vector when transmission to all users is turned on or off simultaneously. When an outage can be declared for each user individually, we have derived the outage probability region and the corresponding power allocation strategy that achieves the outage probability region boundary for the given rate vector. In both cases, iterative algorithms are proposed for obtaining the appropriate power allocation parameters, which determine the optimal decoding order and power allocation for each user such that the minimum common outage probability or the outage probability region boundary is achieved. For a given common outage probability or an outage probability vector of the M users, by applying the derived optimal power allocation strategies, we have also obtained the average power regions that can support a rate vector with the outage probability constraint for each user satisfied.

Appendix A Equivalence of the Conventional Detector with Weighting and the Output-Feedback Detector with Hard-Decision in Chapter 3

We prove here the equivalence under hard-decision decoding of the conventional decoder with weighting and the output-feedback decoder for modulation types that have a decision rule independent of the fade level of the channel.

With hard decision, the received symbol is demodulated to be one of the symbols in the channel input alphabet χ. Denote the size of χ as $|\chi|$ and assume that the symbols in χ are equiprobable. Suppose the demodulation criterion is such that $\forall n = 1, 2, \cdots$, for any fade level α_n, the demodulated symbol y_n can be any of the $|\chi|$ symbols with equal probability[1]. Then in (3.11), since $p(x_n) = \frac{1}{|\chi|}$ and $\forall \gamma$, $\sum_{x_n \in \chi} p(y_n|x_n, \gamma) = 1$, the following holds for all k, $k = 1, 2, \cdots, K$:

$$\begin{aligned} p(y_n|\alpha_n \in S_k) &= \frac{\sum_{x_n \in \chi} \int_{S_k} p(y_n|x_n, \gamma) p_\alpha(\gamma) p(x_n) d\gamma}{\int_{S_k} p_\alpha(\gamma) d\gamma} \\ &= \frac{\frac{1}{|\chi|} \int_{S_k} [\sum_{x_n \in \chi} p(y_n|x_n, \gamma)] p_\alpha(\gamma) d\gamma}{\int_{S_k} p_\alpha(\gamma) d\gamma} \\ &= \frac{1}{|\chi|}, \end{aligned} \tag{A.1}$$

which means that all the diagonal terms of $B(y_n)$ in (3.10) are equal to $\frac{1}{|\chi|}$. Therefore, we can simplify (3.10) as

$$\rho_{n+1} = \rho_n P. \tag{A.2}$$

Since P is the matrix of transition probabilities for the K channel states, which form an irreducible, ergodic Markov chain, if the initial probability vector ρ_1 is chosen to be the stationary distribution of the states, then ρ_n will be the stationary distribution for all n. If ρ_1 is chosen arbitrarily, then ρ_n will converge to the stationary distribution of the channel states as n increases ([80], pp. 393-394). Thus, the metric (3.12) degenerates to the metric

[1] For example, the demodulation criterion for MPSK symbols satisfies this condition but that for MQAM does not.

(3.14) and the equavalence of the output-feedback decoder and the conventional decoder with weighting is proved.

Appendix B Proofs in Chapter 4

B.1 Proof of Theorem 4.1

The achievability and converse of the capacity region $\mathcal{C}(\bar{P})$ in (4.1) is proved in Section B.1.1 and Section B.1.2, respectively, while the convexity of this capacity region is shown in Section B.1.3. Since in the converse proof, the capacity of a probabilistic broadcast channel is used, we prove the capacity formula of a probabilistic broadcast channel in Section B.1.4.

B.1.1 Achievability of the Capacity Region

We prove the achievability of the capacity region $\mathcal{C}(\bar{P})$ in (4.1) by proving the achievability of $\mathcal{C}_{CD}(\mathcal{P})$ in (4.2) for each given power allocation policy $\mathcal{P} \in \mathcal{F}$.

$\forall \mathcal{P} \in \mathcal{F}$, for $j = 1, 2, \cdots, M$, since $P_j(\mathbf{n})$ is the transmit power allocated to User j for the joint fading state $\mathbf{n}$, by denoting

$$C_j \triangleq E_{\mathbf{n}}\left[B \log\left(1 + \frac{P_j(\mathbf{n})}{n_j B + \sum_{i=1}^{M} P_i(\mathbf{n}) \mathbf{1}[n_j > n_i]}\right)\right], \tag{B.1}$$

we need to prove that for any given $\mathbf{R} = (R_1, R_2, \ldots, R_M)$ satisfying $R_j < C_j$, $j = 1, 2, \cdots, M$, there exists a sequence of $\left((2^{nR_{1,n}}, 2^{nR_{2,n}}, \cdots, 2^{nR_{M,n}}), n\right)$ codes with rate $R_{j,n} \to R_j$ $(j = 1, 2, \cdots, M)$ and the probability of error $\epsilon_n \to 0$ as $n \to \infty$. The proof is similar to that of the achievability of the single-user fading channel capacity [72].

Fix any $\epsilon > 0$, and let $R_j = C_j - 3\epsilon$, $1 \leq j \leq M$. Define

$$v_i = \frac{i}{m}, \quad i = 0, 1, 2, \cdots, mI. \tag{B.2}$$

Since the time-varying noise density $n_j = \frac{\nu_j}{g_j}$ of each user ranges from 0 to ∞, we say that a sub-channel is in state $S_i, i = 0, 1, 2, \cdots, mI$ if $v_i \leq n_j < v_{i+1}$, where $v_{mI+1} = \infty$. The set $\{v_i\}_{i=0}^{mI}$ discretizes the fading range of each sub-channel into $mI + 1$ states. Thus there are

$$(mI + 1)^M = N \tag{B.3}$$

discrete joint channel states. We denote the kth $(0 \leq k \leq N-1)$ of these N states as $\mathbf{S}_k = \left[S_{\phi(k,1)}, S_{\phi(k,2)}, \cdots, S_{\phi(k,M)}\right]$, where $[\phi(k,1), \phi(k,2), \cdots, \phi(k,M)]$ is the base-$(mI+1)$ expansion of k, with $\phi(k,1)$ being the least important component. That is, $0 \leq \phi(k,j) \leq mI$ for all $1 \leq j \leq M$ and

$$k = \sum_{j=1}^{M} \phi(k,j) \cdot (mI+1)^{j-1}.$$

Note that a channel state $\mathbf{n} \in \mathbf{S}_k$ if and only if $n_j \in S_{\phi(k,j)}$, $\forall 1 \leq j \leq M$.

Over a given time interval $[0, n]$, let ξ_k be the number of transmissions during which the channel is in the state $\mathbf{S}_k$, and let the transmit power allocated to User j $(j = 1, 2, \cdots, M)$ be

$$P_j(\mathbf{S}_k) = \min_{\mathbf{n} \in \mathbf{S}_k} P_j(\mathbf{n}), \quad k = 0, 1, 2, \cdots, N-1. \tag{B.4}$$

By the stationarity and ergodicity of the channel variation,

$$\frac{\xi_k}{n} \to p(\mathbf{S}_k) \quad \text{as } n \to \infty. \tag{B.5}$$

Consider a time-invariant AWGN broadcast channel with sub-channel noise variances $v_{\phi(k,1)}$, $v_{\phi(k,2)}$, $\cdots$, $v_{\phi(k,M)}$ and transmit power $P_1(\mathbf{S}_k), P_2(\mathbf{S}_k), \cdots, P_M(\mathbf{S}_k)$ of the M users. Note that according to (B.2), $v_{\phi(k,j)} = \frac{\phi(k,j)}{m}$, $\forall 1 \leq j \leq M$. For a given n, let $\xi_k = \lfloor np(\mathbf{S}_k) \rfloor = np(\mathbf{S}_k)$ for n sufficiently large. From [78], we know that for

$$R_{j,\xi_k} = B \log\left(1 + \frac{P_j(\mathbf{S}_k)}{v_{\phi(k,j)}B + \sum_{i=1}^{M} P_i(\mathbf{S}_k)\mathbf{1}[v_{\phi(k,j)} > v_{\phi(k,i)}]}\right), \quad j = 1, 2, \cdots, M,$$

there exists a sequence of $\left((2^{\xi_k R_{1,\xi_k}}, 2^{\xi_k R_{2,\xi_k}}, \ldots, 2^{\xi_k R_{M,\xi_k}}), \xi_k\right)$ codes $\{x_{\mathbf{w}_k}[i]\}_{i=1}^{\xi_k}$, $\mathbf{w}_k = (w_{1,k}, w_{2,k}, \cdots, w_{M,k})$, $w_{j,k} = 1, 2, \cdots, 2^{\xi_k R_{j,\xi_k}}$ with error probability $\epsilon_{n,k} \to 0$ as $\xi_k \to \infty$.

The message index $\mathbf{w} \in \left\{1, 2, \cdots, 2^{nR_{1,n}}\right\} \times \left\{1, 2, \cdots, 2^{nR_{2,n}}\right\} \times \cdots \times \left\{1, 2, \cdots, 2^{nR_{M,n}}\right\}$ is transmitted over the N channel states as follows. We first map $\mathbf{w}$ to the indices $\{\mathbf{w}_k\}_{k=0}^{N-1}$ by dividing the $nR_{j,n}$ bits into sets of $\xi_k R_{j,\xi_k}$ bits for $j = 1, 2, \cdots, M$. We then use the multiplexing strategy [72] to transmit the codeword $x_{\mathbf{w}_k}[\cdot]$ whenever the channel is in state $\mathbf{S}_k$. On the interval [0, n], the kth channel state $\mathbf{S}_k$ is used ξ_k times. We can thus achieve the transmission rates

$$R_{j,n} = \sum_{k=0}^{N-1} R_{j,\xi_k} \frac{\xi_k}{n}$$

$$= \sum_{k=0}^{N-1} B \log\left(1 + \frac{P_j(\mathbf{S}_k)}{v_{\phi(k,j)}B + \sum_{i=1}^{M} P_i(\mathbf{S}_k)\mathbf{1}[v_{\phi(k,j)} > v_{\phi(k,i)}]}\right)\frac{\xi_k}{n},$$
$$j = 1, 2, \cdots, M. \tag{B.6}$$

The average total transmit power for the multiplexed code is

$$P_n = \sum_{k=0}^{N-1}\sum_{j=1}^{M} P_j(\mathbf{S}_k)\frac{\xi_k}{n}.$$

From (B.5) and (B.6), it is easily seen that for $j = 1, 2, \cdots, M$,

$$\lim_{n\to\infty} R_{j,n} = \sum_{k=0}^{N-1} B \log\left(1 + \frac{P_j(\mathbf{S}_k)}{v_{\phi(k,j)}B + \sum_{i=1}^{M} P_i(\mathbf{S}_k)\mathbf{1}[v_{\phi(k,j)} > v_{\phi(k,i)}]}\right)p(\mathbf{S}_k).$$

So for ϵ fixed, we can find n sufficiently large such that

$$R_{j,n} \geq \sum_{k=0}^{N-1} B \log\left(1 + \frac{P_j(\mathbf{S}_k)}{v_{\phi(k,j)}B + \sum_{i=1}^{M} P_i(\mathbf{S}_k)\mathbf{1}[v_{\phi(k,j)} > v_{\phi(k,i)}]}\right)p(\mathbf{S}_k) - \epsilon.$$

Moreover, from (B.4), it is clear that the power control policy satisfies the average total power constraint for asymptotically large n:

$$\begin{aligned}
\lim_{n\to\infty}\sum_{k=0}^{N-1}\sum_{j=1}^{M} P_j(\mathbf{S}_k)\frac{\xi_k}{n} &= \sum_{k=0}^{N-1}\sum_{j=1}^{M} P_j(\mathbf{S}_k)\int_{\mathbf{n}\in\mathbf{S}_k} dF(\mathbf{n}) \\
&\leq \sum_{j=1}^{M}\sum_{k=0}^{N-1}\int_{\mathbf{n}\in\mathbf{S}_k} P_j(\mathbf{n})dF(\mathbf{n}) \\
&\leq \bar{P}.
\end{aligned}$$

The error probability of the multiplexed coding scheme is bounded above by

$$\epsilon_n \leq \sum_{k=0}^{N-1} \epsilon_{n,k} \to 0 \quad \text{as } n \to \infty,$$

since $n \to \infty$ implies $\xi_k \to \infty$, $\forall k = 1, 2, \cdots, N-1$. Thus, it remains to show that for fixed ϵ there exist m and I (note $N = (mI+1)^M$) such that

$$\sum_{k=0}^{N-1} B \log\left(1 + \frac{P_j(\mathbf{S}_k)}{v_{\phi(k,j)}B + \sum_{i=1}^{M} P_i(\mathbf{S}_k)\mathbf{1}[v_{\phi(k,j)} > v_{\phi(k,i)}]}\right)p(\mathbf{S}_k) \geq C_j - 2\epsilon.$$

From the definition of C_j in (B.1), it is easily shown that for $j = 1, 2, \cdots, M$,

$$\begin{aligned} C_j &= E_{\mathbf{n}}\left[B\log\left(1+\frac{P_j(\mathbf{n})}{n_jB+\sum_{i=1}^{M}P_i(\mathbf{n})\mathbf{1}[n_j>n_i]}\right)\right] \\ &\le E_{\mathbf{n}}\left[B\log\left(1+\frac{P_j(\mathbf{n})}{n_jB}\right)\right] \\ &\overset{a}{\le} B\log\left[1+E_{\mathbf{n}}\left(\frac{P_j(\mathbf{n})}{n_jB}\right)\right] \\ &\overset{b}{<} \infty, \end{aligned}$$

where a follows from Jensen's inequality and b follows from the fact that the average SNR for each user is finite. So for fixed ϵ there exists an I_ϵ such that

$$E_{\mathbf{n}\in A(I_\epsilon)}\left[B\log\left(1+\frac{P_j(\mathbf{n})}{n_jB+\sum_{i=1}^{M}P_i(\mathbf{n})\mathbf{1}[n_j>n_i]}\right)\right] < \epsilon, \tag{B.7}$$

where $A(I_\epsilon) \triangleq \{\mathbf{n} : n_j > I_\epsilon, \forall j = 1, 2, \cdots, M\}$. Moreover, by defining $N_1 = (mI)^M$, for I fixed, the monotone convergence theorem [81] implies that

$$\begin{aligned} &\lim_{m\to\infty}\sum_{k=0}^{N_1-1} B\log\left(1+\frac{P_j(\mathbf{S}_k)}{v_{\phi(k,j)}B+\sum_{i=1}^{M}P_i(\mathbf{S}_k)\mathbf{1}[v_{\phi(k,j)}>v_{\phi(k,i)}]}\right)p(\mathbf{S}_k) \\ &= \lim_{m\to\infty}\sum_{\phi(k,1)=0}^{mI-1}\sum_{\phi(k,2)=0}^{mI-1}\cdots\sum_{\phi(k,M)=0}^{mI-1} B\log\left(1+\frac{P_j(\mathbf{S}_k)}{v_{\phi(k,j)}B+\sum_{i=1}^{M}P_i(\mathbf{S}_k)\mathbf{1}[v_{\phi(k,j)}>v_{\phi(k,i)}]}\right)\times \\ &\qquad p(\mathbf{n} : v_{\phi(k,j)} \le n_j < v_{\phi(k,j)+1}, \forall 1 \le j \le M) \\ &= \lim_{m\to\infty}\sum_{\phi(k,1)=0}^{mI-1}\sum_{\phi(k,2)=0}^{mI-1}\cdots\sum_{\phi(k,M)=0}^{mI-1}\int_{v_{\phi(k,1)}}^{v_{\phi(k,1)+1}}\int_{v_{\phi(k,2)}}^{v_{\phi(k,2)+1}}\cdots\int_{v_{\phi(k,M)}}^{v_{\phi(k,M)+1}} \\ &\qquad B\log\left(1+\frac{P_j(\mathbf{n})}{n_jB+\sum_{i=1}^{M}P_i(\mathbf{n})\mathbf{1}[n_j>n_i]}\right)dF(\mathbf{n}) \\ &= E_{\mathbf{n}\notin A(I)}\left[B\log\left(1+\frac{P_j(\mathbf{n})}{n_jB+\sum_{i=1}^{M}P_i(\mathbf{n})\mathbf{1}[n_j>n_i]}\right)\right]. \end{aligned} \tag{B.8}$$

Thus, using the I_ϵ in (B.7) and combining (B.7) and (B.8) we see that for the given ϵ there exists an m sufficiently large such that

$$\sum_{k=0}^{(mI_\epsilon)^M-1} B\log\left(1+\frac{P_j(\mathbf{S}_k)}{v_{\phi(k,j)}B+\sum_{i=1}^M P_i(\mathbf{S}_k)\mathbf{1}[v_{\phi(k,j)}>v_{\phi(k,i)}]}\right)p(\mathbf{S}_k)$$

$$\geq \; E_{\mathbf{n}}\left[B\log\left(1+\frac{P_j(\mathbf{n})}{n_jB+\sum_{i=1}^M P_i(\mathbf{n})\mathbf{1}[n_j>n_i]}\right)\right]-2\epsilon,$$

which completes the proof. □

B.1.2 Converse

Suppose that rate $\mathbf{R}$ is achievable, then we need to prove that any sequence of $\Big((2^{nR_1}, 2^{nR_2}, \cdots, 2^{nR_M}), n\Big)$ codes with average total power $\bar{P}$ and probability of error $\epsilon_n \to 0$ as $n \to \infty$ must have $R_j \leq C_j$, $j = 1, 2, \cdots, M$, where C_j is given in (B.1). We assume that the codes are designed with a priori knowledge of the joint channel state $\mathbf{n}$. Since the transmitter and receivers know state $\mathbf{n}$ up to the current time, this assumption can only results in a higher achievable rate.

As in the proof of the achievability in Section B.1.1, we say that the jth sub-channel is in state S_i, $i = 0, 1, 2, \cdots, mI$ if the time-varying noise density n_j of User j satisfies $v_i \leq n_j < v_{i+1}$, where v_i is defined in (B.2) and $v_{mI+1} = \infty$. Thus, the set $\{v_i\}_{i=0}^{mI}$ discretizes the fading range of each sub-channel into $mI+1$ states and there are $N = (mI+1)^M$ discrete joint channel states. We denote the kth $(0 \leq k \leq N-1)$ of these N states as $\mathbf{S}_k = \Big[S_{\phi(k,1)}, S_{\phi(k,2)}, \cdots, S_{\phi(k,M)}\Big]$, where $[\phi(k,1), \phi(k,2), \cdots, \phi(k,M)]$ is the base-$(mI+1)$ expansion of k, and $S_{\phi(k,j)}$ is the jth sub-channel state. That is, $0 \leq \phi(k,j) \leq mI$ for all $1 \leq j \leq M$ and

$$k = \sum_{j=1}^{M} \phi(k,j) \cdot (mI+1)^{j-1}.$$

Note that a channel state $\mathbf{n} \in \mathbf{S}_k$ if and only if $n_j \in S_{\phi(k,j)}$, $\forall 1 \leq j \leq M$.

Over a given time interval $[0, n]$, let ξ_k be the number of transmissions during which the channel is in the state $\mathbf{S}_k$, let $\Omega(\mathbf{S}_k)$ be the random subset of $[1, 2, \cdots, n]$ at which times the channel is in the state $\mathbf{S}_k$, and let Q be uniformly distributed on $[1, 2, \cdots, n]$. By the stationarity and ergodicity of the channel variation,

$$\frac{\xi_k}{n} \to p(\mathbf{S}_k) \qquad \text{as } n \to \infty.$$

For $i = 1, 2, \cdots, n$, let $P_j^{(n)}(i)$ be the transmit power allocated to User j at time i and define

$$P_j(\mathbf{S}_k, n) = E_Q\left[P_j^{(n)}(Q) | Q \in \Omega(\mathbf{S}_k)\right].$$

Since for any message from the base station to the M users, there is a power constraint on the corresponding codeword, it follows that for each n,

$$\sum_{k=0}^{N-1} \sum_{j=1}^{M} P_j(\mathbf{S}_k, n) p(\mathbf{S}_k) \leq \bar{P}.$$

Therefore, for all $\mathbf{S}_k$ such that $p(\mathbf{S}_k) \neq 0$, $P_j(\mathbf{S}_k, n)$ are bounded sequences in n. Thus, there exist a converging subsequence and a limiting $P_j(\mathbf{S}_k)$ such that

$$P_j(\mathbf{S}_k, n) \to P_j(\mathbf{S}_k) \quad \text{as} \quad n \to \infty.$$

Moreover,

$$\sum_{k=0}^{N-1} \sum_{j=1}^{M} P_j(\mathbf{S}_k) p(\mathbf{S}_k) \leq \bar{P}. \tag{B.9}$$

For a given power allocation policy $\mathcal{P}$, assume that $P_j(\mathbf{S}_k)$ is the transmit power assigned to User j when the time-varying broadcast channel is in channel state $\mathbf{S}_k$. Let $\mathcal{F}_N$ be the set of all the power allocation policies which are piecewise constant in each channel state and which satisfy the average total power constraint (B.9). Assuming that the noise densities of the M users in each channel state $\mathbf{S}_k$ $(0 \leq k \leq N-1)$ are constants and are denoted as $n_1(k), n_2(k), \cdots, n_M(k)$, the N channel states $\mathbf{S}_0, \mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_{N-1}$ can be viewed as N AWGN broadcast channels where at any given time only one of these N channels is in operation and the probability that the jth broadcast channel is in operation is given by $p(\mathbf{S}_j)$, $\forall 1 \leq j \leq N$. We call this the probablistic broadcast channel. We will show in Section B.1.4 that if $\mathbf{R}^{(N)}$ is achievable on the probablistic broadcast channel consisting of the N broadcast channels $\mathbf{S}_0, \mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_{N-1}$ with probabilities $p(\mathbf{S}_0), p(\mathbf{S}_1), p(\mathbf{S}_2), \cdots, p(\mathbf{S}_{N-1})$ under the assumption of perfect transmitter and receiver CSI (i.e., at time i it is known at both the transmitter and receivers which broadcast channel is in operation), then

$$\mathbf{R}^{(N)} \in \bigcup_{\mathcal{P}\in\mathcal{F}_N} \mathcal{C}(\mathcal{P}), \tag{B.10}$$

where

$$\mathcal{C}(\mathcal{P}) = \left\{ R_j \leq \sum_{k=0}^{N-1} p(\mathbf{S}_k) B \log\left(1 + \frac{P_j(\mathbf{S}_k)}{n_j(k)B + \sum_{i=1}^{M} P_i(\mathbf{S}_k)\mathbf{1}[n_j(k) > n_i(k)]}\right), 1 \leq j \leq M \right\}.$$

Define

$$\bar{\mathcal{C}}(\mathcal{P}) = \left\{ R_j \leq \sum_{k=0}^{N-1} p(\mathbf{S}_k) B \log\left(1 + \frac{P_j(\mathbf{S}_k)}{v_{\phi(k,j)}B + \sum_{i=1}^{M} P_i(\mathbf{S}_k)\mathbf{1}[v_{\phi(k,j)} > v_{\phi(k,i)}]}\right), 1 \leq j \leq M \right\}.$$

Since for $0 \leq k \leq N-1$ and $1 \leq j \leq M$,

$$v_{\phi(k,j)} \leq n_j(k) < v_{\phi(k,j)+1},$$

it is clear that

$$\mathcal{C}(\mathcal{P}) \subseteq \bar{\mathcal{C}}(\mathcal{P}). \tag{B.11}$$

$\forall \mathcal{P} \in \mathcal{F}_N$, let

$$P_j(\mathbf{n}) = \sum_{k=0}^{N-1} P_j(\mathbf{S}_k)\mathbf{1}[\mathbf{n} \in \mathbf{S}_k], \quad 1 \leq j \leq M.$$

According to the power constraint (B.9), it is obvious that $\{P_j(\mathbf{n})\}_{j=1}^{M}$ satisfies

$$E_{\mathbf{n}}\left[\sum_{j=1}^{M} P_j(\mathbf{n})\right] \leq \bar{P}.$$

Thus, $\mathcal{P} \in \mathcal{F}$ and

$$\bigcup_{\mathcal{P}\in\mathcal{F}_N} \mathcal{C}_{CD}(\mathcal{P}) \subseteq \bigcup_{\mathcal{P}\in\mathcal{F}} \mathcal{C}_{CD}(\mathcal{P}), \tag{B.12}$$

where $\mathcal{C}_{CD}(\mathcal{P})$ is given in (4.2). Taking the limit of the left-hand side of (B.12), we obtain

$$\lim_{N\to\infty} \bigcup_{\mathcal{P}\in\mathcal{F}_N} \mathcal{C}_{CD}(\mathcal{P}) \subseteq \bigcup_{\mathcal{P}\in\mathcal{F}} \mathcal{C}_{CD}(\mathcal{P}). \tag{B.13}$$

For the time-varying broadcast channel, if $\mathbf{R}$ is achievable, then

$$\mathbf{R} \in \lim_{N\to\infty} \left\{ \mathbf{R}^{(N)} : \mathbf{R}^{(N)} \in \bigcup_{\mathcal{P}\in\mathcal{F}_N} \mathcal{C}(\mathcal{P}) \right\} = \lim_{N\to\infty} \bigcup_{\mathcal{P}\in\mathcal{F}_N} \mathcal{C}(\mathcal{P}). \tag{B.14}$$

From (B.11) and (B.14) we have

$$\mathbf{R} \in \lim_{N\to\infty} \bigcup_{\mathcal{P}\in\mathcal{F}_N} \bar{\mathcal{C}}(\mathcal{P}).$$

That is,

$$\mathcal{C}(\bar{P}) \subseteq \lim_{N\to\infty} \bigcup_{\mathcal{P}\in\mathcal{F}_N} \bar{\mathcal{C}}(\mathcal{P}), \tag{B.15}$$

where $\mathcal{C}(\bar{P})$ denotes the capacity region of the time-varying broadcast channel. Combining (B.13) and (B.15) with the achievability result which indicates that

$$\bigcup_{\mathcal{P}\in\mathcal{F}} \mathcal{C}_{CD}(\mathcal{P}) \subseteq \mathcal{C}(\bar{P}),$$

we obtain

$$\lim_{N\to\infty} \bigcup_{\mathcal{P}\in\mathcal{F}_N} \mathcal{C}_{CD}(\mathcal{P}) \subseteq \bigcup_{\mathcal{P}\in\mathcal{F}} \mathcal{C}_{CD}(\mathcal{P}) \subseteq \mathcal{C}(\bar{P}) \subseteq \lim_{N\to\infty} \bigcup_{\mathcal{P}\in\mathcal{F}_N} \bar{\mathcal{C}}(\mathcal{P}).$$

Since the upper bound equals the lower bound by the monotone convergence theorem [81], it is clear that

$$\mathcal{C}(\bar{P}) = \bigcup_{\mathcal{P}\in\mathcal{F}} \mathcal{C}_{CD}(\mathcal{P}). \quad \square$$

B.1.3 Convexity of the Capacity Region

We use the idea of time-sharing to prove that if $\mathbf{R} \in \mathcal{C}(\bar{P})$ and $\mathbf{R}' \in \mathcal{C}(\bar{P})$, then $\lambda\mathbf{R} + (1-\lambda)\mathbf{R}' \in \mathcal{C}(\bar{P})$ for $0 \leq \lambda \leq 1$. We know from the proof about the achievability of $\mathcal{C}(\bar{P})$ that for a given rate $\mathbf{R} \in \mathcal{C}(\bar{P})$, there exists a set of codebooks for different channel states and $\mathbf{R}$ can be achieved by using the multiplexing strategy. For $\mathbf{R}' \in \mathcal{C}(\bar{P})$, there exists another set of codebooks for different channel states. We can construct a third set of codebooks to achieve the rate $\lambda\mathbf{R} + (1-\lambda)\mathbf{R}'$ by using the codebook from the first set of codebooks for the first $\lambda\xi_k$ symbols and using the codebook from the second set of codebooks for the last $(1-\lambda)\xi_k$ symbols in each channel state $\mathbf{S}_k$, where ξ_k and $\mathbf{S}_k$ are as discussed in Section B.1.1.

Since the new codebook in each state $\mathbf{S}_k$ is constructed with a rate $\lambda R_{j,\xi_k} + (1-\lambda)R'_{j,\xi_k}$, by using the new set of codebooks, the rate of the new code is $\lambda\mathbf{R} + (1-\lambda)\mathbf{R}'$. Because the overall probability of error is less than the sum of the probabilities of error for each of the segments, the probability of error for the new code goes to 0 for n large enough. It is obvious that the new code satisfies the total power constraint, since each of the component codes satisfies this constraint. Therefore, the rate $\lambda\mathbf{R} + (1-\lambda)\mathbf{R}'$ is achievable. □

B.1.4 Capacity of a Probabilistic Broadcast Channel with CSI

In Section B.1.2, while proving the converse of the capacity region in *Theorem 4.1*, we have used the capacity of a probabilistic broadcast channel consisting of N discrete AWGN broadcast channels with given probabilities under the assumption that perfect CSI is available at both the transmitter and the receivers. Now we prove the capacity formula of a probabilistic broadcast channel composed of two AWGN broadcast channels with given probabilities. The result can be easily generalized to N ($N > 2$) channels.

Assume that two discrete degraded memoryless AWGN broadcast channels

$$(\mathcal{X}_1, p(y_1|x_1)p(z_1|y_1), \mathcal{Y}_1 \times \mathcal{Z}_1)$$

and

$$(\mathcal{X}_2, p(z_2|x_2)p(y_2|z_2), \mathcal{Y}_2 \times \mathcal{Z}_2)$$

are as shown in Figure B.1. In the first channel (Channel 1), the Gaussian noises are denoted as ν_{11} and ν_{12}, the noise densities of which are N_{11} and N_{12}, respectively. In the second channel (Channel 2), the Gaussian noises are ν_{21} and ν_{22}, the noise densities of which are N_{21} and N_{22}, respectively. Let $N_1 = N_{11} + N_{12}$ and $N_2 = N_{21} + N_{22}$. We define the probabilistic broadcast channel consisting of two AWGN broadcast channels as a channel $(\mathcal{X}, p(y,z|x), \mathcal{Y} \times \mathcal{Z})$ with two outputs, where $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$, $\mathcal{Y} = \mathcal{Y}_1 \cup \mathcal{Y}_2$, $\mathcal{Z} = \mathcal{Z}_1 \cup \mathcal{Z}_2$, and

$$p(y,z|x) = \begin{cases} p(y_1|x_1)p(z_1|y_1), & \text{with probability } p_c, \\ p(z_2|x_2)p(y_2|z_2), & \text{with probability } \bar{p}_c \triangleq 1 - p_c. \end{cases}$$

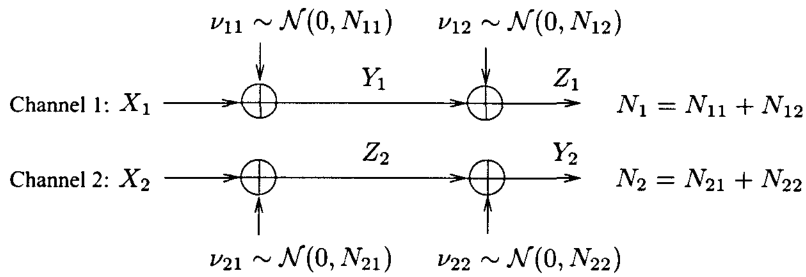


Figure B.1: Probabilistic broadcast channel: Channel 1 in operation with probability p_c, and Channel 2 in operation with probability $\bar{p}_c$.

Denote $C(P/N)$ as the capacity of an AWGN channel with signal-to-noise ratio P/N, i.e., $C(P/N) = \frac{1}{2}\ln(1 + P/N)$. Let R_1 and R_2 be the transmission rates of the particular information to Y and Z respectively, and let R_0 be the transmission rate of the common information. $\forall 0 \leq \alpha_1, \alpha_2, \beta \leq 1$, let $\bar{\alpha}_1 = 1 - \alpha_1$, $\bar{\alpha}_2 = 1 - \alpha_2$, $\bar{\beta} = 1 - \beta$.

Theorem B.1 *The capacity region of the probabilistic broadcast channel in Figure B.1 is defined by:*

$$
\begin{aligned}
C_G \;=\; & \{(R_0, R_1, R_2) : R_0 \leq p_c C(\alpha_1 \beta \bar{P}/(N_{11} + \bar{\alpha}_1 \beta \bar{P})) + \bar{p}_c C(\alpha_2 \bar{\beta} \bar{P}/(N_2 + \bar{\alpha}_2 \bar{\beta} \bar{P})), \\
& R_0 \leq p_c C(\alpha_1 \beta \bar{P}/(N_1 + \bar{\alpha}_1 \beta \bar{P})) + \bar{p}_c C(\alpha_2 \bar{\beta} \bar{P}/(N_{21} + \bar{\alpha}_2 \bar{\beta} \bar{P})), \\
& R_0 + R_1 \leq p_c C(\beta \bar{P}/N_{11}) + \bar{p}_c C(\alpha_2 \bar{\beta} \bar{P}/(N_2 + \bar{\alpha}_2 \bar{\beta} \bar{P})), \\
& R_0 + R_2 \leq p_c C(\alpha_1 \beta \bar{P}/(N_1 + \bar{\alpha}_1 \beta \bar{P})) + \bar{p}_c C(\bar{\beta} \bar{P}/N_{21}), \\
& R_0 + R_1 + R_2 \leq p_c C(\beta \bar{P}/N_{11}) + \bar{p}_c [C(\alpha_2 \bar{\beta} \bar{P}/(N_2 + \bar{\alpha}_2 \bar{\beta} \bar{P})) + C(\bar{\alpha}_2 \bar{\beta} \bar{P}/N_{21})], \\
& R_0 + R_1 + R_2 \leq p_c [C(\alpha_1 \beta \bar{P}/(N_1 + \bar{\alpha}_1 \beta \bar{P})) + C(\bar{\alpha}_1 \beta \bar{P}/N_{11})] + \bar{p}_c C(\bar{\beta} \bar{P}/N_{21}), \\
& \quad \forall 0 \leq \alpha_1, \alpha_2, \beta \leq 1\}, \qquad \text{(B.16)}
\end{aligned}
$$

assuming that the total average power is $\bar{P}$ and perfect CSI is available at both the transmitter and the receivers.

Note that this capacity region is similar to that of a parallel broadcast channel composed of two broadcast channels [82, 83], except that in the parallel case, both p_c and $\bar{p}_c$ equal 1.

In the case of independent rates (i.e., $R_0 = 0$), (B.16) becomes

$$\begin{aligned} C_G &= \{(R_1, R_2) : R_1 \le p_c C(\beta\bar{P}/N_{11}) + \bar{p}_c C(\alpha_2\bar{\beta}\bar{P}/(N_2 + \bar{\alpha}_2\bar{\beta}\bar{P})), \\ & R_2 \le p_c C(\alpha_1\beta\bar{P}/(N_1 + \bar{\alpha}_1\beta\bar{P})) + \bar{p}_c C(\bar{\beta}\bar{P}/N_{21}), \\ & R_1 + R_2 \le p_c C(\beta\bar{P}/N_{11}) + \bar{p}_c[C(\alpha_2\bar{\beta}\bar{P}/(N_2 + \bar{\alpha}_2\bar{\beta}\bar{P})) + C(\bar{\alpha}_2\bar{\beta}\bar{P}/N_{21})], \\ & R_1 + R_2 \le p_c[C(\alpha_1\beta\bar{P}/(N_1 + \bar{\alpha}_1\beta\bar{P})) + C(\bar{\alpha}_1\beta\bar{P}/N_{11})] + \bar{p}_c C(\bar{\beta}\bar{P}/N_{21}), \\ & \quad \forall 0 \le \alpha_1, \alpha_2, \beta \le 1\}, \end{aligned} \tag{B.17}$$

which is equivalent to

$$\begin{aligned} C_G &= \{(R_1, R_2) : R_1 \le p_c C(\bar{\alpha}_1\beta\bar{P}/N_{11}) + \bar{p}_c C(\alpha_2\bar{\beta}\bar{P}/(N_2 + \bar{\alpha}_2\bar{\beta}\bar{P})), \\ & R_2 \le p_c C(\alpha_1\beta\bar{P}/(N_1 + \bar{\alpha}_1\beta\bar{P})) + \bar{p}_c C(\bar{\alpha}_2\bar{\beta}\bar{P}/N_{21}), \\ & \quad \forall 0 \le \alpha_1, \alpha_2, \beta \le 1\}. \end{aligned} \tag{B.18}$$

The equivalence of (B.17) and (B.18) can be similarly shown as in [82, 83].

Proof of Theorem B.1:

Achievability: For fixed α_1, α_2, β we can achieve any triple $(R_0, R_1, R_2) \in C_G$ by first dividing the total power $\bar{P}$ into the part $\beta\bar{P}$ used in the first broadcast channel and $\bar{\beta}\bar{P}$ used in the second. Then $\beta\bar{P}$ is divided into the power $\alpha_1\beta\bar{P}$ used to transmit the common information to Y and $\bar{\alpha}_1\beta\bar{P}$ used to transmit the particular information to Y; $\bar{\beta}\bar{P}$ is divided into the power $\alpha_2\bar{\beta}\bar{P}$ used to transmit the common information to Z and $\bar{\alpha}_2\bar{\beta}\bar{P}$ used to transmit the particular information to Z. The details of the proof of achievability are standard and therefore omitted.

Converse: We know that a $((2^{nR_0}, 2^{nR_1}, 2^{nR_2}), n)$ code for a broadcast channel with perfect CSI at both the transmitter and the receivers consists of three sets of integers

$$\mathcal{M}_0 = \{1, 2, \cdots, 2^{nR_0}\}, \mathcal{M}_1 = \{1, 2, \cdots, 2^{nR_1}\}, \mathcal{M}_2 = \{1, 2, \cdots, 2^{nR_2}\},$$

a coding mapping

$$\mathbf{X} : \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{X}^n,$$

and two decoding mappings

$$g_1 : \mathcal{Y}^n \rightarrow \mathcal{M}_0 \times \mathcal{M}_1, g_1(\mathbf{Y}|\mathbf{V}) = (\hat{W}_0, \hat{W}_1),$$

$$g_2 : \mathcal{Z}^n \rightarrow \mathcal{M}_0 \times \mathcal{M}_2, g_2(\mathbf{Z}|\mathbf{V}) = (\hat{\hat{W}}_0, \hat{W}_2),$$

where $\mathbf{V} = (V_1, V_2, \cdots, V_n)$ represents the CSI, i.e., for $1 \leq i \leq n$, if $V_i = 0$, the first AWGN broadcast channel in Figure B.1 is the current channel state; if $V_i = 1$, the second one in Figure B.1 is the current channel state. Therefore, by denoting $n^* \triangleq \sum_{i=1}^n V_i$, we have

$$\frac{n^*}{n} \rightarrow \bar{p}_c \quad \text{as } n \rightarrow \infty.$$

The set of codewords is $\{\mathbf{X}(\mathbf{w}|\mathbf{V}) : \mathbf{w} \triangleq (w_0, w_1, w_2) \in \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2\}$. Integer W_0 represents the common part of the message and W_i $(i = 1, 2)$ represents the independent part of the message. Assuming that the distribution on (W_0, W_1, W_2) is uniform, we define the mean error probabilities for decoders g_1 and g_2 as follows:

$$P_{e,1}^n = \frac{1}{2^{n(R_0+R_1+R_2)}} \sum_{w_0,w_1,w_2 \in \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2} p\{g_1(\mathbf{Y}|\mathbf{V}) \neq (w_0, w_1)|(w_0, w_1, w_2) \text{ transmitted}\},$$

$$P_{e,2}^n = \frac{1}{2^{n(R_0+R_1+R_2)}} \sum_{w_0,w_1,w_2 \in \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2} p\{g_2(\mathbf{Z}|\mathbf{V}) \neq (w_0, w_2)|(w_0, w_1, w_2) \text{ transmitted}\}.$$

Therefore, we say that a rate triple (R_0, R_1, R_2) is achievable for the broadcast channel if there exists a sequence of $((2^{nR_0}, 2^{nR_1}, 2^{nR_2}), n)$ codes with $\max\{P_{e,1}^n, P_{e,2}^n\} \rightarrow 0$ as $n \rightarrow \infty$.

Assume that we are given an arbitrary sequence of codes that leads to the triple of rates (R_0, R_1, R_2). We will show that there exist α_1, α_2, and β such that $(R_0, R_1, R_2) \in C_G$ in (B.16). The proof follows similar steps as that of the spectral Gaussian broadcast channel [82, 83][1]. First define random vectors $\mathbf{X_1}$, $\mathbf{X_2}$, $\mathbf{Y_1}$, and $\mathbf{Y_2}$ as follows: $\mathbf{X_1} = (X_{1,1}, X_{1,2}, \cdots, X_{1,n-n^*})$ is a random vector consisting of the $(n-n^*)$ components of $\mathbf{X}(\mathbf{w}|\mathbf{V})$ corresponding to the case where the first AWGN broadcast channel in Figure B.1 is the current channel state; $\mathbf{X_2} = (X_{2,1}, X_{2,2}, \cdots, X_{2,n^*})$ is a random vector consisting of the other n^* components of $\mathbf{X}(\mathbf{w}|\mathbf{V})$. The random vectors $\mathbf{Y_1}$, $\mathbf{Y_2}$, $\mathbf{Z_1}$, and $\mathbf{Z_2}$ are similarly

[1]In particular, the derivation of equations (B.19), (B.26), (B.27), and (B.33) are identical to those in [82, 83]. We include them for completeness since [82] and [83] are not readily available.

defined. From Fano's inequality we have

$$H(W_0, W_1|\mathbf{Y}, \mathbf{V}) \leq P_{e,1}^n n(R_0 + R_1) + 1 \triangleq n\epsilon_{n,1},$$

$$H(W_0, W_2|\mathbf{Y}, \mathbf{V}) \leq P_{e,2}^n n(R_0 + R_2) + 1 \triangleq n\epsilon_{n,2}.$$

Thus,

$$\begin{aligned}
nR_0 &= H(W_0|\mathbf{V}) \\
&= I(W_0; \mathbf{Y_1}, \mathbf{Y_2}|\mathbf{V}) + H(W_0|\mathbf{Y_1}, \mathbf{Y_2}, \mathbf{V}) \\
&\leq I(W_0; \mathbf{Y_1}, \mathbf{Y_2}|\mathbf{V}) + n\epsilon_{n,1} \\
&= I(W_0; \mathbf{Y_2}|\mathbf{V}) + I(W_0; \mathbf{Y_1}|\mathbf{Y_2}, \mathbf{V}) + n\epsilon_{n,1} \\
&\leq I(W_0, W_1; \mathbf{Y_2}|\mathbf{V}) + I(W_0, W_2, \mathbf{Y_2}; \mathbf{Y_1}|\mathbf{V}) + n\epsilon_{n,1} \\
&\leq I(W_0, W_1, \mathbf{Z_1}; \mathbf{Y_2}|\mathbf{V}) + I(W_0, W_2, \mathbf{Y_2}; \mathbf{Y_1}|\mathbf{V}) + n\epsilon_{n,1} \\
&= H(\mathbf{Y_2}|\mathbf{V}) - H(\mathbf{Y_2}|W_0, W_1, \mathbf{Z_1}, \mathbf{V}) \\
&\quad + H(\mathbf{Y_1}|\mathbf{V}) - H(\mathbf{Y_1}|W_0, W_2, \mathbf{Y_2}, \mathbf{V}) + n\epsilon_{n,1}. \qquad \text{(B.19)}
\end{aligned}$$

Let

$$\bar{\beta}\bar{P} \triangleq \frac{1}{n}\sum_{i=1}^{n^*} E_{\mathbf{w}}\{|X_{2,i}(\mathbf{w})|^2\}.$$

Then since

$$\begin{aligned}
\bar{P} &\geq \frac{1}{n}\sum_{i=1}^{n} E_{\mathbf{w}}\{|X_i(\mathbf{w}|\mathbf{V})|^2\} \\
&= \frac{1}{n}\sum_{i=1}^{n-n^*} E_{\mathbf{w}}\{|X_{1,i}(\mathbf{w})|^2\} + \frac{1}{n}\sum_{i=1}^{n^*} E_{\mathbf{w}}\{|X_{2,i}(\mathbf{w})|^2\} \\
&= \frac{1}{n}\sum_{i=1}^{n-n^*} E_{\mathbf{w}}\{|X_{1,i}(\mathbf{w})|^2\} + \bar{\beta}\bar{P},
\end{aligned}$$

we obtain

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n-n^*} E_{\mathbf{w}}\{|X_{1,i}(\mathbf{w})|^2\} &\leq \bar{P} - \bar{\beta}\bar{P} \\
&\triangleq \beta\bar{P}.
\end{aligned}$$

Therefore,

$$H(\mathbf{Y_2}|\mathbf{V}) \leq n^* \cdot \frac{1}{2}\ln[2\pi e(\bar{\beta}\bar{P} + N_2)], \tag{B.20}$$

and

$$H(\mathbf{Y_1}|\mathbf{V}) \leq (n - n^*) \cdot \frac{1}{2}\ln[2\pi e(\beta\bar{P} + N_{11})]. \tag{B.21}$$

It is easy to see that there exists $\alpha_1, \alpha_2 \in [0, 1]$ for which

$$\exp\left[\frac{2}{n^*}H(\mathbf{Z_2}|W_0, W_1, \mathbf{Z_1}, \mathbf{V})\right] = 2\pi e(\bar{\alpha}_2\bar{\beta}\bar{P} + N_{21}), \tag{B.22}$$

$$\exp\left[\frac{2}{n - n^*}H(\mathbf{Y_1}|W_0, W_2, \mathbf{Y_2}, \mathbf{V})\right] = 2\pi e(\bar{\alpha}_1\beta\bar{P} + N_{11}). \tag{B.23}$$

Combining (B.22) and (B.23) with the conditional entropy inequalities [78]

$$H(\mathbf{Y_2}|W_0, W_1, \mathbf{Z_1}, \mathbf{V}) \geq \frac{n^*}{2}\ln\left\{\exp\left[\frac{2}{n^*}H(\mathbf{Z_2}|W_0, W_1, \mathbf{Z_1}, \mathbf{V})\right] + 2\pi e N_{22}\right\},$$

$$H(\mathbf{Z_1}|W_0, W_2, \mathbf{Y_2}, \mathbf{V}) \geq \frac{n - n^*}{2}\ln\left\{\exp\left[\frac{2}{n - n^*}H(\mathbf{Y_1}|W_0, W_2, \mathbf{Y_2}, \mathbf{V})\right] + 2\pi e N_{12}\right\},$$

we have

$$H(\mathbf{Y_2}|W_0, W_1, \mathbf{Z_1}, \mathbf{V}) \geq \frac{n^*}{2}\ln[2\pi e(\bar{\alpha}_2\bar{\beta}\bar{P} + N_2)], \tag{B.24}$$

$$H(\mathbf{Z_1}|W_0, W_2, \mathbf{Y_2}, \mathbf{V}) \geq \frac{n - n^*}{2}\ln[2\pi e(\bar{\alpha}_1\beta\bar{P} + N_1)]. \tag{B.25}$$

Substituting (B.20), (B.21), (B.23), and (B.24) into (B.19), dividing both sides by n and then taking the limit as $n \to \infty$ we obtain

$$R_0 \leq p_c C(\alpha_1\beta\bar{P}/(N_{11} + \bar{\alpha}_1\beta\bar{P})) + \bar{p}_c C(\alpha_2\bar{\beta}\bar{P}/(N_2 + \bar{\alpha}_2\bar{\beta}\bar{P})).$$

Similarly, we can show that

$$R_0 \leq p_c C(\alpha_1\beta\bar{P}/(N_1 + \bar{\alpha}_1\beta\bar{P})) + \bar{p}_c C(\alpha_2\bar{\beta}\bar{P}/(N_{21} + \bar{\alpha}_2\bar{\beta}\bar{P})).$$

Furthermore,

$$\begin{aligned} n(R_0 + R_1) &= H(W_0, W_1|\mathbf{V}) \\ &= I(W_0, W_1; \mathbf{Y_1}, \mathbf{Y_2}|\mathbf{V}) + H(W_0, W_1|\mathbf{Y_1}, \mathbf{Y_2}, \mathbf{V}) \end{aligned}$$

$$
\begin{aligned}
&\leq I(W_0, W_1; \mathbf{Y_1}, \mathbf{Y_2}|\mathbf{V}) + n\epsilon_{n,1} \\
&= I(W_0, W_1; \mathbf{Y_2}|\mathbf{V}) + I(W_0, W_1; \mathbf{Y_1}|\mathbf{Y_2}, \mathbf{V}) + n\epsilon_{n,1} \\
&\leq I(W_0, W_1, \mathbf{Z_1}; \mathbf{Y_2}|\mathbf{V}) + I(\mathbf{X_1}; \mathbf{Y_1}|\mathbf{V}) + n\epsilon_{n,1}, \qquad \text{(B.26)}
\end{aligned}
$$

where the last inequality follows from the fact that

$$
I(W_0, W_1; \mathbf{Y_2}|\mathbf{V}) \leq I(W_0, W_1, \mathbf{Z_1}; \mathbf{Y_2}|\mathbf{V})
$$

and

$$
\begin{aligned}
& I(W_0, W_1; \mathbf{Y_1}|\mathbf{Y_2}, \mathbf{V}) \\
\leq\ & I(W_0, W_1; \mathbf{Y_1}|\mathbf{Y_2}, \mathbf{V}) + I(W_2; \mathbf{Y_1}|W_0, W_1, \mathbf{Y_2}, \mathbf{V}) \\
=\ & H(\mathbf{Y_1}|\mathbf{Y_2}, \mathbf{V}) - H(\mathbf{Y_1}|W_0, W_1, \mathbf{Y_2}, \mathbf{V}) \\
& + H(\mathbf{Y_1}|W_0, W_1, \mathbf{Y_2}, \mathbf{V}) - H(\mathbf{Y_1}|W_0, W_1, W_2, \mathbf{Y_2}, \mathbf{V}) \\
=\ & H(\mathbf{Y_1}|\mathbf{Y_2}, \mathbf{V}) - H(\mathbf{Y_1}|W_0, W_1, W_2, \mathbf{Y_2}, \mathbf{V}) \\
\leq\ & H(\mathbf{Y_1}|\mathbf{V}) - H(\mathbf{Y_1}|\mathbf{X_1}, \mathbf{V}) \\
=\ & I(\mathbf{X_1}; \mathbf{Y_1}|\mathbf{V}). \qquad \text{(B.27)}
\end{aligned}
$$

Since

$$
\begin{aligned}
I(\mathbf{X_1}; \mathbf{Y_1}|\mathbf{V}) &= H(\mathbf{Y_1}|\mathbf{V}) - H(\mathbf{Y_1}|\mathbf{X_1}, \mathbf{V}) \\
&\leq \frac{n-n^*}{2}\ln[2\pi e(\beta\bar{P} + N_{11})] - \frac{n-n^*}{2}\ln[2\pi e N_{11}] \\
&= (n-n^*)C(\beta\bar{P}/N_{11}), \qquad \text{(B.28)}
\end{aligned}
$$

substituting (B.20), (B.24), and (B.28) into (B.26), we obtain

$$
n(R_0 + R_1) \leq (n-n^*)C(\beta\bar{P}/N_{11}) + n^*C(\alpha_2\bar{\beta}\bar{P}/(N_2 + \bar{\alpha}_2\bar{\beta}\bar{P})) + n\epsilon_{n,1}. \qquad \text{(B.29)}
$$

Thus, dividing both sides of (B.29) by n and taking the limit as $n \to \infty$, we have

$$
R_0 + R_1 \leq p_c C(\beta\bar{P}/N_{11}) + \bar{p}_c C(\alpha_2\bar{\beta}\bar{P}/(N_2 + \bar{\alpha}_2\bar{\beta}\bar{P})).
$$

Similarly, we can show that

$$\begin{aligned} n(R_0+R_2) &\leq I(W_0,W_2,\mathbf{Y_2};\mathbf{Z_1}|\mathbf{V})+I(\mathbf{X_2};\mathbf{Z_2}|\mathbf{V})+n\epsilon_{n,2} \\ &\leq (n-n^*)C(\alpha_1\beta\bar{P}/(N_1+\bar{\alpha}_1\beta\bar{P}))+n^*C(\bar{\beta}\bar{P}/N_{21})+n\epsilon_{n,2}, \quad \text{(B.30)} \end{aligned}$$

and consequently,

$$R_0+R_2 \leq p_c C(\alpha_1\beta\bar{P}/(N_1+\bar{\alpha}_1\beta\bar{P}))+\bar{p}_c C(\bar{\beta}\bar{P}/N_{21}).$$

Finally, let us bound the sum of rates from above. Assume that for some $\delta > 0$, we have either

$$R_0+R_1+R_2 > p_c C(\beta\bar{P}/N_{11})+\bar{p}_c[C(\alpha_2\bar{\beta}\bar{P}/(N_2+\bar{\alpha}_2\bar{\beta}\bar{P}))+C(\bar{\alpha}_2\bar{\beta}\bar{P}/N_{21})]+\delta, \quad \text{(B.31)}$$

or

$$R_0+R_1+R_2 > p_c[C(\alpha_1\beta\bar{P}/(N_1+\bar{\alpha}_1\beta\bar{P}))+C(\bar{\alpha}_1\beta\bar{P}/N_{11})]+\bar{p}_c C(\bar{\beta}\bar{P}/N_{21})+\delta. \quad \text{(B.32)}$$

we have

$$\begin{aligned} & n(R_0+R_1+R_2) \\ = \;& H(W_0,W_1|\mathbf{V})+H(W_2|\mathbf{V}) \\ = \;& I(W_0,W_1;\mathbf{Y_1},\mathbf{Y_2}|\mathbf{V})+H(W_0,W_1|\mathbf{Y_1},\mathbf{Y_2},\mathbf{V}) \\ & +I(W_2;\mathbf{Z_1},\mathbf{Z_2}|\mathbf{V})+H(W_2|\mathbf{Z_1},\mathbf{Z_2},\mathbf{V}) \\ \overset{a}{\leq} \;& I(W_0,W_1;\mathbf{Y_1},\mathbf{Y_2}|\mathbf{V})+I(W_2;\mathbf{Z_1},\mathbf{Z_2}|\mathbf{V})+n(\epsilon_{n,1}+\epsilon_{n,2}) \\ \overset{b}{\leq} \;& I(W_0,W_1;\mathbf{Y_1},\mathbf{Y_2}|\mathbf{V})+I(W_2;\mathbf{Z_1},\mathbf{Z_2}|W_0,W_1,\mathbf{V})+n(\epsilon_{n,1}+\epsilon_{n,2}) \\ = \;& I(W_0,W_1;\mathbf{Y_2}|\mathbf{V})+I(W_0,W_1;\mathbf{Y_1}|\mathbf{Y_2},\mathbf{V})+I(W_2;\mathbf{Z_1}|W_0,W_1,\mathbf{V}) \\ & +I(W_2;\mathbf{Z_2}|W_0,W_1,\mathbf{Z_1},\mathbf{V})+n(\epsilon_{n,1}+\epsilon_{n,2}) \\ = \;& [I(W_0,W_1,\mathbf{Z_1};\mathbf{Y_2}|\mathbf{V})-I(\mathbf{Z_1};\mathbf{Y_2}|W_0,W_1,\mathbf{V})]+I(W_0,W_1;\mathbf{Y_1}|\mathbf{Y_2},\mathbf{V}) \\ & +I(W_2;\mathbf{Z_1}|W_0,W_1,\mathbf{V})+I(W_2;\mathbf{Z_2}|W_0,W_1,\mathbf{Z_1},\mathbf{V})+n(\epsilon_{n,1}+\epsilon_{n,2}) \\ \overset{c}{\leq} \;& I(W_0,W_1,\mathbf{Z_1};\mathbf{Y_2}|\mathbf{V})-I(\mathbf{Z_1};\mathbf{Y_2}|W_0,W_1,\mathbf{V})+I(W_0,W_1;\mathbf{Y_1}|\mathbf{Y_2},\mathbf{V}) \\ & +I(W_2,\mathbf{Y_2};\mathbf{Z_1}|W_0,W_1,\mathbf{V})+I(W_2;\mathbf{Z_2}|W_0,W_1,\mathbf{Z_1},\mathbf{V})+n(\epsilon_{n,1}+\epsilon_{n,2}) \end{aligned}$$

$$
\begin{aligned}
&= I(W_0, W_1, \mathbf{Z_1}; \mathbf{Y_2}|\mathbf{V}) + I(W_2; \mathbf{Z_2}|W_0, W_1, \mathbf{Z_1}, \mathbf{V}) + I(W_0, W_1; \mathbf{Y_1}|\mathbf{Y_2}, \mathbf{V}) \\
&\quad + [I(W_2, \mathbf{Y_2}; \mathbf{Z_1}|W_0, W_1, \mathbf{V}) - I(\mathbf{Z_1}; \mathbf{Y_2}|W_0, W_1, \mathbf{V})] + n(\epsilon_{n,1} + \epsilon_{n,2}) \\
&= I(W_0, W_1, \mathbf{Z_1}; \mathbf{Y_2}|\mathbf{V}) + I(W_2; \mathbf{Z_2}|W_0, W_1, \mathbf{Z_1}, \mathbf{V}) + I(W_0, W_1; \mathbf{Y_1}|\mathbf{Y_2}, \mathbf{V}) \\
&\quad + I(W_2; \mathbf{Z_1}|W_0, W_1, \mathbf{Y_2}, \mathbf{V}) + n(\epsilon_{n,1} + \epsilon_{n,2}) \\
&\stackrel{d}{\leq} I(W_0, W_1, \mathbf{Z_1}; \mathbf{Y_2}|\mathbf{V}) + I(W_2; \mathbf{Z_2}|W_0, W_1, \mathbf{Z_1}, \mathbf{V}) \\
&\quad + [I(W_0, W_1; \mathbf{Y_1}|\mathbf{Y_2}, \mathbf{V}) + I(W_2; \mathbf{Y_1}|W_0, W_1, \mathbf{Y_2}, \mathbf{V})] + n(\epsilon_{n,1} + \epsilon_{n,2}) \\
&\stackrel{e}{\leq} I(W_0, W_1, \mathbf{Z_1}; \mathbf{Y_2}|\mathbf{V}) + I(W_2; \mathbf{Z_2}|W_0, W_1, \mathbf{Z_1}, \mathbf{V}) \\
&\quad + I(\mathbf{X_1}; \mathbf{Y_1}|\mathbf{V}) + n(\epsilon_{n,1} + \epsilon_{n,2}),
\end{aligned}
\tag{B.33}
$$

where a follows from Fano's inequality; b is due to the independence of W_0, W_1, and W_2 so that

$$I(W_2; \mathbf{Z_1}, \mathbf{Z_2}|\mathbf{V}) \leq I(W_2; \mathbf{Z_1}, \mathbf{Z_2}|W_0, W_1, \mathbf{V});$$

c is because

$$I(W_2; \mathbf{Z_1}|W_0, W_1, \mathbf{V}) \leq I(W_2, \mathbf{Y_2}; \mathbf{Z_1}|W_0, W_1, \mathbf{V});$$

d is a result of the degraded nature of the channel so that

$$I(W_2; \mathbf{Z_1}|W_0, W_1, \mathbf{Y_2}, \mathbf{V}) \leq I(W_2; \mathbf{Y_1}|W_0, W_1, \mathbf{Y_2}, \mathbf{V});$$

and e follows from (B.27). Similarly, we can show that

$$
\begin{aligned}
& n(R_0 + R_1 + R_2) \\
\leq\ & I(W_0, W_2, \mathbf{Y_2}; \mathbf{Z_1}|\mathbf{V}) + I(W_1; \mathbf{Y_1}|W_0, W_2, \mathbf{Y_2}, \mathbf{V}) \\
& + I(\mathbf{X_2}; \mathbf{Z_2}|\mathbf{V}) + n(\epsilon_{n,1} + \epsilon_{n,2}).
\end{aligned}
\tag{B.34}
$$

Combining (B.26), (B.29), and (B.33) we obtain

$$
\begin{aligned}
& n(R_0 + R_1 + R_2) \\
\leq\ & (n - n^*)C(\beta\bar{P}/N_{11}) + n^*C(\alpha_2\bar{\beta}\bar{P}/(N_2 + \bar{\alpha}_2\bar{\beta}\bar{P})) \\
& + I(W_2; \mathbf{Z_2}|W_0, W_1, \mathbf{Z_1}, \mathbf{V}) + n(\epsilon_{n,1} + \epsilon_{n,2}) \\
=\ & (n - n^*)C(\beta\bar{P}/N_{11}) + n^*C(\alpha_2\bar{\beta}\bar{P}/(N_2 + \bar{\alpha}_2\bar{\beta}\bar{P}))
\end{aligned}
$$

$$+ H(\mathbf{Z_2}|W_0, W_1, \mathbf{Z_1}, \mathbf{V}) - H(\mathbf{Z_2}|W_0, W_1, W_2, \mathbf{Z_1}, \mathbf{V}) + n(\epsilon_{n,1} + \epsilon_{n,2})$$
$$= (n - n^*)C(\beta\bar{P}/N_{11}) + n^*C(\alpha_2\bar{\beta}\bar{P}/(N_2 + \bar{\alpha}_2\bar{\beta}\bar{P}))$$
$$+ H(\mathbf{Z_2}|W_0, W_1, \mathbf{Z_1}, \mathbf{V}) - \frac{n^*}{2}\ln[2\pi e N_{21}] + n(\epsilon_{n,1} + \epsilon_{n,2}).$$

Now if we assume that inequality (B.31) is satisfied, we obtain

$$\lim_{n\to\infty} \frac{1}{n} H(\mathbf{Z_2}|W_0, W_1, \mathbf{Z_1}, \mathbf{V}) > \frac{\bar{p}_c}{2}\ln[2\pi e(N_{21} + \bar{\alpha}_2\bar{\beta}\bar{P})] + \delta.$$

However, this contradicts (B.22) as $n \to \infty$. Similarly, assuming that inequality (B.32) is satisfied, we obtain from (B.30) and (B.34) that

$$\lim_{n\to\infty} \frac{1}{n} H(\mathbf{Y_1}|W_0, W_2, \mathbf{Y_2}, \mathbf{V}) > \frac{p_c}{2}\ln[2\pi e(N_{11} + \bar{\alpha}_1\beta\bar{P})] + \delta,$$

and this contradicts (B.23) as $n \to \infty$. Thus the triple (R_0, R_1, R_2) could satisfy the six inequalities in *Theorem B.1* for some α_1, α_2, and β, i.e., $(R_0, R_1, R_2) \in C_G$. □

B.2 Interpretation of the "Water-Filling" Procedure for CD with Theorem 4.2

In the two-step water-filling power allocation procedure described in Section 4.3.1, we assume that

$$n_{\pi(1)} < n_{\pi(2)} < \cdots < n_{\pi(M)}.$$

For any user $\pi(i)$ for which $\exists j$ such that $i < j \leq M$ and $\frac{\mu_{\pi(i)}}{n_{\pi(i)}} \leq \frac{\mu_{\pi(j)}}{n_{\pi(j)}}$, we know that $\mu_{\pi(i)} < \mu_{\pi(j)}$. Thus, $\forall z > 0$, $u_{\pi(i)}(z) < u_{\pi(j)}(z)$ and according to *Theorem 4.2*, no power should be assigned to these users $\pi(i)$ and we will remove them from further consideration.

After this initialization step, for the remaining K users, since $\forall 1 < i \leq K$, $n_{\rho(1)} < n_{\rho(i)}$ and $\frac{\mu_{\rho(1)}}{n_{\rho(1)}} > \frac{\mu_{\rho(i)}}{n_{\rho(i)}}$, for the users $\rho(i)$ with $\mu_{\rho(i)} \leq \mu_{\rho(1)}$, it is clear that $\forall z > 0$, $u_{\rho(i)}(z) < u_{\rho(1)}(z)$ and we do not assign power to these users. For other users $\rho(i)$ with $\mu_{\rho(i)} > \mu_{\rho(1)}$, $u_{\rho(1)}(z)$ and $u_{\rho(i)}$ will cross each other once at

$$z_{c_i} = \frac{\mu_{\rho(1)} n_{\rho(i)} B - \mu_{\rho(i)} n_{\rho(1)} B}{\mu_{\rho(i)} - \mu_{\rho(1)}}.$$

Therefore, $u^*(z) = u_{\rho(1)}(z)$ if and only if $z \in [0, z_c]$, where

$$z_c \triangleq \min_{\mu_{\rho(i)} > \mu_{\rho(1)}} \{z_{c_i}\}.$$

Note that $u_{\rho(1)}(z)$ equals zero at $z = \frac{\mu_{\rho(1)}}{\lambda} - n_{\rho(1)}B$ and, as shown in the proof of *Theorem 4.2*, $u^*(\cdot)$ and $u_{\rho(i)}(\cdot)$ ($\forall 1 \leq i \leq K$) are all decreasing functions. Thus, according to *Theorem 4.2*, if

$$P^*_{\rho(1)}(\mathbf{n}) = \begin{cases} z_c, & \text{if } u_{\rho(1)}(z_c) > 0, \text{ i.e., } z_c < \left[\frac{\mu_{\rho(1)}}{\lambda} - n_{\rho(1)}B\right]_+, \\ \left[\frac{\mu_{\rho(1)}}{\lambda} - n_{\rho(1)}B\right]_+, & \text{otherwise.} \end{cases}$$

That is, $P^*_{\rho(1)}(\mathbf{n})$ is as given in (4.7). Moreover, if

$$P^*_{\rho(1)}(\mathbf{n}) = \left[\frac{\mu_{\rho(1)}}{\lambda} - n_{\rho(1)}B\right]_+, \quad \text{i.e., } z_c \geq \left[\frac{\mu_{\rho(1)}}{\lambda} - n_{\rho(1)}B\right]_+,$$

then

$$u^*(z) < 0 \quad \text{for all } z > \left[\frac{\mu_{\rho(1)}}{\lambda} - n_{\rho(1)}B\right]_+,$$

which means that no more power should be allocated to any user in the state $\mathbf{n}$. Otherwise if $P^*_{\rho(1)}(\mathbf{n}) = z_c$, then at least for $z_c < z < \left[\frac{\mu_{\rho(1)}}{\lambda} - n_{\rho(1)}B\right]_+$, $u^*(z) > 0$ and more power should be allocated to state $\mathbf{n}$. However, in this case, since for all $z > z_c$, $u^*(z) > u_{\rho(1)}(z)$, the power for User $\rho(1)$ in state $\mathbf{n}$ has been allocated and $P^*_{\rho(1)}(\mathbf{n})$ becomes the interfering power for each of the remaining $K-1$ users $\{\rho(i)\}_{i=2}^{K}$. Now our maximization problem is reduced to decide which one of the $K-1$ users should be chosen to transmit to at an interference level $z = P^*_{\rho(1)}(\mathbf{n})$. This new problem is equivalent to assigning power for the $K-1$ users with their corresponding noise levels $n_{\rho(i)}B$ all increased by $P^*_{\rho(1)}(\mathbf{n})$. Therefore, the iterative water-filling power allocation procedure follows.

B.3 Proof of Theorem 4.3

Achievability of the Rate Region: The proof of the achievability follows along the same lines as that for the capacity region of CD given in Appendix B.1 and is therefore omitted.

Convexity of the Rate Region: $\forall \tau > 0$, $\alpha > 0$, $n > 0$, let

$$R(\tau, \alpha) \triangleq \tau \ln\left(1 + \frac{\alpha}{\tau\ n}\right).$$

Since

$$\frac{\partial^2 R}{\partial \tau^2} = -\frac{\alpha^2/\tau}{(\tau\ n + \alpha)^2} < 0, \tag{B.35}$$

$$\frac{\partial^2 R}{\partial \alpha^2} = -\frac{\tau}{(\tau\ n + \alpha)^2} < 0, \tag{B.36}$$

$$\frac{\partial^2 R}{\partial \tau \partial \alpha} = \frac{\partial^2 R}{\partial \alpha \partial \tau} = \frac{\alpha}{(\tau\ n + \alpha)^2}, \tag{B.37}$$

from (B.35)-(B.37), we obtain the Hessian matrix

$$\begin{vmatrix} \frac{\partial^2 R}{\partial \tau^2} & \frac{\partial^2 R}{\partial \tau \partial \alpha} \\ \frac{\partial^2 R}{\partial \alpha \partial \tau} & \frac{\partial^2 R}{\partial \alpha^2} \end{vmatrix} = 0.$$

Therefore, $R(\tau, \alpha)$ is a concave function of (τ, α). That is, $\forall 0 \le \lambda \le 1$, $\forall \tau_i > 0$, $\alpha_i > 0$, $i = 1, 2$,

$$R(\tau_0, \alpha_0) > \lambda R(\tau_1, \alpha_1) + (1 - \lambda) R(\tau_2, \alpha_2), \tag{B.38}$$

where $\tau_0 = \lambda \tau_1 + (1 - \lambda)\tau_2$, $\alpha_0 = \lambda \alpha_1 + (1 - \lambda)\alpha_2$. This result will be used in the following to prove the convexity of the capacity region.

$\forall \mathbf{R}, \mathbf{R}' \in \mathcal{C}(\bar{P})$, $0 \le \lambda \le 1$, we need to show that $\lambda \mathbf{R} + (1 - \lambda)\mathbf{R}' \in \mathcal{C}(\bar{P})$. Let $\mathcal{P}$ and $\mathcal{P}'$ be the two power policies corresponding to the rates $\mathbf{R}$ and $\mathbf{R}'$, respectively. In a given channel state $\mathbf{n}$, according to the two policies, the transmit power and fractions of transmission time allocated to User j ($j = 1, 2, \ldots, M$) are $P_j(\mathbf{n})$ and $P'_j(\mathbf{n})$, $\tau_j(\mathbf{n})$ and $\tau'_j(\mathbf{n})$, respectively. Therefore, for the two power policies, the achievable rates for User j in the state $\mathbf{n}$ are

$$R_j(\mathbf{n})[\tau_j(\mathbf{n}), P_j(\mathbf{n})] = \tau_j(\mathbf{n}) B \log\left(1 + \frac{P_j(\mathbf{n})}{n_j B}\right), \tag{B.39}$$

and

$$R_j(\mathbf{n})[\tau'_j(\mathbf{n}), P'_j(\mathbf{n})] = \tau'_j(\mathbf{n}) B \log\left(1 + \frac{P'_j(\mathbf{n})}{n_j B}\right), \tag{B.40}$$

respectively. (B.39) and (B.40) can also be expressed as [84]:

$$R_j(\mathbf{n})[\tau_j(\mathbf{n}), A_j(\mathbf{n})] = \tau_j(\mathbf{n}) B \log\left(1 + \frac{A_j(\mathbf{n})}{\tau_j(\mathbf{n}) n_j B}\right),$$

$$R_j(\mathbf{n})[\tau_j'(\mathbf{n}), A_j'(\mathbf{n})] = \tau_j'(\mathbf{n}) B \log\left(1 + \frac{A_j'(\mathbf{n})}{\tau_j'(\mathbf{n}) n_j B}\right),$$

where $A_j(\mathbf{n}) = P_j(\mathbf{n})\tau_j(\mathbf{n})$ and $A_j'(\mathbf{n}) = P_j'(\mathbf{n})\tau_j'(\mathbf{n})$, $j = 1, 2, \cdots, M$.

If we define a third power policy $\mathcal{P}''$ such that in each channel state $\mathbf{n}$, the transmit power $P_j''(\mathbf{n})$ and fraction of transmission time $\tau_j''(\mathbf{n})$ allocated to User j ($j = 1, 2, \ldots, M$) are

$$P_j''(\mathbf{n}) = \frac{\lambda\tau_j(\mathbf{n})P_j(\mathbf{n}) + (1-\lambda)\tau_j'(\mathbf{n})P_j'(\mathbf{n})}{\lambda\tau_j(\mathbf{n}) + (1-\lambda)\tau_j'(\mathbf{n})},$$

and

$$\tau_j''(\mathbf{n}) = \lambda\tau_j(\mathbf{n}) + (1-\lambda)\tau_j'(\mathbf{n}),$$

respectively, then

$$\begin{aligned} A_j''(\mathbf{n}) &\triangleq P_j''(\mathbf{n})\tau_j''(\mathbf{n}) \\ &= \lambda\tau_j(\mathbf{n})P_j(\mathbf{n}) + (1-\lambda)\tau_j'(\mathbf{n})P_j'(\mathbf{n}) \\ &= \lambda A_j(\mathbf{n}) + (1-\lambda)A_j'(\mathbf{n}). \end{aligned}$$

Since

$$\begin{aligned} E_{\mathbf{n}}\left[\sum_{j=1}^{M} P_j''(\mathbf{n})\tau_j''(\mathbf{n})\right] &= E_{\mathbf{n}}\left[\sum_{j=1}^{M} \lambda\tau_j(\mathbf{n})P_j(\mathbf{n}) + (1-\lambda)\tau_j'(\mathbf{n})P_j'(\mathbf{n})\right] \\ &\leq \bar{P}, \end{aligned}$$

we know that $\mathcal{P}'' \in \mathcal{C}(\bar{P})$. Moreover, it is proved in (B.38) that for $j = 1, 2, \ldots, M$,

$$R_j(\mathbf{n})[\tau, \alpha] \triangleq \tau B \log\left(1 + \frac{\alpha}{\tau\, n_j B}\right)$$

is a concave function of (τ, α), i.e., $\forall \mathbf{n} \in \mathcal{N}$,

$$R_j(\mathbf{n})[\tau_j''(\mathbf{n}), A_j''(\mathbf{n})] > \lambda R_j(\mathbf{n})[\tau_j(\mathbf{n}), A_j(\mathbf{n})] + (1-\lambda)R_j(\mathbf{n})[\tau_j'(\mathbf{n}), A_j'(\mathbf{n})].$$

Therefore, $\lambda\mathbf{R} + (1-\lambda)\mathbf{R}' \in \mathcal{C}(\bar{P})$. $\square$

B.4 Proof of Lemma 4.1

In the following, P, P_j, τ_j ($j = 1, 2$), P_a, P_b, P_0, λ_0 and τ are all functions of $\mathbf{n}$, which is not shown for simplicity.

Let $\tau_1 = \tau$ and $\tau_2 = 1 - \tau$. Define

$$f_0(P) = f_2(P) - f_1(P),$$

where $f_j(x)$ ($j = 1, 2$) is given in (4.22), then

$$\begin{aligned} f_0'(P) &= \frac{\mu_2}{P + n_2 B} - \frac{\mu_1}{P + n_1 B} \\ &= \frac{(\mu_2 - \mu_1)P + (\mu_2 n_1 B - \mu_1 n_2 B)}{(P + n_1 B)(P + n_2 B)}. \end{aligned} \tag{B.41}$$

1. If $\frac{\mu_1}{n_1} \le \frac{\mu_2}{n_2}$, then $\mu_2 n_1 B \ge \mu_1 n_2 B$. Since $\mu_1 < \mu_2$ by assumption, the numerator of (B.41) is non-negative for any $P > 0$. Thus, $\forall P > 0$, $f_0'(P) > 0$. Therefore, if $P > 0$, $f_0(P) > f_0(0) = 0$, i.e., $f_2(P) > f_1(P)$. For $P_1 \ge 0$, $P_2 \ge 0$ and $\tau P_1 + (1 - \tau)P_2 = P$, we have

$$\begin{aligned} \tau f_1(P_1) + (1 - \tau) f_2(P_2) &\le \tau f_2(P_1) + (1 - \tau) f_2(P_2) \\ &\le f_2(P). \end{aligned} \tag{B.42}$$

The last inequality in (B.42) is due to the concavity of $f_2(P)$ and the equality is achieved when $\tau = 0, P_1 = 0$ and $P_2 = P$. Thus, in this case, the solution to (4.21) is $J(P) = f_2(P)$.

2. If $\frac{\mu_1}{n_1} > \frac{\mu_2}{n_2}$, from (B.41) we know that

$$\begin{cases} f_0'(P) < 0 & \text{if } 0 \le P < \frac{\mu_1 n_2 B - \mu_2 n_1 B}{\mu_2 - \mu_1}, \\ f_0'(P) > 0 & \text{if } P > \frac{\mu_1 n_2 B - \mu_2 n_1 B}{\mu_2 - \mu_1}. \end{cases} \tag{B.43}$$

Since $\mu_1 < \mu_2$, for P large enough, $(1 + \frac{P}{n_1 B})^{\mu_1} < (1 + \frac{P}{n_2 B})^{\mu_2}$. Thus, for large P, $f_1(P) < f_2(P)$, i.e., $f_0(P) > 0$. However, $f_0(0) = 0$. Using this and (B.43), it is clear that $f_1(P)$ and $f_2(P)$ will cross each other once at some positive value P_0, as shown in Figure 4.5. In this figure, P_a and P_b are the points that satisfy (4.30), i.e.,

$$f_1'(P_a) = f_2'(P_b) = \lambda_0,$$

where λ_0 is the slope of the common tangent line in the figure. Since

$$f_1'(P_a) = \frac{\mu_1}{P_a + n_1 B} = \lambda_0,$$

$$f_2'(P_b) = \frac{\mu_2}{P_b + n_2 B} = \lambda_0,$$

and λ_0 can also be expressed as

$$\lambda_0 = \frac{f_2(P_b) - f_1(P_a)}{P_b - P_a},$$

we have

$$P_a = \frac{\mu_1}{\lambda_0} - n_1 B, \tag{B.44}$$

$$P_b = \frac{\mu_2}{\lambda_0} - n_2 B, \tag{B.45}$$

and λ_0 satisfies

$$\lambda_0 = \frac{\mu_2 \ln(\frac{\mu_2}{\lambda_0 n_2 B}) - \mu_1 \ln(\frac{\mu_1}{\lambda_0 n_1 B})}{(\frac{\mu_2}{\lambda_0} - n_2 B) - (\frac{\mu_1}{\lambda_0} - n_1 B)}, \quad \text{i.e., } h(\lambda_0, \mathbf{n}) = 0, \tag{B.46}$$

where $h(x, \mathbf{n})$ is given in (4.24). Therefore, if $P_a < P < P_b$, by time-sharing, $J(P)$ in (4.21) can achieve the values between $f_1(P_a)$ and $f_2(P_b)$ on the straight line; if $0 < P \leq P_a$ or $P \geq P_b$, $J(P)$ is simply $f_1(P)$ or $f_2(P)$, respectively. That is, in this case, the solution to (4.21) is (4.25). □

B.5 Proof of Theorem 4.4

For a given fading state $\mathbf{n}$, from (4.23) we know that the optimal power $P^*(\mathbf{n})$ satisfies (4.31). Let $P_j^*(\mathbf{n})$ and $\tau_j^*(\mathbf{n})$ be the optimal power and fraction of transmission time allocated to User j $(j = 1, 2)$ at state $\mathbf{n}$, respectively. From Lemma 4.1, we know that

1. if $\frac{\mu_1}{n_1} \leq \frac{\mu_2}{n_2}$, then $J(P^*(\mathbf{n})) = f_2(P^*(\mathbf{n}))$. Thus,

$$J'(P^*(\mathbf{n})) = \frac{\mu_2}{P^*(\mathbf{n}) + n_2 B} \tag{B.47}$$

and $\tau^*(\mathbf{n}) = 0$, $P_1^*(\mathbf{n}) = 0$, $P_2^*(\mathbf{n}) = P^*(\mathbf{n})$. Substituting (B.47) into (4.31), we have

$$P_2^*(\mathbf{n}) = P^*(\mathbf{n}) = \frac{\mu_2}{\lambda} - n_2 B.$$

Since $P_2^*(\mathbf{n})$ must be non-negative,

$$P_2^*(\mathbf{n}) = \left[\frac{\mu_2}{\lambda} - n_2 B\right]_+ ;$$

2. if $\frac{\mu_1}{n_1} > \frac{\mu_2}{n_2}$, as shown in Figure 4.5, it is clear that

$$\begin{cases} 0 < P^*(\mathbf{n}) \le P_a & \text{if } J'(P^*(\mathbf{n})) > \lambda_0 \ , \\ P_a < P^*(\mathbf{n}) < P_b & \text{if } J'(P^*(\mathbf{n})) = \lambda_0 \ , \\ P^*(\mathbf{n}) \ge P_b & \text{if } J'(P^*(\mathbf{n})) < \lambda_0 \ , \end{cases} \tag{B.48}$$

where P_a, P_b and λ_0 are all functions of $\mathbf{n}$ and they are given in (B.44), (B.45) and (B.46), respectively. Since we know from (4.31) that $J'(P^*(\mathbf{n})) = \lambda$, from (4.25), (4.27), (4.28), (4.29) and (B.48) we have:

a) if $\lambda > \lambda_0$ then $J(P^*(\mathbf{n})) = f_1(P^*(\mathbf{n}))$ and $f_1'(P^*(\mathbf{n})) = \lambda$. Thus, $P^*(\mathbf{n}) = [\frac{\mu_1}{\lambda} - n_1 B]_+$. Consequently,

$$\begin{cases} \tau_1^*(\mathbf{n}) = 1, \\ \tau_2^*(\mathbf{n}) = 0, \end{cases} \text{and} \quad \begin{cases} P_1^*(\mathbf{n}) = [\frac{\mu_1}{\lambda} - n_1 B]_+ , \\ P_2^*(\mathbf{n}) = 0; \end{cases}$$

b) if $\lambda < \lambda_0$ then $J(P^*(\mathbf{n})) = f_2(P^*(\mathbf{n}))$ and $f_2'(P^*(\mathbf{n})) = \lambda$. Thus, $P^*(\mathbf{n}) = [\frac{\mu_2}{\lambda} - n_2 B]_+$. Consequently,

$$\begin{cases} \tau_1^*(\mathbf{n}) = 0, \\ \tau_2^*(\mathbf{n}) = 1, \end{cases} \text{and} \quad \begin{cases} P_1^*(\mathbf{n}) = 0, \\ P_2^*(\mathbf{n}) = [\frac{\mu_2}{\lambda} - n_2 B]_+ ; \end{cases}$$

c) if $\lambda = \lambda_0$, then $J(P^*(\mathbf{n})) = f_1(Pa) + \lambda_0[P^*(\mathbf{n}) - P_a]$ and $P^*(\mathbf{n})$ can be any value between P_a and P_b. Thus,

$$\begin{cases} \tau_1^*(\mathbf{n}) = \tau_0^*, \\ \tau_2^*(\mathbf{n}) = 1 - \tau_0^*, \end{cases} \text{and} \quad \begin{cases} P_1^*(\mathbf{n}) = [\frac{\mu_1}{\lambda} - n_1 B]_+ , \\ P_2^*(\mathbf{n}) = [\frac{\mu_2}{\lambda} - n_2 B]_+ , \end{cases}$$

where τ_0^* can be any value between 0 and 1. λ and τ_0^* satisfy the total average power constraint (4.32).

Because $P_a < P_b$, from (B.44) and (B.45), we have

$$\lambda_0 < \frac{\mu_2 - \mu_1}{n_2 B - n_1 B}. \tag{B.49}$$

Furthermore, since $h(\lambda_0, \mathbf{n}) = 0$ and

$$\begin{cases} h'(x, \mathbf{n}) > 0 & \text{if } x > \frac{\mu_2 - \mu_1}{n_2 B - n_1 B}, \\ h'(x, \mathbf{n}) < 0 & \text{if } 0 < x < \frac{\mu_2 - \mu_1}{n_2 B - n_1 B}, \end{cases}$$

where $h'(x, \mathbf{n}) = (n_2 B - n_1 B) - (\mu_2 - \mu_1)/x$, we have:

A) if $\lambda \geq \frac{\mu_2 - \mu_1}{n_2 B - n_1 B}$, or if $\lambda < \frac{\mu_2 - \mu_1}{n_2 B - n_1 B}$ and $h(\lambda, \mathbf{n}) < 0$, then $\lambda > \lambda_0$;

B) if $\lambda < \frac{\mu_2 - \mu_1}{n_2 B - n_1 B}$ and $h(\lambda, \mathbf{n}) > 0$, then $\lambda < \lambda_0$;

C) if $\lambda < \frac{\mu_2 - \mu_1}{n_2 B - n_1 B}$ and $h(\lambda, \mathbf{n}) = 0$, then $\lambda = \lambda_0$.

Therefore, the second part of Theorem 4.4 is proved by combining a), b) and c) with A), B) and C). □

B.6 Decision Region Comparison for Optimal and Sub-Optimal TD Schemes

For the sub-optimal TD policy, define

$$g(x, \mathbf{n}) \triangleq g_2(x, \mathbf{n}) - g_1(x, \mathbf{n}),$$

where $g_i(x, \mathbf{n})$ $(i = 1, 2)$ is given in (4.42). Let λ_0^{sub} satisfy

$$g(\lambda_0^{sub}, \mathbf{n}) = 0,$$

then

$$\lambda_0^{sub} = \left(\frac{\mu_1}{n_1 B}\right)\left(\frac{\mu_2}{n_2 B} / \frac{\mu_1}{n_1 B}\right)^{\frac{\mu_2}{\mu_2 - \mu_1}} \tag{B.50}$$

$$= \left(\frac{\mu_2}{n_2 B}\right)\left(\frac{\mu_2}{n_2 B} / \frac{\mu_1}{n_1 B}\right)^{\frac{\mu_1}{\mu_2 - \mu_1}}. \tag{B.51}$$

Assuming that $\mu_1 < \mu_2$, from (B.50) and (B.51) we know that:

1. if $\frac{\mu_1}{n_1} \leq \frac{\mu_2}{n_2}$, then

$$\lambda_0^{sub} \geq \frac{\mu_1}{n_1 B} \quad \text{and} \quad \lambda_0^{sub} \geq \frac{\mu_2}{n_2 B};$$

2. if $\frac{\mu_1}{n_1} > \frac{\mu_2}{n_2}$, then

$$\lambda_0^{sub} < \frac{\mu_1}{n_1 B} \quad \text{and} \quad \lambda_0^{sub} < \frac{\mu_2}{n_2 B}.$$

Since for $x > 0$, $g'(x, \mathbf{n}) = \frac{\mu_1 - \mu_2}{x} < 0$, where $g'(x, \mathbf{n})$ denotes the derivative of $g(x, \mathbf{n})$ with respect to x, $g(x, \mathbf{n})$ is a decreasing function. Thus, when the c.d.f. $F(\mathbf{n})$ is continuous, the sub-optimal policy is equivalent to:

a) if $\frac{\mu_1}{n_1} \leq \frac{\mu_2}{n_2}$ and $\lambda < \frac{\mu_2}{n_2 B}$ or if $\frac{\mu_1}{n_1} > \frac{\mu_2}{n_2}$ and $\lambda < \lambda_0^{sub}$, then

$$\begin{cases} \tau_1^*(\mathbf{n}) = 0, \\ \tau_2^*(\mathbf{n}) = 1, \end{cases} \quad \text{and} \quad \begin{cases} P_1^*(\mathbf{n}) = 0, \\ P_2^*(\mathbf{n}) = \left[\frac{\mu_2}{\lambda} - n_2 B\right]_+; \end{cases}$$

b) if $\frac{\mu_1}{n_1} > \frac{\mu_2}{n_2}$ and $\lambda_0^{sub} < \lambda < \frac{\mu_1}{n_1 B}$, then

$$\begin{cases} \tau_1^*(\mathbf{n}) = 1, \\ \tau_2^*(\mathbf{n}) = 0, \end{cases} \quad \text{and} \quad \begin{cases} P_1^*(\mathbf{n}) = \left[\frac{\mu_1}{\lambda} - n_1 B\right]_+, \\ P_2^*(\mathbf{n}) = 0; \end{cases}$$

From Theorem 4.4, it is clear that the optimal policy is equivalent to:

a) if $\frac{\mu_1}{n_1} \leq \frac{\mu_2}{n_2}$ and $\lambda < \frac{\mu_2}{n_2 B}$ or if $\frac{\mu_1}{n_1} > \frac{\mu_2}{n_2}$ and $\lambda < \lambda_0$, then

$$\begin{cases} \tau_1^*(\mathbf{n}) = 0, \\ \tau_2^*(\mathbf{n}) = 1, \end{cases} \quad \text{and} \quad \begin{cases} P_1^*(\mathbf{n}) = 0, \\ P_2^*(\mathbf{n}) = \left[\frac{\mu_2}{\lambda} - n_2 B\right]_+; \end{cases}$$

b) if $\frac{\mu_1}{n_1} > \frac{\mu_2}{n_2}$ and $\lambda_0 < \lambda < \frac{\mu_1}{n_1 B}$, then

$$\begin{cases} \tau_1^*(\mathbf{n}) = 1, \\ \tau_2^*(\mathbf{n}) = 0, \end{cases} \quad \text{and} \quad \begin{cases} P_1^*(\mathbf{n}) = \left[\frac{\mu_1}{\lambda} - n_1 B\right]_+, \\ P_2^*(\mathbf{n}) = 0; \end{cases}$$

Since if $\frac{\mu_2}{n_2} < \frac{\mu_1}{n_1}$, $h(\lambda_0, \mathbf{n}) = 0$. From the definition of $h(x, \mathbf{n})$ in (4.24), we have

$$h(\lambda_0, \mathbf{n}) = g(\lambda_0, \mathbf{n}) + (n_2 B - n_1 B)\lambda_0 - (\mu_2 - \mu_1) = 0. \tag{B.52}$$

According to (B.49) and (B.52), it is clear that $g(\lambda_0, \mathbf{n}) > 0$. Consequently, $\lambda_0 < \lambda_0^{sub}$. Therefore, for a given $\lambda > 0$, the only difference between the optimal and sub-optimal policy is that when $\mathbf{n} \in L_D$, where L_D is given in (4.43), the sub-optimal scheme transmits the information of User 2, while the optimal scheme transmits the information of User 1. This sub-optimal allocation of resources occurs only rarely. Note that the values of λ in the two schemes satisfying the two-user power constraint (4.32) are not the same, though they may be very close to each other. □

B.7 Proof of Lemma 4.2

For simplicity, we denote

$$N_i \triangleq n_i B, \quad i = 1, 2, \cdots, M.$$

When $M = 2$, let $P \triangleq P_1 + P_2$, then $P_2 = P - P_1$, $0 \le P_1 \le P$. Let

$$q(P_1) \triangleq \mu_1 \ln\left(1 + \frac{P_1}{N_1 + P - P_1}\right) + \mu_2 \ln\left(1 + \frac{P - P_1}{N_2 + P_1}\right),$$

then

$$\begin{aligned} q'(P_1) &= \frac{\mu_1}{N_1 + P - P_1} - \frac{\mu_2}{N_2 + P_1} \\ &= \frac{(\mu_1 + \mu_2)P_1 + [\mu_1 N_2 - \mu_2(N_1 + P)]}{(N_1 + P - P_1)(N_2 + P_1)}. \end{aligned}$$

If $\mu_1 N_2 \ge \mu_2(N_1 + P)$, then for $0 \le P_1 \le P$, we have $q'(P_1) \ge 0$ and

$$q(P_1) \le q(P) = \mu_1 \ln(1 + P/N_1); \tag{B.53}$$

if $\mu_1 N_2 < \mu_2(N_1 + P)$ and $\frac{\mu_2(N_1+P)-\mu_1 N_2}{\mu_1+\mu_2} \le P$, then

$$\begin{cases} q'(P_1) < 0 & \text{for } 0 \le P_1 < \frac{\mu_2(N_1+P)-\mu_1 N_2}{\mu_1+\mu_2}, \\ q'(P_1) > 0 & \text{for } \frac{\mu_2(N_1+P)-\mu_1 N_2}{\mu_1+\mu_2} < P_1 \le P, \end{cases}$$

thus,

$$q(P_1) \le \max[q(P), q(0)]$$

$$= \max\left[\mu_1 \ln\left(1+\frac{P}{N_1}\right), \mu_2 \ln\left(1+\frac{P}{N_2}\right)\right]; \tag{B.54}$$

if $\mu_1 N_2 < \mu_2(N_1+P)$ and $\frac{\mu_2(N_1+P)-\mu_1 N_2}{\mu_1+\mu_2} > P$, then for $0 \le P_1 \le P$, we have $q'(P_1) < 0$ and

$$q(P_1) \le q(0) = \mu_2 \ln\left(1+\frac{P}{N_2}\right). \tag{B.55}$$

From (B.53), (B.54) and (B.55), we conclude that

$$q(P_1) \le \max\left[\mu_1 \ln\left(1+\frac{P}{N_1}\right), \mu_2 \ln\left(1+\frac{P}{N_2}\right)\right].$$

That is,

$$\begin{aligned} & \mu_1 \ln\left(1+\frac{P_1}{N_1+P_2}\right) + \mu_2 \ln\left(1+\frac{P_2}{N_1+P_1}\right) \\ \le\ & \max\left[\mu_1 \ln\left(1+\frac{P_1+P_2}{N_1}\right), \mu_2 \ln\left(1+\frac{P_1+P_2}{N_2}\right)\right]. \end{aligned}$$

Now assume that (4.45) is true for $M = K$:

$$\sum_{i=1}^{K} \mu_i \log\left(1+\frac{P_i}{N_i+\sum_{j=1,j\neq i}^{K} P_j}\right) \le \max_{i\in\{1,2,\dots,K\}}\left[\mu_i \log\left(1+\frac{\sum_{j=1}^{K} P_j}{N_i}\right)\right],$$

then for $M = K+1$,

$$\begin{aligned} \sum_{i=1}^{K+1} \mu_i \log\left(1+\frac{P_i}{N_i+\sum_{j=1,j\neq i}^{K+1} P_j}\right) &= \sum_{i=1}^{K} \mu_i \log\left(1+\frac{P_i}{(N_i+P_{K+1})+\sum_{j=1,j\neq i}^{K} P_j}\right) \\ &\quad + \mu_{K+1} \log\left(1+\frac{P_{K+1}}{N_{K+1}+\sum_{j=1}^{K} P_j}\right) \\ &\le \max_{i\in\{1,2,\dots,K\}}\left[\mu_i \log\left(1+\frac{\sum_{j=1}^{K} P_j}{N_i+P_{K+1}}\right)\right] \\ &\quad + \mu_{K+1} \log\left(1+\frac{P_{K+1}}{N_{K+1}+\sum_{j=1}^{K} P_j}\right) \\ &\le \max_{i\in\{1,2,\dots,K+1\}}\left[\mu_i \log\left(1+\frac{\sum_{j=1}^{K+1} P_j}{N_i}\right)\right]. \end{aligned}$$

Therefore, by induction, we know that (4.45) is true. □

Appendix C Proofs in Chapter 5

C.1 Proof of Theorem 5.1

Achievability of the Capacity Region: We prove the achievability of the capacity region $\mathcal{C}_{zero}(\bar{P})$ in (5.1) by proving the achievability of $\mathbf{R} \in \bigcap_{\mathbf{n} \in \mathcal{N}} \mathcal{C}_{CD}(\mathbf{n}, \mathcal{P})$ in (5.2) for each given power allocation policy $\mathcal{P} \in \mathcal{F}$.

$\forall \mathcal{P} \in \mathcal{F}$, for $j = 1, 2, \cdots, M$, since $P_j(\mathbf{n})$ denotes the transmit power for User j in fading state $\mathbf{n}$, $\forall \mathbf{R} \in \bigcap_{\mathbf{n} \in \mathcal{N}} \mathcal{C}_{CD}(\mathbf{n}, \mathcal{P})$, $\mathbf{R} = (R_1, R_2, \cdots, R_M)$, we need to prove that for every $\epsilon > 0$, there exists a sequence of $\left((2^{R_1 T}, 2^{R_2 T}, \cdots, 2^{R_M T}), T\right)$ codes and a coding and decoding scheme with probability of error $P_e^{(T)} < \epsilon$ for every fading process with stationary distribution $\mathcal{Q}$, i.e., a coding delay T which is independent of the correlation structure of the fading. We prove in the following that this is true for the two-user case. The result can be easily generalized to the M-user case ($M > 2$). Note that with the availability of CSI at both the transmitter and the receivers, the codewords can be chosen based on the realization of the fading process.

Let

$$\Phi_1 \triangleq \{\mathbf{n} : n_1 < n_2, \mathbf{n} \in \mathcal{N}\},$$

$$\Phi_2 \triangleq \{\mathbf{n} : n_1 > n_2, \mathbf{n} \in \mathcal{N}\}.$$

Recall that we assume $Pr\{n_i = n_j\} = 0, \forall i \neq j$. Thus

$$\mathcal{N} = \Phi_1 \cup \Phi_2.$$

When the channel fading states are in Φ_1, let $\mathbf{N}^{(T)} = (\mathbf{N}(1), \mathbf{N}(2), \cdots, \mathbf{N}(T))$ denote the realization of the slowly time-varying fading process. Let $M_1 = 2^{TR_1}$ and $M_2 = 2^{TR_2}$. Generate M_2 independent codewords $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_{M_2}$ of length T according to the normal distribution $N(0, 1)$, scaled by $\sqrt{P_2(\mathbf{N}(n))}$, $n = 1, 2, \cdots, T$. For each codeword $\mathbf{u}_i = (u_{i,1}, u_{i,2}, \cdots, u_{i,T})$ ($\mathbf{u}_i$ is called a cluster center, $1 \leq i \leq M_2$), generate M_1 independent codewords $\mathbf{x}_{1,i}, \mathbf{x}_{2,i}, \cdots, \mathbf{x}_{M_1,i}$ of length T according to the conditional normal distribution

$N(u_{i,n}, 1)$, scaled by $\sqrt{P_1(\mathbf{N}(n))}$, $n = 1, 2, \cdots, T$.

Assuming that the source for User 1 produces integer m $(1 \le m \le M_1)$ and the source for User 2 produces integer i $(1 \le 1 \le M_2)$, the encoder maps the pair (m, i) into a codeword $\mathbf{x}_{m,i} = (x_{m,i,1}, x_{m,i,2}, \cdots, x_{m,i,T})$, which is then transmitted. Let $\mathbf{y}$ and $\mathbf{z}$ be the received sequences for User 1 and User 2, respectively. We use the decoding rule in [85]. That is, the decoder of User 2 decodes that i for which $p(\mathbf{z}|\mathbf{u}_i)$ is maximized (a decoding failure occurs when there is a tie for the maximum). Let $P_{e,2}^{(T)}$ be the probability of decoding error for User 2. The decoder for User 1 first decodes the cluster center $\mathbf{u}_i$ in the same way as the decoder of User 2 does, and then uses its estimate of i to choose the m for which $p(\mathbf{y}|\mathbf{x}_{m,i})$ is maximized. For User 1, let $P_{e,12}^{(T)}$ be the probability of decoding error for index i and let $P_{e,11}^{(T)}$ be the probability of decoding error for index m. Thus, by denoting $P_{e,1}^{(T)}$ as the probability of decoding error for User 1, we have

$$P_{e,1}^{(T)} \le P_{e,11}^{(T)} + P_{e,12}^{(T)}. \tag{C.1}$$

Based on the above encoding and decoding rules, the probability of decoding error for User 2 is bounded by[85][1]

$$\begin{aligned} P_{e,2}^{(T)} \;\le\; & \exp(\rho T R_2/(2B)) \sum_{\mathbf{n}\in\Phi_1} f(\mathbf{n}) \cdot \\ & \sum_{\mathbf{z}} \left(\sum_{\mathbf{u}} Q_1(\mathbf{u}|\mathbf{n}) [p(\mathbf{z}|\mathbf{u},\mathbf{n})]^{\frac{1}{1+\rho}} \right)^{1+\rho}, \end{aligned} \tag{C.2}$$

for any $\rho > 0$, where $f(\mathbf{n})$ denotes the probability density function of $\mathbf{n}$, $Q_1(\mathbf{u}|\mathbf{n})$ is the conditional probability density function of $\mathbf{u}$, conditional on the fading being $\mathbf{n}$, and $p(\mathbf{z}|\mathbf{u},\mathbf{n})$ is the conditional probability density function of the received sequence $\mathbf{z}$, conditional on the codeword being $\mathbf{u}$ and the fading being $\mathbf{n}$. Since

$$Q_1(\mathbf{u}|\mathbf{n}) = \prod_{n=1}^{T} Q_1(u_n|\mathbf{n}), \tag{C.3}$$

[1]Note that the unit for R_1 and R_2 in this paper is "bits per second," while the unit in [85] is "bits per sample." This is why R_2 in (C.2) is divided by $2B$, the number of samples per second for the band-limited channel.

$$\begin{aligned} p(\mathbf{z}|\mathbf{u},\mathbf{n}) &= \prod_{n=1}^{T} p(z_n|u_n,\mathbf{n}) \\ &= \prod_{n=1}^{T} \left[\sum_x Q_2(x|u_n,\mathbf{n}) p(z_n|x,\mathbf{n}) \right], \end{aligned} \tag{C.4}$$

where $\forall 1 \leq n \leq T$, $Q_1(u_n|\mathbf{n}) \sim N(0, P_2(\mathbf{n}))$, $Q_2(x|u_n,\mathbf{n}) \sim N(u_n, P_1(\mathbf{n}))$, and $p(z_n|x,\mathbf{n}) \sim N(x, n_2 B)$. Therefore, in (C.4), $p(z_n|u_n,\mathbf{n}) \sim N(u_n, n_2 B + P_1(\mathbf{n}))$. Substituting (C.3) and (C.4) into (C.2), it is easy to verify that

$$\begin{aligned} P_{e,2}^{(T)} &\leq \sum_{\mathbf{n}\in\Phi_1} f(\mathbf{n}) \cdot \\ &\exp\left\{ -\rho \left[-T\frac{R_2}{2B} + \frac{1}{2}\sum_{n=1}^{T} \log\left(1 + \frac{P_2(\mathbf{n})}{(1+\rho)(n_2 B + P_1(\mathbf{n}))}\right)\right]\right\}. \end{aligned}$$

By assumption, $\exists \delta_2 > 0$ such that

$$R_2 \leq B \log\left(1 + \frac{P_2(\mathbf{n})}{n_2 B + P_1(\mathbf{n})}\right) - B\delta_2, \quad \forall \mathbf{n} \in \Phi_1. \tag{C.5}$$

Thus,

$$P_{e,2}^{(T)} \leq \exp\left\{ -\frac{\rho T}{2}\left[\delta_2 - \log(1+\rho)\right]\right\}, \tag{C.6}$$

since $\sum_{\mathbf{n}\in\Phi_1} f(\mathbf{n}) \leq 1$.

Similarly, we can show that

$$\begin{aligned} P_{e,12}^{(T)} &\leq \exp(\rho T R_2/(2B)) \sum_{\mathbf{n}\in\Phi_1} f(\mathbf{n}) \cdot \\ &\sum_{\mathbf{y}} \left(\sum_{\mathbf{u}} Q_1(\mathbf{u}|\mathbf{n}) [p(\mathbf{y}|\mathbf{u},\mathbf{n})]^{\frac{1}{1+\rho}} \right)^{1+\rho}, \end{aligned} \tag{C.7}$$

for any $\rho > 0$, where $Q_1(\mathbf{u}|\mathbf{n})$ is given by (C.3) and $p(\mathbf{y}|\mathbf{u},\mathbf{n})$ is the conditional probability density function of the received sequence $\mathbf{y}$, conditional on the codeword being $\mathbf{u}$ and the fading being $\mathbf{n}$. Since

$$\begin{aligned} p(\mathbf{y}|\mathbf{u},\mathbf{n}) &= \prod_{n=1}^{T} p(y_n|u_n,\mathbf{n}) \\ &= \prod_{n=1}^{T} \left[\sum_x Q_2(x|u_n,\mathbf{n}) p(y_n|x,\mathbf{n}) \right], \end{aligned} \tag{C.8}$$

where $\forall 1 \leq n \leq T$, $Q_2(x|u_n, \mathbf{n}) \sim N(u_n, P_1(\mathbf{n}))$ as given above, and $p(y_n|x, \mathbf{n}) \sim N(x, n_1 B)$. Therefore, in (C.8), $p(y_n|u_n, \mathbf{n}) \sim N(u_n, n_1 B + P_1(\mathbf{n}))$. Substituting (C.3) and (C.8) into (C.7), it is easily shown that

$$\begin{aligned} P_{e,12}^{(T)} &\leq \sum_{\mathbf{n}\in\Phi_1} f(\mathbf{n}) \cdot \\ &\exp\left\{-\rho\left[-T\frac{R_2}{2B} + \frac{1}{2}\sum_{n=1}^{T}\log\left(1 + \frac{P_2(\mathbf{n})}{(1+\rho)(n_1 B + P_1(\mathbf{n}))}\right)\right]\right\}. \end{aligned}$$

Since $\forall \mathbf{n} \in \Phi_1$, $n_1 < n_2$, from (C.5) we obtain

$$R_2 \leq B\log\left(1 + \frac{P_2(\mathbf{n})}{n_1 B + P_1(\mathbf{n})}\right) - B\delta_2, \quad \forall \mathbf{n} \in \Phi_1.$$

Thus,

$$P_{e,12}^{(T)} \leq \exp\left\{-\frac{\rho T}{2}\left[\delta_2 - \log(1+\rho)\right]\right\}. \tag{C.9}$$

Further moreover, for User 1, the probability of decoding error $P_{e,11}^{(T)}$ is bounded by[85]

$$\begin{aligned} P_{e,11}^{(T)} &\leq \exp(\rho T R_1/(2B)) \sum_{\mathbf{n}\in\Phi_1} f(\mathbf{n}) \cdot \\ &\sum_{\mathbf{y}}\sum_{\mathbf{u}} Q_1(\mathbf{u}|\mathbf{n}) \left(Q_2(\mathbf{x}|\mathbf{u}, \mathbf{n})[p(\mathbf{y}|\mathbf{x}, \mathbf{n})]^{\frac{1}{1+\rho}}\right)^{1+\rho}, \end{aligned} \tag{C.10}$$

for any $\rho > 0$, where $Q_1(\mathbf{u}|\mathbf{n})$ is given in (C.3),

$$Q_2(\mathbf{x}|\mathbf{u}, \mathbf{n}) = \prod_{n=1}^{T} Q_2(x_n|u_n, \mathbf{n}), \tag{C.11}$$

$$p(\mathbf{y}|\mathbf{x}, \mathbf{n}) = \prod_{n=1}^{T} p(y_n|x_n, \mathbf{n}), \tag{C.12}$$

and $\forall 1 \leq n \leq T$, $Q_2(x_n|u_n, \mathbf{n}) \sim N(u_n, P_1(\mathbf{n}))$, $p(y_n|x_n, \mathbf{n}) \sim N(x_n, n_1 B)$. Therefore, by substituting (C.3), (C.11), and (C.12) into (C.10), it is easy to verify that

$$\begin{aligned} P_{e,11}^{(T)} &\leq \sum_{\mathbf{n}\in\Phi_1} f(\mathbf{n}) \cdot \\ &\exp\left\{-\rho\left[-T\frac{R_1}{2B} + \frac{1}{2}\sum_{n=1}^{T}\log\left(1 + \frac{P_1(\mathbf{n})}{(1+\rho)n_1 B}\right)\right]\right\}. \end{aligned}$$

By assumption, $\exists \delta_1 > 0$ such that

$$R_1 \leq B \log\left(1 + \frac{P_1(\mathbf{n})}{n_1 B}\right) - B\delta_1, \quad \forall \mathbf{n} \in \Phi_1.$$

Thus,

$$P_{e,11}^{(T)} \leq \exp\left\{-\frac{\rho T}{2}\left[\delta_1 - \log(1+\rho)\right]\right\}. \tag{C.13}$$

Denoting $\delta = \min\{\delta_1, \delta_2\}$, then by (C.6) we obtain

$$P_{e,2}^{(T)} \leq \exp\left\{-\frac{\rho T}{2}\left[\delta - \log(1+\rho)\right]\right\},$$

and by (C.1), (C.9) and (C.13) we obtain

$$P_{e,1}^{(T)} \leq 2 \cdot \exp\left\{-\frac{\rho T}{2}\left[\delta - \log(1+\rho)\right]\right\}.$$

Therefore, when the channel fading states are in Φ_1, the overall probability of decoding error $P_e^{(T)}(\Phi_1)$ for the two users is

$$\begin{aligned} P_e^{(T)}(\Phi_1) &\leq P_{e,1}^{(T)} + P_{e,2}^{(T)} \\ &\leq 3 \cdot \exp\left\{-\frac{\rho T}{2}\left[\delta - \log(1+\rho)\right]\right\}. \end{aligned}$$

By taking ρ sufficiently small, we have $\delta - \log(1+\rho) > 0$ and it follows that the probability of error $P_e^{(T)}(\Phi_1)$ decreases exponentially with T, i.e., $\forall \epsilon > 0$, $\exists T_d(\Phi_1) > 0$, $\forall T > T_d(\Phi_1)$, $P_e^{(T)}(\Phi_1) < \epsilon$.

It can be similarly shown that when the channel fading states are in Φ_2, there exists a sequence of $\left((2^{R_1 T}, 2^{R_2 T}, \cdots, 2^{R_M T}), T\right)$ codes and a coding and decoding scheme for which the probability of error $P_e^{(T)}(\Phi_2)$ decays exponentially with T, i.e., $\forall \epsilon > 0$, $\exists T_d(\Phi_2) > 0$, $\forall T > T_d(\Phi_2)$, $P_e^{(T)}(\Phi_2) < \epsilon$. Thus, $\forall \mathbf{n} \in \mathcal{N}$, there exists a sequence of $\left((2^{R_1 T}, 2^{R_2 T}, \cdots, 2^{R_M T}), T\right)$ codes and a coding and decoding scheme for which the probability of error $P_e^{(T)} \to 0$ as $T \to \infty$. Moreover, $P_e^{(T)}$ decreases in T at a rate independent of the correlation character of the fading, i.e., by denoting $T_d = \max\{T_d(\Phi_1), T_d(\Phi_2)\}$, $\forall \epsilon > 0$, $\forall T > T_d$, we have $P_e^{(T)} < \epsilon$ for every fading process with stationary distribution $\mathcal{Q}$. □

<u>Converse</u>: Suppose that rate $\mathbf{R}$ is achievable, i.e., $\mathbf{R} \in \mathcal{C}_{zero}(\bar{P})$. We need to prove that

$\mathbf{R}$ cannot be outside of the region defined in (5.1). The proof is similar to that of the MAC capacity region [3].

Define $v_i = \frac{i}{m}$, $i = 0, 1, 2, \cdots, mI$. Since the time-varying noise density n_j of each user ranges from 0 to ∞, we say that a subchannel is in state $S_i, i = 0, 1, 2, \cdots, mI$ if $v_i \leq n_j < v_{i+1}$, where $v_{mI+1} = \infty$. Therefore, there are $(mI+1)^M = N$ discrete joint channel states. We denote the kth $(0 \leq k \leq N-1)$ of these N states as $\mathbf{S}_k = \left[S_{\phi(k,1)}, S_{\phi(k,2)}, \cdots, S_{\phi(k,M)}\right]$, where $[\phi(k,1), \phi(k,2), \cdots, \phi(k,M)]$ is the base-$(mI+1)$ expansion of k, with $\phi(k,1)$ being the least important component. That is, $0 \leq \phi(k,j) \leq mI$ for all $1 \leq j \leq M$ and

$$k = \sum_{j=1}^{M} \phi(k,j) \cdot (mI+1)^{j-1}.$$

Note that a channel state $\mathbf{n} \in \mathbf{S}_k$ if and only if $n_j \in S_{\phi(k,j)}$, $\forall 1 \leq j \leq M$.

Consider a sequence of Markov processes defined on $\mathcal{N}$ by using a Markov chain which is composed of the above mentioned N channel states with transition probabilities $t(\mathbf{S}_j, \mathbf{S}_k)$. The process remains in a state $\mathbf{S}_j$ for an exponential time $\tau(\mathbf{S}_j) \triangleq \text{Exponential}(\lambda(\mathbf{S}_j))$ and then selects a new state according to $t(\mathbf{S}_j, \mathbf{S}_k)$. By choosing the appropriate $\{\tau(\mathbf{S}_j)\}_{j=0}^{N-1}$ and transition probabilities, we assume that the Markov process has the required stationary distribution $\mathcal{Q}$ of the fading channel.

For each $T = 1, 2, \cdots$, let $\mathbf{N}^{(T)}$ be a fading process starting with $\mathbf{N}^{(T)}(0) = \mathbf{N}(0)$, where $\mathbf{N}(0)$ is a random variable with the stationary distribution $\mathcal{Q}$. The initial sojourn time in state $\mathbf{N}(0)$ of fading $\mathbf{N}^{(T)}$ is given by $\tau_T(\mathbf{N}(0))$, where $\tau_T(\mathbf{S}_j) \asymp \text{Exponential}(r_T \lambda(\mathbf{S}_j))$, $j = 0, 1, \cdots, N-1$. The scaling constant r_T determines the fading speed for process $\mathbf{N}^{(T)}$. $\forall \delta > 0$ fixed, by selecting an appropriate decreasing sequence $\{r_T\}_{T=1}^{\infty}$ where $r_T \to 0$ as $T \to \infty$, we can have

$$Pr(\forall T, \tau_T(\mathbf{S}_j) > T) > 1 - \delta, \quad 0 \leq j \leq N-1. \tag{C.14}$$

Since $\mathbf{R} \in \mathcal{C}_{zero}(\bar{P})$, we can choose for each T a code of size $2^{TR_1} \cdot 2^{TR_2} \cdots 2^{TR_M}$ for which the probability of error $p(T)$ under fading process $\mathbf{N}^{(T)}$ goes to zero as $T \to \infty$. Let $\left\{X^{(T)}(n)\right\}_{n=1}^{T}$ denote a random selection of codewords from the codebook for the M users and let $P_j^{(T)}(n)$ be the transmit power for User j $(1 \leq j \leq M)$. Note that $\left\{X^{(T)}(n)\right\}_{n=1}^{T}$ can be chosen according to the fading process $\mathbf{N}^{(T)}$. For $0 \leq k \leq N-1$, let $\Omega(\mathbf{S}_k)$ be the subset of the sample space on which $\mathbf{N}(0) \in \mathbf{S}_k$ and $\forall T, \tau_T(\mathbf{S}_k) > T$. Let Q be a random variable uniformly distributed on $[0, T]$. For $1 \le j \le M$, define

$$V_j(\mathbf{S}_k, T) = E_Q\left[P_j^{(T)}(Q)|\Omega(\mathbf{S}_k)\right],$$

$$W_j(\mathbf{S}_k, T) = E_Q\left[P_j^{(T)}(Q)|[\mathbf{N}(0) \in \mathbf{S}_k] - \Omega(\mathbf{S}_k)\right],$$

$$\begin{aligned} Z_j(\mathbf{S}_k, T) &= V_j(\mathbf{S}_k, T) Pr\{\forall T, \tau_T(\mathbf{S}_k) > T | \mathbf{N}(0) \in \mathbf{S}_k\} \\ &\quad + W_j(\mathbf{S}_k, T) Pr\{\exists T : \tau_T(\mathbf{S}_k) \le T | \mathbf{N}(0) \in \mathbf{S}_k\}. \end{aligned}$$

Then the power constraint is that $\forall T$,

$$\sum_{k=0}^{N-1} p(\mathbf{S}_k) \sum_{j=1}^{M} Z_j(\mathbf{S}_k, T) \le \bar{P}.$$

According to (C.14), we have

$$\sum_{k=0}^{N-1} p(\mathbf{S}_k) \sum_{j=1}^{M} V_j(\mathbf{S}_k, T) \le \frac{\bar{P}}{1-\delta}.$$

By the bounded convergence theorem, it is clear that there exists a a convergent subsequence along which if taking the limit we obtain for $1 \le j \le M$ and $0 \le k \le N-1$

$$V_j(\mathbf{S}_k, T) \to V_j(\mathbf{S}_k) \quad \text{as } T \to \infty.$$

Therefore,

$$\sum_{k=0}^{N-1} p(\mathbf{S}_k) \sum_{j=1}^{M} V_j(\mathbf{S}_k) \le \frac{\bar{P}}{1-\delta}. \tag{C.15}$$

Now we define a new fading process $\mathbf{N}$ by

$$\mathbf{N}(n) = \sum_{k=0}^{N-1} 1[\mathbf{N}(0) \in \mathbf{S}_k] \mathbf{n}^l(\mathbf{S}_k),$$

where $\mathbf{n}^l(\mathbf{S}_k) = [n_1^l(\mathbf{S}_k), n_2^l(\mathbf{S}_k), \cdots, n_M^l(\mathbf{S}_k)]$ and $n_j^l(\mathbf{S}_k) \triangleq v_{\phi(k,j)}$, $1 \le j \le M$. Note that conditioned on $\mathbf{N}(0)$, the fading process is deterministic. Let $q(T|\Omega(\mathbf{S}))$ be the conditional probability of error for code $X^{(T)}$ in the new fading channel. Then, obviously,

$$p(T) \geq p(\mathbf{S}_k)q(T|\Omega(\mathbf{S}_k)).$$

By assumption, $p(T) \to 0$ as $T \to \infty$. Thus, $q(T|\Omega(\mathbf{S}_k)) \to 0$ as $T \to \infty$. But conditional on $\Omega(\mathbf{S}_k)$, we have a constant channel, and a sequence of codes satisfying the power constraint (C.15). It follows that $\forall 0 \leq k \leq N-1$, $\mathbf{R} \in \mathcal{C}_{CD}(\mathbf{n}^l(\mathbf{S}_k), \mathcal{V})$, where $\mathcal{V}$ denotes the power allocation policy that assigns power $V_j(\mathbf{S}_k)$ to User j ($1 \leq j \leq M$) in channel state $\mathbf{S}_k$, and $\mathcal{C}_{CD}(\cdot,\cdot)$ is as given in (5.2).

$\forall \mathcal{P} \in \mathcal{F}$, define

$$\bar{\mathcal{C}}_{CD}^{(N)}(\mathbf{n}, \mathcal{P}) = \mathcal{C}_{CD}(\mathbf{n}^l(\mathbf{S}_k), \mathcal{P}), \quad \text{if } \mathbf{n} \in \mathbf{S}_k,\ 0 \leq k \leq N-1.$$

Denote $\mathcal{F}_{N,\delta}$ as the set of all power control policies that satisfy the power constraint $\frac{\bar{P}}{1-\delta}$ and are piecewise constant on each fading state $\mathbf{S}_k$, $0 \leq k \leq N-1$. Denote $\mathcal{P}_{N,\delta}$ as the power allocation policy that assigns power $P_j^{N,\delta}(\mathbf{n})$ to User j ($1 \leq j \leq M$) in each fadng state $\mathbf{n}$, where

$$P_j^{N,\delta}(\mathbf{n}) = \sum_{k=0}^{N-1} V_j(\mathbf{S}_k) 1[\mathbf{n} \in \mathbf{S}_k].$$

Thus, $\mathcal{P}_{N,\delta} \in \mathcal{F}_{N,\delta}$. We have shown that for any $\delta > 0$,

$$\forall \mathbf{n} \in \mathcal{N}, \mathbf{R} \in \bar{\mathcal{C}}_{CD}^{(N)}(\mathbf{n}, \mathcal{P}_{N,\delta}).$$

It follows that

$$\mathbf{R} \in \bigcup_{\mathcal{P} \in \mathcal{F}_N} \bigcap_{\mathbf{n} \in \mathcal{N}} \bar{\mathcal{C}}_{CD}^{(N)}(\mathbf{n}, \mathcal{P}),$$

where $\mathcal{F}_N = \mathcal{F}_{N,0}$. Thus,

$$\mathcal{C}_{zero}(\bar{P}) \subseteq \bigcup_{\mathcal{P} \in \mathcal{F}_N} \bigcap_{\mathbf{n} \in \mathcal{N}} \bar{\mathcal{C}}_{CD}^{(N)}(\mathbf{n}, \mathcal{P}). \tag{C.16}$$

Now combining (C.16) with the achievability result which indicates that

$$\bigcup_{\mathcal{P} \in \mathcal{F}_N} \bigcap_{\mathbf{n} \in \mathcal{N}} \mathcal{C}_{CD}(\mathbf{n}, \mathcal{P}) \subseteq \bigcup_{\mathcal{P} \in \mathcal{F}} \bigcap_{\mathbf{n} \in \mathcal{N}} \mathcal{C}_{CD}(\mathbf{n}, \mathcal{P}) \subseteq \mathcal{C}_{zero}(\bar{P}).$$

we obtain

$$\bigcup_{\mathcal{P}\in\mathcal{F}_N}\bigcap_{\mathbf{n}\in\mathcal{N}}\mathcal{C}_{CD}(\mathbf{n},\mathcal{P})\subseteq\mathcal{C}_{zero}(\bar{P})\subseteq\bigcup_{\mathcal{P}\in\mathcal{F}_N}\bigcap_{\mathbf{n}\in\mathcal{N}}\bar{\mathcal{C}}_{CD}^{(N)}(\mathbf{n},\mathcal{P}).$$

Since the lower and upper bounds converge as $N\to\infty$, it is clear that

$$\mathcal{C}_{zero}(\bar{P})=\bigcup_{\mathcal{P}\in\mathcal{F}}\bigcap_{\mathbf{n}\in\mathcal{N}}\mathcal{C}_{CD}(\mathbf{n},\mathcal{P}).\quad\square$$

C.2 Necessary and Sufficient Condition for Equations in (5.8) to Have Positive Solutions

We show that $\det A>0$ is the necessary and sufficient condition for the M linear equations in (5.8) to have positive solutions for all $P_i^{min}(\mathbf{n})$ $(1\le i\le M)$ in every fading state $\mathbf{n}$.

<u>**Necessary Condition:**</u> Since the equations in (5.8) is equivalent to

$$A\cdot\mathbf{P}^{min}(\mathbf{n})=B\mathbf{n},$$

if (5.8) has a solution for each state $\mathbf{n}$, then $\det A\neq 0$ so that the inverse of A exists and the solution to $\mathbf{P}^{min}(\mathbf{n})$ is as given in (5.10). Furthermore, if $\mathbf{P}^{min}(\mathbf{n})>0$ (i.e., $\forall 1\le i\le M$, $P_i^{min}(\mathbf{n})>0$) for any given $\mathbf{n}>0$, then every component of A^{-1} must be non-negative.

Define the cofactor matrix

$$\text{cof}A=[(-1)^{i+j}\det A_{ij}],\quad i,j=1,2,\cdots,M,$$

where A_{ij} is the $(M-1)\times(M-1)$ matrix formed by deleting the ith row and the jth column of A. Then

$$A^{-1}=\frac{1}{\det A}\cdot(\text{cof}A)^T. \tag{C.17}$$

Given $i<j$ fixed, since each component of the $(j-1)$th row of A_{ij} is -1, by subtracting the $(j-1)$th row from all other rows of A_{ij}, it is easily seen that the expansion by the ith column of A_{ij} is

$$\det A_{ij}=(-1)^{i+j}\prod_{k\neq i,j}(a_{kk}+1).$$

Thus

$$(-1)^{i+j}\det A_{ij} = \prod_{k\neq i,j}(a_{kk}+1) > 0, \tag{C.18}$$

since $\forall 1 \le k \le M$, $a_{kk} > 0$. It can be similarly shown that (C.18) holds for $i > j$ as well. Therefore, if every component of A^{-1} must be non-negative, from (C.17) we know that $\det A > 0$.

Sufficient Condition: Denote the cofactor matrix of A as $\text{cof}A = C = (c_{ij})$, $i, j = 1, 2, \cdots, M$, then as shown above,

$$c_{ij} = (-1)^{i+j}\det A_{ij} > 0, \quad \forall i \neq j. \tag{C.19}$$

Since $\forall 1 \le k \le M$, $a_{kk} > 0$, and the expansion by the kth column of A is

$$\det A = a_{kk}\cdot c_{kk} - \sum_{i=1, i\neq k}^{M} c_{ik},$$

if $\det A > 0$, then

$$c_{kk} = \frac{1}{a_{kk}}\cdot\left(\det A + \sum_{i=1,i\neq k}^{M} c_{ik}\right) > 0, \quad \forall 1 \le k \le M. \tag{C.20}$$

Therefore, combining (C.19) and (C.20) we get

$$c_{ij} > 0, \quad \forall 1 \le i, j \le M.$$

By (C.17) it is clear that every component of A^{-1} must be positive. Consequently, from the expression for $\mathbf{P}^{min}(\mathbf{n})$ in (5.10) we conclude that $\mathbf{P}^{min}(\mathbf{n}) > 0$ in every fading state $\mathbf{n}$ ($\mathbf{n} > 0$). □

C.3 Proof of Lemma 5.1

Since for $k > 0$, $i \ge 0$, $\sum_{n=0}^{i+k-1}\binom{i+k-1}{n} = 2^{i+k-1}$, we have

$$\binom{i+k-1}{i} \le 2^{i+k-1}.$$

For $0 < p < 1/2$,

$$\begin{aligned}\sum_{i=k+r+1}^{\infty} \binom{i+k-1}{i} p^i (1-p)^k &\leq \sum_{i=k+r+1}^{\infty} 2^{i+k-1} p^i (1-p)^k \\ &= \frac{(2p)^{k+r+1}}{1-2p} 2^{k-1} (1-p)^k \\ &= \frac{(2p)^{r+1}}{2(1-2p)} [4p(1-p)]^k.\end{aligned}$$

Because $4p(1-p) < [p + (1-p)]^2 = 1$,

$$\lim_{k\to\infty} [4p(1-p)]^k = 0.$$

Therefore, for fixed integer r,

$$\lim_{k\to\infty} \sum_{i=k+r+1}^{\infty} \binom{i+k-1}{i} p^i (1-p)^k = 0. \tag{C.21}$$

For a given positive integer k, we know that [80]

$$\sum_{i=0}^{\infty} \binom{i+k-1}{i} p^i (1-p)^k = 1. \tag{C.22}$$

Thus, from (C.21) and (C.22) we obtain

$$\begin{aligned}\lim_{k\to\infty} \sum_{i=0}^{k+r} \binom{i+k-1}{i} p^i (1-p)^k &= 1 - \lim_{k\to\infty} \sum_{i=k+r+1}^{\infty} \binom{i+k-1}{i} p^i (1-p)^k \\ &= 1. \qquad \square\end{aligned}$$

C.4 Proof of Lemma 5.2

Since for $n = 1, 2, \cdots$,

$$\int_x^{\infty} t^{n-1} e^{-t} dt = \begin{cases} e^{-x}(n-1)! \sum_{i=0}^{n-1} \frac{x^i}{i!}, & \text{if } x > 0, \\ (n-1)!, & \text{if } x = 0, \end{cases} \tag{C.23}$$

we have

$$\int_0^{\infty} \frac{p_i(\gamma_i)}{\gamma_i} d\gamma_i = \frac{m}{m-1} \bar{n}_i B, \quad i = 1, 2,$$

where $p_i(\gamma_i)$ is given in (5.19). Thus,

$$D_a(m) + D_d(m) = \frac{m}{m-1}\bar{n}_1 B, \tag{C.24}$$

$$D_b(m) + D_c(m) = \frac{m}{m-1}\bar{n}_2 B. \tag{C.25}$$

In (5.19), let $v_i = m\bar{n}_i B\gamma_i$, then,

$$p_i(v_i) = \frac{1}{(m-1)!} v_i^{m-1} e^{-v_i}, \qquad i = 1, 2. \tag{C.26}$$

Substituting (C.26) into (5.21) and (5.22) and using (C.23), we have

$$\begin{aligned} D_a(m) &= \int_0^\infty \int_{\frac{\bar{n}_1 B}{\bar{n}_2 B} v_2}^\infty \frac{m\bar{n}_1 B}{v_1} p_1(v_1) p_2(v_2) dv_1 dv_2 \\ &= \int_0^\infty \left\{ \frac{m}{m-1}\bar{n}_1 B e^{-\frac{\bar{n}_1 B}{\bar{n}_2 B} v_2} \sum_{i=0}^{m-2} \frac{\left(\frac{\bar{n}_1 B}{\bar{n}_2 B} v_2\right)^i}{i!} \right\} p_2(v_2) dv_2 \\ &= \frac{m}{m-1}\bar{n}_1 B \sum_{i=0}^{m-2} \frac{(i+m-1)!}{(m-1)!\, i!} \left(\frac{\bar{n}_1 B}{\bar{n}_1 B + \bar{n}_2 B}\right)^i \left(\frac{\bar{n}_2 B}{\bar{n}_1 B + \bar{n}_2 B}\right)^m, \end{aligned} \tag{C.27}$$

$$\begin{aligned} D_b(m) &= \int_0^\infty \int_{\frac{\bar{n}_1 B}{\bar{n}_2 B} v_2}^\infty \frac{m\bar{n}_2 B}{v_2} p_1(v_1) p_2(v_2) dv_1 dv_2 \\ &= \int_0^\infty \left\{ e^{-\frac{\bar{n}_1 B}{\bar{n}_2 B} v_2} \sum_{i=0}^{m-1} \frac{\left(\frac{\bar{n}_1 B}{\bar{n}_2 B} v_2\right)^i}{i!} \right\} \frac{m\bar{n}_2 B}{v_2} p_2(v_2) dv_2 \\ &= \frac{m}{m-1}\bar{n}_2 B \sum_{i=0}^{m-1} \frac{(i+m-2)!}{(m-2)!\, i!} \left(\frac{\bar{n}_1 B}{\bar{n}_1 B + \bar{n}_2 B}\right)^i \left(\frac{\bar{n}_2 B}{\bar{n}_1 B + \bar{n}_2 B}\right)^{(m-1)}. \end{aligned} \tag{C.28}$$

When $\bar{n}_1 B < \bar{n}_2 B$, applying *Lemma 5.1* to (C.27) and (C.28), we have

$$\lim_{m\to\infty} D_a(m) = \bar{n}_1 B,$$

$$\lim_{m\to\infty} D_b(m) = \bar{n}_2 B.$$

Combining these limiting expressions with (C.24) and (C.25) we obtain

$$\lim_{m\to\infty} D_c(m) = 0,$$

$$\lim_{m \to \infty} D_d(m) = 0.$$

C.5 Proof of Lemma 5.3

To prove the convexity of the region $\bar{\mathcal{O}}(\bar{P}, \mathbf{R})$, we need to show that if two usage probability vectors $\mathbf{Pr}^{on}$ and $\mathbf{Pr}^{on'}$ are achievable, then $\forall 0 \leq \lambda \leq 1$, $\lambda \mathbf{Pr}^{on} + (1-\lambda)\mathbf{Pr}^{on'}$ is also achievable.

Let $\bar{P}$ be the total average transmit power. For each fading state $\mathbf{n}$, let $w_i(\mathbf{n})$ and $P_i(\mathbf{n})$ be the probability and transmit power that the broadcast channel transmits the information of User i ($1 \leq i \leq M$) which results in the usage probability vector $\mathbf{Pr}^{on}$, and let $w_i'(\mathbf{n})$ and $P_i'(\mathbf{n})$ be the probability and transmit power for User i ($1 \leq i \leq M$) corresponding to the usage probability vector $\mathbf{Pr}^{on'}$. That is, in each fading state $\mathbf{n}$, $\sum_{i=1}^{M} w_i(\mathbf{n}) = 1$ and $\sum_{i=1}^{M} w_i'(\mathbf{n}) = 1$. Moreover,

$$Pr_i^{on} = E_{\mathbf{n}}[w_i(\mathbf{n})], \quad 1 \leq i \leq M,$$

$$Pr_i^{on'} = E_{\mathbf{n}}[w_i'(\mathbf{n})], \quad 1 \leq i \leq M,$$

$$E_{\mathbf{n}}\left[\sum_{i=1}^{M} w_i(\mathbf{n}) P_i(\mathbf{n})\right] \leq \bar{P},$$

$$E_{\mathbf{n}}\left[\sum_{i=1}^{M} w_i'(\mathbf{n}) P_i'(\mathbf{n})\right] \leq \bar{P}.$$

We construct a third transmission strategy in each fading state $\mathbf{n}$ by transmitting the information of User i ($1 \leq i \leq M$) with probability $w_i(\mathbf{n})$ and transmit power $P_i(\mathbf{n})$ in the first λ fraction of time and with probability $w_i'(\mathbf{n})$ and transmit power $P_i'(\mathbf{n})$ in the last $1-\lambda$ fraction of time. Therefore, in each fading state $\mathbf{n}$, $\sum_{i=1}^{M} [\lambda w_i(\mathbf{n}) + (1-\lambda) w_i'(\mathbf{n})] = 1$. Under the new strategy, the average usage probability $Pr_i^{on''}$ for User i ($1 \leq i \leq M$) is:

$$\begin{aligned} Pr_i^{on''} &= E_{\mathbf{n}}\left[\lambda w_i(\mathbf{n}) + (1-\lambda) w_i'(\mathbf{n})\right] \\ &= \lambda Pr_i^{on} + (1-\lambda) Pr_i^{on'}, \end{aligned}$$

and the total average transmit power satisfies

$$\begin{aligned}
& E_{\mathbf{n}}\left\{\sum_{i=1}^{M}\left[\lambda w_i(\mathbf{n})P_i(\mathbf{n})+(1-\lambda)w_i'(\mathbf{n})P_i'(\mathbf{n})\right]\right\} \\
= \; & \lambda E_{\mathbf{n}}\left[\sum_{i=1}^{M} w_i(\mathbf{n})P_i(\mathbf{n})\right]+(1-\lambda)E_{\mathbf{n}}\left[\sum_{i=1}^{M} w_i'(\mathbf{n})P_i'(\mathbf{n})\right] \\
\leq \; & \bar{P}.
\end{aligned}$$

Thus, $\forall 0 \leq \lambda \leq 1$, $\lambda \mathbf{Pr}^{on} + (1-\lambda)\mathbf{Pr}^{on'}$ is achievable. □

C.6 Proof of Theorem 5.3

In the following, $\forall 1 \leq m_0 \leq N$, $\forall 1 \leq j \leq m_0$, for simplicity we denote $\lambda_j(\mathbf{n})$, $v_j(\mathbf{n})$, $z_j(\mathbf{n})$, and $u_j(\mathbf{n})$ as λ_j, v_j, z_j, and u_j, respectively.

For $m_0 = 1, 2, \cdots, N$, $\forall \mathbf{n} \in \Omega_{m_0}$, let $\mathbf{u}^* \triangleq [u_1^*, u_2^*, \cdots, u_{m_0}^*]$, $\mathbf{u} \triangleq [u_1, u_2, \cdots, u_{m_0}]$. Since $\mathbf{u}^*$ satisfies

$$\begin{aligned}
\sum_{m_0=1}^{N} E_{\mathbf{n}\in\Omega_{m_0}}\left[\sum_{i=1}^{m_0} v_i u_i^*\right] &= P(s^*) + \sum_{m_0=1}^{N}\sum_{i=1}^{m_0}\left[\tau^* v_j + (1-\tau^*)v_{j-1}\right] Pr(\tilde{L}_j(m_0, s^*)) \\
&= \bar{P},
\end{aligned}$$

$\forall \mathbf{u} \neq \mathbf{u}^*$, by denoting $A(m_0) \triangleq \{\mathbf{n} : \mathbf{n} \in \Omega_{m_0}, \frac{1}{s^*} > z_1\}$, we have

$$\begin{aligned}
& \sum_{m_0=1}^{N} E_{\mathbf{n}\in\Omega_{m_0}}\left[\sum_{i=1}^{m_0} v_i u_i\right] - \bar{P} \\
= \; & \sum_{m_0=1}^{N} E_{\mathbf{n}\in\Omega_{m}}\left[\sum_{i=1}^{m_0} v_i u_i\right] - \sum_{m_0=1}^{N} E_{\mathbf{n}\in\Omega_{m_0}}\left[\sum_{i=1}^{m_0} v_i u_i^*\right] \\
= \; & \sum_{m_0=1}^{N} \int_{\mathbf{n}\in A(m_0)} \sum_{i=1}^{m_0} v_i u_i dF(\mathbf{n}) + \sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\int_{\mathbf{n}\in L_j(m_0,s^*)}\left(\sum_{i=1}^{m_0} v_i u_i - v_j\right) dF(\mathbf{n}) \\
& + \sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\left[\sum_{i=1}^{m_0} v_i u_i - v_j\tau^* - v_{j-1}(1-\tau^*)\right] Pr(\tilde{L}_j(m_0, s^*)).
\end{aligned} \tag{C.29}$$

We now show from (5.58)-(5.61) that $\forall \mathbf{n} \in \Omega_{m_0}$,

$$\frac{\lambda_i}{v_i} < \frac{\lambda_{i-1}}{v_{i-1}}, \quad i = 2, 3, \cdots, m_0. \tag{C.30}$$

Since $z_1 > z_2$, i.e., $\frac{\lambda_1}{v_1} > \frac{\lambda_2 - \lambda_1}{v_2 - v_1}$, for $i = 2$, (C.30) is true. Assuming that when $i = k$, (C.30) is true, or equivalently, $\frac{\lambda_k - \lambda_{k-1}}{v_k - v_{k-1}} < \frac{\lambda_k}{v_k}$, then because $z_{k+1} < z_k$, we have

$$\frac{\lambda_{k+1} - \lambda_k}{v_{k+1} - v_k} < \frac{\lambda_k - \lambda_{k-1}}{v_k - v_{k-1}} < \frac{\lambda_k}{v_k}.$$

That is,

$$\frac{\lambda_{k+1}}{v_{k+1}} < \frac{\lambda_k}{v_k}.$$

Thus, when $i = k + 1$, (C.30) is also true. So (C.30) is true for all $i = 2, 3, \cdots, m_0$. Continuing the derivation in (C.29), we have:

a) $\forall \mathbf{n} \in A(m_0)$, since $\frac{1}{s^*} > z_1 = \frac{\lambda_1}{v_1}$, from (C.30) we know that $\forall i = 1, 2, \cdots, m_0$, $\frac{1}{s^*} > \frac{\lambda_i}{v_i}$, i.e., $v_i > \lambda_i s^*$. Therefore, the first term in (C.29) will be:

$$\begin{aligned} \sum_{m_0=1}^{N} \int_{\mathbf{n} \in A(m_0)} \sum_{i=1}^{m_0} v_i u_i dF(\mathbf{n}) &> \sum_{m_0=1}^{N} \int_{\mathbf{n} \in A(m_0)} \sum_{i=1}^{m_0} \lambda_i s^* u_i dF(\mathbf{n}) \\ &= s^* \sum_{m_0=1}^{N} \int_{\mathbf{n} \in A(m_0)} \sum_{i=1}^{m_0} \lambda_i u_i dF(\mathbf{n}). \qquad \text{(C.31)} \end{aligned}$$

b) $\forall \mathbf{n} \in L_j(m_0, s^*)$, $j = 1, 2, \cdots, m$, since $z_j > \frac{1}{s^*} > z_{j+1}$, we have

$$v_{j+1} > v_j + (\lambda_{j+1} - \lambda_j)s^*,$$

$$v_{j-1} > v_j + (\lambda_{j-1} - \lambda_j)s^*.$$

For $i \neq j, 1 \leq i \leq m_0$, denoting

$$x_i \triangleq \frac{\lambda_j - \lambda_i}{v_j - v_i},$$

it is easy to verify that $x_i > x_{i+1}$ by using (C.30). Since $x_{j-1} = z_j$, $x_{j+1} = z_{j+1}$, $\forall i = 1, 2, \cdots, j-1$, $\forall k = j+1, j+2, \cdots, m_0$, we have

$$x_i > \frac{1}{s^*} > x_k.$$

Therefore, $\forall i \neq j, 1 \leq i \leq m_0$,

$$v_i > v_j + (\lambda_i - \lambda_j)s^*.$$

In (C.29), the second term will be:

$$
\begin{aligned}
& \sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\int_{\mathbf{n}\in L_j(m_0,s^*)}\left(\sum_{i=1}^{m_0} v_i u_i - v_j\right) dF(\mathbf{n}) \\
> & \sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\int_{\mathbf{n}\in L_j(m_0,s^*)}\left(v_j u_j + \sum_{i\neq j,i=1}^{m_0} [v_j + (\lambda_i - \lambda_j)s^*]u_i - v_j\right) dF(\mathbf{n}) \\
= & \sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\int_{\mathbf{n}\in L_j(m_0,s^*)}\left(\sum_{i\neq j,i=1}^{m_0} (\lambda_i - \lambda_j)s^* u_i\right) dF(\mathbf{n}) \\
= & \sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\int_{\mathbf{n}\in L_j(m_0,s^*)}\left(\sum_{i\neq j,i=1}^{m_0} \lambda_i s^* u_i - \lambda_j s^*[1-u_j]\right) dF(\mathbf{n}) \\
= & \; s^*\sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\int_{\mathbf{n}\in L_j(m_0,s^*)}\left(\sum_{i\neq j,i=1}^{m} \lambda_i u_i - \lambda_j[1-u_j]\right) dF(\mathbf{n}) \qquad \text{(C.32)}
\end{aligned}
$$

c) $\forall \mathbf{n} \in \tilde{L}_j(m_0, s^*)$, $j = 1, 2, \cdots, m_0$, since $\frac{1}{s^*} = z_j = x_{j-1}$, we have

$$v_{j-1} = v_j + (\lambda_{j-1} - \lambda_j)s^*.$$

Because $\forall i = 1, 2, \cdots, j-2$, $\forall k = j+1, j+2, \cdots, m$,

$$x_i > \frac{1}{s^*} > x_k,$$

$\forall i \neq j-1, j$ and $1 \leq i \leq m_0$, we have

$$v_i > v_j + (\lambda_i - \lambda_j)s^*.$$

Therefore, in (C.29), the third term will be:

$$
\begin{aligned}
& \sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\left[\sum_{i=1}^{m_0} v_i u_i - v_j\tau^* - v_{j-1}(1-\tau^*)\right] Pr(\tilde{L}_j(m_0,s^*)) \\
> & \sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\left[v_j u_j + \sum_{i\neq j,i=1}^{m_0} \{v_j + (\lambda_i - \lambda_j)s^*\} u_i - v_j\tau^* - v_{j-1}(1-\tau^*)\right] Pr(\tilde{L}_j(m_0,s^*)) \\
= & \sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\left[(v_j - v_{j-1})(1-\tau^*) + \sum_{i\neq j,i=1}^{m_0} (\lambda_i - \lambda_j)s^* u_i\right] Pr(\tilde{L}_j(m_0,s^*)) \\
= & \sum_{m_0=1}^{N}\sum_{j=1}^{m_0}\left[(\lambda_j - \lambda_{j-1})(1-\tau^*)s^* + \left(\sum_{i\neq j,i=1}^{m_0} \lambda_i s^* u_i\right) - \lambda_j s^*[1-u_j]\right] Pr(\tilde{L}_j(m_0,s^*))
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{m_0=1}^{N} \sum_{j=1}^{m_0} \left[\left(\sum_{i \neq j-1, j, i=1}^{m_0} \lambda_i s^* u_i \right) + \lambda_{j-1} s^* [u_{j-1} - (1 - \tau^*)] + \lambda_j s^* (u_j - \tau^*) \right] \\
&\quad \cdot Pr(\tilde{L_j}(m_0, s^*)) \\
&= s^* \sum_{m_0=1}^{N} \sum_{j=1}^{m_0} \left[\left(\sum_{i \neq j-1, j, i=1}^{m_0} \lambda_i u_i \right) + \lambda_{j-1} [u_{j-1} - (1 - \tau^*)] + \lambda_j (u_j - \tau^*) \right] \\
&\quad \cdot Pr(\tilde{L_j}(m_0, s^*)).
\end{aligned} \tag{C.33}
$$

Therefore, substituting (C.31), (C.32) and (C.33) into (C.29), we have

$$
\begin{aligned}
&\sum_{m_0=1}^{N} E_{\mathbf{n} \in \Omega_{m_0}} \left[\sum_{i=1}^{m_0} v_i u_i \right] - \bar{P} \\
&> s^* \sum_{m_0=1}^{N} \left\{ \int_{\mathbf{n} \in A(m_0)} \sum_{i=1}^{m_0} \lambda_i u_i dF(\mathbf{n}) + \sum_{j=1}^{m_0} \int_{\mathbf{n} \in L_j(m_0, s^*)} \left(\sum_{i \neq j, i=1}^{m_0} \lambda_i u_i - \lambda_j [1 - u_j] \right) dF(\mathbf{n}) \right. \\
&\quad \left. + \sum_{j=1}^{m_0} \left[\left(\sum_{i \neq j-1, j, i=1}^{m_0} \lambda_i u_i \right) + \lambda_{j-1} [u_{j-1} - (1 - \tau^*)] + \lambda_j (u_j - \tau^*) \right] Pr(\tilde{L_j}(m_0, s^*)) \right\} \\
&= s^* \left\{ \sum_{m_0=1}^{N} E_{\mathbf{n} \in \Omega_{m_0}} \left[\sum_{i=1}^{m_0} \lambda_i u_i \right] - \sum_{m_0=1}^{N} E_{\mathbf{n} \in \Omega_{m_0}} \left[\sum_{i=1}^{m_0} \lambda_i u_i^* \right] \right\}.
\end{aligned}
$$

Thus, if

$$
\sum_{m_0=1}^{N} E_{\mathbf{n} \in \Omega_{m_0}} \left[\sum_{i=1}^{m_0} \lambda_i u_i \right] > \sum_{m_0=1}^{N} E_{\mathbf{n} \in \Omega_{m_0}} \left[\sum_{i=1}^{m_0} \lambda_i u_i^* \right],
$$

then

$$
\sum_{m_0=1}^{N} E_{\mathbf{n} \in \Omega_{m_0}} \left[\sum_{i=1}^{m_0} v_i u_i \right] > \bar{P},
$$

which means that $\mathbf{u}$ does not satisfy the total power constraint. Therefore, $\forall \mathbf{u} \neq \mathbf{u}^*$,

$$
\sum_{m_0=1}^{N} E_{\mathbf{n} \in \Omega_{m_0}} \left[\sum_{i=1}^{m_0} \lambda_i u_i \right] \leq \sum_{m_0=1}^{N} E_{\mathbf{n} \in \Omega_{m_0}} \left[\sum_{i=1}^{m_0} \lambda_i u_i^* \right]. \quad \square
$$

Appendix D Proofs in Chapter 6

D.1 Proof of Lemma 6.1

In the following we prove that the set $\mathcal{Q}_C$ defined in (6.9) is convex. For a given rate vector $\mathbf{R}$ and a power allocation policy $\mathcal{P}^{(i)}$ ($\forall i = 1, 2$), let $w^{(i)}(\mathbf{R}, \mathbf{h})$ denote the probability that the M users transmit information at rate vector $\mathbf{R}$ in fading state $\mathbf{h}$, $\forall \mathbf{h} \in \mathcal{H}_{all}$, and let $\mathbf{P}^{(i)}(\mathbf{R}, \mathbf{h})$ denote the transmit power vector of the M users that can support rate vector $\mathbf{R}$ in state $\mathbf{h}$. Then the average common usage probability $Pr^{on(i)}(\mathbf{R})$ for power allocation policy $\mathcal{P}^{(i)}$ is:

$$Pr^{on(i)}(\mathbf{R}) = E_{\mathbf{h}}\left[w^{(i)}(\mathbf{R}, \mathbf{h})\right], \quad i = 1, 2,$$

and the average transmit power vector $\bar{\mathbf{P}}^{(i)}(\mathbf{R})$ of the M users is:

$$\bar{\mathbf{P}}^{(i)}(\mathbf{R}) = E_{\mathbf{h}}\left[w^{(i)}(\mathbf{R}, \mathbf{h})\mathbf{P}^{(i)}(\mathbf{R}, \mathbf{h})\right], \quad i = 1, 2.$$

Therefore, $Pr^{on(i)}(\mathbf{R}) \in \overline{\mathcal{O}}_C\left(\bar{\mathbf{P}}^{(i)}(\mathbf{R}), \mathbf{R}\right)$, $\forall i = 1, 2$. That is, $\left(Pr^{on(i)}(\mathbf{R}), \bar{\mathbf{P}}^{(i)}(\mathbf{R})\right) \in \mathcal{Q}_C$, $\forall i = 1, 2$.

Now we construct a third power allocation policy $\mathcal{P}^{(3)}$ as follows. In each fading state $\mathbf{h} \in \mathcal{H}_{all}$, $\forall 0 < \alpha < 1$, let the M users transmit information at rate vector $\mathbf{R}$ in a fraction α of the time with probability $w^{(1)}(\mathbf{R}, \mathbf{h})$ and transmit power $\mathbf{P}^{(1)}(\mathbf{R}, \mathbf{h})$, and in the remaining fraction $1 - \alpha$ of time with probability $w^{(2)}(\mathbf{R}, \mathbf{h})$ and transmit power $\mathbf{P}^{(2)}(\mathbf{R}, \mathbf{h})$. Then obviously the average common usage probability $Pr^{on(3)}(\mathbf{R})$ for power allocation policy $\mathcal{P}^{(3)}$ is:

$$\begin{aligned} Pr^{on(3)}(\mathbf{R}) &= E_{\mathbf{h}}\left[\alpha w^{(1)}(\mathbf{R}, \mathbf{h}) + (1-\alpha) w^{(2)}(\mathbf{R}, \mathbf{h})\right] \\ &= \alpha Pr^{on(1)}(\mathbf{R}) + (1-\alpha) Pr^{on(2)}(\mathbf{R}), \end{aligned}$$

and the average transmit power vector $\bar{\mathbf{P}}^{(3)}(\mathbf{R})$ of the M users for power allocation policy

$\mathcal{P}^{(3)}$ is:

$$\begin{aligned}\bar{\mathbf{P}}^{(3)}(\mathbf{R}) &= E_{\mathbf{h}}\left[\alpha w^{(1)}(\mathbf{R},\mathbf{h})\mathbf{P}^{(1)}(\mathbf{R},\mathbf{h}) + (1-\alpha)w^{(2)}(\mathbf{R},\mathbf{h})\mathbf{P}^{(2)}(\mathbf{R},\mathbf{h})\right] \\ &= \alpha\bar{\mathbf{P}}^{(1)}(\mathbf{R}) + (1-\alpha)\bar{\mathbf{P}}^{(2)}(\mathbf{R}).\end{aligned}$$

Therefore, $Pr^{on(3)}(\mathbf{R}) \in \overline{\mathcal{O}}_C\left(\bar{\mathbf{P}}^{(3)}(\mathbf{R}),\mathbf{R}\right)$, i.e., $\left(Pr^{on(3)}(\mathbf{R}),\bar{\mathbf{P}}^{(3)}(\mathbf{R})\right) \in \mathcal{Q}_C$, which means that the set $\mathcal{Q}_C$ is convex.

D.2 Proof of Lemma 6.2

Here we prove that both the usage probability region $\overline{\mathcal{O}}_I(\bar{\mathbf{P}}^*,\mathbf{R})$ given in *Definition 6.8* and the set $\mathcal{Q}_I$ defined in (6.44) are convex.

Since independent outage declaration for each user is allowed, for a given rate vector $\mathbf{R}$ and a power allocation policy $\mathcal{P}^{(j)}$ ($\forall j = 1,2$), let $w^{(j)}(\mathbf{R},\mathbf{h},S_k)$ denote the probability that only all the users in subset S_k ($1 \le k \le 2^M-1$) transmit information with rate vector $\mathbf{R}_{S_k}$ in fading state $\mathbf{h}$, $\forall \mathbf{h} \in \mathcal{H}_{all}$, and let $\mathbf{P}^{(j)}(\mathbf{R},\mathbf{h},S_k)$ denote the transmit power vector of the M users when only users in subset S_k transmit information. Then obviously, for $j = 1,2$, $P_i^{(j)}(\mathbf{R},\mathbf{h},S_k) = 0$, $\forall i \notin S_k$. From (6.41), we know that the average usage probability $Pr_i^{on(j)}(\mathbf{R})$ of each user i ($1 \le i \le M$) for power allocation policy $\mathcal{P}^{(j)}$ ($j = 1,2$) is

$$Pr_i^{on(j)}(\mathbf{R}) = E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w_j(\mathbf{R},\mathbf{h},S_k)\mathbf{1}[i \in S_k]\right], \quad \forall 1 \le i \le M,$$

and from (6.42), it is clear that the average transmit power vector $\bar{\mathbf{P}}^{(j)}(\mathbf{R})$ of the M users for power allocation policy $\mathcal{P}^{(j)}$ ($j = 1,2$) is:

$$\bar{\mathbf{P}}^{(j)}(\mathbf{R}) = E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w_j(\mathbf{R},\mathbf{h},S_k)\mathbf{P}^{(j)}(\mathbf{R},\mathbf{h},S_k)\right].$$

Let $\mathbf{Pr}^{on(j)}(\mathbf{R}) = [Pr_1^{on(j)}(\mathbf{R}), Pr_2^{on(j)}(\mathbf{R}),\cdots,Pr_M^{on(j)}(\mathbf{R})]$. Then

$$\mathbf{Pr}^{on(j)}(\mathbf{R}) \in \overline{\mathcal{O}}_I\left(\bar{\mathbf{P}}^{(j)}(\mathbf{R}),\mathbf{R}\right), \quad j = 1,2.$$

That is, $\left(\mathbf{Pr}^{on(j)}(\mathbf{R}),\bar{\mathbf{P}}^{(j)}(\mathbf{R})\right) \in \mathcal{Q}_I$, $j = 1,2$.

Now we construct a third power allocation policy $\mathcal{P}^{(3)}$ as follows. In each fading state $\mathbf{h} \in \mathcal{H}_{all}$, $\forall 0 < \alpha < 1$, let all the users in subset S_k ($\forall 1 \leq k \leq 2^M - 1$) transmit information with probability $w^{(1)}(\mathbf{R}, \mathbf{h}, S_k)$ in a fraction α of the time and let the corresponding transmit power vector of the M users be $\mathbf{P}^{(1)}(\mathbf{R}, \mathbf{h}, S_k)$. In the remaining fraction $1 - \alpha$ of time let all the users in subset S_k ($\forall 1 \leq k \leq 2^M - 1$) transmit information with probability $w^{(2)}(\mathbf{R}, \mathbf{h}, S_k)$ in a fraction α of the time and let the corresponding transmit power vector of the M users be $\mathbf{P}^{(2)}(\mathbf{R}, \mathbf{h}, S_k)$. Then obviously the average usage probability $Pr_i^{on(3)}(\mathbf{R})$ of each user i ($1 \leq i \leq M$) for power allocation policy $\mathcal{P}_3$ is

$$
\begin{aligned}
Pr_i^{on(3)}(\mathbf{R}) &= E_{\mathbf{h}}\left[\alpha\left(\sum_{k=1}^{2^M-1} w^{(1)}(\mathbf{R}, \mathbf{h}, S_k)\mathbf{1}[i \in S_k]\right)\right. \\
&\qquad \left. + (1-\alpha)\left(\sum_{k=1}^{2^M-1} w^{(2)}(\mathbf{R}, \mathbf{h}, S_k)\mathbf{1}[i \in S_k]\right)\right] \\
&= \alpha Pr_i^{on(1)}(\mathbf{R}) + (1-\alpha)Pr_i^{on(2)}(\mathbf{R}), \qquad \forall 1 \leq i \leq M,
\end{aligned}
$$

and the average transmit power vector $\bar{\mathbf{P}}^{(3)}(\mathbf{R})$ of the M users for power allocation policy $\mathcal{P}^{(3)}$ is:

$$
\begin{aligned}
\bar{\mathbf{P}}^{(3)}(\mathbf{R}) &= E_{\mathbf{h}}\left[\alpha\left(\sum_{k=1}^{2^M-1} w^{(1)}(\mathbf{R}, \mathbf{h}, S_k)\mathbf{P}^{(1)}(\mathbf{R}, \mathbf{h}, S_k)\right)\right. \\
&\qquad \left. + (1-\alpha)\left(\sum_{k=1}^{2^M-1} w^{(2)}(\mathbf{R}, \mathbf{h}, S_k)\mathbf{P}^{(2)}(\mathbf{R}, \mathbf{h}, S_k)\right)\right] \\
&= \alpha\bar{\mathbf{P}}^{(1)}(\mathbf{R}) + (1-\alpha)\bar{\mathbf{P}}^{(2)}(\mathbf{R}).
\end{aligned}
$$

Therefore, $\mathbf{Pr}^{on(3)}(\mathbf{R}) \in \overline{\mathcal{O}}_I\left(\bar{\mathbf{P}}^{(3)}(\mathbf{R}), \mathbf{R}\right)$, i.e., $\left(\mathbf{Pr}^{on(3)}(\mathbf{R}), \bar{\mathbf{P}}^{(3)}(\mathbf{R})\right) \in \mathcal{Q}_I$, which means that the set $\mathcal{Q}_I$ is convex.

Now if both power allocation policy $\mathcal{P}^{(1)}$ and $\mathcal{P}^{(2)}$ satisfy a given average power constraint vector $\bar{\mathbf{P}}^*$, i.e., $\bar{\mathbf{P}}^{(j)} \leq \bar{\mathbf{P}}^*$, $j = 1, 2$, then $\mathbf{Pr}^{on(j)}(\mathbf{R}) \in \overline{\mathcal{O}}_I\left(\bar{\mathbf{P}}^*, \mathbf{R}\right)$. For power allocation policy $\mathcal{P}^{(3)}$, since

$$
\begin{aligned}
\bar{\mathbf{P}}^{(3)}(\mathbf{R}) &= \alpha\bar{\mathbf{P}}^{(1)}(\mathbf{R}) + (1-\alpha)\bar{\mathbf{P}}^{(2)}(\mathbf{R}) \\
&\leq \alpha\bar{\mathbf{P}}^* + (1-\alpha)\bar{\mathbf{P}}^* \\
&= \bar{\mathbf{P}}^*, \qquad \text{(D.1)}
\end{aligned}
$$

it is obvious that $\mathbf{Pr}^{on(3)}(\mathbf{R}) \in \overline{\mathcal{O}}_I(\bar{\mathbf{P}}^*, \mathbf{R})$, which means that the set $\overline{\mathcal{O}}_I(\bar{\mathbf{P}}^*, \mathbf{R})$ is convex.

D.3 Proof of Lemma 6.3

In an M-user system, for a given power price vector $\boldsymbol{\lambda}$ and channel usage reward vector $\boldsymbol{\mu}$, $\left\{w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_k)\right\}_{k=1}^{2^M-1}$ in (6.64) is determined once the threshold power s^* in the *Independent Outage Transmission Policy* is fixed to be 1. From (6.64) we know that the average usage probability of each user is

$$\begin{aligned} Pr^{on}_{i,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}) &= 1 - Pr_{i,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}) \\ &= E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_k)\mathbf{1}[i \in S_k]\right], \quad 1 \le i \le M. \end{aligned} \tag{D.2}$$

Therefore, for fixed power price vector $\boldsymbol{\lambda}$ and threshold power $s^* = 1$, the average usage probability $Pr^{on}_{i,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R})$ of each user i ($1 \le i \le M$) varies with the channel usage reward vector $\boldsymbol{\mu}$. We need to show that if μ_i increases while μ_j remains unchanged for all $j \ne i$, then $Pr^{on}_{i,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R})$ will increase or remain unchanged, and $Pr^{on}_{j,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R})$ ($\forall j \ne i$) will decrease or remain unchanged. In the following for simplicity we give the proof for a two-user system. The M-user case can be similarly proved.

In a two-user system, let the power price vector of the two users be $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$, and let the channel usage reward vector be $\boldsymbol{\mu} = (\mu_1, \mu_2)$. For a given rate vector $\mathbf{R} = (R_1, R_2)$ and in each fading state $\mathbf{h} = (h_1, h_2)$, let $w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_i)$ ($i = 1, 2$) denote the probability that only User i is transmitting information, and let $w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_3)$ denote the probability that both users are transmitting through the channel. Therefore, the transmission (usage) probability $Pr^{on}_{i,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})$ of each user i ($i = 1, 2$) is:

$$Pr^{on}_{i,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}) = w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_i) + w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_3), \quad \forall i = 1, 2. \tag{D.3}$$

Note that according to the *Independent Outage Transmission Policy* in Section 6.5.3, at most one subset of the two users is chosen for transmission in each fading state $\mathbf{h}$, i.e., at most one probability in $\left\{w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_j)\right\}_{j=1}^{3}$ will be one and the other two will be zero. For a given state $\mathbf{h}$, let $P_{\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_i)$ ($i = 1, 2$) denote the required weighted power for User i when the channel is used by User i alone, and let $P_{\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_3)$ denote the required total weighted power of the two users when both users are transmitting information. That is,

$$P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_1)=\lambda_1\cdot\frac{\sigma^2}{h_1}\cdot[\exp(2R_1)-1],$$

$$P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_2)=\lambda_2\cdot\frac{\sigma^2}{h_2}\cdot[\exp(2R_2)-1],$$

and if $\frac{\lambda_1}{h_1}>\frac{\lambda_2}{h_2}$, then

$$P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_3)=\lambda_1\cdot\frac{\sigma^2}{h_1}\cdot[\exp(2R_1)-1]+\lambda_2\cdot\frac{\sigma^2}{h_2}\cdot[\exp(2R_1+2R_2)-\exp(2R_1)];$$

if $\frac{\lambda_1}{h_1}<\frac{\lambda_2}{h_2}$, then

$$P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_3)=\lambda_1\cdot\frac{\sigma^2}{h_1}\cdot[\exp(2R_1+2R_2)-\exp(2R_2)]+\lambda_2\cdot\frac{\sigma^2}{h_2}\cdot[\exp(2R_2)-1].$$

Since $P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_3)>P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_1)+P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_2)$, there are four possible cases:

Case (1): $P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_2)>P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_1)$ and $\frac{\mu_1}{P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_1)}>\frac{\mu_2}{P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_2)}>\frac{\mu_1+\mu_2}{P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_3)}$,

Case (2): $P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_2)<P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_1)$ and $\frac{\mu_2}{P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_2)}>\frac{\mu_1}{P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_1)}>\frac{\mu_1+\mu_2}{P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_3)}$,

Case (3): $P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_2)>P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_1)$ and $\frac{\mu_2}{P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_2)}>\frac{\mu_1}{P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_1)}$, and $\frac{\mu_2}{P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_2)}>\frac{\mu_1+\mu_2}{P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_3)}$,

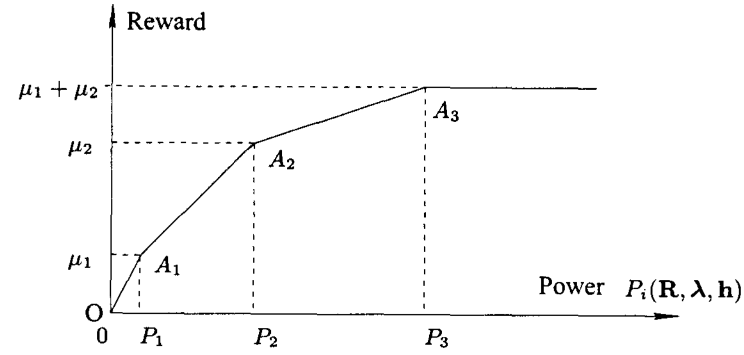
Case (4): $P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_2)<P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_1)$ and $\frac{\mu_2}{P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_2)}<\frac{\mu_1}{P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_1)}$, and $\frac{\mu_1}{P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_1)}>\frac{\mu_1+\mu_2}{P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_3)}$.

These four cases are shown in Figures D.1 and Figures D.2, where $P_{\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_i)$ $(i=1,2,3)$ is represented by P_i for simplicity.

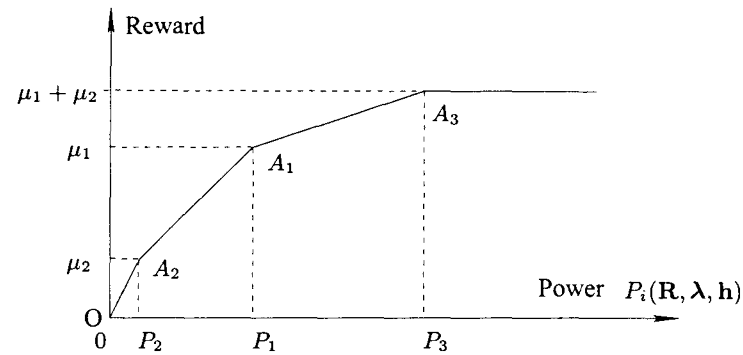
In Case (1), for a given fading state $\mathbf{h}$, if the subset S_2 is chosen for transmission, i.e.,

$$w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h},S_i)=\begin{cases}0, & \text{if } i=1,3\\ 1, & \text{if } i=2,\end{cases}\tag{D.4}$$

then by the *Independent Outage Transmission Policy*, it must be true that in Figure D.1, the slope of the line $\overline{A_1A_2}$ (we will refer to it as t_{12}) is larger than $\frac{1}{s^*}=1$, and the slope of the line $\overline{A_2A_3}$ (we will refer to it as t_{23}) is less than $\frac{1}{s^*}=1$. Therefore, if μ_2 increases while μ_1 remains unchanged, it will still be true that $t_{12}>1$ and $t_{23}<1$, since t_{12} will increase and t_{23} will remain unchanged. Consequently, subset S_2 will still be chosen for transmission,



Case (1)



Case (2)

Figure D.1: Four possible cases of the two-user system: Case (1) and (2).

i.e., $w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_i)$ $(i = 1, 2, 3)$ will be the same as in (D.4). Thus, $Pr^{on}_{1,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})$ and $Pr^{on}_{2,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})$ in (D.3) will both remain unchanged.

Similarly, if the subset S_3 is chosen for transmission, i.e.,

$$w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_i) = \begin{cases} 0, & \text{if } i = 1, 2 \\ 1, & \text{if } i = 3, \end{cases} \tag{D.5}$$

then it must be true that $t_{23} > 1$. Note that the slope of the line on the right side of point A_3 is always zero. Therefore, if μ_2 increases while μ_1 remains unchanged, it will still be true that $t_{23} > 1$, since t_{23} will remain unchanged. Consequently, subset S_3 will still be chosen for transmission, i.e., $w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_i)$ $(i = 1, 2, 3)$ will be the same as in (D.5). Thus, $Pr^{on}_{1,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})$ and $Pr^{on}_{2,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})$ in (D.3) will both remain unchanged.

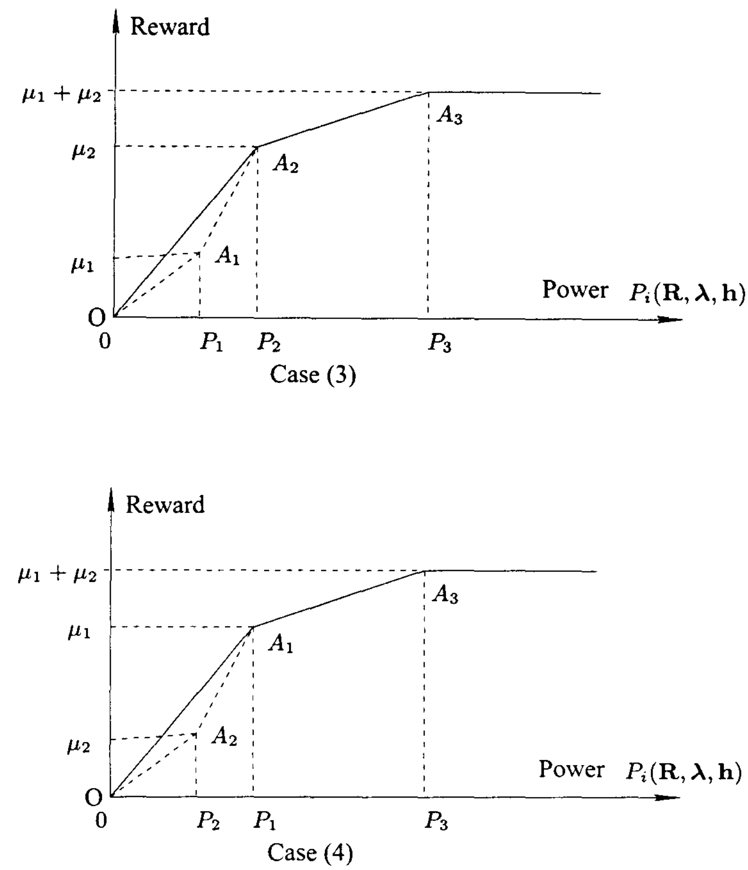


Figure D.2: Four possible cases of the two-user system: Case (3) and (4).

Now if the subset S_1 is chosen for transmission, i.e.,

$$w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_i) = \begin{cases} 0, & \text{if } i = 2, 3 \\ 1, & \text{if } i = 1, \end{cases}$$

then it must be true that the slope of the line $\overline{OA_1}$ (we will refer to it as t_{01}) is larger than $\frac{1}{s^*} = 1$, and $t_{23} < t_{12} < 1$. This might not be true if μ_2 increases and μ_1 is fixed, since t_{01} and t_{23} will remain unchanged and t_{12} will increase. Once t_{12} becomes larger than $\frac{1}{s^*} = 1$, the subset S_2 will be chosen for transmission, i.e., $w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_2)$ will equal one and $w_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h}, S_i)$ will equal zero, $i = 1, 3$. Thus, $Pr^{on}_{2,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})$ in (D.3) will either remain unchanged or increase, and $Pr^{on}_{1,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R}, \mathbf{h})$ in (D.3) will either remain unchanged or decrease. The same argument can be applied to the case where, before μ_2 increases, none of the subsets S_1, S_2, and S_3 is chosen for transmission.

Similarly, it is easily verified that in Case (2)-(4), it is also true that if μ_2 increases while μ_1 is fixed, then for each fading state $\mathbf{h}$, $Pr^{on}_{2,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h})$ will increase or remain unchanged, and $Pr^{on}_{1,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R},\mathbf{h})$ will decrease or remain unchanged. Therefore, the resulting average usage probability $Pr^{on}_{2,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R})$ will increase or remain unchanged, and the average usage probability $Pr^{on}_{1,\boldsymbol{\mu},\boldsymbol{\lambda}}(\mathbf{R})$ will decrease or remain unchanged.

D.4 The *HT* Algorithm

Given an average power constraint vector $\bar{\mathbf{P}}^*$ and rate vector $\mathbf{R}$, the *HT Algorithm* described below provides the optimal power allocation parameter pair $(\boldsymbol{\lambda}^*, \mathcal{L}^*)$ that achieves the infimum in (6.75). Note that in this algorithm, the required average power vector $\bar{\mathbf{P}}(\mathbf{R},\boldsymbol{\lambda},\mathcal{L})$ of the M users for supporting rate vector $\mathbf{R}$ without any outage is given in (6.74).

Initialization: For the M users in set $G = \{1, 2, \cdots, M\}$, start with an arbitrary power price vector $\boldsymbol{\lambda}(1) \in \Re^M_+$. Set $n = 1$.

Step n: Increase the power price of the user i with the largest normalized (normalized by its given average power constraint $\bar{P}_i^*$) average power $\frac{\bar{P}_i(\mathbf{R},\boldsymbol{\lambda},\{G\})}{\bar{P}_i^*}$ until its normalized average power equals that of another user, keeping the power prices of other users fixed. Then increase the power prices of *both* users by the same factor until the normalized average power of one of them equals that of a third user. Repeat the process and consider two cases:

(1) The process continues until there are no more users left. In this case, let the final value of the power price vector be $\boldsymbol{\lambda}(n+1)$ and go to Step $n+1$.

(2) The process terminates when the powers of a subset U of users whose prices are being increased do not meet the power of any of the other users, even when the prices of that subset are increased to infinity. In this case, perfect balancing of normalized average powers between the two subsets is impossible, even when absolute priority is given to the users in subset U. Partition the users into U and L, the subset of remaining users. The users in U from this step on will always be given absolute priority over users in L. The power prices of each user i in L will be fixed at $\lambda_i(n)$ and will

not be further adjusted in the algorithm. The algorithm is now recursively applied to users in U as if users in L did not exist. The current power prices of users in U will be used as initialization. □

After a finite number of iterations of this algorithm, the users will be partitioned into subsets $L_1, L_2, \cdots$, and L_I, where users in L_i are given absolute priority over users in L_j for $i > j$ and users in L_I are given the highest priority, and such that no further partitioning of L_I will take place. It is shown in [3] that the optimal absolute priority nesting $\mathcal{L}^*$ that solves the optimization problem (6.75) is

$$\begin{aligned} \mathcal{L}^* &= \{L_I, L_I \cup L_{I-1}, \cdots, \cup_{1\le i\le I} L_i\} \\ &= \{G_1, G_2, \cdots, G_I \equiv G\}, \end{aligned} \tag{D.6}$$

where $\forall 1 \le j \le I$,

$$G_j \triangleq \bigcup_{I-j+1\le i\le I} L_i.$$

Moreover, assuming that there exists a positive lower bound ϵ to the fading gains of all the users, if $\boldsymbol{\lambda}(n)$ is the power price vector of the M users at iteration n, then $\lim_{n\to\infty} \frac{\bar{P}_i(\mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}^*)}{\bar{P}_i^*}$ exists for each user i, $1 \le i \le M$, and the infimum in (6.75) is:

$$\begin{aligned} \inf_{(\boldsymbol{\lambda},\mathcal{L})} \max_{1\le i\le M} \frac{\bar{P}_i(\mathbf{R}, \boldsymbol{\lambda}, \mathcal{L})}{\bar{P}_i^*} &= \max_{1\le i\le M} \lim_{n\to\infty} \frac{\bar{P}_i(\mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}^*)}{\bar{P}_i^*} \\ &= \lim_{n\to\infty} \frac{\bar{P}_i(\mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}^*)}{\bar{P}_i^*}, \quad \begin{cases} \forall i \in G_1, & \text{if } \mathcal{L}^* \neq \{G\}, \\ \forall i \in G, & \text{if } \mathcal{L}^* = \{G\}. \end{cases} \end{aligned} \tag{D.7}$$

D.5 Proof of Lemma 6.4

Given two average common outage probabilities $Pr^{(1)}$ and $Pr^{(2)}$, we wish to prove that if $Pr^{(1)} < Pr(2)$, then $Inf(Pr^{(1)}) > Inf(Pr^{(2)})$.

When the common outage probability Pr in (6.76) equals $Pr^{(1)}$, denote the power allocation parameter pair that achieves the infimum in (6.76) as $(\boldsymbol{\lambda}^{(1)}, \mathcal{L}^{(1)})$, and denote the corresponding average transmit power vector of the M users as $\bar{\mathbf{P}}(Pr^{(1)}, \mathbf{R}, \boldsymbol{\lambda}^{(1)}, \mathcal{L}^{(1)})$. Then $\bar{\mathbf{P}}(Pr^{(1)}, \mathbf{R}, \boldsymbol{\lambda}^{(1)}, \mathcal{L}^{(1)}) \in APV_{out}(Pr^{(1)}, \mathbf{R})$. Since $Pr^{(1)} < Pr^{(2)}$, by definition of $APV_{out}(Pr^{(2)}, \mathbf{R})$ given in (6.25) with Pr^* replaced by $Pr^{(2)}$, it is clear that $\bar{\mathbf{P}}(Pr^{(1)}$, $\mathbf{R}$,

$\boldsymbol{\lambda}^{(1)}$, $\mathcal{L}^{(1)}) \in APV_{out}(Pr^{(2)}, \mathbf{R})$. Moreover, since obviously $\bar{\mathbf{P}}(Pr^{(1)}, \mathbf{R}, \boldsymbol{\lambda}^{(1)}, \mathcal{L}^{(1)})$ is not on the boundary surface of region $APV_{out}(Pr^{(2)}, \mathbf{R})$, there must exist a boundary average power vector $\bar{\mathbf{P}}(Pr^{(2)}, \mathbf{R}, \boldsymbol{\lambda}^*, \mathcal{L}^*)$ of $APV_{out}(Pr^{(2)}, \mathbf{R})$ that satisfies:

$$\bar{\mathbf{P}}(Pr^{(2)}, \mathbf{R}, \boldsymbol{\lambda}^*, \mathcal{L}^*) < \bar{\mathbf{P}}(Pr^{(1)}, \mathbf{R}, \boldsymbol{\lambda}^{(1)}, \mathcal{L}^{(1)}).$$

Therefore,

$$\begin{aligned} Inf(Pr^{(2)}) &\leq \max_{1 \leq i \leq M} \frac{\bar{P}_i(, Pr^{(2)}, \mathbf{R}, \boldsymbol{\lambda}^*, \mathcal{L}^*)}{\bar{P}_i^*} \\ &< \max_{1 \leq i \leq M} \frac{\bar{P}_i(Pr^{(1)}, \mathbf{R}, \boldsymbol{\lambda}^{(1)}, \mathcal{L}^{(1)})}{\bar{P}_i^*} \\ &= Inf(Pr^{(1)}). \end{aligned}$$

That is, if $Pr^{(1)} < Pr^{(2)}$, then $Inf(Pr^{(1)}) > Inf(Pr^{(2)})$.

D.6 Proof of Lemma 6.5

Given two average outage probability vector $\mathbf{Pr}^{(1)}$ and $\mathbf{Pr}^{(2)}$, we wish to prove that if $\mathbf{Pr}^{(1)} \leq \mathbf{Pr}^{(2)}$ and $\mathbf{Pr}^{(1)} \neq \mathbf{Pr}^{(2)}$, then $Inf\left(\mathbf{Pr}^{(1)}\right) > Inf\left(\mathbf{Pr}^{(2)}\right)$.

When the average outage probability vector $\mathbf{Pr}$ in (6.77) equals $\mathbf{Pr}^{(1)}$, denote the power allocation parameter pair that achieves the infimum in (6.77) as $(\boldsymbol{\lambda}^{(1)}, \mathcal{L}^{(1)})$, and denote the corresponding average transmit power vector of the M users as $\bar{\mathbf{P}}(\mathbf{Pr}^{(1)}, \mathbf{R}, \boldsymbol{\lambda}^{(1)}, \mathcal{L}^{(1)})$. Then $\bar{\mathbf{P}}(\mathbf{Pr}^{(1)}, \mathbf{R}, \boldsymbol{\lambda}^{(1)}, \mathcal{L}^{(1)}) \in APV_{out}\left(\mathbf{Pr}^{(1)}, \mathbf{R}\right)$. Since $\mathbf{Pr}^{(1)} \leq \mathbf{Pr}^{(2)}$, by definition of $APV_{out}\left(\mathbf{Pr}^{(2)}, \mathbf{R}\right)$ given in (6.66) with $\mathbf{Pr}^*$ replaced by $\mathbf{Pr}^{(2)}$, it is clear that $\bar{\mathbf{P}}(\mathbf{Pr}^{(1)}, \mathbf{R},$ $\boldsymbol{\lambda}^{(1)}$, $\mathcal{L}^{(1)}) \in APV_{out}\left(\mathbf{Pr}^{(2)}, \mathbf{R}\right)$. Moreover, since obviously $\bar{\mathbf{P}}(\mathbf{Pr}^{(1)}, \mathbf{R}, \boldsymbol{\lambda}^{(1)}, \mathcal{L}^{(1)})$ is not on the boundary surface of region $APV_{out}\left(\mathbf{Pr}^{(2)}, \mathbf{R}\right)$, there must exist a boundary average power vector $\bar{\mathbf{P}}(\mathbf{Pr}^{(2)}, \mathbf{R}, \boldsymbol{\lambda}^*, \mathcal{L}^*)$ of $APV_{out}\left(\mathbf{Pr}^{(2)}, \mathbf{R}\right)$ that satisfies:

$$\bar{\mathbf{P}}(\mathbf{Pr}^{(2)}, \mathbf{R}, \boldsymbol{\lambda}^*, \mathcal{L}^*) < \bar{\mathbf{P}}(\mathbf{Pr}^{(1)}\mathbf{R}, \boldsymbol{\lambda}^{(1)}, \mathcal{L}^{(1)})$$

Therefore,

$$Inf\left(\mathbf{Pr}^{(2)}\right) \leq \max_{1 \leq i \leq M} \frac{\bar{P}_i(\mathbf{Pr}^{(2)}, \mathbf{R}, \boldsymbol{\lambda}^*, \mathcal{L}^*)}{\bar{P}_i^*}$$

$$< \max_{1\le i\le M} \frac{\bar{P}_i(\mathbf{Pr}^{(1)}, \mathbf{R}, \boldsymbol{\lambda}^{(1)}, \mathcal{L}^{(1)})}{\bar{P}_i^*}$$
$$= Inf\left(\mathbf{Pr}^{(1)}\right).$$

That is, if $\mathbf{Pr}^{(1)} \le \mathbf{Pr}^{(2)}$ and $\mathbf{Pr}^{(1)} \ne \mathbf{Pr}^{(2)}$, then $Inf\left(\mathbf{Pr}^{(1)}\right) > Inf\left(\mathbf{Pr}^{(2)}\right)$.

D.7 The HT^* Algorithm

Given an average power constraint vector $\bar{\mathbf{P}}^*$, a rate vector $\mathbf{R}$, and a common outage probability Pr, the *HT^* Algorithm* provides the optimal power allocation parameter pair $(\boldsymbol{\lambda}^*, \mathcal{L}^*)$ that achieves the infimum in (6.76). The description of this algorithm is the same as that for the *HT Algorithm* given in Appendix D.4, except that the average power vector $\bar{\mathbf{P}}(\mathbf{R}, \boldsymbol{\lambda}, \mathcal{L})$ in the *HT Algorithm* is now replaced by $\bar{\mathbf{P}}(Pr, \mathbf{R}, \boldsymbol{\lambda}, \mathcal{L})$, which can be calculated from (6.32) for $\mathcal{L} = \{G\}$ and from (6.38) for $\mathcal{L} \ne \{G\}$ as shown in Section 6.4.4. After a finite number of iterations of this algorithm, the users will be partitioned into subsets $L_1, L_2, \cdots$, and L_I, where users in L_i are given absolute priority over users in L_j for $i > j$ and users in L_I are given the highest priority, and such that no further partitioning of L_I will take place. We will show that in this case, the optimal absolute priority nesting $\mathcal{L}^*$ that solves the optimization problem (6.76) is still as given in (D.6) and the infimum in (6.76) is similar to that of (D.7).

For $1 \le j \le I-1$, let $(\lambda_i^*)_{i\in L_j}$ be the power prices of the users in L_j when L_j is created and they are the components of vector $\boldsymbol{\lambda}^{*(L_j)}$, and let $H_j \triangleq L_{j+1} \cup \cdots \cup L_I$, which represents the subset of users who are given higher priority than users in L_j at all fading states. The users in $L_1 \cup \cdots \cup L_{j-1}$ will be given lower priority than users in L_j at all fading states. Therefore, there is no interference from any user in $L_1 \cup \cdots \cup L_{j-1}$ to users in L_j. Since the set of fading states $\mathcal{H}_{tran}$ where the M users transmit information simultaneously is determined by users in H_j, let

$$\mathcal{H}_{tran}^{(H_j)} \triangleq \{\mathbf{h}^{(H_j)} : \mathbf{h} \in \mathcal{H}_{tran}\},$$

where $\mathbf{h}^{(H_j)}$ denotes the subvector of $\mathbf{h}$ corresponding to users in H_j. Then $\mathbf{h} \in \mathcal{H}_{tran}$ if and only if $\mathbf{h}^{(H_j)} \in \mathcal{H}_{tran}^{(H_j)}$, and

$$\begin{aligned} Prob\{\mathbf{h} \in \mathcal{H}_{tran}\} &= Prob\{\mathbf{h}^{(H_j)} \in \mathcal{H}_{tran}^{(H_j)}\} \\ &= 1 - Pr. \end{aligned}$$

Since $\forall \mathbf{h} \in \mathcal{H}_{tran}$, as shown in (6.35), the optimal power allocation for each user $i \in L_j$ is not a function of $\mathbf{h}^{(H_j)}$ but just a function of $\mathbf{h}^{(L_j)}$ and $\boldsymbol{\lambda}^{*(L_j)}$ (note that $\mathbf{h}^{(L_j)}$ denotes the subvector of $\mathbf{h}$ corresponding to users in L_j), i.e.,

$$P_{\pi_{L_j}(i),\boldsymbol{\lambda}^{*(L_j)}}(\mathbf{R},\mathbf{h}) = \begin{cases} \frac{\sigma^2}{h_{\pi_{L_j}(1)}} \cdot \exp\left(2\sum_{k\in H_j} R_k\right) \cdot \left[\exp(2R_{\pi_{L_j}(1)}) - 1\right], & \text{if } i = 1, \\ \frac{\sigma^2}{h_{\pi_{L_j}(i)}} \cdot \exp\left(2\sum_{k\in H_j} R_k\right) \cdot \left[\exp\left(2\sum_{k=1}^{i} R_{\pi_{L_j}(k)}\right) \right. \\ \qquad \left. - \exp\left(2\sum_{k=1}^{i-1} R_{\pi_{L_j}(k)}\right)\right], & \forall 2 \le i \le |L_j|, \end{cases} \tag{D.8}$$

where $|L_j|$ denotes the number of users in set L_j and the permutation $\pi_{L_j}(\cdot)$ of the $|L_j|$ users satisfies:

$$\frac{\lambda^*_{\pi_{L_j}(1)}}{h_{\pi_{L_j}(1)}} \ge \frac{\lambda^*_{\pi_{L_j}(2)}}{h_{\pi_{L_j}(2)}} \ge \cdots \ge \frac{\lambda^*_{\pi_{L_j}(|L_j|)}}{h_{\pi_{L_j}(|L_j|)}},$$

the required average power $\bar{P}_{\pi_{L_j}(i)}\left(Pr, \mathbf{R}, \boldsymbol{\lambda}^{*(L_j)}, \{L_j\}\right)$ of each user in set L_j is:

$$\begin{aligned} \bar{P}_{\pi_{L_j}(i)}\left(Pr, \mathbf{R}, \boldsymbol{\lambda}^{*(L_j)}, \{L_j\}\right) &= E_{\mathbf{h}\in\mathcal{H}_{tran}}\left[P_{\pi_{L_j}(i),\boldsymbol{\lambda}^{*(L_j)}}(\mathbf{R},\mathbf{h})\right] \\ &= E_{\mathbf{h}^{(H_j)}\in\mathcal{H}_{tran}^{(H_j)}}\left\{E_{\mathbf{h}^{(L_j)}}\left[P_{\pi_{L_j}(i),\boldsymbol{\lambda}^{*(L_j)}}(\mathbf{R},\mathbf{h})\right]\right\} \\ &= (1 - Pr) \cdot E_{\mathbf{h}^{(L_j)}}\left[P_{\pi_{L_j}(i),\boldsymbol{\lambda}^{*(L_j)}}\left(\mathbf{R},\mathbf{h}^{(L_j)}\right)\right]. \end{aligned}$$

Therefore, regardless of the power allocation policy for users in H_j, the required average power of each user in L_j will remain fixed once L_j is created. Moreover, we see from the description of the algorithm that when the partitioning into H_j and L_j occurs, the minimum of the normalized average powers in H_j must be greater than the maximum of the normalized average powers in L_j. It can also be seen that for each j $(1 \le j \le I-1)$, the minimum of the normalized average powers in H_j (high priority users) must monotonically increase after the iteration when L_j is formed. Therefore, at any iteration after L_j is created, the normalized average power of any user in H_j must be greater than that of any user in $L_1 \cup \cdots \cup L_j$. In other words, the normalized average power of any user in L_j must be greater than that of any user in L_i for all $j > i$. Therefore, for the nesting $\mathcal{L}^*$ given in (D.6), it is clear that if $\boldsymbol{\lambda}(n)$ is the power price vector of the M users at iteration n, then $\lim_{n\to\infty} \frac{\bar{P}_i(Pr, \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}^*)}{\bar{P}_i^*}$ exists for each user $i \in L_1 \cup \cdots \cup L_{I-1}$.

As for the limiting behavior of the normalized average powers of users in $L_I \equiv G_1$, the final set of users for which no further splitting occurs, it can be shown by using exactly the same argument as in [3] that they all converge to a common limit. Therefore, $\lim_{n\to\infty} \frac{\bar{P}_i(Pr, \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}^*)}{\bar{P}_i^*}$ exists for each user i, $1 \le i \le M$, and the infimum in (6.76) is similar to that of (D.7), i.e.,

$$\inf_{(\boldsymbol{\lambda},\mathcal{L})} \max_{1\le i\le M} \frac{\bar{P}_i(Pr, \mathbf{R}, \boldsymbol{\lambda}, \mathcal{L})}{\bar{P}_i^*} \tag{D.9}$$

$$= \max_{1\le i\le M} \lim_{n\to\infty} \frac{\bar{P}_i(Pr, \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}^*)}{\bar{P}_i^*}$$

$$= \lim_{n\to\infty} \frac{\bar{P}_i(Pr, \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}^*)}{\bar{P}_i^*}, \quad \begin{cases} \forall i \in G_1, & \text{if } \mathcal{L}^* \neq \{G\}, \\ \forall i \in G, & \text{if } \mathcal{L}^* = \{G\}. \end{cases} \tag{D.10}$$

Thus, $\mathcal{L}^*$ given in (D.6) is the optimal absolute priority nesting that solves the optimization problem (6.76).

D.8 The HT^{**} Algorithm

Given an average power constraint vector $\bar{\mathbf{P}}^*$, a rate vector $\mathbf{R}$, and an average outage probability vector $\mathbf{Pr}$, the HT^{**} *Algorithm* provides the optimal power allocation parameter pair $(\boldsymbol{\lambda}^*, \mathcal{L}^*)$ that achieves the infimum in (6.77). The description of this algorithm is the same as that for the *HT Algorithm* given in Appendix D.4, except that the average power vector $\bar{\mathbf{P}}(\mathbf{R}, \boldsymbol{\lambda}, \mathcal{L})$ in the *HT Algorithm* is now replaced by $\bar{\mathbf{P}}(\mathbf{Pr}, \mathbf{R}, \boldsymbol{\lambda}, \mathcal{L})$, which can be calculated from (6.71) for $\mathcal{L} = \{G\}$ and from (6.72) for $\mathcal{L} \neq \{G\}$ as shown in Section 6.5.4. After a finite number of iterations of this algorithm, the users will be partitioned into subsets $L_1, L_2, \cdots$, and L_I, where users in L_i are given absolute priority over users in L_j for $i > j$ and users in L_I are given the highest priority, and such that no further partitioning of L_I will take place. We will show that in this case, the optimal absolute priority nesting $\mathcal{L}^*$ that solves the optimization problem (6.77) is still as given in (D.6) and the infimum in (6.77) is similar to that of (D.7).

For $1 \le j \le I-1$, let $(\lambda_i^*)_{i \in L_j}$ be the power prices of the users in L_j when L_j is created and they are the components of vector $\boldsymbol{\lambda}^{*(L_j)}$, and let $H_j \triangleq L_{j+1} \cup \cdots \cup L_I$, which represents the subset of users who are given higher priority than users in L_j at all fading states. The users in $L_1 \cup \cdots \cup L_{j-1}$ will be given lower priority than users in L_j at all fading states. Therefore, there is no interference from any user in $L_1 \cup \cdots \cup L_{j-1}$ to users in L_j. $\forall \mathbf{h} \in \mathcal{H}_{all}$, the optimal power allocation for each subset $S \in \{S_k\}_{k=1}^{2^{|L_j|}-1}$ of the users in L_j is given in (6.73), i.e.,

$$P_{\hat{\rho}_s(i), \boldsymbol{\lambda}^{*(L_j)}}(\mathbf{R}, \mathbf{h}, S) = \begin{cases} \frac{\sigma^2}{h_{\hat{\rho}_s(1)}} \cdot \exp\left(2 \sum_{k \in S_j} R_k\right) \cdot \left[\exp(2R_{\hat{\rho}_s(1)}) - 1\right], & \text{if } i = 1, \\ \frac{\sigma^2}{h_{\hat{\rho}_s(i)}} \cdot \exp\left(2 \sum_{k \in S_j} R_k\right) \cdot \left[\exp\left(2 \sum_{k=1}^{i} R_{\hat{\rho}_s(k)}\right) \right. \\ \qquad \left. - \exp\left(2 \sum_{k=1}^{i-1} R_{\hat{\rho}_s(k)}\right)\right], & \forall 2 \le i \le |L_j|, \end{cases}$$

where S_j denotes the subset of users in H_j who are transmitting information in state $\mathbf{h}$, and the permutation $\hat{\rho}_s(\cdot)$ of the $|S|$ users satisfies:

$$\frac{\lambda^*_{\hat{\rho}_s(1)}}{h_{\hat{\rho}_s(1)}} \ge \frac{\lambda^*_{\hat{\rho}_s(2)}}{h_{\hat{\rho}_s(2)}} \ge \cdots \ge \frac{\lambda^*_{\hat{\rho}_s(|S|)}}{h_{\hat{\rho}_s(|S|)}}.$$

Thus, the required average power $\bar{P}_{\hat{\rho}_s(i)}\left(\mathbf{Pr}, \mathbf{R}, \boldsymbol{\lambda}^{*(L_j)}, \{L_j\}\right)$ of each user in set L_j is:

$$\bar{P}_{\hat{\rho}_s(i)}\left(\mathbf{Pr}, \mathbf{R}, \boldsymbol{\lambda}^{*(L_j)}, \{L_j\}\right) = E_{\mathbf{h}}\left[\sum_{k=1}^{2^{|L_j|}-1} w_{\boldsymbol{\mu}^{*(L_j)}, \boldsymbol{\lambda}^{*(L_j)}}(\mathbf{R}, \mathbf{h}, S_k) \cdot P_{\hat{\rho}_s(i), \boldsymbol{\lambda}^{*(L_j)}}(\mathbf{R}, \mathbf{h}, S_k)\right],$$

where, as discussed in Section 6.5.4, $\forall \mathbf{h} \in \mathcal{H}_{all}$, $\left\{w_{\boldsymbol{\mu}^{*(L_j)}, \boldsymbol{\lambda}^{*(L_j)}}(\mathbf{R}, \mathbf{h}, S_k)\right\}_{k=1}^{2^{|L_j|}-1}$ is obtained by applying the *Independent Outage Transmission Policy* to the $|L_j|$ users instead of to the M users with $s^* = 1$. Therefore, the required average power of each user in L_j depends on the power allocation policy for users in H_j. However, the required average power of each user in L_j will not increase when the power allocation policy for users in H_j is adjusted as described in the algorithm. This is because the transmit powers of users in H_j can be viewed as the background noise for each user in L_j and such noise will not increase when the normalized average powers of users in H_j are balanced as much as possible by adjusting the power prices of users in H_j. Moreover, we see from the description of the algorithm that when the partitioning into H_j and L_j occurs, the minimum of the normalized average powers in H_j must be greater than the maximum of the normalized average powers in L_j. It can also be seen that for each j $(1 \leq j \leq I-1)$, the minimum of the normalized average powers in H_j (high priority users) must monotonically increase after the iteration when L_j is formed. Therefore, at any iteration after L_j is created, the normalized average power of any user in H_j must be greater than that of any user in $L_1 \cup \cdots \cup L_j$. That is, the normalized average power of any user in L_j must be greater than that of any user in L_i for all $j > i$. Therefore, it is clear that if $\boldsymbol{\lambda}(n)$ is the power price vector of the M users at iteration n, then $\lim_{n\to\infty} \dfrac{\bar{P}_i(\mathbf{Pr}, \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}^*)}{\bar{P}_i^*}$ exists for each user $i \in L_1 \cup \cdots \cup L_{I-1}$.

As for the limiting behavior of the normalized average powers of users in $L_I \equiv G_1$, the final set of users for which no further splitting occurs, it can be shown by using exactly the same argument as in [3] that they all converge to a common limit. Therefore, $\lim_{n\to\infty} \dfrac{\bar{P}_i(\mathbf{Pr}, \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}^*)}{\bar{P}_i^*}$ exists for each user i, $1 \leq i \leq M$, and the infimum in (6.77) is similar to that of (D.7), i.e.,

$$\inf_{(\boldsymbol{\lambda},\mathcal{L})} \max_{1\leq i\leq M} \frac{\bar{P}_i(\mathbf{Pr}, \mathbf{R}, \boldsymbol{\lambda}, \mathcal{L})}{\bar{P}_i^*} \tag{D.11}$$

$$= \max_{1\leq i\leq M} \lim_{n\to\infty} \frac{\bar{P}_i(\mathbf{Pr}, \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}^*)}{\bar{P}_i^*}$$

$$= \lim_{n\to\infty} \frac{\bar{P}_i(\mathbf{Pr}, \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}^*)}{\bar{P}_i^*}, \begin{cases} \forall i \in G_1, & \text{if } \mathcal{L}^* \neq \{G\}, \\ \forall i \in G, & \text{if } \mathcal{L}^* = \{G\}. \end{cases} \tag{D.12}$$

Thus, $\mathcal{L}^*$ given in (D.6) is the optimal absolute priority nesting that solves the optimization problem (6.77).

D.9 Determination of Users with Zero-Outage

When independent outage declaration is allowed for each user, given the average power constraint vector $\bar{\mathbf{P}}^*$ and rate vector $\mathbf{R}$, we now show how to determine those users in *Algorithm 6.3* who can support their given rates without any outage if the infimum in (6.75) is larger than 1.

Assuming that the infimum in (6.75) is larger than 1, if the optimal $(\boldsymbol{\lambda}^*, \mathcal{L}^*)$ obtained from the *HT* algorithm satisfies $\mathcal{L}^* = \{G\}$, then as shown in Appendix D.4, $\frac{\bar{P}_i(\mathbf{R},\boldsymbol{\lambda}^*,\{G\})}{\bar{P}_i^*}$ equals the infimum for each user i, $1 \le i \le M$. Thus,

$$\frac{\bar{P}_i(\mathbf{R},\boldsymbol{\lambda}^*,\{G\})}{\bar{P}_i^*} > 1, \quad \forall 1 \le i \le M.$$

However, if $\mathcal{L}^* \neq \{G\}$ and $\mathcal{L}^* = \{G_1, G_2, \cdots, G_I \equiv G\}$ $(I \geq 2)$ as given in Appendix D.4, then $\frac{\bar{P}_i(\mathbf{R},\boldsymbol{\lambda}^*,\mathcal{L}^*)}{\bar{P}_i^*}$ equals the infimum in (6.75) for each user $i \in G_1$ and $\frac{\bar{P}_j(\mathbf{R},\boldsymbol{\lambda}^*,\mathcal{L}^*)}{\bar{P}_j^*}$ is less than the infimum for each user $j \notin G_1$, where G_1 denotes the set of users with the highest priority. Therefore, even if the infimum in (6.75) is larger than 1, it is possible that for some user $j \notin G_1$,

$$\frac{\bar{P}_j(\mathbf{R},\boldsymbol{\lambda}^*,\mathcal{L}^*)}{\bar{P}_j^*} \le 1.$$

Thus, user j can support his given rate without any outage under his average power constraint.

As shown in Appendix D.4, for $\mathcal{L}^* \neq \{G\}$, after a finite number of iterations of the *HT Algorithm*, the M users will be partitioned into subsets $L_1, L_2, \cdots$, and L_I, where users in L_i are given absolute priority over users in L_j for $i > j$ and $G_1 = L_I$, $G_j = \cup_{I-j+1 \le i \le I} L_i$, $\forall 2 \le j \le I$. Since power prices $(\lambda_k^*)_{k \in L_i}$ $(1 \le i \le I-1)$ do not necessarily represent the optimal decoding order that minimizes the maximum normalized average power (normalized by each user's given average power constraint)

$$\max_{k \in L_i} \frac{\bar{P}_k(\mathbf{R},\boldsymbol{\lambda},\mathcal{L})}{\bar{P}_k^*},$$

the *HT* algorithm can be applied to the $|L_i|$ users instead of to all M users to obtain the optimal decoding order for the users in set L_i, $1 \le i \le I-1$ [86]. Once these truly optimal power prices are obtained for the users in each set L_i, $1 \le i \le I-1$, it is easy to find out all the users that can support their given rates without any outage under their average power constraints, since they will have a normalized average power no greater than 1 with the truly optimal power prices.

D.10 Derivation of $\mathbf{Pr}(n)$ in Algorithm 6.3

In *Step n* (1) of *Algorithm 6.3*, given the power allocation parameter pair $(\boldsymbol{\lambda}(n-1), \mathcal{L}(n-1))$ and usage reward vector $\boldsymbol{\mu}(n-1)$, we now show how to obtain the average outage probability vector $\mathbf{Pr}(n)$ that satisfies (6.84).

If $n = 1$ or $\mathcal{L}(n-1) = \{G\}$, then the M components of vector $\mathbf{Pr}(n)$ are:

$$Pr_i(n) = 1 - E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w_{\boldsymbol{\mu}(n-1),\boldsymbol{\lambda}(n-1)}(\mathbf{R}, \mathbf{h}, S_k) \cdot \mathbf{1}[i \in S_k]\right], \quad 1 \le i \le M,$$

where $\forall \mathbf{h} \in \mathcal{H}_{all}$, $\left\{w_{\boldsymbol{\mu}(n-1),\boldsymbol{\lambda}(n-1)}(\mathbf{R}, \mathbf{h}, S_k)\right\}_{k=1}^{2^M-1}$ is determined from the *Independent Outage Transmission Policy* given in Section 6.5.3 once the threshold power s^* is known. Now we choose s^* such that (6.84) is satisfied, since $\bar{P}_i(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}(n-1), \{G\})$ $(1 \le i \le M)$ is a function of $w_{\boldsymbol{\mu}(n-1),\boldsymbol{\lambda}(n-1)}(\mathbf{R}, \mathbf{h}, S_k)$ $(1 \le k \le 2^M - 1)$ and therefore a function of s^*, i.e.,

$$\begin{aligned} &\bar{P}_i(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}(n-1), \{G\}) \\ = \; & E_{\mathbf{h}}\left[\sum_{k=1}^{2^M-1} w_{\boldsymbol{\mu}(n-1),\boldsymbol{\lambda}(n-1)}(\mathbf{R}, \mathbf{h}, S_k) \cdot P_{i,\boldsymbol{\lambda}(n-1)}(\mathbf{R}, \mathbf{h}, S_k)\right], \end{aligned}$$

with $P_{i,\boldsymbol{\lambda}(n-1)}(\mathbf{R}, \mathbf{h}, S_k)$ $(1 \le i \le M)$ given in (6.59).

In the case where $n > 1$ and $\mathcal{L}(n-1) \ne \{G\}$, since $\mathcal{L}(n-1)$ is a set of nested subsets of $G = \{1, 2, \cdots, M\}$, let the subsets be denoted as $G_1, G_2, \cdots, G_I$ $(I \ge 2)$, where $G_1 \subseteq G_2 \subseteq \cdots \subseteq G_I \equiv G$, and let $U_1 \triangleq G_1$, $U_j \triangleq G_j \backslash G_{j-1}$, $\forall 2 \le j \le I$. The average outage probability $Pr_i(n)$ of each user i $(1 \le i \le M)$ in different subsets is given as follows.

(a) For each user $i \in G_{I-1}$, where $G_{I-1} = \cup_{1 \le j \le I-1} U_j$, the power allocation policy remains unchanged. Thus,

$$Pr_i(n) = Pr_i(n-1), \quad \forall i \in G_{I-1}, \quad n > 1.$$

(b) For each user $i \in U_j$ with $j = I$,

$$Pr_i(n) = 1 - E_{\mathbf{h}}\left[\sum_{k=1}^{2^{|U_j|}-1} w_{\boldsymbol{\mu}^{(U_j)}(n-1),\boldsymbol{\lambda}^{(U_j)}(n-1)}(\mathbf{R}, \mathbf{h}, S_k) \cdot \mathbf{1}[i \in S_k]\right], \tag{D.13}$$

where $\forall \mathbf{h} \in \mathcal{H}_{all}$, $\left\{ w_{\boldsymbol{\mu}^{(U_j)}(n-1), \boldsymbol{\lambda}^{(U_j)}(n-1)}(\mathbf{R}, \mathbf{h}, S_k) \right\}_{k=1}^{2^{|U_j|}-1}$ is determined by applying the *Independent Outage Transmission Policy* in Section 6.5.3 to the $|U_j|$ users instead of to the M users once the threshold power s^* is known. Now we choose s^* such that

$$\max_{i \in U_j} \frac{\bar{P}_i(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}(n-1), \mathcal{L}(n-1))}{\bar{P}_i^*} = 1, \tag{D.14}$$

where $\bar{P}_i(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}(n-1), \mathcal{L}(n-1))$ ($\forall i \in U_j$) is a function of

$$\left\{ w_{\boldsymbol{\mu}^{(U_j)}(n-1), \boldsymbol{\lambda}^{(U_j)}(n-1)}(\mathbf{R}, \mathbf{h}, S_k) \right\}_{k=1}^{2^{|U_j|}-1}$$

and therefore a function of s^* and is given by

$$\begin{aligned} & \bar{P}_i(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}(n-1), \mathcal{L}(n-1)) \\ = \; & E_{\mathbf{h}} \left[\sum_{k=1}^{2^{|U_j|}-1} w_{\boldsymbol{\mu}^{(U_j)}(n-1), \boldsymbol{\lambda}^{(U_j)}(n-1)}(\mathbf{R}, \mathbf{h}, S_k) \cdot P_{i,(\boldsymbol{\lambda}(n-1), \mathcal{L}(n-1))}(\mathbf{R}, \mathbf{h}, S_k) \right], \end{aligned}$$

with $P_{i,(\boldsymbol{\lambda}(n-1), \mathcal{L}(n-1))}(\mathbf{R}, \mathbf{h}, S_k)$ ($\forall i \in U_j$) given in (6.73).

D.11 Proof of Theorem 6.2

Here we first prove that the sequence $\{Inf(\mathbf{Pr}(n))\}$ ($n \geq 1$) obtained from *Algorithm 6.3* is monotonically increasing and it will converge to $Inf(\mathbf{Pr}^*) = 1$.

$\forall n \geq 1$, we know from *Step n (1)* that for the given power allocation parameter pair $(\boldsymbol{\lambda}(n-1), \mathcal{L}(n-1))$ and channel usage reward vector $\boldsymbol{\mu}(n-1)$, (6.84) holds when the average outage probability vector is $\mathbf{Pr}(n)$. Given the average outage probability vector $\mathbf{Pr}(n)$, it is clear from *Step n (2)* that $(\boldsymbol{\lambda}(n), \mathcal{L}(n))$ satisfies

$$\begin{aligned} \max_{1 \leq i \leq M} \frac{\bar{P}_i(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}(n))}{\bar{P}_i^*} &= Inf(\mathbf{Pr}(n)) \\ &= \inf_{(\boldsymbol{\lambda}, \mathcal{L})} \max_{1 \leq i \leq M} \frac{\bar{P}_i(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}, \mathcal{L})}{\bar{P}_i^*} \\ &\leq \frac{\bar{P}_i(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}(n-1), \mathcal{L}(n-1))}{\bar{P}_i^*} \\ &= 1, \end{aligned} \tag{D.15}$$

where the last equality is due to (6.84). Note that in *Step n (2)* , given the average outage probability vector $\mathbf{Pr}(n)$ and the power allocation parameter pair $(\boldsymbol{\lambda}(n), \mathcal{L}(n))$, $\bar{\mathbf{P}}(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}(n))$ is a boundary average power vector of region $APV_{out}(\mathbf{Pr}(n), \mathbf{R})$. Therefore, when $\mathcal{L}(n) = \{G\}$, components of $\bar{\mathbf{P}}(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}(n), \{G\})$ are implicit functions of a channel usage reward vector $\boldsymbol{\mu}(n)$ and a threshold power s^* (we will refer to it as $s^*(G, n)$) as given in (6.71), with $\boldsymbol{\mu}^*$ and $\boldsymbol{\lambda}$ in (6.71) replaced by $\boldsymbol{\mu}(n)$ and $\boldsymbol{\lambda}(n)$[1]. By fixing $s^*(G, n) = 1$, $\boldsymbol{\mu}(n)$ is chosen such that the resulting average outage probability vector equals $\mathbf{Pr}(n)$. When $\mathcal{L}(n) \neq \{G\}$, as stated in Appendix D.10, we denote the nested subsets of G in $\mathcal{L}(n)$ as $G_1, G_2, \cdots, G_I$ $(I \geq 2)$, where $G_1 \subseteq G_2 \subseteq \cdots \subseteq G_I \equiv G$, and define $U_1 = G_1$, $U_j = G_j \backslash G_{j-1}$, $2 \leq j \leq I$. Then components of subvector $\bar{\mathbf{P}}^{(U_j)}(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}(n))$ are implicit functions of a channel usage reward vector $\boldsymbol{\mu}^{(U_j)}(n)$ and a threshold power s^* (we will refer to it as $s^*(U_j, n)$) as given in (6.72), with $\boldsymbol{\mu}^{*(U_j)}$ and $(\boldsymbol{\lambda}, \mathcal{L})$ in (6.72) replaced by $\boldsymbol{\mu}^{(U_j)}(n)$ and $(\boldsymbol{\lambda}(n), \mathcal{L}(n))$[2]. By fixing $s^*(U_j, n) = 1$, $\boldsymbol{\mu}^{(U_j)}(n)$ is chosen such that the resulting average outage probability of each user $i \in U_j$ equals Pr_i^*.

Now in *Step* $n+1$ *(1)*, an average outage probability vector $\mathbf{Pr}(n+1)$ is to be chosen such that (6.84) holds with n replaced by $n+1$, i.e., the required average power vector $\bar{\mathbf{P}}(\mathbf{Pr}(n+1), \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}(n))$ satisfies

$$\max_{1 \leq i \leq M} \frac{\bar{P}_i(\mathbf{Pr}(n+1), \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}(n))}{\bar{P}_i^*} = 1. \tag{D.16}$$

That is, for the given rate vector $\mathbf{R}$ and the same power allocation parameter pair $(\boldsymbol{\lambda}(n), \mathcal{L}(n))$ and channel usage reward vector $\boldsymbol{\mu}(n)$, if $\mathcal{L}(n) = \{G\}$, we choose a new threshold power $s^*(G, n+1)$ (vector $\mathbf{Pr}(n+1)$ is then determined) for the M users such that (D.16) is satisfied. Comparison of (D.15) and (D.16) implies that the threshold power $s^*(G, n+1)$ must be no less than $s^*(G, n)$ such that the M users will transmit in more fading states based on the *Independent Outage Transmission Policy* in Section 6.5.3. That is, $s^*(G, n+1) \geq s^*(G, n) = 1$ and if $s^*(G, n+1) > 1$, the resulting average outage probability vector $\mathbf{Pr}(n+1)$ must satisfy

$$\mathbf{Pr}(n+1) \notin \mathcal{O}\left(\bar{\mathbf{P}}(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}(n)), \mathbf{R}\right). \tag{D.17}$$

[1]In (6.71) $\bar{\mathbf{P}}_{\boldsymbol{\lambda}}(\mathbf{R})$ is used to denote the boundary power vector of region $APV_{out}(\mathbf{Pr}, \mathbf{R})$ instead of $\bar{\mathbf{P}}(\mathbf{Pr}, \mathbf{R}, \boldsymbol{\lambda}, \{G\})$.

[2]In (6.72) $\bar{\mathbf{P}}^{(U_j)}_{(\boldsymbol{\lambda}, \mathcal{L})}(\mathbf{R})$ is used to denote the boundary power vector of region $APV_{out}\left(\mathbf{Pr}^{(U_j)}, \mathbf{R}^{(U_j)}\right)$ instead of $\bar{\mathbf{P}}^{(U_j)}(\mathbf{Pr}, \mathbf{R}, \boldsymbol{\lambda}, \mathcal{L})$ for simplicity.

Since $\mathcal{L}(n) = \{G\}$, we know from the *HT*** *Algorithm* in Appendix D.8 that

$$f_n \triangleq Inf(\mathbf{Pr}(n)) = \frac{\bar{P}_i(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}(n))}{\bar{P}_i^*}, \quad \forall 1 \leq i \leq M,$$

which implies that $\bar{\mathbf{P}}(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}(n)) = f_n \cdot \bar{\mathbf{P}}^*$ and (D.17) is equivalent to

$$\mathbf{Pr}(n+1) \notin \mathcal{O}\left(f_n \cdot \bar{\mathbf{P}}^*, \mathbf{R}\right). \tag{D.18}$$

By denoting the infimum in (6.77) as $Inf_0(\mathbf{Pr}(n+1))$ when $\bar{P}_i^*$ and Pr_i $(1 \leq i \leq M)$ in (6.77) are replaced by $f_n \cdot \bar{P}_i^*$ and $Pr_i(n+1)$, respectively, it is clear from (D.18) that

$$Inf_0(\mathbf{Pr}(n+1)) > 1. \tag{D.19}$$

Since

$$Inf_0(\mathbf{Pr}(n+1)) = \frac{Inf(\mathbf{Pr}(n+1))}{f_n}, \tag{D.20}$$

substituting (D.20) into (D.19) we obtain $Inf(\mathbf{Pr}(n+1)) > f_n$, i.e.,

$$Inf(\mathbf{Pr}(n+1)) > Inf(\mathbf{Pr}(n)).$$

If $s^*(G, n+1) = s^*(G, n)$, then obviously $\mathbf{Pr}(n+1) = \mathbf{Pr}(n)$ and $Inf(\mathbf{Pr}(n+1)) = Inf(\mathbf{Pr}(n))$. Thus, for $s^*(G, n+1) \geq s^*(G, n)$,

$$Inf(\mathbf{Pr}(n+1)) \geq Inf(\mathbf{Pr}(n)). \tag{D.21}$$

If $\mathcal{L}(n) \neq \{G\}$, as stated in Appendix D.10, suppose $\mathcal{L}(n) = \{G_1, G_2, \cdots, G_I\}$ $(I \geq 2)$, $U_1 = G_1$, and $U_j = G_j \backslash G_{j-1}$, $2 \leq j \leq I$. Then users in U_i are given absolute priority over users in U_j for $i < j$ at all fading states and users in U_I are given the lowest priority. That is, signals from users in U_I are decoded before that of users in $G_{I-1} = U_1 \cup \cdots \cup U_{I-1}$ in every fading state $\mathbf{h}$ and are subtracted out from the total received signal. Therefore, there will be no interference from users in U_I to any user in G_{I-1} and they can be viewed as nonexistent to users in G_{I-1}. For the same power allocation parameter pair $(\boldsymbol{\lambda}(n), \mathcal{L}(n))$ and channel usage reward vector $\boldsymbol{\mu}(n)$, since we keep the threshold power for users in G_{I-1} unchanged, i.e., $s^*(U_j, n+1) = s^*(U_j, n)$, $\forall 1 \leq j \leq I-1$, it must be true that

$\mathbf{Pr}^{(U_j)}(n+1) = \mathbf{Pr}^{(U_j)}(n)$, $\forall 1 \leq j \leq I-1$. Now $\mathbf{Pr}^{(U_I)}(n+1)$ is determined by choosing a new threshold power $s^*(U_I, n+1)$ for the users in U_I such that

$$\max_{i \in U_I} \frac{\bar{P}_i(\mathbf{Pr}(n+1), \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}(n))}{\bar{P}_i^*} = 1. \tag{D.22}$$

Since the *HT*** *Algorithm* given in Appendix D.8 indicates that

$$\max_{i \in U_I} \frac{\bar{P}_i(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}(n))}{\bar{P}_i^*} < Inf(\mathbf{Pr}(n)) \leq 1, \tag{D.23}$$

where the last inequality is due to (D.15), comparison of (D.22) and (D.23) implies that the threshold power $s^*(U_I, n+1)$ must be larger than $s^*(U_I, n)$ such that users in U_I will transmit in more fading states based on the *Independent Outage Transmission Policy* applied to the $|U_I|$-user system. Therefore, given $s^*(U_I, n+1) > s^*(U_I, n)$, the resulting average outage probability subvector $\mathbf{Pr}^{(U_I)}(n+1)$ must satisfy

$$\mathbf{Pr}^{(U_I)}(n+1) \notin \mathcal{O}_I\left(\bar{\mathbf{P}}^{(U_I)}(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}(n)), \mathbf{R}^{(U_I)}\right).$$

That is,

$$\mathbf{Pr}(n+1) \notin \mathcal{O}_I\left(\bar{\mathbf{P}}(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}(n)), \mathbf{R}\right),$$

and it must be true that $Inf(\mathbf{Pr}(n+1)) \geq Inf(\mathbf{Pr}(n))$, which will be shown below.

From the *HT*** *Algorithm* given in Appendix D.8, we know that $(\boldsymbol{\lambda}(n), \mathcal{L}(n))$ solves the problem (6.77) for the given average outage probability vector $\mathbf{Pr}(n)$ and the infimum in (6.77) is:

$$Inf(\mathbf{Pr}(n)) = \frac{\bar{P}_i(\mathbf{Pr}(n), \mathbf{R}, \boldsymbol{\lambda}(n), \mathcal{L}(n))}{\bar{P}_i^*}, \quad \forall i \in U_1.$$

The users in U_1 are given the highest priority, i.e., there is no interference from any user in U_1^c to users in U_1. Thus,

$$Inf(\mathbf{Pr}(n)) = \inf_{\boldsymbol{\lambda}^{(U_1)}} \max_{i \in U_1} \frac{\bar{P}_i\left(\mathbf{Pr}^{(U_1)}(n), \mathbf{R}^{(U_1)}, \boldsymbol{\lambda}^{(U_1)}\right)}{\bar{P}_i^*}. \tag{D.24}$$

where $\bar{\mathbf{P}}^{(U_1)}\left(\mathbf{Pr}^{(U_1)}(n), \mathbf{R}^{(U_1)}, \boldsymbol{\lambda}^{(U_1)}\right)$ denotes the corresponding boundary power vector of the $|U_1|$-dimensional set $APV_{out}\left(\mathbf{Pr}^{(U_1)}(n), \mathbf{R}^{(U_1)}\right)$ for the power price vector $\boldsymbol{\lambda}^{(U_1)}$.

Therefore, for any given average outage probability vector $\mathbf{Pr}(n+1)$, so long as $Pr_i(n+$

$1) = Pr_i(n)$, $\forall i \in U_1$, it is always true that

$$\begin{aligned} Inf(\mathbf{Pr}(n+1)) &= \inf_{(\boldsymbol{\lambda},\mathcal{L})} \max_{1\le i\le M} \frac{\bar{P}_i(\mathbf{Pr}(n+1),\mathbf{R},\boldsymbol{\lambda},\mathcal{L})}{\bar{P}_i^*} \\ &\ge \inf_{\boldsymbol{\lambda}^{(U_1)}} \max_{i\in U_1} \frac{\bar{P}_i(\mathbf{Pr}^{(U_1)}(n+1),\mathbf{R}^{(U_1)},\boldsymbol{\lambda}^{(U_1)})}{\bar{P}_i^*} \\ &= \inf_{\boldsymbol{\lambda}^{(U_1)}} \max_{i\in U_1} \frac{\bar{P}_i(\mathbf{Pr}^{(U_1)}(n),\mathbf{R}^{(U_1)},\boldsymbol{\lambda}^{(U_1)})}{\bar{P}_i^*} \\ &= Inf(\mathbf{Pr}(n)), \end{aligned} \tag{D.25}$$

where the last equality is due to (D.24). From (D.21) and (D.25) it is clear that $\{Inf(\mathbf{Pr}(n))\}$ $(n \ge 1)$ is a monotonically increasing sequence.

Since each iteration in *Algorithm 6.3* can be represented as a map

$$\begin{aligned} T : \Re_+ &\rightarrow \Re_+ \\ Inf(\mathbf{Pr}(n)) &\mapsto Inf(\mathbf{Pr}(n+1)), \end{aligned}$$

any fixed point of T will be the solution $Inf(\mathbf{Pr}^*)$ satisfying $Inf(\mathbf{Pr}^*) = 1$. From (D.15) it is clear that $\forall n \ge 1$, $Inf(\mathbf{Pr}(n)) \le 1$. Therefore, $\{Inf(\mathbf{Pr}(n)\}$ $(n \ge 1)$ is a monotonically increasing sequence bounded from above and must converge to a limit. The limit has to be a fixed point of T, and hence the solution $Inf(\mathbf{Pr}^*)$ that satisfies $Inf(\mathbf{Pr}^*) = 1$. This implies that sequence $\{\mathbf{Pr}(n)\}$ $(n \ge 1)$ converges to an average outage probability vector $\mathbf{Pr}^*$ and the sequence $\{(\boldsymbol{\lambda}(n), \mathcal{L}(n))\}$ converges to $(\boldsymbol{\lambda}^*, \mathcal{L}^*)$ that solves (6.77) with the average outage probability vector equal to $\mathbf{Pr}^*$. We now show that vector $\mathbf{Pr}^*$ is a regular point on the boundary surface of region $\mathcal{O}(\bar{\mathbf{P}}^*, \mathbf{R})$, i.e., $\mathcal{L}^* = \{G\}$ and (6.85) holds. We prove this by contradiction.

Suppose $\mathcal{L}^* \ne \{G\}$ and denote the set of users with the lowest priority as U_I. Then for the given average outage probability vector $\mathbf{Pr}^*$, from the HT^{**} *Algorithm* given in Appendix D.8 we know that

$$\max_{i\in U_I} \frac{\bar{P}_i(\mathbf{Pr}^*,\mathbf{R},\boldsymbol{\lambda}^*,\mathcal{L}^*)}{\bar{P}_i^*} < Inf(\mathbf{Pr}^*) = 1.$$

Let $s^*(U_I)$ denote the corresponding threshold for the $|U_I|$ users. Now in the next iteration of *Algorithm 6.3*, while the power allocation policy remains the same for all other users, a larger threshold will be chosen for the $|U_I|$ users such that the resulting average outage probability vector $\mathbf{Pr}^{**}$ satisfies

$$\max_{i \in U_I} \frac{\bar{P}_i(\mathbf{Pr}^{**}, \mathbf{R}, \boldsymbol{\lambda}^*, \mathcal{L}^*)}{\bar{P}_i^*} = 1.$$

Therefore, $\mathbf{Pr}^{**} \neq \mathbf{Pr}^*$ and $\mathbf{Pr}^*$ is not the fixed point to which *Algorithm 6.3* converges. This contradicts the fact that the sequence $\{\mathbf{Pr}(n)\}$ in *Algorithm 6.3* converges to $\mathbf{Pr}^*$. Thus, it must be true that $\mathcal{L}^* = \{G\}$ and, consequently, (6.85) holds according to the HT^{**} *Algorithm* given in Appendix D.8. □

Bibliography

[1] R. H. Katz, "Adaptation and mobility in wireless information systems," *IEEE Personal Commun.*, pp. 6–17, First Quarter 1994.

[2] D. N. Tse and S. V. Hanly, "Multiple-access fading channels: Part I: Polymatroidal structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2796–2815, Nov. 1998.

[3] S. V. Hanly and D. N. Tse, "Multiple-access fading channels: Part II: Delay-limited capacities," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2816–2831, Nov. 1998.

[4] J. W. C. Jakes, *Microwave Mobile Communications.* New York: Wiley, 1974.

[5] K. Pahlavan and A. H. Levesque, *Wireless Information Networks.* New York : Wiley, 1995.

[6] G. L. Stüber, *Mobile Communication Systems.* Boston: Kluwer Academic, 1996.

[7] T. S. Rappaport, *Wireless Communications : Principles and Practice.* New Jersey: Prentice Hall, 1996.

[8] E. K. Wesel, *Wireless Multimedia Communications: Networking Video, Voice, and Data.* Massachusetts: Addison-Wesley, 1998.

[9] M.-S. Alouini, "Adaptive and diversity techniques for wireless digital communications over fading channels." Ph.D. Thesis, California Institute of Technology, May 1998.

[10] M. K. Simon and M.-S. Alouini, "Digital communications over generalized fading channels: A unified approach to performance analysis." New York, NY: John Wiley & Sons, to be published.

[11] H. Suzuki, "A statistical model for urban multipath propagation," *IEEE Trans. on Commun.*, vol. 25, pp. 673-680, July 1977.

[12] T. Aulin, "Characteristics of a digital mobile radio channel," *IEEE Trans. Vehic. Technol.*, vol. 30, pp. 45–53, May 1981.

[13] W. R. Braun and U. Dersch, "A physical mobile radio channel model," *IEEE Trans. Vehic. Technol.*, vol. 40, pp. 472–482, May 1991.

[14] A. U. Sheikh, M. Handforth, and M. Abdi, "Indoor mobile radio channel at 946 MHz: Measurements and modeling," in *Proc. IEEE Vehic. Technol. Conf. (VTC'93), Secaucus, NJ*, pp. 73–76, May 1993.

[15] M. Nakagami, "The m-distribution - A general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation*, pp. 3–36, Pergamon Press, Oxford, England, 1960.

[16] M. L. Moher and J. H. Lodge, "TCMP - a modulation and coding strategy for Rician fading channels," *IEEE J. Sel. Area on Commun.*, vol. 7, pp. 1347–1355, Dec. 1989.

[17] S. Sampei and T. Sunaga, "Rayleigh fading compensation for QAM in land mobile radio-communications," *IEEE Trans. Vehic. Technol.*, vol. 42, pp. 137–147, May 1993.

[18] A. Aghamohammadi, H. Meyr, and G. Ascheid, "A new method for phase synchronization and automatic gain-control of linearly modulated signals on frequency-flat fading channels," *IEEE Trans. Commun.*, vol. 39, pp. 25–29, Jan. 1991.

[19] D. Subasinghe-Dias and K. Feher, "A coded 16-QAM scheme for fast fading mobile radio channels," *IEEE Trans. Commun.*, vol. 43, pp. 1906–1916, Feb-Apr. 1995.

[20] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Vehic. Technol.*, vol. 40, pp. 686–693, Nov. 1991.

[21] G. T. Irvine and P. J. Mclane, "Symbol-aided plus decision-directed reception for PSK TCM modulation on shadowed mobile satellite fading channels," *IEEE J. Sel. Area Commun.*, vol. 10, pp. 1289–1299, Oct. 1992.

[22] Y. Liu and S. D. Blostein, "Identification of frequency nonselective fading channels using decision-feedback and adaptive linear prediction," *IEEE Trans. Commun.*, vol. 43, pp. 1484–1492, Feb.-Apr. 1995.

[23] J. P. Seymour and M. P. Fitz, "Two-stage carrier synchronization techniques for nonselective fading," *IEEE Trans. Vehic. Technol.*, vol. 44, pp. 103–110, Feb. 1995.

[24] A. N. D'Andrea, A. Diglio, and U. Mengali, "Symbol-aided channel estimation with nonselective Rayleigh fading channels," *IEEE Trans. Vehic. Technol.*, vol. 44, pp. 41–49, Feb. 1995.

[25] R. Raheli, A. Polydoros, and C.-K. Tzou, "Per-survivor processing (MLSE algorithm approximation)," *Digital Signal Processing: a Review Journal*, vol. 3, pp. 175–178, July 1993.

[26] J. K. Cavers and M. Liao, "A comparison of pilot tone and pilot symbol techniques for digital mobile communication," in *Proc. IEEE Global Telecommun. Conf. (GLOBE-COM'92)*, pp. 915–921, 1992.

[27] P. Y. Kam, "Optimal detection of digital data over the nonselective Rayleigh fading channel with diversity reception," *IEEE Trans. Commun.*, vol. 39, pp. 214–219, Feb. 1991.

[28] R. Haeb and H. Meyr, "A systematic-approach to carrier recovery and detection of digitally phase modulated signals on fading channels," *IEEE Trans. Commun.*, vol. 37, pp. 748–754, July 1989.

[29] P. Y. Kam and H. M. Ching, "Sequence estimation over the slow nonselective Rayleigh fading channel with diversity reception and its application to Viterbi decoding," *IEEE J. Sel. Area Commun.*, vol. 10, pp. 562–570, Apr. 1992.

[30] J. H. Lodge and M. L. Moher, "Maximum-likelihood sequence estimation of CPM signals transmitted over Rayleigh flat-fading channels," *IEEE Trans. Commun.*, vol. 38, pp. 787–794, June 1990.

[31] X. Yu and S. Pasupathy, "Innovations-based MLSE for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 43, pp. 1534–1544, Feb.-Apr. 1995.

[32] A. Aghamohammadi, H. Meyr, and G. Ascheid, "Adaptive synchronization and channel parameter-estimation using an extended kalman filter," *IEEE Trans. Commun.*, vol. 37, pp. 1212–1219, Nov. 1989.

[33] W. C. Dam and D. P. Taylor, "An adaptive maximum-likelihood receiver for correlated Rayleigh-fading channels," *IEEE Trans. Commun.*, vol. 42, pp. 2684–2692, Sept. 1994.

[34] G. M. Vitetta and D. P. Taylor, "Maximum-likelihood decoding of uncoded and coded PSK signal sequences transmitted over Rayleigh flat-fading channels," *IEEE Trans. Commun.*, vol. 43, pp. 2750–2758, Nov. 1995.

[35] G. M. Vitetta and D. P. Taylor, "Viterbi decoding of differentially encoded PSK signals transmitted over Rayleigh frequency-flat fading channels," *IEEE Trans. Commun.*, vol. 43, pp. 256–1259, Feb.-Apr. 1995.

[36] H. Kong and E. Shwedyk, "Adaptive MLSE of QAM signals over frequency nonselective Rayleigh fading channels," *IEE Proc. - Commun.*, vol. 145, pp. 426–430, Dec. 1998.

[37] H. Kong and E. Shwedyk, "On channel estimation and sequence detection of interleaved coded signals over frequency nonselective Rayleigh fading channels," *IEEE Trans. Vihic. Technol.*, vol. 47, pp. 558–565, May 1998.

[38] D. Makrakis, P. T. Mathiopoulos, and D. P. Bouras, "Optimal decoding of coded PSK and QAM signals in correlated fast fading channels and AWGN - A combined envelope, multiple differential and coherent detection approach," *IEEE Trans. Commun.*, vol. 42, pp. 63–75, Jan. 1994.

[39] D. K. Borah and B. D. Hart, "A robust receiver structure for time-varying, frequency-flat, Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, pp. 360–364, Mar. 1999.

[40] G. M. Vitetta, D. P. Taylor, and U. Mengali, "Double-filtering receivers for PSK signals transmitted over Rayleigh frequency-flat fading channels," *IEEE Trans. Commun.*, vol. 44, pp. 686–695, June 1996.

[41] G. M. Vitetta, U. Mengali, and D. P. Taylor, "Double-filter differential detection of PSK signals transmitted over linearly time-selective Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, pp. 239–247, Feb. 1999.

[42] A. Goldsmith and P. Varaiya, "Capacity, mutual information and coding for finite-state Markov channels," *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 868–886, May 1996.

[43] A. Viterbi and J. Omura, *Principles of Digital Communications and Coding.* New York: McGraw-Hill, 1979.

[44] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert-Elliott channels," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 1277–1290, Nov. 1989.

[45] H. Wang and N. Moayeri, "Finite-state Markov channel - A useful model for radio communication channels," *IEEE Trans. Vehic. Technol.*, vol. VT-44, pp. 163–171, Feb. 1995.

[46] Y. L. Guan and L. F. Turner, "Generalised FSMC model for radio channels with correlated fading," *IEE Proc. Commun.*, vol. 146, pp. 133–137, Apr. 1999.

[47] H. Wang and P.-C. Chang, "On verifying the first-order Markovian assumption for a Rayleigh fading channel model," *IEEE Trans. Vehic. Technol.*, vol. VT-45, pp. 353–357, May. 1996.

[48] F. Babich, O. E. Kelly, and G. Lombardi, "A context-tree based model for quantized fading," *IEEE Commun. Lett.*, vol. 3, pp. 46–48, Feb. 1999.

[49] F. Babich, G. Lombardi, and E. Valentinuzzi, "Variable order Markov modelling for LEO mobile satellite channels," *Electron. Lett.*, vol. 35, pp. 621–623, Apr. 1999.

[50] P. Y. Kam and C. H. Teh, "Reception of PSK signals over fading channels via quadrature amplitude estimation," *IEEE Trans. Commun.*, vol. 31, pp. 1024–1027, Aug. 1983.

[51] P. Y. Kam, "An adaptive receiver with memory for slowly fading channels," *IEEE Trans. Commun.*, vol. 32, pp. 654–659, June 1984.

[52] F. Gini and G. Giannakis, "Generalized differential encoding: A nonlinear signal processing perspective," *IEEE Trans. Signal Proces.*, vol. 46, pp. 2967–2974, Nov. 1998.

[53] M. K. Simon, G. K. Huth, and A. Polydoros, "Differentially coherent detection of QASK for frequency-hopping systems - part I: Performance in the presence of a Gaussian-noise environment," *IEEE Trans. Commun.*, vol. 30, pp. 158–164, Jan. 1982.

[54] M. A. Herro, L. Hu, and J. M. Nowack, "Bit error-probability calculations for convolutional-codes with short constraint lengths on very noisy channels," *IEEE Trans. Commun.*, vol. 36, pp. 885–888, July 1988.

[55] B. D. Hart and D. P. Taylor, "Extended MLSE receiver for the frequency-flat, fast-fading channel," *IEEE Trans. Vehic. Technol.*, vol. 46, pp. 381–389, May 1997.

[56] D. Divsalar and M. K. Simon, "Trellis coded modulation for 4800-9600 bits/s transmission over a fading mobile satellite channel," *IEEE J. Select. Areas Commun.*, vol. SAC-5, pp. 162–175, Feb. 1987.

[57] Y.-L. Chen and C.-H. Wei, "On the performance of rate 1/2 convolutional-codes with QPSK on Rician fading channels," *IEEE Trans. Vehic. Technol.*, vol. 39, pp. 161–170, May 1990.

[58] Y.-J. Liu, I. Oka, and E. Biglieri, "Error-probability for digital transmission over nonlinear channels with application to TCM," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1101–1110, Sept. 1990.

[59] R. G. McKay, P. J. McLane, and E. Biglieri, "Error-bounds for trellis-coded MPSK on a fading mobile satellite channel," *IEEE Trans. Commun.*, vol. 39, pp. 1750–1761, Dec. 1991.

[60] J. K. Cavers and P. Ho, "Analysis of the error performance of trellis-coded modulations in Rayleigh-fading channels," *IEEE Trans. Commun.*, vol. 40, pp. 74–83, Jan. 1992.

[61] R. V. Nobelen and D. P. Taylor, "Analysis of the pairwise error probability of non-interleaved codes on the Rayleigh-fading channel," *IEEE Trans. Commun.*, vol. 44, pp. 456–462, Apr. 1996.

[62] J. W. Modestino and S. Y. Mui, "Convolutional code performance in the Rician fading channel," *IEEE Trans. Commun.*, vol. 24, pp. 592–606, June 1976.

[63] Synopsys, Inc., *COSSAP Model Libraries*, Dec. 1994.

[64] E. Biglieri, D. Divsalar, P. J. McLane, and M. K. Simon, *Introduction to Trellis-Coded Modulation with Applications.* New York: Macmillan, 1991.

[65] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55–67, Jan. 1982.

[66] J. Du and B. Vucetic, "Trellis coded 16-QAM for fading channels," *Europ. Trans. Telecomm. Rel. Technol.*, vol. 4, pp. 335–341, May-June 1993.

[67] C. Valdez and H. Yamamoto, "On the performance of TCM with channel state information in frequency flat Rayleigh mobile channels," *IEICE Trans. Commun.*, vol. E77-B, pp. 501–510, Apr. 1994.

[68] S.-G. Chua and A. J. Goldsmith, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, pp. 1218–1230, Oct. 1997.

[69] J. H. Hays, "Adaptive feedback communications," *IEEE Trans. Commun. Tech.*, vol. 16, pp. 29–34, Feb. 1968.

[70] J. K. Cavers, "Variable-rate transmission for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 20, pp. 15–22, Feb. 1972.

[71] B. Vucetic, "An adaptive coding scheme for time-varying channels," *IEEE Trans. Commun.*, vol. 39, pp. 653–663, May 1991.

[72] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1986–1992, Nov. 1997.

[73] D. Hughes-Hartogs, "The capacity of a degraded spectral Gaussian broadcast channel." Ph.D. Thesis and Technical Report No. 7002-2, Information Systems Laboratory, Center for Systems Research, Stanford University, July 1975.

[74] D. N. Tse, "Optimal power allocation over parallel Gaussian broadcast channels," in *Proc. of International Symp. Inform. Theory, Ulm, Germany*, p. 27, June 1997. Also submitted to IEEE Trans. Inform. Theory.

[75] A. J. Goldsmith, "The capacity of downlink fading channels with variable-rate and power," *IEEE Trans. Vehic. Technol.*, vol. 46, pp. 569–580, Aug. 1997.

[76] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1468–1489, July 1999.

[77] S. H. Jamali and T. Le-Ngoc, *Coded-Modulation Techniques for Fading Channels.* Kluwer Academic, 1994.

[78] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. 19, pp. 197–207, Mar. 1973.

[79] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 20, pp. 279–280, Mar. 1974.

[80] W. Feller, *An Introduction to Probability Theory and Its Application.* New York: Wiley, 3rd ed., 1968.

[81] P. Billingsley, *Probability and Measure.* New York : Wiley, 2nd ed., 1986.

[82] A. A. E. Gamal, "Capacity of the product and sum of two unmatched broadcast channels," *Problemy Peredachi Informatsii*, vol. 16, pp. 3–23, Jan.-Mar. 1980.

[83] A. A. E. Gamal, "Results in multiple user channel capacity." Ph.D. Thesis, Stanford University, May 1978.

[84] P. P. Bergmans and T. A. Cover, "Cooperative broadcasting," *IEEE Trans. Inform. Theory*, vol. 20, pp. 317–324, May 1974.

[85] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredachi Informatsii*, vol. 10, pp. 3–14, July-Sept. 1974.

[86] S. V. Hanly and D. N. Tse, "Min-max power allocation for successive decoding," in *Proc. of IEEE Inform. Theory Workshop*, pp. 56–57, 1998.