SERVICE OUTAGE BASED ADAPTIVE TRANSMISSION IN FADING CHANNELS

BY JIANGHONG LUO

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Electrical and Computer Engineering

Written under the direction of

Prof. Roy Yates, Prof. Predrag Spasojević

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2004

ABSTRACT OF THE DISSERTATION

Service Outage Based Adaptive Transmission in Fading Channels

by Jianghong Luo

Dissertation Director: Prof. Roy Yates, Prof. Predrag Spasojević

The service outage based allocation problem explores variable rate transmission schemes and combines the concepts of ergodic capacity and outage capacity for fading channels. The ergodic capacity determines the maximum achievable rate for non real-time applications, and the outage capacity is developed for constant rate real-time applications. Neither is completely appropriate for variable rate multimedia applications for next generation wireless networks. In this context, the service outage based allocation problem is proposed. A service outage occurs when the instantaneous transmission rate is smaller than a basic rate specified by an application. The service outage allocation problem is to find the optimum power allocation that maximizes the average rate subject to a service outage probability constraint and an average power constraint. The optimum power allocation is derived for a single flat fading channel, and is generalized to M parallel fading channels. Two near optimum schemes are also derived for a small outage probability. The minimum outage based near optimum scheme significantly reduces the computational complexity. The allocation problem with respect to the energy efficiency is also examined for $M = 1$ fading channel. In the application of the transmission of mixed real-time and non real-time services in fading channels, the

optimum service outage based allocation can be implemented using an adaptive channel partition approach. The optimum fixed channel partition scheme is derived and compared to the adaptive scheme. In addition, a suboptimum fixed partition scheme called proportional average power partition is studied and observed to be close to the optimum fixed partition scheme in the Rayleigh fading channel.

We also study the performance of variable-rate turbo bit-interleaved coded modulation (Turbo-BICM) with random puncturing. A union-Bhattacharyya rate threshold for the Turbo-BICM based on a reliable channel region for turbo codes transmitted over parallel-channel is derived. A closed form approximation of this rate threshold is determined for an AWGN channel. This rate threshold is shown to predict the Turbo-BICM iterative decoding performance very well. Adaptation of Turbo-BICM in a slow fading channel is studied. The optimum power and modulation allocations are described. A dual problem solution which achieves a rate close to the optimum solution with significantly reduced computational complexity is described. Two simple schemes: water-filling with optimum modulation and equal power allocation with optimum modulation are also presented and shown to achieve a good performance. Proposed adaptive schemes are shown to achieve a rate within $2-3$ db of the ergodic capacity of a Rayleigh fading channel.

Acknowledgements

Many people have been a part of my graduate education, as friends, teachers and colleagues. First and foremost, I would like to gratefully acknowledge the enthusiastic supervision and guidance of Prof. Roy Yates and Prof. Predrag Spasojević throughout the duration of this thesis. It is a genuine pleasure for me to learn, to discuss, and to discover new ideas in our weekly meetings. I cann't tell how many times they helped me out when I was stuck in a problem. I have been greatly learned from their broad knowledge and exceptional insight on the technical research area. I am also truly indebted to Prof. Roy Yates and Prof. Predrag Spasojevć for helping me to improve my technical writing and oral presentation.

I would like to express my thanks to Prof. Chris Rose and Prof. Giuseppe Caire for taking time to read my thesis and participate in my dissertation committee, and for their valuable comments and suggestions. I would like to thank Prof. Andrzej Ruszczynski and Prof. Zoran Gajic for the advice and suggestions on the optimization problems. I would like to thank Dr. Leo Razoumov for the discussion on CDMA systems. Thank Dr. Siun-chuon Mau for the discussion on adhoc network.

I would like to thank all of my colleagues at WINLAB for the enjoyable time and many useful technical discussions. I would like to specifically thank Lang Lin and Ruoheng Liu as valuable colleagues, for their research contribution to this thesis, as well as all the discussions and suggestions to my research topics. Thanks to all WINLAB graduates, especially, Feng Nan, Heng Wang, Ivana Maric, Aylin Yener, Sennur Ulukus, Rajnish Sinha, Wenfeng Zhang, Xiaohua Chen, Hongbo Liu, Nanyan Jiang, Lalitha Sankaranarayanan, Furuzan Atay for their friendship and help. I would like to thank Ivan Seskar and Kevin Wine for helping me in all problems regarding the computers.

I would like to thank my parents for their endless support and love in my life.

Without their help in taking care of my son, I could not have finished my graduate study and thesis. Thank my husband for his love, encouragement, and care during my difficult time. Thank my parents in law and my sister in law for their help during all these years. Thank my son, Joseph Chen, for all the happiness he brings to my life.

Dedication

To my parents

Table of Contents

List of Tables

List of Figures

Chapter 1

introduction

1.1 Background and Motivation

Wireless communication channels vary with time due to multipath, user mobility, and changes in the environment. Channel variations are characterized by two types of fading: large-scale pathloss and shadow fading, and small-scale multipath fading [73]. The large-scale path loss and shadowing specify signal attenuation as a function of distance, which is affected by prominent terrain contours (buildings, hills, forests, etc.). They describe the mean signal attenuation as a function of distance in the form of nth-power law as well as the statistical variation about the mean. Small-scale fading refers to the dramatic changes in signal amplitude due to the small changes of distance (as small as a half-wavelength) in the space. Small-scale fading can be flat or frequency-selective, depending on the transmission bandwidth of the radio relative to the coherence bandwidth of the channel [73].

Wireless communication systems are characterized by limited resources such as spectrum and battery energy. Efficient use of limited resources in the wireless channel is of significant interest. For a time varying channel, adaptive transmission is one approach to achieve this goal. In the case of nonadaptive transmission, a fixed link margin is required to maintain acceptable performance when the channel quality is poor. Thus, nonadaptive systems are effectively designed for the worst case channel conditions, resulting in an insufficient utilization of the full channel capacity. In contrast, adaptive transmission systems dynamically allocate according to channel conditions one or more of the following: power, modulation and coding scheme, spreading gain (in a spread spectrum system), packet length, and bandwidth occupancy [71]. Adaptation can take advantage of favorable channel conditions, and, thus, allow for a more efficient usage

of energy and spectrum. Adaptive transmissions have been employed in EDGE [33], GPRS [31], and HDR [4], and are proposed as standards for next generation cellular systems. A survey of adaptation techniques for various wireless systems can be found in [65].

Wireless communication systems are expected to provide a wide variety of services, including voice, video, and data. The system performance criterion is usually application specific, therefore, different classes of applications will result in different adaptive transmission schemes. Most of the services provided by wireless systems can be divided into four different QoS classes based on their ability to tolerate delay and some other requirements [47]. These classes are as follows:

- *Conversational class* - for voice traffic
- *Streaming class* - for audio and video traffic
- *Interactive class* - for web browsing and database access traffic
- *Background class* - for best effort traffic

In order to meet or exceed the QoS requirements of these applications in a time varying wireless environment, adaptive transmission and resource allocation schemes can be employed. Indeed, limited adaptation has already been adopted in the second-generation CDMA systems for an acceptable voice quality. In IS-95 terminals, transmitted power is adjusted to achieve a fixed target E_b/N_0 level at the base station [78]. In the third-generation wireless networks such as Universal Mobile Telecommunications System (UMTS), adaptive mechanisms are employed at the air interface, radio access network, and core network levels to provide a required quality of service [30].

So far, we can see two important objectives: efficient spectrum and battery energy utilization, and maintenance of the QoS requirements of applications. Two approaches have been taken in the literature to study adaptive schemes for accomplishing these objectives. In several papers, for example [11, 23, 72], a queuing model was set up and adaptive schemes based on channel states, buffer state, and packet arrival statistics were studied. The analytical development in this direction requires combining queuing

theory and information theory; the initial investigation in [77] formulated problems that remain open. Another approach is to apply information theory while using simple parameters to characterize the QoS requirements. The delay limited capacity and capacity versus outage are examples of this approach [18, 42]. In this work we follow the second approach.

In order to differentiate real-time from non real-time service, three capacity measures have been defined in the literature: ergodic capacity [38], delay limited capacity [42], and capacity versus outage probability [18, 68]. A comprehensive survey of these concepts can be found in [13]. The ergodic capacity [38] was developed for non real-time data services. It determines the maximum achievable rate averaged over all fading states. The corresponding optimum power allocation is the well known water filling allocation [24, 34]. The ergodic capacity is not necessarily relevant for constant-rate real-time applications, since the ergodic time needed to average out all fading states is usually much longer than the delay constraint of these applications. Delay limited capacity [42] and the capacity versus outage probability [18,68] were developed for constant-rate real-time applications. They determine the highest achievable rate or ϵ-rate [82] within an application defined delay time interval. Some of the results regarding to these capacity notions are summarized below.

- **Ergodic capacity**. Ergodic capacity was first derived by Goldsmith [38]. In [38], multiple codebooks with adaptive rate and adaptive power are used, and the resulting maximum achievable time average rate is the ergodic capacity. Ergodic capacity can also be achieved by using a single constant rate codebook in conjunction with adaptive power control [13, 18]. Ergodic capacity for multi-access fading channels is examined in [79] and for broadcast fading channels in [51].

- **Delay limited capacity**. Delay limited capacity was introduced by Hanly and Tse [42]. In [42], the instantaneous mutual information is kept constant over all states of the fading process. This maximum achievable instantaneous mutual information is called the delay limited capacity. In a single user flat fading channel, the corresponding optimum power allocation policy is channel inversion.

- **Capacity versus outage probability**. The concept of capacity versus outage probability was introduced by Ozarow, Shamai, and Wyner for block interference fading channels [68]. An information outage occurs if the instantaneous mutual information is less than a target transmission code rate. The capacity-versus-outage performance is determined by the information outage probability for a given rate. In [18], Caire, Taricco, and Biglieri identify the optimum power allocation scheme which minimizes the information outage probability for M parallel block fading channels. In [18], the zero-outage capacity is defined as the delay limited capacity using the terminology of [42]. In [66], Negi and Cioffi introduced an optimum causal power adaptation strategy for fading blocks separated in time. In [50], Li and Goldsmith derived the optimum power allocation which minimizes the outage probability for fading multiple access channels. The capacity versus outage probability for the broadcast fading channel can be found in [52].

1.2 Service Outage Based Allocation Problem

We notice that for some variable-rate real-time applications, neither the ergodic capacity nor the outage capacity is appropriate. For example, for applications with simultaneous voice and data transmissions, as soon as a basic rate r_{o} for the voice service has been guaranteed, any excess rate can be used to transmit data in a best effort fashion. For some video or audio applications, the source rate can be adapted according to the fading channel conditions to provide multiple quality of service levels. Typically, a nonzero basic rate r_{o} is required to achieve a minimum acceptable service quality. For these applications, maximizing the long term average rate while meeting a basic rate requirement for the instantaneous rate allocation is a desirable property. However, neither the ergodic capacity nor the outage capacity can achieve this goal, since the ergodic capacity offers no guarantee on the instantaneous rate while the outage capacity achieves a low long term average rate. Therefore, in this paper we combine the notion of ergodic capacity and outage capacity, and formulate the service outage based allocation problem, which maximizes the long term average rate subject to basic rate and average power constraints.

In a Rayleigh fading channel, infinite average power is needed to achieve any nonzero rate at all times. Hence, we impose the basic service rate requirement in a probabilistic way to obviate a need for infinite average power. The service is said to be in an outage when the instantaneous rate is smaller than the basic service rate r_o. A service outage constraint dictates that the probability of a service outage be less than ϵ, a parameter indicating the outage tolerance of the application. Unlike the information outage in the capacity versus outage problem [18, 66], the bits transmitted during the service outage may still be valuable in that they will be transmitted reliably and will contribute to the average rate. The service outage based allocation problem is to find the optimum power allocation that maximizes the average rate subject to the service outage probability constraint and the average power constraint.

Although the service outage based allocation problem has been motivated by variable-rate real-time applications, it also characterizes coverage versus capacity tradeoffs in cellular systems. Mobility in cellular systems results in channel variations due to changes in distance attenuation. An important objective of a cellular system is to provide a basic service rate over as much of the service area as possible. In this case, the service outage constraint characterizes the spatial coverage requirement of the system. In the allocation problem of this work, the objective is then to maximize the average rate over all geographic locations subject to meeting the service outage constraint.

Similar to the service outage concept, a minimum rate requirement has recently been stressed for the fading broadcast channel in [45, 46]. In [46], the minimum rate requirement is imposed on each channel state. In [45] the minimum rate with outage is further discussed. The minimum rate problem with outage for a single user channel in [45] is the same as our service outage problem in [58, 59].

1.3 Thesis Outline

In this work, we assume that perfect channel state information is available at both transmitter and receiver and that the fading process is ergodic within the whole communication session. Furthermore, we assume that the channel is slowly fading relative

to the codeword length; that is, the channel state is constant during the transmission of a codeword.

The thesis is organized as follows. In Chapter 2, we study the service outage based allocation problem in a single flat fading channel model. Numerical results for a Rayleigh fading channel are given. The coverage versus capacity tradeoffs in cellular systems are also examined for a geographic attenuation due to path loss. In Chapter 3, energy efficiency under a service outage constraint are examined for a single flat fading channel model. In Chapter 4, we generalize the service outage approach to an M-parallel fading channel model. In Chapter 5, we apply the service outage approach to applications with simultaneous data and voice transmissions. In Chapter 6, we study variable rate turbo bit-interleaved code modulation.

Chapter 2

Service Outage Based Capacity in a Single Flat Fading Channel

In this chapter, the service outage based allocation problem is studied in a single flat fading channel model. In Section 2.1, the system model and the optimization problem are presented. In Section 2.2, the optimum allocation policy is derived for continuous channel distribution. In Section 2.3, a supporting theorem for the optimum allocation policy is proved. Further discussion of the optimum solution is presented in Section 2.4. In this chapter, we only consider a continuous channel distribution. The allocation problem for the discrete channel distribution in the class of deterministic schemes requires to determine for each channel state whether it is in outage or not, and can be formulated as a mixed integer programming problem, which in general does not have a closed form solution. However, it will be studied in the class of probabilistic schemes in Chapter 4 for the more general M-parallel fading channel model.

2.1 Channel Model and Allocation Problem

In this work, we employ the block flat fading Gaussian channel (BF-AWGN) model [68]. In the BF-AWGN channel, a block of N symbols experiences the same channel state, which is constant over the whole block, but may vary from block to block. The value of N is related to the product of the coherence time and the coherence bandwidth of the wireless channel. We make the following assumptions:

- *The channel state information is known perfectly at both transmitter and receiver.*

  Within each block we have the time-invariant Gaussian channel

$$y = \sqrt{h}x + n. \tag{2.1}$$

Here x is the channel input, y is the channel output, n is white Gaussian noise with variance σ^2, and h is the channel state.

- *One codeword spans one fading block and the block size N is sufficiently large for reliable communication.*

  This assumption is reasonable as long as the product of the coherence bandwidth and the block duration is large enough. Since each codeword spans only one block, the decoding delay is fixed and independent of the correlation structure of the fading process.

- *The fading process is ergodic over the time scale of the application.*

  In this work, our objective is to maximize the average rate while meeting the service outage constraint. Under this assumption, the time average rate is equal to the expected rate. Depending on the system and application, there can be great variation in the time scale over which this average would be achieved. In the context of Rayleigh fading, we may observe the time average over a large number of fade durations within a single communication session. In the context of cellular coverage, the average rate would characterize performance over multiple communication sessions in a large number of geographic locations.

Let $f(h)$ denote the probability density function of the channel state h and $F(h)$ denote the corresponding cumulative distribution function. Here, we only consider the case where h is a continuous random variable. Let $p(h)$ denote the power allocation for a channel state h and $r[hp(h)]$ denote the capacity of a Gaussian channel with received power $hp(h)$, where

$$r[p] = \log\left(1 + \frac{p}{\sigma^2}\right). \tag{2.2}$$

In the thesis, for the convenience of derivation (and without loss of generality), we drop the usual factor $1/(2\log(2))$ in the capacity expression 2.2. We take into account this factor in all the simulation results and the unit for the rate is bits/symbol. The assigned code rate for a channel state h is always equal to the capacity of the Gaussian channel $r[hp(h)]$ with received power $hp(h)$. Under a service outage constraint, we allow for the

transmission code rate to be only occasionally (with a probability less than or equal to ϵ) below an application specified rate $r_{\rm o}$. Therefore, given the average power $P_{\rm av}$, the basic service rate $r_{\rm o}$, and the allowable service outage probability ϵ, we wish to maximize the expected code rate, as follows:

$$R^* = \max_{p(h)} \mathrm{E}_h \left\{ r[hp(h)] \right\} \tag{2.3}$$

$$\text{subject to:} \quad \mathrm{E}_h \left\{ p(h) \right\} \leq P_{\rm av} \tag{2.3a}$$

$$p(h) \geq 0 \tag{2.3b}$$

$$\Pr\{ r[hp(h)] < r_{\rm o} \} \leq \epsilon \, . \tag{2.3c}$$

In the absence of the service outage constraint (2.3c), R^* would be the ergodic capacity for the fading channel, and the well known water filling allocation [24,34] would be the corresponding optimum power assignment.

2.2 Optimum Power and Rate Allocation

In this section, we derive an optimum power allocation $p^*(h)$ for problem (2.3). The difficulty in deriving $p^*(h)$ is primarily due to the probabilistic constraint (2.3c). Since the probabilistic constraint is not a continuous function of $p(h)$, many well known optimization theorems, such as Kuhn-Tucker conditions, can not be applied directly. Our approach first solves an optimization problem with a basic rate requirement on an arbitrary set $\mathcal{H}_{\rm a}$. Next, we show that for the optimum power allocation scheme the service outage must occur in bad channel states below a certain threshold. Finally, we show that an optimum power allocation can be derived based on the optimization problem with a particular $\mathcal{H}_{\rm a}$.

Given a basic service rate $r_{\rm o}$ and an arbitrary power policy $p(h)$, the corresponding *service set* is defined as $\mathcal{H}_{\rm s}(p(h)) = \{h | r[hp(h)] \geq r_{\rm o}\}$, and its complement is the *outage set* $\mathcal{H}_{\rm o}(p(h)) = \{h | r[hp(h)] < r_{\rm o}\}$. In the following optimization problem, we require

that the service set contains an arbitrary set $\mathcal{H}_a$.

$$R^*(\mathcal{H}_a) = \max_{p(h)} \mathrm{E}_h \{r[hp(h)]\} \tag{2.4}$$

$$\text{subject to:} \quad \mathrm{E}_h \{p(h)\} \leq P_{\mathrm{av}} \tag{2.4a}$$

$$p(h) \geq 0 \tag{2.4b}$$

$$r[hp(h)] \geq r_{\mathrm{o}} \qquad h \in \mathcal{H}_a. \tag{2.4c}$$

Clearly, constraint (2.4c) implies that $\mathcal{H}_a$ is a subset of all service sets corresponding to feasible policies for problem (2.4). Let $p^*(h, \mathcal{H}_a)$ denote an optimum solution to problem (2.4). Therefore, $p^*(h, \mathcal{H}_a)$ achieves the highest average rate among all the schemes whose service set contains $\mathcal{H}_a$.

Problem (2.4) does not necessarily have a solution for a given $(P_{av}, r_{\mathrm{o}}, \mathcal{H}_a)$. Constraint (2.4c) implies that a feasible allocation $p(h)$ must satisfy

$$p(h) \geq \frac{\sigma^2(e^{r_{\mathrm{o}}} - 1)}{h} \qquad h \in \mathcal{H}_a. \tag{2.5}$$

This implies that the minimum average power needed to meet the constraint (2.4c) for a given $(r_{\mathrm{o}}, \mathcal{H}_a)$ is

$$P_{\min}(r_{\mathrm{o}}, \mathcal{H}_a) = \int_{\mathcal{H}_a} \frac{\sigma^2(e^{r_{\mathrm{o}}} - 1)}{h} f(h)\, dh. \tag{2.6}$$

Consequently, problem (2.4) has a solution only if $P_{\mathrm{av}} \geq P_{\min}(r_{\mathrm{o}}, \mathcal{H}_a)$. When $P_{\mathrm{av}} = P_{\min}(r_{\mathrm{o}}, \mathcal{H}_a)$ the corresponding power allocation is

$$p^*(h, \mathcal{H}_a) = \begin{cases} \dfrac{\sigma^2(e^{r_{\mathrm{o}}} - 1)}{h} & h \in \mathcal{H}_a \\ 0 & \text{otherwise} \end{cases} \tag{2.7}$$

When $P_{\mathrm{av}} > P_{\min}(r_{\mathrm{o}}, \mathcal{H}_a)$ the corresponding power allocation is given by the following theorem. We use the notation $(x)^+ = \max(x, 0)$.

Theorem 1 *When $P_{\mathrm{av}} > P_{\min}(r_o, \mathcal{H}_a)$ the optimum solution for problem (2.4) is:*

$$p^*(h, \mathcal{H}_a) = \begin{cases} \dfrac{\sigma^2(e^{r_o} - 1)}{h} & h \in \mathcal{H}_a \cap \{h \leq h_0 e^{r_o}\} \\ \sigma^2 \left[\dfrac{1}{h_0} - \dfrac{1}{h}\right]^+ & \text{otherwise} \end{cases}, \tag{2.8}$$

where h_0 is the solution of $\mathrm{E}_h \{p^*(h, \mathcal{H}_a)\} = P_{\mathrm{av}}$.

Theorem 1 follows from standard variational arguments [22] and Kuhn-Tucker conditions [3]. Note that when $P_{\rm av} = P_{\rm min}(r_{\rm o}, \mathcal{H}_a)$ the resulting power allocation (2.7) can be viewed as a limiting case of expression (2.8) as $h_0 \to \infty$. The power allocation $p^*(h, \mathcal{H}_a)$ is a combination of channel inversion and water filling allocations. To obtain a high average rate, we would like to allocate power in the form of the water filling allocation, while to meet the service constraint (2.4c), we must allocate power no less than the channel inversion allocation within the set $\mathcal{H}_a$. The solution $p^*(h, \mathcal{H}_a)$ balances these two factors.

Given the distribution $F(h)$ on channel states, we define h_ϵ such that $F(h_\epsilon) = \epsilon$. Note that h_ϵ is well defined if h is a continuous random variable. The threshold h_ϵ partitions the channels into a set $\mathcal{H}_\epsilon = \{h \geq h_\epsilon\}$ of good channels and the complementary set $\overline{\mathcal{H}_\epsilon} = \{h < h_\epsilon\}$ of bad channels. Since an efficient policy ought to meet the service constraint on the good channel states, it seems likely that $\mathcal{H}_\epsilon$ should be a subset of the service set of an optimum policy. In the following, we show that problem (2.4) with $\mathcal{H}_a = \mathcal{H}_\epsilon$ generates an optimum solution of problem (2.3) $p^*(h) = p^*(h, \mathcal{H}_\epsilon)$. In order to prove this, we need to define the partial ordering $\prec$ and a number of preliminary results.

Definition 1 $\mathcal{H}_1 \prec \mathcal{H}_2$ if $h_1 < h_2$ for all $h_1 \in \mathcal{H}_1$ and $h_2 \in \mathcal{H}_2$.

Theorem 2 *Problem (2.3) has an optimal solution $p^*(h)$ with outage set $\mathcal{H}_{\rm o}(p^*(h))$ and service set $\mathcal{H}_{\rm s}(p^*(h))$ satisfying $\mathcal{H}_{\rm o}(p^*(h)) \prec \mathcal{H}_{\rm s}(p^*(h))$.*

Proof of Theorem 2 involves a somewhat complicated two-step construction and is deferred to Section 2.3. Using Theorem 2 and the fact that $\Pr\{\mathcal{H}_{\rm s}(p^*(h))\} \geq 1 - \epsilon$ by constraint (2.3c), it is easy to show the following corollary.

Corollary 1 *Problem (2.3) has an optimal solution $p^*(h)$ such that $\mathcal{H}_\epsilon \subseteq \mathcal{H}_{\rm s}(p^*(h))$.*

Now we can prove $p^*(h) = p^*(h, \mathcal{H}_\epsilon)$ by showing that $R^* = R^*(\mathcal{H}_\epsilon)$. With $\mathcal{H}_a = \mathcal{H}_\epsilon$ in the outage constraint (2.4c), the service set of $p^*(h, \mathcal{H}_\epsilon)$ must contain $\mathcal{H}_\epsilon$. Thus $p^*(h, \mathcal{H}_\epsilon)$ satisfies the outage constraint (2.3c) and is a feasible power allocation scheme for problem (2.3), implying $R^*(\mathcal{H}_\epsilon) \leq R^*$. On the other hand, Corollary 1 implies that

problem (2.3) has an optimal solution $p^*(h)$ achieving an average rate of R^* that satisfies constraint (2.4c) with $\mathcal{H}_a = \mathcal{H}_\epsilon$. Thus, $p^*(h)$ is a feasible power allocation scheme for problem (2.4) and $R^* \leq R^*(\mathcal{H}_\epsilon)$. Therefore, $R^* = R^*(\mathcal{H}_\epsilon)$. In conclusion, an optimum solution is $p^*(h) = p^*(h, \mathcal{H}_\epsilon)$ and the following conclusions apply to problem (2.3).

- Problem (2.3) is feasible only if $(P_{\text{av}}, r_{\text{o}}, \epsilon)$ satisfies

$$P_{\text{av}} \geq P_{\min}(r_{\text{o}}, \mathcal{H}_\epsilon) = \int_{h_\epsilon}^{\infty} \frac{\sigma^2(e^{r_{\text{o}}} - 1)}{h} f(h)\, dh, \tag{2.9}$$

- When $P_{\text{av}} = P_{\min}(r_{\text{o}}, \mathcal{H}_\epsilon)$ we have

$$p^*(h) = \begin{cases} \dfrac{\sigma^2(e^{r_{\text{o}}} - 1)}{h} & h \geq h_\epsilon \\ 0 & h < h_\epsilon \end{cases}. \tag{2.10}$$

- When $P_{\text{av}} > P_{\min}(r_{\text{o}}, \mathcal{H}_\epsilon)$, we can apply Theorem 1 with $\mathcal{H}_a = \mathcal{H}_\epsilon$ yielding an optimum solution to problem (2.3) of the form

$$p^*(h) = \begin{cases} \dfrac{\sigma^2(e^{r_{\text{o}}} - 1)}{h} & h \in \{h \geq h_\epsilon\} \cap \{h < h_0^* e^{r_{\text{o}}}\} \\ \sigma^2 \left[\dfrac{1}{h_0^*} - \dfrac{1}{h}\right]^+ & \text{otherwise} \end{cases}, \tag{2.11}$$

  where h_0^* is the solution of $\mathrm{E}_h\{p^*(h)\} = P_{\text{av}}$. As P_{av} approaches $P_{\min}(r_{\text{o}}, \mathcal{H}_\epsilon)$, $h_0^* \to \infty$ and the power allocation (2.11) will reduce to the allocation (2.10).

2.3 Optimum Service Sets

In this section, we will prove Theorem 2, which implies that we can find an optimal solution whose service set $\mathcal{H}_{\text{s}}(p^*(h))$ includes the good channel states $\mathcal{H}_\epsilon$.

Before the proof of Theorem 2, let us first examine the following discrete channel distribution case with equal probability. This simple example can provide some intuition behind Theorem 2. Let h_1 and h_2 be two channel states with equal probability. WIthout loss of generality, we assume $h_1 < h_2$. Let r_1 and r_2 be two rates assigned to h_1 and h_2 respectively. Then the power for the two channel states are $\frac{e^{r_1}-1}{h_1}$ and $\frac{e^{r_2}-1}{h_2}$ respectively. If $r_1 > r_2$, it is easy to show that

$$\frac{e^{r_1} - 1}{h_1} + \frac{e^{r_2} - 1}{h_2} > \frac{e^{r_2} - 1}{h_1} + \frac{e^{r_1} - 1}{h_2}. \tag{2.12}$$

Therefore, if we switch the rates for h_1 and h_2, the average power is reduced, while the average rate and the outage probability remain the same. This implies that in the optimum scheme the rate function should be a non-decreasing function of channel states.

Although it is easy to prove Theorem 2 in the case of discrete channel distribution with equal probability, the switching approach used in the above example is hard to extend to continuous channel distribution and discrete channel distribution with unequal probability. Thus, in this section, we adopt a different approach. Our approach will be to show that given an arbitrary feasible power allocation scheme $\hat{p}(h)$, we can always construct a better scheme $p''(h)$ which satisfies $\mathcal{H}_{\rm o}(p''(h)) \prec \mathcal{H}_{\rm s}(p''(h))$. This implies that there is an optimum power allocation scheme $p^*(h)$ with $\mathcal{H}_{\rm o}(p^*(h)) \prec \mathcal{H}_{\rm s}(p^*(h))$.

Let $\hat{\mathcal{H}}_{\rm s}$ denote the service set and $\hat{R}$ the average rate for the policy $\hat{p}(h)$. Feasibility of $\hat{p}(h)$ implies $\mathrm{E}_h\{\hat{p}(h)\} \le P_{\rm av}$ and $\Pr\{\hat{\mathcal{H}}_{\rm s}\} \ge 1-\epsilon$. We use a two-step construction. First, we construct $p'(h)$ from $\hat{p}(h)$ by setting $\mathcal{H}_a = \hat{\mathcal{H}}_{\rm s}$ in problem (2.4), yielding the solution

$$p'(h) = p^*(h, \hat{\mathcal{H}}_{\rm s}) = \begin{cases} \dfrac{\sigma^2(e^{r_{\rm o}}-1)}{h} & h \in \hat{\mathcal{H}}_{\rm s} \cap \{h < h_0' e^{r_{\rm o}}\} \\ \sigma^2\left[\dfrac{1}{h_0'} - \dfrac{1}{h}\right]^+ & \text{otherwise} \end{cases}, \tag{2.13}$$

where h_0' is the solution of $E_h\{p^*(h, \hat{\mathcal{H}}_{\rm s})\} = P_{\rm av}$. Here in the case of $P_{\rm av} = P_{\min}(r_{\rm o}, \hat{\mathcal{H}}_{\rm s})$, $p'(h)$ can be expressed by (2.13) as $h_0' \to \infty$. Clearly, $p'(h)$ is feasible and achieves a higher average rate than $\hat{p}(h)$. Second, we construct $p''(h)$ by decomposing $p'(h)$ into a water filling component and a residual power component. Given h_0', we define the following functions over the whole channel state space:

$$p_{\rm wf}(h) = \sigma^2\left[\frac{1}{h_0'} - \frac{1}{h}\right]^+ \qquad 0 \le h \le \infty \tag{2.14}$$

$$p'_{\rm res}(h) = \left(\frac{\sigma^2(e^{r_{\rm o}}-1)}{h} - p_{\rm wf}(h)\right)^+ \qquad 0 \le h \le \infty. \tag{2.15}$$

The function $p_{\rm wf}(h)$ is a water filling allocation over the whole channel space. The function $p'_{\rm res}(h)$ is the nonnegative difference of channel inversion and water filling allocations. Note that $p'_{\rm res}(h)$ depends only on the water filling cutoff h_0' and $r_{\rm o}$.

From (2.13), we observe that water filling alone meets the service condition $R(hp(h)) \geq r_{\mathrm{o}}$ over the set of channel states

$$\mathcal{H}'_{\mathrm{wf}} = \left\{h | h \geq h'_0 e^{r_{\mathrm{o}}}\right\} \tag{2.16}$$

In particular, $p'_{\mathrm{res}}(h) = 0$ for $h \in \mathcal{H}'_{\mathrm{wf}}$ while residual power $p'_{\mathrm{res}}(h) > 0$ is needed to meet the service condition over the channel inversion set

$$\mathcal{H}'_{\mathrm{inv}} = \hat{\mathcal{H}}_{\mathrm{s}} \backslash \mathcal{H}'_{\mathrm{wf}}. \tag{2.17}$$

Thus, $p'(h)$ can be rewritten in the form

$$p'(h) = p_{\mathrm{wf}}(h) + \mathit{1}(h \in \mathcal{H}'_{\mathrm{inv}}) p'_{\mathrm{res}}(h) \tag{2.18}$$

where $\mathit{1}(x)$ denotes the indicator function such that $\mathit{1}(x) = 1$ when x is true, and 0 otherwise. Here, we call $\mathit{1}(h \in \mathcal{H}'_{\mathrm{inv}}) p'_{\mathrm{res}}(h)$ the residual power allocation for $p'(h)$. As shown in Figure 2.1, $p'(h)$ can be viewed as a two-layer allocation: the first layer is the water filling allocation over the whole channel space and the second layer is the residual power allocation over $\mathcal{H}'_{\mathrm{inv}}$.

Based on $p'(h)$, we construct $p''(h)$ by preserving the first layer water filling allocation and redistributing the residual power. Intuitively, the best allocation scheme for the residual power is to allocate it to the good channel states. Since $p'_{\mathrm{res}}(h)$ is strictly positive within $0 \leq h < h'_0 e^{r_{\mathrm{o}}}$, we will allocate the residual power over the interval $[h'_{\mathrm{b}}, h'_0 e^{r_{\mathrm{o}}}]$ where h'_{b} is chosen to consume the total residual power. As shown in Figure 2.1, we have

$$p''(h) = p_{\mathrm{wf}}(h) + \mathit{1}(h \in \mathcal{H}''_{\mathrm{inv}}) p'_{\mathrm{res}}(h) \tag{2.19}$$

where

$$\mathcal{H}''_{\mathrm{inv}} = \left\{h'_{\mathrm{b}} \leq h < h'_0 e^{r_{\mathrm{o}}}\right\}, \tag{2.20}$$

and h'_{b} is the solution to

$$\int_{\mathcal{H}''_{\mathrm{inv}}} p'_{\mathrm{res}}(h) f(h)\, dh = \int_{\mathcal{H}'_{\mathrm{inv}}} p'_{\mathrm{res}}(h) f(h)\, dh. \tag{2.21}$$

Note that (2.18), (2.19), and (2.21) imply that $p''(h)$ has the same total power as $p'(h)$.

Let R' and R'' denote the average rates for $p'(h)$ and $p''(h)$, respectively. The following lemma gives us the properties of $p''(h)$.

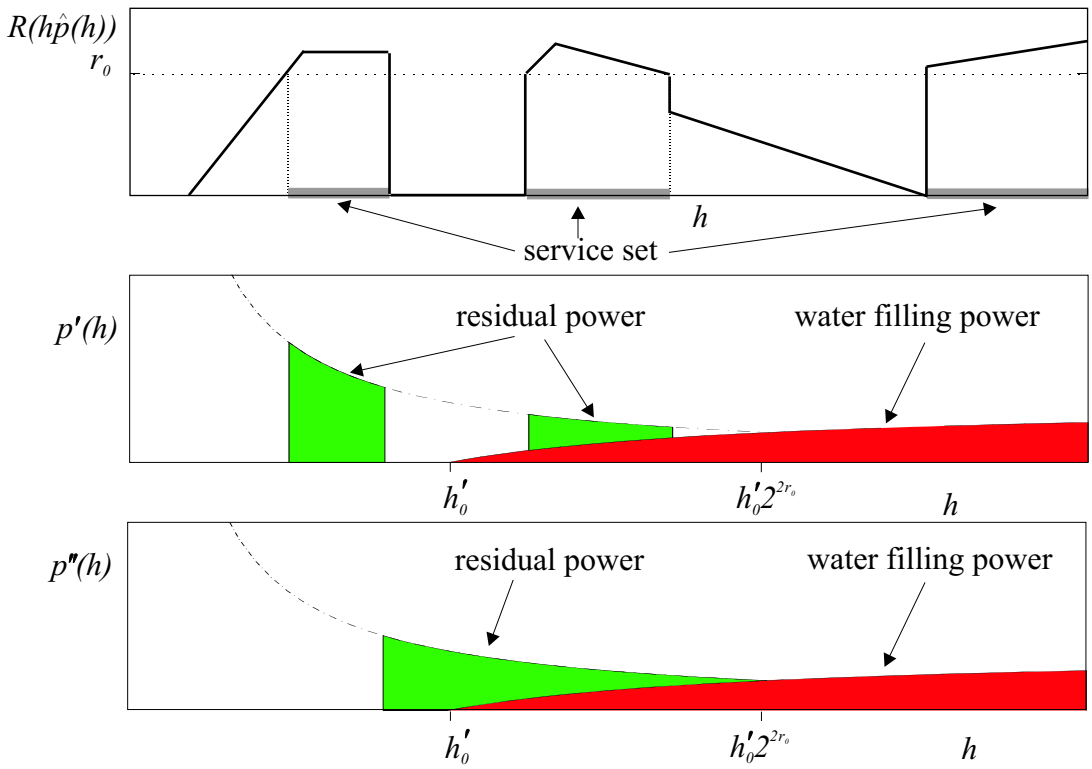


Figure 2.1: (a) Rate allocation $R(h\hat{p}(h))$ for policy $\hat{p}(h)$, (b) the improved policy $p'(h)$ given by (2.18) with water filling $p_{\rm wf}(h)$ and residual power $1(h \in \mathcal{H}'_{\rm inv})p'_{\rm res}(h)$, (c) the new policy $p''(h)$ given by (2.19) with water filling $p_{\rm wf}(h)$ and residual power $1(h \in \mathcal{H}''_{\rm inv})p'_{\rm res}(h)$.

Lemma 1 *The power scheme $p''(h)$ has the following properties:*

(a) $\mathrm{E}_h\{p''(h)\} = \mathrm{E}_h\{p'(h)\} = P_{\rm av}$

(b) $\mathcal{H}_{\rm o}(p''(h)) \prec \mathcal{H}_{\rm s}(p''(h))$

(c) $R'' \geq R'$

(d) $Pr\{\mathcal{H}_{\rm s}(p''(h))\} \geq Pr\{\mathcal{H}_{\rm s}(p'(h))\}$.

Proof of Lemma 1 is given in Appendix 3.A. At this point, it may be instructive to review our proof:

1. Start with arbitrary $\hat{p}(h)$ with average rate $\hat{R}$ and service set $\hat{\mathcal{H}}_{\rm s}$.

2. Set $\mathcal{H}_a = \hat{\mathcal{H}}_{\rm s}$ and solve (2.4) yielding $p'(h)$ with rate $R' \geq \hat{R}$ and service set $\mathcal{H}_{\rm s}(p'(h))$ containing $\hat{\mathcal{H}}_{\rm s}$. Thus $p'(h)$ is a better power allocation scheme than

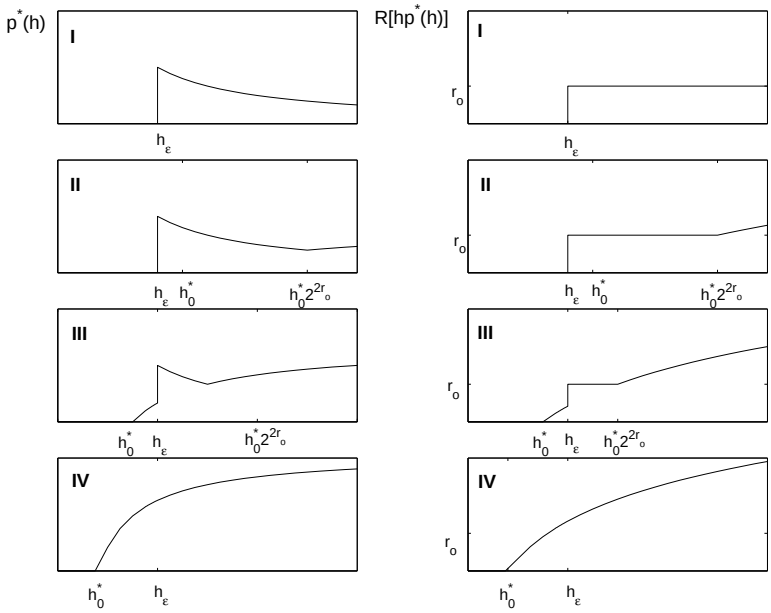


Figure 2.2: For optimum solution types I-IV, power policies are given on the left and corresponding rate allocation are on the right.

$\hat{p}(h)$ for problem (2.3).

3. Decompose $p'(h)$ into water filling $p_{\rm wf}(h)$ and residual power components $1(h \in \mathcal{H}'_{\rm inv})p'_{\rm res}(h)$.

4. Fix the water filling component $p_{\rm wf}(h)$ and reallocate the residual power to generate $p''(h)$. The power policy $p''(h)$ satisfies $\Pr\{\mathcal{H}_{\rm s}(p''(h))\} \geq \Pr\{\mathcal{H}_{\rm s}(p'(h))\}$ and $R'' \geq R'$. Hence, $p''(h)$ is a better power allocation scheme than $p'(h)$ for problem (2.3).

We can conclude that from any feasible $\hat{p}(h)$ we can obtain a better power allocation $p''(h)$ in which $\mathcal{H}_{\rm o}(p''(h)) \prec \mathcal{H}_{\rm s}(p''(h))$ holds. This implies that problem (2.3) has an optimum solution $p^*(h)$ satisfying $\mathcal{H}_{\rm o}(p^*(h)) \prec \mathcal{H}_{\rm s}(p^*(h))$.

2.4 Properties of the Optimum Policy

In Section 2.2, we derived the optimum allocation scheme for problem (2.3). In this section, we will discuss this optimum solution, and show how problem (2.3) in this paper is related to the capacity versus outage probability problem.

The optimum power allocation scheme (2.11) includes a combination of channel inversion and water filling. For a given probability distribution $f(h)$, the optimum solution belongs to one of the following possible types depending on the value of $(P_{\text{av}}, r_{\text{o}}, \epsilon)$:[1]

I When $P_{\text{av}} = P_{\min}(r_{\text{o}}, \mathcal{H}_\epsilon)$, $p^*(h)$ includes no transmission for $h < h_\epsilon$ and channel inversion for $h \geq h_\epsilon$.

II When $P_{\text{av}} > P_{\min}(r_{\text{o}}, \mathcal{H}_\epsilon)$ but $h_\epsilon \leq h_0^*$, $p^*(h)$ includes no transmission for $h < h_\epsilon$, channel inversion for $h_\epsilon \leq h < h_0^* e^{r_{\text{o}}}$, and water filling for $h \geq h_0^* e^{r_{\text{o}}}$.

III When P_{av} is sufficiently high such that $h_\epsilon e^{-r_{\text{o}}} < h_0^* < h_\epsilon$, $p^*(h)$ includes no transmission for $h < h_0^*$, water filling for $h_0^* \leq h < h_\epsilon$, channel inversion for $h_\epsilon \leq h < h_0^* e^{r_{\text{o}}}$, and water filling for $h \geq h_0^* e^{r_{\text{o}}}$.

IV When P_{av} is high enough for $h_0^* \leq h_\epsilon e^{-r_{\text{o}}}$, $p^*(h)$ is just the water filling allocation.

These four types of power allocation schemes as well as the corresponding rate allocations are depicted in Figure 2.2. For solution types I, II, and III, the optimum service set is $\mathcal{H}_{\text{s}}(p^*(h)) = \mathcal{H}_\epsilon$ and the resulting outage probability is ϵ, while for type IV solution $\mathcal{H}_\epsilon \subseteq \mathcal{H}_{\text{s}}(p^*(h))$ and the resulting outage probability is less than ϵ. The Type I solution is the on-off channel inversion allocation. In this case, we have just enough average power to satisfy the service outage constraint. When we have extra power beyond $P_{\min}(r_{\text{o}}, \mathcal{H}_\epsilon)$, we can allocate the power in a more efficient way to obtain a higher average rate and, at the same time, to meet the service outage constraint. When P_{av} is sufficiently high for the water filling allocation to satisfy the service outage constraint, then it must also be the optimum solution for problem (2.3). Thus, for a given pair (r_{o}, ϵ), the optimum power allocation scheme gradually changes from the on-off channel inversion allocation to the water filling allocation as P_{av} increases.

Now we examine the connection of the service outage problem with the outage capacity in [18] and the ergodic capacity in [38]. The outage capacity $C_\epsilon(P_{\text{av}})$ in [18] specifies the maximum supportable rate for a given average power P_{av} with outage

[1] In the case of $r_{\text{o}} = 0$ or $\epsilon = 1$, the solution types II and III will degenerate into solution type IV, which is the pure water filling allocation.

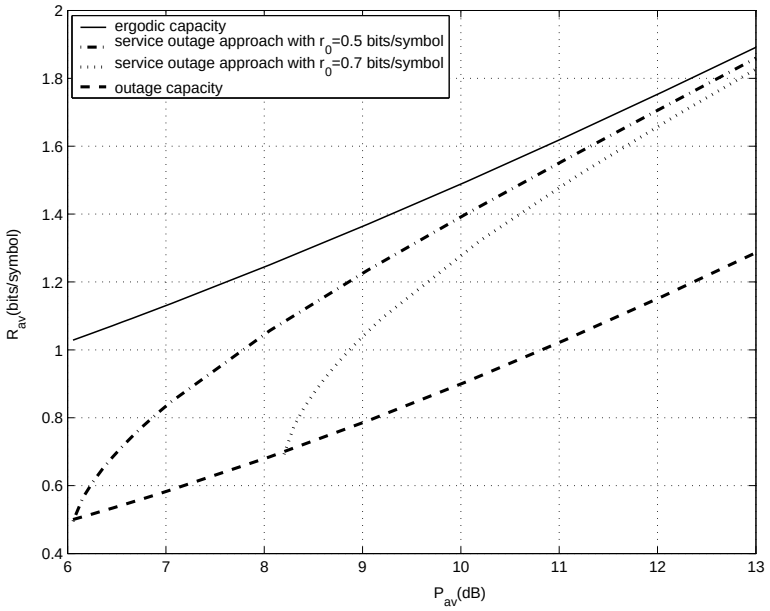


Figure 2.3: Comparison of service outage approach with other capacity notions in the Rayleigh fading channel, for a fixed $\epsilon = 0.01$.

probability ϵ, which implies that the basic service rate in this work must satisfy $r_{\rm o} \leq C_\epsilon(P_{\rm av})$. It is easy to see that the above condition is equivalent to the feasibility condition (2.9), that is, $P_{\rm av} \geq P_{\min}(r_{\rm o}, \mathcal{H}_\epsilon)$. Furthermore, we can see that the resulting average rate R^* changes from $C_\epsilon(P_{\rm av})(1-\epsilon)$ to the ergodic capacity with increasing $P_{\rm av}$ for a given $(r_{\rm o}, \epsilon)$. In Fig 2.3, the expected rate achieved by the service outage approach is plotted against the ergodic capacity and the outage capacity in Rayleigh fading channel with normalized mean for channel gain and normalized noise variance. As we can see, for a given outage probability $\epsilon = 0.01$, the outage capacity has nearly a 5 dB loss in average power compared to the ergodic capacity for a given rate. Between the outage capacity and the ergodic capacity, a number of service outage approaches with different basic rates exist. The outage probability for different r_0 achieved by the water filling allocation is also plotted against the service outage approach with a given $\epsilon = 0.01$ in Fig 2.4. It can be observed that, for a range of $P_{\rm av}$, the service outage approach achieves a rate very close to the ergodic capacity, and at the same time significantly reduces the outage probability. Hence, the service outage approach strikes good balance between average rate and outage probability.

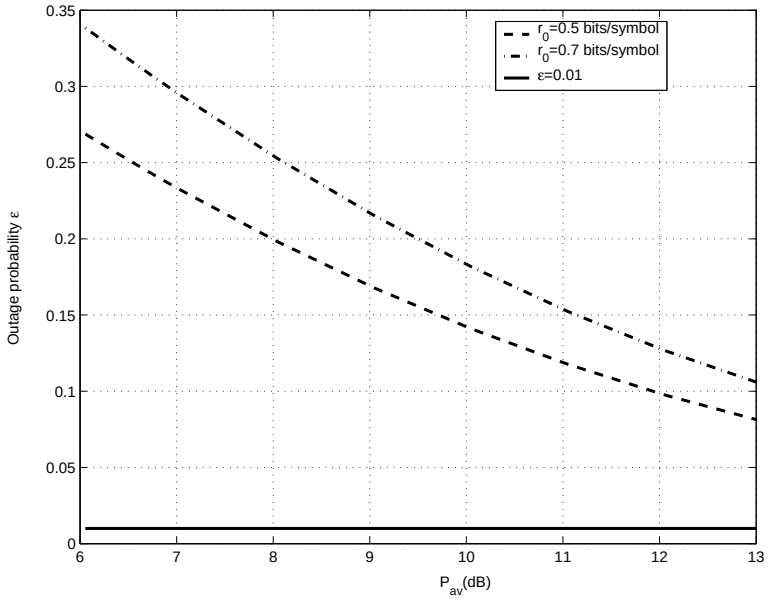


Figure 2.4: Outage probability achieved by the water filling allocation for different basic rate r_0.

2.5 Coverage versus Capacity Tradeoff in Cellular System

As we said before, problem (2.3) can also characterize the coverage versus capacity tradeoffs in cellular systems. In this case, the objective is to maximize the average rate over all geographic locations subject to the spatial coverage requirement of the system. Now we apply our service outage concept to the cellular system and demonstrate the coverage versus capacity tradeoff.

We assume a cellular system where the cochannel interference is significantly mitigated through cell isolation, sectorization, and smart antennas. We only consider large scale geographic fading due to path loss. In this case, problem (2.3) can characterize the long term single user performance in a cellular system. It can be assumed that, over a long time period, one user travels around the cellular system and experiences an ergodic geographic attenuation process. Usually, the user expects to get at least a basic service rate r_{o} almost everywhere (coverage requirement), and at the same time he would like to be provided a high average rate. The results will be somewhat optimistic in that co-channel interference will further degrade the channel, especially, for users near the edge of the cell. Using a simple path loss model [49,73], the channel gain

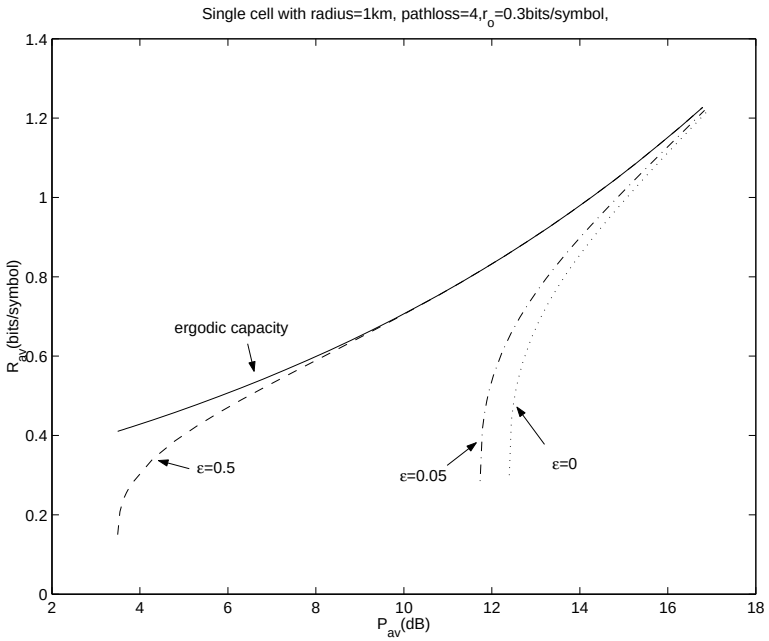


Figure 2.5: Average rate versus average power

h is a function of distance d to the base station as $h(d) = A\left(\frac{d_0}{d}\right)^{\beta}$ for $d_0 \leq d \leq d_{\max}$. Here A is a constant parameter, $d_{\max}$ is the radius of the cell, and d_0 is some close-in reference distance beyond which the path loss model holds. In this paper, we choose the path loss exponent $\beta = 4$, the cell radius $d_{\max} = 1$ km, and $d_0 = 100$ m, and only concentrate on the area from d_0 to $d_{\max}$. We also choose $A = 1/E_d[(d_0/d)^{\beta}]$ so that $h(d)$ is normalized as $E[h(d)] = 1$. Since the mobile station is uniformly distributed within the cell, the pdf of the distance d is

$$f_{\mathrm{D}}(d) = \frac{2d}{d_{\max}^2 - d_0^2} \quad d_0 \leq d \leq d_{\max}. \tag{2.22}$$

Since $h(d)$ is a one to one mapping from distance d to channel gain h, we will present our results in terms of distance d instead of channel gain h.

The average rate versus the average power is plotted in Figure 2.5 with $r_0 = 0.3$ bits/symbol. Four curves are included in this figure depicting the cases of an ergodic capacity, $\epsilon = 0.5$, $\epsilon = 0.05$, and a full coverage $\epsilon = 0$. The average rate versus the outage probability is plotted in Figure 2.5 with $r_0 = 0.3$ bits/symbol and $P_{\mathrm{av}} = 12.39$ dB for

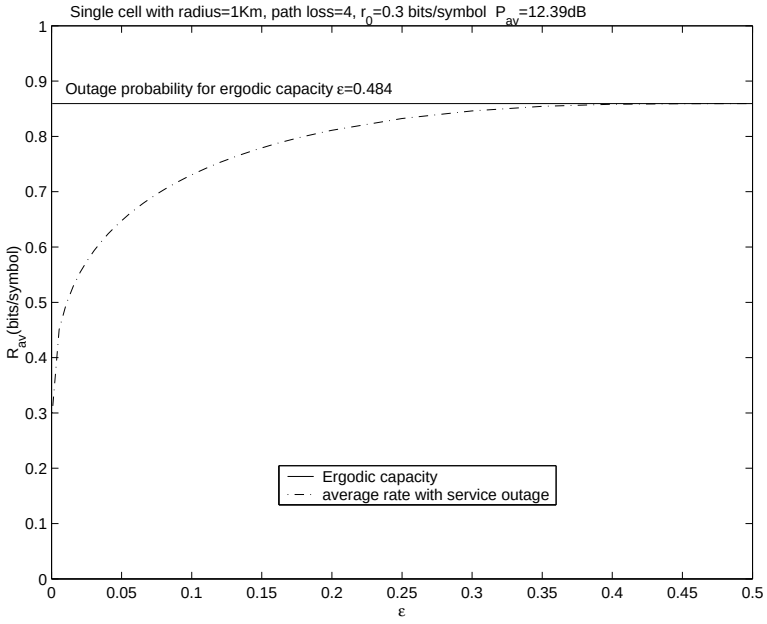


Figure 2.6: Average rate versus service outage probability

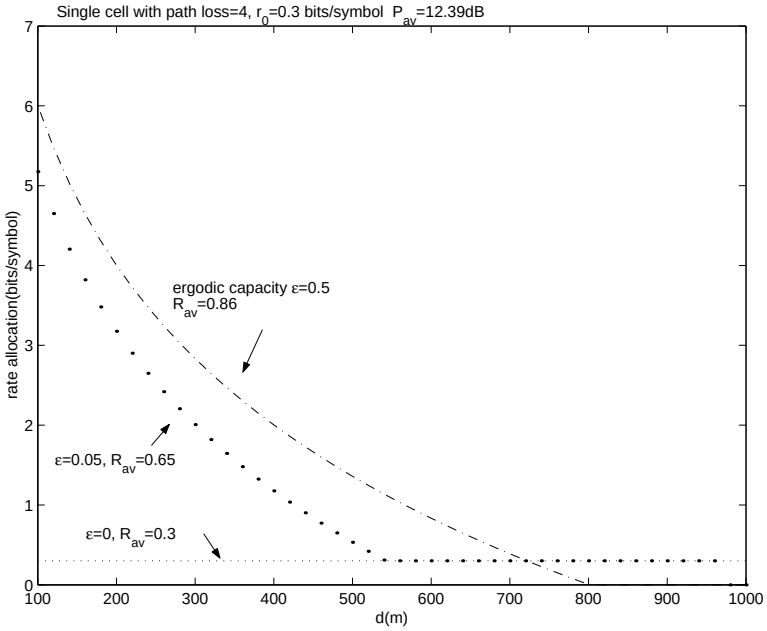


Figure 2.7: Optimum rate allocation

the coverage versus capacity tradeoff. From these two figures, we can see that the average rate increases rapidly with the average power and the outage probability. Similar to Figure 2.5, a small increase in the average power above $P_{\min}(r_o, h_\epsilon)$ results in a large increase of the average rate. For a given (r_o, P_{av}) we can also get a much higher average rate with a small sacrifice in coverage. Figure 2.5 gives the optimum rate allocation for full coverage, $\epsilon = 0$, as well as $\epsilon = 0.05$ and $\epsilon = 0.5$. With $r_0 = 0.3$ bits/symbol and $P_{av} = 12.39$ dB, if we provide full coverage in the cell, we get 0.3 bits/symbol everywhere and the average rate is 0.3 bits/symbol. If we allow an $\epsilon = 0.05$ outage probability, the average rate is 0.65 bits/symbol. The ergodic capacity of 0.86 bits/symbol is achieved when $\epsilon = 0.5$.

2.A Proof of Theorem 1

When $P_{av} \geq P_{\min}(r_o, \mathcal{H}_a)$, problem (2.4) is feasible and can be, equivalently, translated into the following problem:

$$\max_{p(h)} \mathrm{E}_h \{r[hp(h)]\} \tag{2.23}$$

$$\text{subject to} \quad \mathrm{E}_h \{p(h)\} = P_{av} \tag{2.23a}$$

$$p(h) \geq 0 \tag{2.23b}$$

$$p(h) \geq \frac{\sigma^2(e^{r_o} - 1)}{h} \qquad h \in \mathcal{H}_a. \tag{2.23b}$$

This is a standard variational optimization problem [22]. Using a Lagrange multiplier $\frac{h_0}{\sigma^2}$, we define

$$g(p(h), h, h_0) = \left[r[hp(h)] - \frac{h_0}{\sigma^2} p(h) \right] f(h). \tag{2.24}$$

For nonboundary points $h \in \overline{\mathcal{H}}$, where

$$\overline{\mathcal{H}} = \left\{ h \in \mathcal{H}_a | p^*(h, \mathcal{H}_a) > \frac{\sigma^2(e^{r_o} - 1)}{h} \right\} \cup \{ h \notin \mathcal{H}_a | p^*(h, \mathcal{H}_a) > 0 \}, \tag{2.25}$$

the optimum solution $p^*(h, \mathcal{H}_a)$ must satisfy

$$\frac{dg(p^*(h, \mathcal{H}_a), h, h_0)}{dp^*(h, \mathcal{H}_a)} = 0. \tag{2.26}$$

This implies that for $h \in \overline{\mathcal{H}}$,

$$p^*(h, \mathcal{H}_a) = \sigma^2 \left(\frac{1}{h_0} - \frac{1}{h} \right). \tag{2.27}$$

For boundary points, the optimum solution must satisfy the Kuhn-Tucker conditions [3]:

$$\frac{dg(p^*(h,\mathcal{H}_a),h,h_0)}{dp^*(h,\mathcal{H}_a)} \leq 0. \tag{2.28}$$

We then obtain

$$p^*(h,\mathcal{H}_a) = \begin{cases} \dfrac{\sigma^2(e^{r_o}-1)}{h} & h \leq h_0 e^{r_o} \text{ and } h \in \mathcal{H}_a \\ 0 & h \leq h_0 \text{ and } h \notin \mathcal{H}_a \end{cases} \tag{2.29}$$

Combining (2.27) and (2.29), we obtain the solution given in Theorem 1.

2.B Proof of Lemma 1

2.B.1 Propositions

Power schemes $p'(h)$ and $p''(h)$ differ in the allocation of the residual power. In order to show $p''(h)$ allocates the residual power in a better way than $p'(h)$, we define the following power efficiency function for $p'_{\text{res}}(h)$ over its strictly positive space.

Definition 2 The power efficiency function $\eta(h)$ for $p'_{\text{res}}(h)$ is

$$\eta(h) = \frac{r_o - r[hp_{\text{wf}}(h)]}{p'_{\text{res}}(h)} \qquad 0 \leq h < h'_0 e^{r_o}. \tag{2.30}$$

The power efficiency function indicates the rate increment corresponding to a unit power assigned from $p'_{\text{res}}(h)$. We have the following property for $\eta(h)$.

Proposition 1 *The power efficiency function $\eta(h)$ is a strictly increasing function of h over the interval $0 \leq h < h'_0 e^{r_o}$.*

Proof: Proposition 1 We consider the cases $h \leq h'_0$ and $h \geq h'_0$ separately. For $h \leq h'_0$, we have $p_{\text{wf}}(h) = 0$ and

$$\eta(h) = \frac{hr_o}{\sigma^2(e^{r_o}-1)}, \tag{2.31}$$

which is an increasing function of h.

For $h \geq h'_0$, (2.14), (2.15), and (2.30) imply

$$\eta(h) = \frac{r_o - \log(h/h'_0)}{\sigma^2\left(\frac{e^{r_o}}{h} - \frac{1}{h'_0}\right)} \tag{2.32}$$

We define $u(h) = r_{\rm o} - \log(h/h'_0)$, so that $\eta(h) = \hat{\eta}(u(h))$ where

$$\hat{\eta}(u) = \frac{h'_0}{\sigma^2}\frac{u}{e^u - 1} \tag{2.33}$$

It is straightforward to verify that $\hat{\eta}(u)$ is a strictly decreasing function of u for $u \geq 0$. Since $u(h)$ is a strictly decreasing function of h and $u(h) \geq 0$ when $h \leq h'_0 e^{r_{\rm o}}$, it follows that $\eta(h)$ is an increasing function of h for $h'_0 \leq h \leq h'_0 e^{r_{\rm o}}$. $\square$

We also employ the following proposition for the proof of Lemma 1.

Proposition 2 *For disjoint sets Ψ' and Ψ'', let $f(x)$ be an arbitrary function such that $f(x'') \geq f(x')$ for all $x'' \in \Psi''$ and $x' \in \Psi'$. For any nonnegative function $g(x)$ satisfying $\int_{\Psi''} g(x)dx = \int_{\Psi'} g(x)dx$, we have $\int_{\Psi''} f(x)g(x)dx \geq \int_{\Psi'} f(x)g(x)dx$.*

2.B.2 Proof of Lemma 1

With these preliminaries, we now verify the claims of Lemma 1.

(a) Equations (2.18), (2.19), and (2.21) imply $\mathrm{E}_h\{p''(h)\} = \mathrm{E}_h\{p'(h)\} = P_{\rm av}$.

(b) From equations (2.19) and (2.20), the service and outage sets of $p''(h)$ are $\mathcal{H}_{\rm s}(p''(h)) = \{h|h \geq h'_{\rm b}\}$ and $\mathcal{H}_{\rm o}(p''(h)) = \{h|h < h'_{\rm b}\}$ respectively. Therefore, $\mathcal{H}_{\rm o}(p''(h)) \prec \mathcal{H}_{\rm s}(p''(h))$.

(c) Let $\Psi = \mathcal{H}'_{\rm inv} \cap \mathcal{H}''_{\rm inv}$ so that $\Psi' = \mathcal{H}'_{\rm inv} \backslash \Psi$ and $\Psi'' = \mathcal{H}''_{\rm inv} \backslash \Psi$ are two disjoint sets and nonempty when $p''(h) \neq p'(h)$. The average rate of $p'(h)$ can be expressed as

$$R' = \int_0^\infty r[hp_{\rm wf}(h)]f(h)\,dh + \int_{\mathcal{H}'_{\rm inv}} (r_{\rm o} - r[hp_{\rm wf}(h)])f(h)\,dh \tag{2.34}$$

The rate contribution of the water filling component is

$$R_{\rm wf} = \int_0^\infty r[hp_{\rm wf}(h)]f(h)\,dh \tag{2.35}$$

Since $\mathcal{H}'_{\rm inv} = \Psi \cup \Psi'$, Definition 2 for the efficiency function $\eta(h)$ allows us to write

$$R' = R_{\rm wf} + \int_{\mathcal{H}'_{\rm inv}} \eta(h)p'_{\rm res}(h)f(h)\,dh \tag{2.36}$$

$$= R_{\rm wf} + \int_{\Psi} \eta(h)p'_{\rm res}(h)f(h)\,dh + \int_{\Psi'} \eta(h)p'_{\rm res}(h)f(h)\,dh \tag{2.37}$$

Similarly, $\mathcal{H}''_{\text{inv}} = \Psi \cup \Psi''$, so the average rate for $p''(h)$ can be expressed as

$$R'' = R_{\text{wf}} + \int_{\Psi} \eta(h) p'_{\text{res}}(h) f(h)\, dh + \int_{\Psi''} \eta(h) p'_{\text{res}}(h) f(h)\, dh \tag{2.38}$$

Thus,

$$R'' - R' = \int_{\Psi''} \eta(h) p'_{\text{res}}(h) f(h)\, dh - \int_{\Psi'} \eta(h) p'_{\text{res}}(h) f(h)\, dh. \tag{2.39}$$

Note that the construction of $\mathcal{H}''_{\text{inv}}$ implies $\Psi' \prec \Psi''$. That is, $h' \leq h''$ for any $h' \in \Psi'$ and $h'' \in \Psi''$. By Proposition 1, $\eta(h)$ is a strictly increasing function of h for $0 \leq h < h'_0 e^{r_{\text{o}}}$. Thus, $\eta(h'') \geq \eta(h')$. Furthermore, equation (2.21) implies

$$\int_{\Psi''} p'_{\text{res}}(h) f(h)\, dh = \int_{\Psi'} p'_{\text{res}}(h) f(h)\, dh. \tag{2.40}$$

Therefore, the conditions of Proposition 2 are satisfied and we have $R'' \geq R'$.

(d) From equations (2.13), (2.17), (2.19) and (2.20), the service sets $\mathcal{H}_{\text{s}}(p'(h))$ and $\mathcal{H}_{\text{s}}(p''(h))$ are disjoint unions given by

$$\mathcal{H}_{\text{s}}(p'(h)) = \mathcal{H}'_{\text{wf}} \cup \mathcal{H}'_{\text{inv}} = \mathcal{H}'_{\text{wf}} \cup \Psi \cup \Psi' \tag{2.41}$$

$$\mathcal{H}_{\text{s}}(p''(h)) = \mathcal{H}'_{\text{wf}} \cup \mathcal{H}''_{\text{inv}} = \mathcal{H}'_{\text{wf}} \cup \Psi \cup \Psi'' \tag{2.42}$$

This implies

$$\Pr\{\mathcal{H}_{\text{s}}(p''(h))\} - \Pr\{\mathcal{H}_{\text{s}}(p'(h))\} \tag{2.43}$$

$$= \Pr\{\Psi''\} - \Pr\{\Psi'\} \tag{2.44}$$

$$= \int_{\Psi''} \frac{1}{p'_{\text{res}}(h)} p'_{\text{res}}(h) f(h)\, dh - \int_{\Psi'} \frac{1}{p'_{\text{res}}(h)} p'_{\text{res}}(h) f(h)\, dh \tag{2.45}$$

From equations (2.14) and (2.15), we observe that $1/p'_{\text{res}}(h)$ is a increasing function of h. Since $\Psi' \prec \Psi''$, Proposition 2 implies $\Pr\{\mathcal{H}_{\text{s}}(p''(h))\} \geq \Pr\{\mathcal{H}_{\text{s}}(p'(h))\}$.

Chapter 3

Energy Efficient Allocation in a Flat Fading Channel

In Chapter 2 the service outage based achievable rate was derived. The service outage based achievable rate in bits/symbol represents the spectral efficiency since the symbol rate is determined by the bandwidth. In this chapter, we study the service outage based allocation from an energy efficiency point of view. Since higher average rate can be achieved at the expense of higher average power, the objectives of high average rate and low power consumption are contradictory. We will use an energy efficiency measure to balance these two objectives. A good measure for the energy efficiency is the ratio of average rate and average power in bits per Joule. This energy efficiency measure is an example of capacity per unit cost [83] and can be found in, for example, [20, 89]. The goal is to identify the power allocation that maximizes the energy efficiency under a service outage constraint. We show that the optimum power allocation assigns the most energy efficient average power in a most spectrally efficient manner [63].

3.1 Allocation Problem

We employ the same block flat fading channel model as in Chapter 2. For convenience of derivation, the noise variance σ^2 is normalized to unity in this chapter. Then we have

$$r[hp(h)] = \log(1 + hp(h)) \,. \tag{3.1}$$

We can see that the expected rate $\mathrm{E}_h\{r[hp(h)]\}$ represents the average number of information bits carried by each transmission symbol in bits/symbol, and the expected power $\mathrm{E}_h\{p(h)\}$ represents the average energy consumed by each transmission symbol. Therefore, for a given power policy $p(h)$, its energy efficiency in bits/Joule is just the ratio of expected rate and the expected power, as $\mathrm{E}_h\{r[hp(h)]\}/\mathrm{E}_h\{p(h)\}$.

Given the basic service rate $r_{\rm o}$ and the allowable service outage probability ϵ, we wish to maximize the following energy efficiency measure:

$$\eta^* = \max_{p(h)} \frac{\mathrm{E}_h\{r[hp(h)]\}}{\mathrm{E}_h\{p(h)\}} \tag{3.2}$$

$$\text{subject to} \quad \Pr\{r[hp(h)] < r_{\rm o}\} \le \epsilon \tag{3.2a}$$

$$p(h) \ge 0\,. \tag{3.2b}$$

The energy efficiency problem (3.2) can be solved by the following two optimization steps.

(I) For a given average power $P_{\rm av}$, find the power allocation scheme that maximizes the expected rate under the service outage constraint. This is the allocation problem studied in chapter 2. The resulting average rate is denoted as $R(P_{\rm av})$.

In this chapter, we use $P_{\min}$ to denote the minimum average power when the allocation problem is feasible, and $p(h, h_0)$ to denote the corresponding optimum power allocation. When $P_{\rm av} \ge P_{\min}$, as shown in Chapter 2, we have

$$p(h,h_0) = \begin{cases} \dfrac{(e^{r_{\rm o}}-1)}{h} & h \in \{h \ge h_\epsilon\} \cap \{h < h_0 e^{r_{\rm o}}\} \\ \left[\dfrac{1}{h_0} - \dfrac{1}{h}\right]^+ & \text{otherwise} \end{cases}, \tag{3.3}$$

where the water filling cutoff h_0 is the solution to $\mathrm{E}_h\{p(h,h_0)\} = P_{\rm av}$.

(II) Find the most energy efficient average power P_{av}^* as follows:

$$\eta^* = \max_{P_{\rm av}} \frac{R(P_{\rm av})}{P_{\rm av}}\,. \tag{3.4}$$

3.2 The Most Energy Efficient Average Power $P_{\mathbf{a}v}^*$

In this section we determine the optimum average power P_{av}^*.

For simplicity, we define $P(h_0)$ as the function mapping from h_0 to the corresponding average power $P_{\rm av}$, that is $P(h_0) = \mathrm{E}_h\{p(h,h_0)\}$. Let us define the maximum channel state with nonzero probability as $h_m = \sup\{h : f(h) \ne 0\}$. In a Rayleigh channel model, we have $h_m = \infty$. The $P(h_0)$ is a strictly decreasing function of h_0 when $h_0 \le h_m$, and

$P(h_0) = P_{\min}$ when $h_0 \geq h_m$. Therefore, we can define an inverse function mapping $h_0(P_{\text{av}})$ as follows:

$$h_0(P_{\text{av}}) = \begin{cases} P^{-1}(P_{\text{av}}) & P_{\text{av}} > P_{\min} \\ h_m & P_{\text{av}} = P_{\min} \end{cases} . \tag{3.5}$$

Let $\eta(P_{\text{av}})$ denote the achievable energy efficiency at P_{av}, as follows:

$$\eta(P_{\text{av}}) = \frac{R(P_{\text{av}})}{P_{\text{av}}} . \tag{3.6}$$

In this section we determine the optimum P^*_{av} that maximizes $\eta(P_{\text{av}})$. We show that P^*_{av} is either a stationary point of $\eta(P_{\text{av}})$ or an end point of the interval. The computation of P^*_{av} is based on a line search technique that first finds the corresponding h^*_0.

The derivative of $\eta(P_{\text{av}})$ can be expressed as follows:

$$\eta'(P_{\text{av}}) \quad = \quad \frac{1}{P_{\text{av}}}[R'(P_{\text{av}}) - \frac{R(P_{\text{av}})}{P_{\text{av}}}] \quad P_{\text{av}} \geq P_{\min} . \tag{3.7}$$

In (3.7), $R(P_{\text{av}})$ is an implicit function of P_{av}. Its first derivative is given in the following lemma.

Lemma 2 *$R(P_{\text{av}})$ is a concave increasing function of P_{av} and its first derivative $R'(P_{\text{av}}) = h_0(P_{\text{av}})$ for all $P_{\text{av}} \geq P_{\min}$.*

Lemma 2 follows from the sensitivity theorem in constrained convex optimization problems [12]. In Chapter 2, it has been shown that the service outage based allocation problem is equivalent to a problem with deterministic constraints. This new problem has a concave objective function over a convex set. The $h_0(P_{\text{av}})$ is the Lagrange multiplier associated with the average power P_{av} constraint in the equivalent problem. Lagrange multiplier in constrained convex optimization problems determines the derivative of the optimum objective with respect to the constraint parameter. Therefore, the water filling cutoff $h_0(P_{\text{av}})$ is a measure of the sensitivity of $R(P_{\text{av}})$ with respect to the average power P_{av}.

Although function $\eta(P_{\text{av}})$ is not concave in P_{av}, it can be shown that $\eta(P_{\text{av}})$ is unimodal in the interval $[P_{\min}, \infty)$ using the following Lemma 3. Let $\hat{P}_{av}$ denote the solution to $R'(P_{\text{av}}) - R(P_{\text{av}})/P_{\text{av}} = 0$, then $\hat{P}_{av}$ is a stationary point for $\eta'(P_{\text{av}})$ if it exists.

Lemma 3 *$\eta'(P_{\rm av})$ has the following properties:*

(a) *when $\eta'(P_{\min}) > 0$, $\hat{P}_{av}$ exists and is unique.*

(b) *when $\eta'(P_{\min}) > 0$, $\eta'(P_{\rm av}) > 0$ for $P_{\rm av} < \hat{P}_{av}$ and $\eta'(P_{\rm av}) < 0$ for $P_{\rm av} > \hat{P}_{av}$.*

(c) *when $\eta'(P_{\min}) < 0$, $\eta'(P_{\rm av}) < 0$ for $P_{\rm av} \geq P_{\min}$.*

From Lemma 3, we can see that $\eta(P_{\rm av})$ is unimodal in the interval $[P_{\min}, \infty)$. When $\eta'(P_{\min}) > 0$, $\eta(P_{\rm av})$ achieves it maximum at its stationary point, that is, $P^*_{\rm av} = \hat{P}_{\rm av}$. When $\eta'(P_{\min}) < 0$, $\eta(P_{\rm av})$ achieves its maximum at $P^*_{\rm av} = P_{\min}$. These two qualitatively different cases are illustrated in Fig 3.1(a) and Fig 3.1(b), respectively. Combining above results and replacing $R'(P_{\rm av})$ with $h_0(P_{\rm av})$, we obtain the following theorem.

Theorem 3 *When $\eta'(P_{\min}) > 0$, P^*_{av} is the solution to $\eta(P_{\rm av}) = h_0(P_{\rm av})$. Otherwise, $P^*_{av} = P_{\min}$.*

Note that both $h_0(P_{\rm av})$ and $R(P_{\rm av})$ are implicit functions of $P_{\rm av}$, nevertheless, they can be explicitly expressed in terms of $p(h, h_0)$. Therefore, in order to calculate $P^*_{\rm av}$ for the case of $\eta'(P_{\min}) > 0$, we first find the optimum h^*_0 and then compute $P^*_{\rm av}$ by using the explicit function mapping $P(h^*_0)$. The optimum h^*_0 is the solution of the following equation:

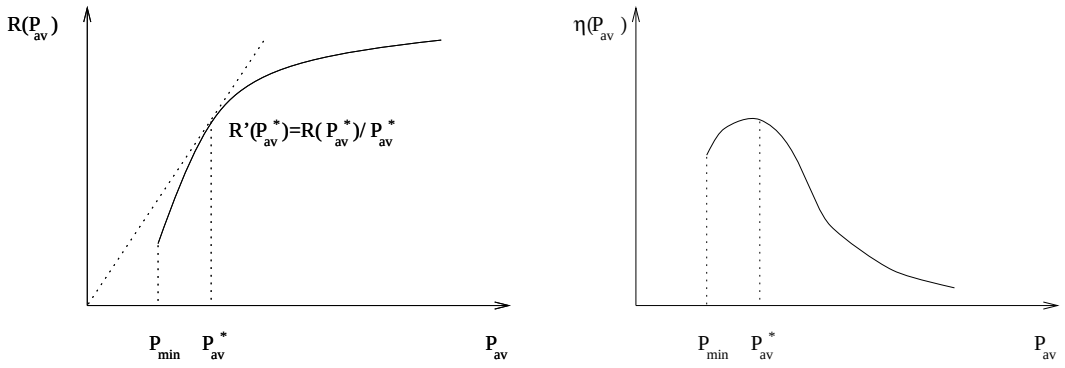
$$\frac{\mathrm{E}_h\left\{r[hp(h,h_0)]\right\}}{\mathrm{E}_h\left\{p(h,h_0)\right\}} - h_0 = 0, \quad 0 < h_0 \leq h_m\,, \tag{3.8}$$

and can be obtained using a line search method.

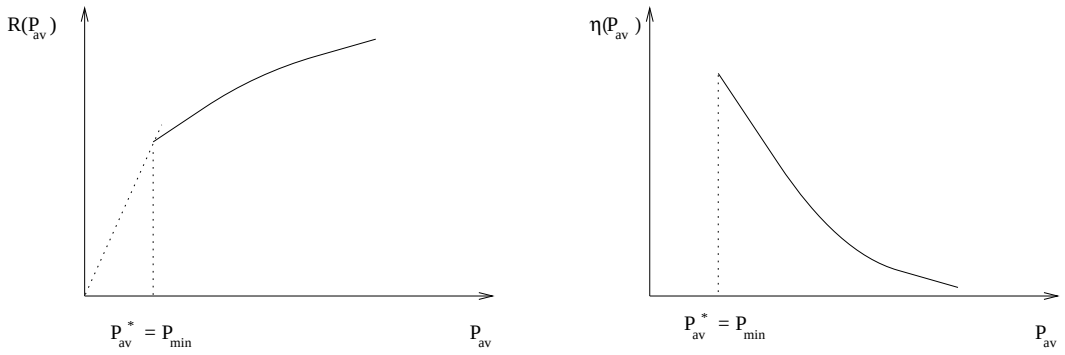
3.3 Discussion

Before we discuss the energy efficiency under the service outage constraint for flat fading channels, we first examine the energy efficiency for AWGN channels and that for flat fading channels without the service outage constraint. Several numerical results given at the end of this section illustrate the analysis given in this work as applied to a Rayleigh fading channel.

The energy efficiency versus the spectral efficiency in AWGN channels is studied in [70]. Since maximizing energy efficiency in bits/Joule is equivalent to minimizing the



(a) when $\eta'(P_{\min}) > 0$



(b) when $\eta'(P_{\min}) < 0$

Figure 3.1: Optimum average rate and the corresponding efficiency

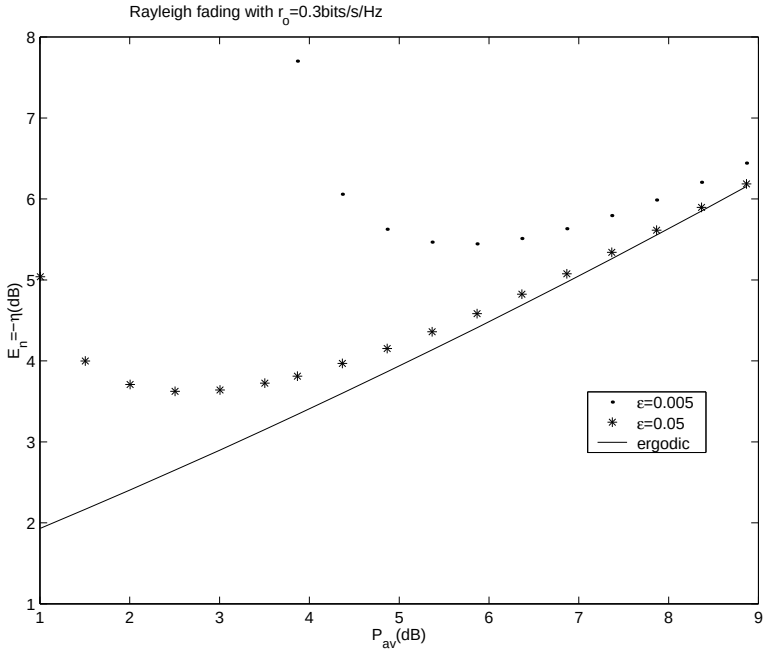


Figure 3.2: E_b versus average power in Rayleigh fading channel

bit energy $\mathcal{E}_{\rm b}$ in Joule/bit, $\mathcal{E}_{\rm b}/N_0$ is used as the energy efficiency measure in [70]. $\mathcal{E}_{\rm b}/N_0$ approaches its lower limit when the spectral efficiency approaches zero. In this case, the corresponding average power goes to zero for finite bandwidth systems. Therefore, in AWGN channels the energy efficiency and spectral efficiency are two competing objectives. To achieve a high energy efficiency, the system should operate at a low average power, resulting in a low spectral efficiency.

The energy efficiency for the ergodic capacity problem in flat fading channels can be regarded as a special case of problem (3.2) with $\epsilon = 1$ or $r_{\rm o} = 0$. In this case, the corresponding optimum power allocation is always the water filling allocation for any $P_{\rm av} \geq 0$. It can be shown that the energy efficiency $\eta(P_{\rm av})$ decreases with $P_{\rm av}$ and achieves its maximum value at $P_{\rm min} = 0$. In the same manner as for the AWGN channel, a high energy efficiency for fading channels without a service outage constraint requires a low average power and results in a low average rate, hence, a low spectral efficiency.

For flat fading channels with a service outage constraint, the optimum average power is not necessarily the minimum required average power $P_{\rm min}$. The reason is as follows:

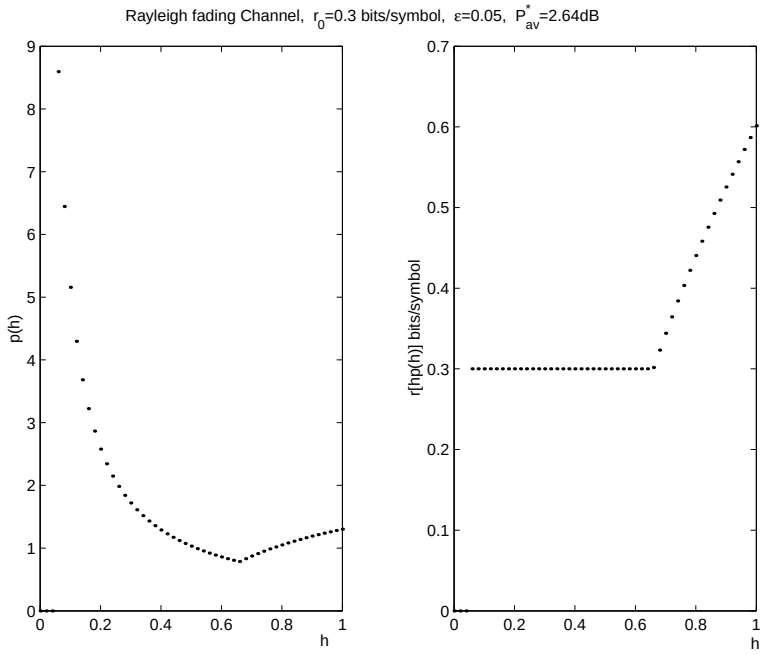


Figure 3.3: At the optimum average power $P^*_{av} = 2.64$ dB, the power allocation is given on the left and corresponding rate allocation is on the right

under the service outage constraint the optimum power allocation is a combination of channel inversion and water filling allocations. With an increase in average power, the optimum power allocation includes a larger water filling component and a smaller channel inversion component. Since the water filling allocation achieves the highest average rate without the service outage constraint, it is more energy efficient than the channel inversion allocation. On the other hand, the energy efficiency for pure water filling allocation decreases with an increase in the average power. Therefore, the energy efficiency may increase with $P_{\rm av}$ initially when the first factor is dominant, and then decrease with $P_{\rm av}$ when the second factor is dominant. Relative to $P_{\rm min}$, the optimum average power P^*_{av} achieves both higher energy efficiency and higher spectral efficiency.

In the following, we apply the analysis given in this paper to Rayleigh fading channels. For the Rayleigh fading channel with normalized mean, we have

$$f(h) = \begin{cases} e^{-h} & h \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.9}$$

and, thus, $h_0(P_{\rm min}) = h_m = \infty$. For an active service outage constraint, that is, when $r_{\rm o} > 0$ and $\epsilon < 1$, we have $P_{\rm min} > 0$ and $R(P_{\rm min}) = r_{\rm o}(1-\epsilon)$. Applying (3.7) and

P_{av}(dB)	$R(P_{\text{av}})$(bits/symbol)	$\mathcal{E}_{\text{b}}(P_{\text{av}})$(dB)
$P_{\min} = 3.8668$	0.2985	9.1173
$P^*_{av} = 5.7053$	0.7663	6.8610

Table 3.1: Comparison of the spectral efficiency and the energy efficiency at P^*_{av} with that at $P_{\min}$ for $r_{\text{o}} = 0.3$ bits/symbol and $\epsilon = 0.005$.

P_{av}(dB)	$R(P_{\text{av}})$(bits/symbol)	$\mathcal{E}_{\text{b}}(P_{\text{av}})$(dB)
$P_{\min} = 1.004$	0.285	6.4559
$P^*_{av} = 2.6404$	0.5758	5.0378

Table 3.2: Comparison of the spectral efficiency and the energy efficiency at P^*_{av} with that at $P_{\min}$ for $r_{\text{o}} = 0.3$ bits/symbol and $\epsilon = 0.05$.

Lemma 2, it follows that

$$\eta'(P_{\min}) = \frac{1}{P_{\min}}\left(R'(P_{\min}) - \frac{R(P_{\min})}{P_{\min}}\right) \tag{3.10}$$

$$= \frac{1}{P_{\min}}\left(h_0(P_{\min}) - \frac{r_{\text{o}}(1-\epsilon)}{P_{\min}}\right). \tag{3.11}$$

Since $h_0(P_{\min}) = \infty$, we have $\eta'(P_{\min}) > 0$. Consequently, the optimum $P^*_{av} > P_{\min}$ and can be obtained as a solution to $\eta(P_{\text{av}}) = h_0(P_{\text{av}})$.

In the preceding mathematical development, $\eta(P_{\text{av}})$ is used as a measure for the energy efficiency. Now for the purpose of presenting numerical results, we will use its reciprocal $\mathcal{E}_{\text{b}}(P_{\text{av}}) = 1/\eta(P_{\text{av}})$ to represent the energy efficiency as in [70]. A low value in $\mathcal{E}_{\text{b}}(P_{\text{av}})$ indicates a high energy efficiency. $\mathcal{E}_{\text{b}}(P_{\text{av}})$ in dB is plotted in Figure 3.2 for Rayleigh fading channels. We consider the following three service outage probabilities $\epsilon = 0.005$, $\epsilon = 0.05$, and $\epsilon = 1$ for fixed basic service rate $r_{\text{o}} = 0.3$ bits/symbol. In the case of $\epsilon = 1$ there is no service outage constraint and the ergodic capacity is achieved for a given P_{av}. In this case, required $\mathcal{E}_{\text{b}}(P_{\text{av}})$ increases with P_{av} and the highest power efficiency is achieved at zero power. When $\epsilon = 0.005$ or 0.05, the optimum average power P^*_{av} is no longer the corresponding minimum average power $P_{\min}$. We compare the spectral efficiency (average rate) and the energy efficiency at P^*_{av} with that at $P_{\min}$ in table 3.1 and 3.2. Observe that at the optimum average power both spectral efficiency and energy efficiency are higher than at $P_{\min}$. Compared to that at $P_{\min}$, the average rate at P^*_{av} is almost doubled and the bit energy is about $1-3$ dB less.

The power and rate allocation for the optimum average power, $\epsilon = 0.05$, and $r_o = 0.3$ bits/symbol is plotted in Figure 3.3. We can see that the optimum energy efficient power allocation is a combination of water filling and channel inversion.

3.A Proofs

In order to prove Lemma 3, we need the following definition and proposition. Define function $g(P_{\text{av}})$ as follows:

$$g(P_{\text{av}}) = R'(P_{\text{av}})P_{\text{av}} - R(P_{\text{av}}) \quad P_{\text{av}} \geq P_{\min}\,. \tag{3.12}$$

Then from (3.7) we have $\eta'(P_{\text{av}}) = g(P_{\text{av}})/P_{\text{av}}^2$. The following Proposition 3 gives the properties of $g(P_{\text{av}})$.

Proposition 3 *$g(P_{\text{av}})$ is*

(a) *a decreasing function of P_{av}.*

(b) *negative for any P_{av} when the corresponding $p(h, h_0)$ is a water filling allocation.*

Proof: Proposition 3

(a) The first derivative of $g(P_{\text{av}})$ is

$$g'(P_{\text{av}}) = R''(P_{\text{av}})P_{\text{av}} + R'(P_{\text{av}}) - R'(P_{\text{av}}) = R''(P_{\text{av}})P_{\text{av}}\,. \tag{3.13}$$

Since $R(P_{\text{av}})$ is a concave function by Lemma 2 we that $R''(P_{\text{av}}) < 0$ and, consequently, $g'(P_{\text{av}}) < 0$. Therefore, $g(P_{\text{av}})$ is a decreasing function of P_{av}.

(b) By Lemma 2, we have

$$g(P_{\text{av}}) = R'(P_{\text{av}})P_{\text{av}} - R(P_{\text{av}}) = h_0(P_{\text{av}})P_{\text{av}} - R(P_{\text{av}})\,. \tag{3.14}$$

When $p(h, h_0)$ is a water filling allocation, we have

$$P_{\text{av}} = \mathrm{E}_h\left\{p(h, h_0)\right\} = \int_{h_0}^{\infty} (\frac{1}{h_0} - \frac{1}{h}) f(h) dh \tag{3.15}$$

$$R(P_{\text{av}}) = \mathrm{E}_h\left\{r[hp(h, h_0)]\right\} = \int_{h_0}^{\infty} \log(\frac{h}{h_0}) f(h) dh \tag{3.16}$$

$$g(P_{\text{av}}) = \int_{h_0}^{\infty} \left[1 - \frac{h_0}{h} - \log(\frac{h}{h_0})\right] f(h) dh\,. \tag{3.17}$$

Let $q(x) = 1 - 1/x - \log(x)$, then $g(P_{\text{av}}) = \int_{h_0}^{\infty} q(h/h_0) f(h) dh$. It is easy to verify $q(x)|_{x=1} = 0$ and $q'(x) < 0$ for any $x > 1$. Therefore, $q(h/h_0) < 0$ for any $h > h_0$ and $g(P_{\text{av}}) < 0$.

□

We prove Lemma 3 as follows:

Proof:

(a) Since $\eta'(P_{\text{av}}) = g(P_{\text{av}})/P_{\text{av}}^2$, $\eta'(P_{\min}) > 0$ indicates $g(P_{\min}) > 0$. By Proposition 3 property (b) we also have $g(P_{\text{av}}) < 0$ for any $P_{\text{av}} \geq P(h_\epsilon e^{-r_{\text{o}}})$ since the corresponding $p(h, h_0)$ is the water filling allocation. From Proposition 3 (a), $g(P_{\text{av}})$ is a decreasing function of P_{av}, therefore, $g(P_{\text{av}}) = 0$ must exist a unique solution when $\eta'(P_{\min}) > 0$. We can see that $g(\hat{P}_{\text{a}v}) = 0$, hence $\hat{P}_{\text{a}v}$ exists and is unique when $\eta'(P_{\min}) > 0$.

(b) When $\eta'(P_{\min}) > 0$, we have $g(P_{\min}) > 0$ and $g(\hat{P}_{\text{a}v}) = 0$. Since $g(P_{\text{av}})$ decreases with P_{av} and $\eta'(P_{\text{av}}) = g(P_{\text{av}})/P_{\text{av}}^2$, we have $\eta'(P_{\text{av}}) > 0$ for $P_{\text{av}} > \hat{P}_{\text{a}v}$ and $\eta'(P_{\text{av}}) < 0$ for $P_{\text{av}} < \hat{P}_{\text{a}v}$.

(c) When $\eta'(P_{\min}) < 0$, we have $g(P_{\min}) < 0$. Since $g(P_{\text{av}})$ decreases with P_{av} and $\eta'(P_{\text{av}}) = g(P_{\text{av}})/P_{\text{av}}^2$, we have $\eta'(P_{\text{av}}) < 0$ for $P_{\text{av}} > P_{\min}$.

□

Chapter 4

Service Outage Based Allocations in an M-Parallel Fading Channel

In this chapter, we generalize the service outage based allocations to an M-parallel flat fading channel model. In Section 4.1, the channel model and the service outage based allocation problem are presented. The generalized Kuhn-Tucker conditions for functional optimization problems are reviewed in Section 4.2. The optimum power allocation is derived in Section 4.3. Two near optimum power allocation policies are derived in Section 4.4. Numerical results are given in Section 4.5 and conclusions are drawn in Section 4.6. All proofs in this paper are provided in Appendix A.

4.1 System Model and Allocation Problem

The M-parallel flat fading channel model can characterize a number of diversity systems, including an OFDM system with frequency selective fading and the multiple antenna signal model when the perfect channel state information is available at transmitter and singular value decomposition is employed. In an M-parallel flat fading channel model, each fading block consists of M subchannels of the form

$$y_i = \sqrt{h_i} x_i + n_i \quad i = 1, 2, \ldots, M \,. \tag{4.1}$$

For a subchannel i, x_i is the channel input, y_i is the channel output, and h_i is the channel state. The noise components $n_1, \ldots, n_M$ are independent Gaussian random variables with normalized unit variance. It is assumed that the channel state vector $\mathbf{h} = (h_1, \ldots, h_M)$ stays the same within one fading block but may vary from block to block. For a typical slow fading environment, it is also assumed that block length $N \to \infty$ so that the information theoretic results can be applied. One codeword spans

M subchannels in one fading block and perfect channel state information is available at both the transmitter and the receiver. The vector fading process is ergodic within the communication session.

Throughout this paper, we use the following notation:

- For a vector of channel states $\mathbf{h} = (h_1, \ldots, h_M)$, the power allocation vector is $\mathbf{p}(\mathbf{h}) = (p_1(\mathbf{h}), \ldots, p_M(\mathbf{h}))$. Here $p_i(\mathbf{h})$, the power allocated to subchannel i, depends on the current channel state vector $\mathbf{h}$.

- Given a vector $\mathbf{a}$ of length M, we denote its arithmetic mean by $\langle \mathbf{a} \rangle = M^{-1} \sum_{i=1}^{M} a_i$.

- The maximum mutual information of an M-parallel Gaussian channel $\mathbf{h}$ with power allocation $\mathbf{p}(\mathbf{h})$ is

$$r(\mathbf{h}, \mathbf{p}(\mathbf{h})) = \frac{1}{M} \sum_{i=1}^{M} \log(1 + h_i p_i(\mathbf{h})). \tag{4.2}$$

  To simplify the derivations, we use the natural logarithm and drop the usual factor 1/2 in the Gaussian capacity expression. The rate unit is nats/subchannel.

- For a scalar x, $[x]^+ = \max(x, 0)$. For a vector $\mathbf{x} = (x_1, \ldots, x_M)$, $[\mathbf{x}]^+ = ([x_1]^+, \ldots, [x_M]^+)$.

- The indicator function $1(x)$ is equal to 1 if x is true and is equal to 0 otherwise.

- For two vectors $\mathbf{a}$ and $\mathbf{b}$ of length M, we write $\mathbf{a} \geq \mathbf{b}$ if $a_i \geq b_i$ for all $i = 1, \ldots, M$.

It can be seen that with perfect channel state information at the transmitter and receiver, the maximum achievable rate of a given power vector $\mathbf{p}(\mathbf{h})$ at fading block $\mathbf{h}$ is given by (4.2). Thus, we only need to identify the optimum power allocation scheme.

Although we could formulate the allocation problem for M-parallel fading channels in the class of deterministic schemes, as we did for $M = 1$ fading channel in [62], the deterministic allocation problem turns out to be difficult to solve. For example, in the case of discrete channel distribution, the allocation problem will become an integer programming problem, which in general does not have a closed form solution. Moreover, as shown in [18], the optimum allocation for the outage capacity is a probabilistic policy

for discrete channel distribution, suggesting that deterministic schemes are likely to be suboptimal. Therefore, this paper formulates the allocation problem using the more general class of probabilistic schemes.

In a probabilistic scheme, multiple power vectors can be assigned to the same channel state vector with a conditional pdf $f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h})$, and each realization of the power allocation is associated with a coding scheme. A service outage occurs when the code rate is less than the basic rate r_{o} specified by the application. Since multiple codes are employed in a probabilistic manner, at each channel state we can have the situation where some code rates are less than r_{o} while others are greater or equal to r_{o}. Thus, at each channel state, a service outage occurs with some probability. In order to simplify the derivation, we use $\mathbf{P}(\mathbf{h})$ to indicate a probabilistic power allocation scheme with conditional PDF $f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h})$, while using $\mathbf{p}(\mathbf{h})$ to indicate a deterministic scheme. Due to the assumptions of ergodicity and perfect channel state information, the power allocation only depends on the current channel state vector. We use $F(\mathbf{h})$ to represent the cdf of channel state vector $\mathbf{h}$.

For a given probabilistic power allocation $\mathbf{P}(\mathbf{h})$, the average rate, average power, and outage probability are given by

$$E\left[r(\mathbf{h},\mathbf{P}(\mathbf{h}))\right] = \int\int r(\mathbf{h},\mathbf{p}) f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h})\, d\mathbf{p}\, dF(\mathbf{h}) \tag{4.3}$$

$$E\left[\langle\mathbf{P}(\mathbf{h})\rangle\right] = \int\int \langle\mathbf{p}\rangle f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h})\, d\mathbf{p}\, dF(\mathbf{h}) \tag{4.4}$$

$$\Pr\{r(\mathbf{h},\mathbf{P}(\mathbf{h})) < r_{\mathrm{o}}\} = \int\int 1(r(\mathbf{h},\mathbf{p}) < r_{\mathrm{o}}) f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h})\, d\mathbf{p}\, dF(\mathbf{h})\,. \tag{4.5}$$

The service outage based allocation problem is to identify the optimum conditional PDF $f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h})$ as follows:

$$R^* = \max_{f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h})} E\left[r(\mathbf{h},\mathbf{P}(\mathbf{h}))\right] \tag{4.6}$$

$$\text{subject to} \quad E\left[\langle\mathbf{P}(\mathbf{h})\rangle\right] \le p_{\mathrm{av}} \tag{4.6a}$$

$$\Pr\{r(\mathbf{h},\mathbf{P}(\mathbf{h})) < r_{\mathrm{o}}\} \le \epsilon\,, \tag{4.6b}$$

where the conditional PDF $f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h})$ is a set of functions for each $\mathbf{h}$ satisfying

$$\int f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h}) d\,\mathbf{p} = 1, \qquad f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h}) \ge 0 \quad \text{for all } \mathbf{h}\,. \tag{4.7}$$

The resulting maximum average rate R^* is called the service outage achievable rate. This work can be extended to other rate expressions besides the Shannon capacity in (4.2), which may depend on decoding error probability, and the set of modulation and coding schemes in a practical system. Problem (4.6) may seem to be more complicated than the corresponding deterministic allocation problem, but in fact it will be easier to solve. In later sections, we will see that this problem can be simplified and solved using generalized Kuhn-Tucker conditions [56].

4.2 Functional Optimization

In this section, we briefly review the Kuhn-Tucker conditions for functional optimization, since in this work the optimization variables are functions instead of vectors in an Euclidean space. Readers are referred to texts [43, 56] for comprehensive results on optimization theory in a general vector space and [3] on optimization theory in an Euclidean space.

Specifically, we are interested in the following type of functional optimization problem in a Lebesgue L_p space with measure m:

$$\min_{x(t)} \int y_1(t, x(t)) dm(t) \tag{4.8}$$

$$\text{subject to} \quad \int y_2(t, x(t)) dm(t) \leq 0 \tag{4.8a}$$

$$\int y_3(t, x(t)) dm(t) = 0 \tag{4.8b}$$

$$y_4(t, x(t)) \leq 0 \tag{4.8c}$$

$$a \leq x(t) \leq b \tag{4.8d}$$

where functions $x(t)$ and $y_i(t, x(t))$ belong to the L_p space with measure m. The L_p space consists of those real-valued measurable functions x for which $\int |x(t)|^p dm(t)$ is finite [56,67]. It is shown that the Lagrange multiplier associated with constraint (4.8a) and (4.8b) are scalers denoted as u and λ, while the Lagrange multiplier associated with constraint (4.8c) is a function $v(t) \in L_q$ where $1/p + 1/q = 1$ [15, 56]. Usually, no Lagrange multipliers are employed for simple constraints such as (4.8d), instead it is

absorbed in the Kuhn-Tucker conditions as shown below. Let

$$l(x(t), u, \lambda, v(t)) = y_1(t, x(t)) + u y_2(t, x(t)) + \lambda y_3(t, x(t)) + v(t) y_4(t, x(t)) \,.$$

The Lagrangian of problem (4.8) is $L(x(t), u, \lambda, v(t)) = \int l(x(t), u, \lambda, v(t)) dm(t)$. The variation of $L(x(t), u, \lambda, v(t))$ with respect to $x(t)$ is equal to 0 iff the derivative of $l(x, u, \lambda, v(t))$ with respect to x at $x = x(t)$ is equal to zero. Thus, according to the generalized Kuhn-Tucker necessary conditions theorem [56], if the optimum solution $x^*(t)$ is a regular point (constraint qualification), it must satisfy the following conditions:

$$\left.\frac{dl(x, u, \lambda, v(t))}{dx}\right|_{x=x^*(t)} = 0 \quad a < x^*(t) < b \tag{4.9}$$

$$\left.\frac{dl(x, u, \lambda, v(t))}{dx}\right|_{x=x^*(t)} \geq 0 \quad x^*(t) = a \tag{4.10}$$

$$\left.\frac{dl(x, u, \lambda, v(t))}{dx}\right|_{x=x^*(t)} \leq 0 \quad x^*(t) = b \tag{4.11}$$

$$u \int y_2(t, x^*(t)) dm(t) = 0 \tag{4.12}$$

$$v(t) y_4(t, x^*(t)) = 0 \tag{4.13}$$

$$u \geq 0, \quad v(t) \geq 0 \tag{4.14}$$

In addition, $x^*(t)$ must also satisfy the constraints (4.8a), (4.8b), and (4.8c). Condition (4.13) is obtained from the usual condition $\int v(t) y_4(t, x^*(t)) dm(t) = 0$ by applying $y_4(t, x(t)) \leq 0$ for all t and $v(t) \geq 0$.

Furthermore, if $y_1(t, x(t))$, $y_2(t, x(t))$, and $y_4(t, x(t))$ are convex functionals with respect to $x(t)$, and $y_3(t, x(t))$ is a linear functional with respect to $x(t)$, conditions (4.9)-(4.14) and constraints (4.8a)-(4.8d) are sufficient for the global optimum solution of (4.8).

A similar approach can be applied to the more general case where $y_i(t, x(t))$ is replaced by $y_i(t, x_1(t), \ldots, x_n(t))$. In this case, we just replace the derivation with respect to $x(t)$ in (4.9) with the partial derivatives with respect to $x_j(t)$ for all $j = 1, \ldots, n$.

4.3 The Optimum Service Outage Based Allocation

4.3.1 Allocations for an M-parallel Fading Channel

In this section, we introduce two deterministic power allocation schemes: the multi-dimensional water-filling allocation $\mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)$ and the basic-rate power allocation $\mathbf{p}_{\mathrm{r}_0}(\mathbf{h})$. These two allocations will be used to characterize the optimum solution in later sections.

The multi-dimensional water-filling allocation is the optimum allocation achieving the ergodic capacity in M-parallel fading channels as

$$\mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0) = \arg\max_{\mathbf{p}(\mathbf{h})} E\left[r(\mathbf{h}, \mathbf{p}(\mathbf{h}))\right] \tag{4.15}$$

$$\text{subject to} \quad E\left[\langle \mathbf{p}(\mathbf{h}) \rangle\right] \leq p_{\mathrm{av}} \tag{4.15a}$$

$$\mathbf{p}(\mathbf{h}) \geq 0\,. \tag{4.15b}$$

Applying the generalized Kuhn-Tucker condition in vector spaces [56], we have $\mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0) = (p_{\mathrm{wf},1}(h_1, h_0), \ldots, p_{\mathrm{wf},M}(h_M, h_0))$, where

$$p_{\mathrm{wf},i}(h_i, h_0) = \left[\frac{1}{h_0} - \frac{1}{h_i}\right]^+ \quad i = 1, \ldots, M\,, \tag{4.16}$$

and the water-filling cutoff h_0 is the solution to $E\left[\langle \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0) \rangle\right] = p_{\mathrm{av}}$.

The basic-rate power allocation is the power allocation that requires the minimum average power to maintain a basic rate at each channel state, as follows:

$$\mathbf{p}_{\mathrm{r}_0}(\mathbf{h}) = \arg\min_{\mathbf{p}(\mathbf{h})} \langle \mathbf{p}(\mathbf{h}) \rangle \tag{4.17}$$

$$\text{subject to} \quad r(\mathbf{h}, \mathbf{p}(\mathbf{h})) = r_{\mathrm{o}} \tag{4.17a}$$

$$\mathbf{p}(\mathbf{h}) \geq 0\,. \tag{4.17b}$$

The solution to the above problem is given by Lemma 1 in [18], and is summarized below. The basic-rate allocation is $\mathbf{p}_{\mathrm{r}_0}(\mathbf{h}) = (p_{\mathrm{r}_0,1}(\mathbf{h}), \ldots, p_{\mathrm{r}_0,M}(\mathbf{h}))$ with

$$p_{\mathrm{r}_0,i}(\mathbf{h}) = \left[\lambda(\mathbf{h}) - \frac{1}{h_i}\right]^+ \quad i = 1, \ldots, M\,. \tag{4.18}$$

For a given $\mathbf{h}$, the basic-rate allocation also allocates power in the form of water-filling among subchannels, but $\lambda(\mathbf{h})$ changes with $\mathbf{h}$ to ensure $r(\mathbf{h}, \mathbf{p}_{\mathrm{r}_0}(\mathbf{h})) = r_{\mathrm{o}}$. Let $\pi(i)$ be

the permutation of index i such that $h_{\pi(1)} \geq h_{\pi(2)} \geq \ldots \geq h_{\pi(M)}$. The $\lambda(\mathbf{h})$ is given by

$$\lambda(\mathbf{h}) = \left(\frac{e^{Mr_o}}{\prod_{l=1}^{\mu} h_{\pi(l)}} \right)^{\frac{1}{\mu}} , \tag{4.19}$$

where μ is the unique integer in $\{1, \ldots, M\}$ such that $\lambda(\mathbf{h}) \geq h_{\pi(l)}^{-1}$ for $l \leq \mu$ and $\lambda(\mathbf{h}) < h_{\pi(l)}^{-1}$ for $l > \mu$ [18]. Parameter μ indicates number of sub-channels with non zero power allocation at $\mathbf{h}$. When $M = 1$ the basic-rate allocation $\mathbf{p}_{\mathrm{r}_0}(\mathbf{h})$ becomes channel inversion, and when $M \to \infty$ it converges to the water-filling allocation $\mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)$ [18].

Based on the observation that for a given $\mathbf{h}$ both $\mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)$ and $\mathbf{p}_{\mathrm{r}_0}(\mathbf{h})$ are in the form of 'water-filling' but with different water levels, we have the following proposition.

Proposition 4 *We have*

(a) $r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)) \geq r_o \overset{\text{iff}}{\Longleftrightarrow} h_0^{-1} \geq \lambda(\mathbf{h}) \overset{\text{iff}}{\Longleftrightarrow} \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0) \geq \mathbf{p}_{\mathrm{r}_0}(\mathbf{h})$.

(b) For any $\mathbf{h}$*, either* $\mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0) \geq \mathbf{p}_{\mathrm{r}_0}(\mathbf{h})$ *or* $\mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0) \leq \mathbf{p}_{\mathrm{r}_0}(\mathbf{h})$ *holds.*

An example of $\mathbf{p}_{\mathrm{r}_0}(\mathbf{h})$ for $M = 2$ fading channel is given below.

1. When $h_1/h_2 > e^{2r_o}$, we have

$$P_{1,r_0} = (e^{2r_o} - 1)/h_1 \qquad P_{2,r_o} = 0 \, .$$

2. When $e^{2r_o} \geq h_1/h_2 \geq e^{-2r_o}$, we have

$$P_{1,r_0} = \sqrt{\frac{e^{2r_o}}{h_1 h_2}} - \frac{1}{h_1} \qquad P_{2,r_0} = \sqrt{\frac{e^{2r_o}}{h_1 h_2}} - \frac{1}{h_2} \, .$$

3. When $e^{-2r_o} > h_1/h_2$, we have

$$P_{1,r_o} = 0 \qquad P_{2,r_0} = (e^{2r_o} - 1)/h_2 \, .$$

4.3.2 Feasibility and Outage Capacity

The feasibility of problem (4.6) is directly related to outage capacity in [18]. Let rate $C_\epsilon(p_{\mathrm{av}})$ be the outage capacity for a given p_{av}. The $C_\epsilon(p_{\mathrm{av}})$ is the maximum instantaneous rate which can be transmitted with an outage probability ϵ. Thus, for a

given $p_{\rm av}$ and ϵ, Problem (4.6) is feasible iff $r_{\rm o} \leq C_\epsilon(p_{\rm av})$. For convenience of subsequent derivations, the feasibility condition is expressed in the following equivalent form

$$p_{\rm av} \geq P_{\min}(r_{\rm o}, \epsilon)\,, \tag{4.20}$$

where the $P_{\min}(r_{\rm o}, \epsilon)$ is the minimum average power needed to support $r_{\rm o}$ with an outage probability ϵ. When $p_{\rm av} = P_{\min}(r_{\rm o}, \epsilon)$, we have $C_\epsilon(p_{\rm av}) = r_{\rm o}$ and problem (4.6) shares the same optimum solution, denoted $\mathbf{P}_{\min}(\mathbf{h})$, with the outage capacity problem. For convenience of subsequent derivations, we rewrite Proposition 4 in [18] and express $\mathbf{P}_{\min}(\mathbf{h})$ as follows.

Definition 3 For any $\mathbf{h}$, let $X_w(\mathbf{h})$ be a Bernoulli $w(\mathbf{h})$ random variable: $X_w(\mathbf{h}) = 1$ with probability $w(\mathbf{h})$ and $X_w(\mathbf{h}) = 0$ with probability $1 - w(\mathbf{h})$.

The minimum average power allocation is $\mathbf{P}_{\min}(\mathbf{h}) = X_{w'}(\mathbf{h})\mathbf{p}_{\rm r_0}(\mathbf{h})$, where

$$w'(\mathbf{h}) = \begin{cases} 1 & \langle \mathbf{p}_{\rm r_0}(\mathbf{h}) \rangle < s' \\ v' & \langle \mathbf{p}_{\rm r_0}(\mathbf{h}) \rangle = s' \\ 0 & \langle \mathbf{p}_{\rm r_0}(\mathbf{h}) \rangle > s' \end{cases}, \tag{4.21}$$

and the parameters s' and $0 \leq v' \leq 1$ are solutions to $E\left[w'(\mathbf{h})\right] = 1 - \epsilon$. That is

$$s' = \sup\left\{x : \Pr\{\langle \mathbf{p}_{\rm r_0}(\mathbf{h}) \rangle < x\} < 1 - \epsilon\right\} \tag{4.22}$$

$$v' = \frac{1 - \epsilon - \Pr\{\langle \mathbf{p}_{\rm r_0}(\mathbf{h}) \rangle < s'\}}{\Pr\{\langle \mathbf{p}_{\rm r_0}(\mathbf{h}) \rangle = s'\}}\,. \tag{4.23}$$

$\mathbf{P}_{\min}(\mathbf{h})$ is an on-off transmission policy. If the required sum power $\langle \mathbf{p}_{\rm r_0}(\mathbf{h}) \rangle > s'$, transmission is turned off, while if $\langle \mathbf{p}_{\rm r_0}(\mathbf{h}) \rangle \leq s'$, transmission is turned on and the power is allocated according to $\mathbf{p}_{\rm r_0}(\mathbf{h})$. $\mathbf{P}_{\min}(\mathbf{h})$ for an $M = 2$ fading channel is plotted in Figure 4.1. The off region may or may not be a convex set depending on $r_{\rm o}$ and ϵ.

4.3.3 Derivation of the Optimum Allocation Scheme

In this section, we derive the optimum solution for the service outage based allocation problem (4.6). We first show that an optimum power allocation in (4.6) is a scheme which is randomized between two deterministic schemes.

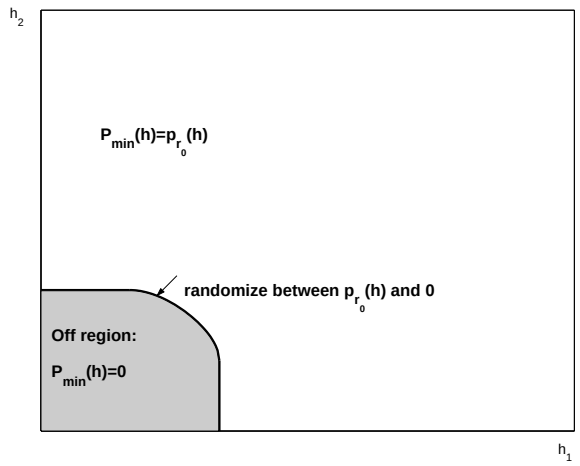


Figure 4.1: The $\mathbf{P}_{\min}(\mathbf{h})$ in an $M = 2$ parallel fading channel.

Lemma 4 *There exists an optimum solution of problem (4.6) of the following form*

$$\mathbf{P}^*(\mathbf{h}) = X_w(\mathbf{h})\mathbf{p}_{\mathrm{a}}(\mathbf{h}) + (1 - X_w(\mathbf{h}))\mathbf{p}_{\mathrm{b}}(\mathbf{h})\,, \tag{4.24}$$

where $r(\mathbf{h}, \mathbf{p}_{\mathrm{a}}(\mathbf{h})) \geq r_o$ *for all* $\mathbf{h}$, $E\left[w(\mathbf{h})\right] \geq 1 - \epsilon$ *and* $E\left[\langle \mathbf{P}^*(\mathbf{h})\rangle\right] = p_{\mathrm{av}}$.

Proof of Lemma 4 is based on the concavity of the rate function $r(\mathbf{h}, \mathbf{p})$.

By Lemma 4, we have $\mathbf{P}^*(\mathbf{h}) = \mathbf{p}_{\mathrm{a}}(\mathbf{h})$ with probability $w(\mathbf{h})$, and $\mathbf{P}^*(\mathbf{h}) = \mathbf{p}_{\mathrm{b}}(\mathbf{h})$ with probability $1 - w(\mathbf{h})$. Moreover, the conditions $r(\mathbf{h}, \mathbf{p}_{\mathrm{a}}(\mathbf{h})) \geq r_{\mathrm{o}}$ and $E\left[w(\mathbf{h})\right] \geq 1 - \epsilon$ ensure that the randomized scheme meets the service outage constraint. Thus, problem (4.6) can be simplified into a problem which requires identifying $\mathbf{p}_{\mathrm{a}}(\mathbf{h})$, $\mathbf{p}_{\mathrm{b}}(\mathbf{h})$, and $w(\mathbf{h})$ as follows:

$$\max_{\mathbf{p}_{\mathrm{a}}(\mathbf{h}), \mathbf{p}_{\mathrm{b}}(\mathbf{h}), w(\mathbf{h})} E\left[w(\mathbf{h}) r(\mathbf{h}, \mathbf{p}_{\mathrm{a}}(\mathbf{h})) + (1 - w(\mathbf{h})) r(\mathbf{h}, \mathbf{p}_{\mathrm{b}}(\mathbf{h}))\right] \tag{4.25}$$

$$\text{subject to} \quad E\left[w(\mathbf{h})\langle \mathbf{p}_{\mathrm{a}}(\mathbf{h})\rangle + (1 - w(\mathbf{h}))\langle \mathbf{p}_{\mathrm{b}}(\mathbf{h})\rangle\right] = p_{\mathrm{av}} \tag{4.25a}$$

$$E\left[w(\mathbf{h})\right] \geq 1 - \epsilon \tag{4.25b}$$

$$r(\mathbf{h}, \mathbf{p}_{\mathrm{a}}(\mathbf{h})) \geq r_{\mathrm{o}} \tag{4.25c}$$

$$\mathbf{p}_{\mathrm{a}}(\mathbf{h}) \geq 0 \quad \mathbf{p}_{\mathrm{b}}(\mathbf{h}) \geq 0 \quad 0 \leq w(\mathbf{h}) \leq 1 \tag{4.25d}$$

In the following, we derive the optimum solution of problem (4.25) using the generalized Kuhn-Tucker conditions theorem described in Section 4.2.

Let $\mathbf{p}_{\rm a}^*(\mathbf{h})$, $\mathbf{p}_{\rm b}^*(\mathbf{h})$, and $w^*(\mathbf{h})$ denote the optimum solution of (4.25). Let h_0^*, $s^* \geq 0$, and $u^*(\mathbf{h}) \geq 0$ denote the corresponding Lagrange multipliers for constraints (4.25a), (4.25b), and (4.25c), respectively. Define

$$\begin{aligned} l(\mathbf{h}, \mathbf{p}_{\rm a}(\mathbf{h}), \mathbf{p}_{\rm b}(\mathbf{h}), w(\mathbf{h}), h_0, s, u(\mathbf{h})) =& w(\mathbf{h})[r(\mathbf{h}, \mathbf{p}_{\rm a}(\mathbf{h})) - h_0 \langle \mathbf{p}_{\rm a}(\mathbf{h}) \rangle] \\ &+ (1 - w(\mathbf{h}))[r(\mathbf{h}, \mathbf{p}_{\rm b}(\mathbf{h})) - h_0 \langle \mathbf{p}_{\rm b}(\mathbf{h}) \rangle] \\ &+ s\, w(\mathbf{h}) + u(\mathbf{h}) r(\mathbf{h}, \mathbf{p}_{\rm a}(\mathbf{h})) \end{aligned} \tag{4.26}$$

In following, for simplicity we use the notation

$$l(\cdots) \stackrel{\triangle}{=} l(\mathbf{h}, \mathbf{p}_{\rm a}^*(\mathbf{h}), \mathbf{p}_{\rm b}^*(\mathbf{h}), w^*(\mathbf{h}), h_0^*, s^*, u^*(\mathbf{h})). \tag{4.27}$$

According to the Kuhn-Tucker necessary conditions theorem, the optimum solution must satisfy the following conditions[1]:

$$\frac{\partial l(\cdots)}{\partial p_{a,i}^*(\mathbf{h})} \begin{cases} = 0 & p_{a,i}^*(\mathbf{h}) > 0 \\ \leq 0 & p_{a,i}^*(\mathbf{h}) = 0 \end{cases} \quad \text{for all} \quad i = 1, \ldots, M \tag{4.28}$$

$$\frac{\partial l(\cdots)}{\partial p_{b,i}^*(\mathbf{h})} \begin{cases} = 0 & p_{b,i}^*(\mathbf{h}) > 0 \\ \leq 0 & p_{b,i}^*(\mathbf{h}) = 0 \end{cases} \quad \text{for all} \quad i = 1, \ldots, M \tag{4.29}$$

$$\frac{\partial l(\cdots)}{\partial w^*(\mathbf{h})} \begin{cases} = 0 & 0 < w^*(\mathbf{h}) < 1 \\ \leq 0 & w^*(\mathbf{h}) = 0 \\ \geq 0 & w^*(\mathbf{h}) = 1 \end{cases} \tag{4.30}$$

$$u^*(\mathbf{h})[r(\mathbf{h}, \mathbf{p}_{\rm a}^*(\mathbf{h})) - r_{\rm o}] = 0, \qquad u^*(\mathbf{h}) \geq 0 \tag{4.31}$$

$$s^*[E\,[w^*(\mathbf{h})] - (1 - \epsilon)] = 0, \qquad s^* \geq 0 \tag{4.32}$$

$$E\,[w^*(\mathbf{h}) \langle \mathbf{p}_{\rm a}^*(\mathbf{h}) \rangle + (1 - w^*(\mathbf{h})) \langle \mathbf{p}_{\rm b}^*(\mathbf{h}) \rangle] = p_{\rm av} \tag{4.33}$$

Moreover, the following lemma shows that any solution that satisfies the above conditions is an optimum solution.

Lemma 5 *The Kuhn-Tucker conditions (4.28)-(4.33) are sufficient conditions for the optimum solution of problem (4.25).*

[1]Notation $\frac{\partial l(\cdots)}{\partial p_{a,i}^*(\mathbf{h})}$ is the derivative over $p_{a,i}(\mathbf{h})$ evaluated at $p_{a,i}^*(\mathbf{h})$

The proof of Lemma 5 requires transforming of Problem (4.25) into a convex optimization problem. From Kuhn-Tucker conditions (4.28), (4.29), and (4.31), we have the following lemma.

Lemma 6 *The optimum* $\mathbf{p}_{\mathrm{a}}^*(\mathbf{h})$ *and* $\mathbf{p}_{\mathrm{b}}^*(\mathbf{h})$ *are*

$$\mathbf{p}_{\mathrm{a}}^*(\mathbf{h}) = \begin{cases} \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*) & r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*)) \geq r_o \\ \mathbf{p}_{\mathrm{r}_0}(\mathbf{h}) & \text{otherwise} \end{cases} \tag{4.34}$$

$$\mathbf{p}_{\mathrm{b}}^*(\mathbf{h}) = \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*)\,. \tag{4.35}$$

Proposition 4(b) implies that $\mathbf{p}_{\mathrm{a}}^*(\mathbf{h})$ has an equivalent expression as

$$\mathbf{p}_{\mathrm{a}}^*(\mathbf{h}) = \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*) + [\mathbf{p}_{\mathrm{r}_0}(\mathbf{h}) - \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)]^+\,. \tag{4.36}$$

We define the second term as the *supplemental power allocation*, that is

$$\mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) = [\mathbf{p}_{\mathrm{r}_0}(\mathbf{h}) - \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)]^+\,. \tag{4.37}$$

The supplemental power allocation provides the additional power needed for the water-filling allocation to meet the basic rate requirement. The rate achieved by $\mathbf{p}_{\mathrm{a}}^*(\mathbf{h})$ can be expressed as

$$r(\mathbf{h}, \mathbf{p}_{\mathrm{a}}^*(\mathbf{h})) = r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*)) + r_{\mathrm{s}}(\mathbf{h}, h_0^*) \tag{4.38}$$

with $r_{\mathrm{s}}(\mathbf{h}, h_0) = [r_{\mathrm{o}} - r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0))]^+$ being the additional rate allocation needed for water-filling allocation to meet the basic rate requirement.

Combining Lemma 4, Lemma 6 and expression (4.36), the optimum power allocation $\mathbf{P}^*(\mathbf{h})$ is

$$\mathbf{P}^*(\mathbf{h}) = \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*) + X_{w^*}(\mathbf{h})\mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0^*)\,. \tag{4.39}$$

In the following, we determine $w^*(\mathbf{h})$. Employing (4.35), (4.36), and (4.38) in $l(\cdots)$, we have

$$l(\cdots) = w^*(\mathbf{h})[s^* - g(\mathbf{h}, h_0^*)] + r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*)) - h_0^*\langle \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*)\rangle \tag{4.40}$$

with

$$g(\mathbf{h}, h_0) = h_0\langle \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0)\rangle - r_{\mathrm{s}}(\mathbf{h}, h_0). \tag{4.41}$$

The first term in $g(\mathbf{h}, h_0)$ is the power expense of allocating supplemental power, the second term is the corresponding rate return, and h_0 is the Lagrange multiplier that connects the power with the rate. Thus, function $g(\mathbf{h}, h_0)$ provides a measure for the cost of allocating the supplemental power, and is called the *supplemental cost function*.

Lemma 7 *Properties of $g(\mathbf{h}, h_0)$ are as follows:*

(a) If $\mathbf{h}' \geq \mathbf{h}$, then $g(\mathbf{h}', h_0) \leq g(\mathbf{h}, h_0)$.

(b) If $\mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) > 0$, then $g(\mathbf{h}, h_0) > 0$. If $\mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) = 0$, then $g(\mathbf{h}, h_0) = 0$.

Lemma 7(a) shows that a higher cost is associated with a poorer channel state vector. Based on Lemma 7(b) and Proposition 4, we have the following equivalent statements.

$$g(\mathbf{h}, h_0) = 0 \Longleftrightarrow \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) = 0 \Longleftrightarrow r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)) \geq r_{\mathrm{o}}. \tag{4.42}$$

Taking the derivative of $l(\cdots)$ over $w^*(\mathbf{h})$, we have

$$\frac{\partial l(\cdots)}{\partial w^*(\mathbf{h})} = s^* - g(\mathbf{h}, h_0^*)\,. \tag{4.43}$$

From condition (4.30) and equality (4.43), we obtain

$$w^*(\mathbf{h}) = \begin{cases} 1 & g(\mathbf{h}, h_0^*) < s^* \\ v^*(\mathbf{h}) & g(\mathbf{h}, h_0^*) = s^* \\ 0 & g(\mathbf{h}, h_0^*) > s^* \end{cases}, \tag{4.44}$$

where $0 \leq v^*(\mathbf{h}) \leq 1$ needs to be determined. As we can see, the cost function $g(\mathbf{h}, h_0^*)$ determines the value of $w^*(\mathbf{h})$ and indicates where the supplemental power should be allocated.

Condition (4.32) implies the following two situations:

- When $s^* > 0$, we must have $E\left[w^*(\mathbf{h})\right] = 1 - \epsilon$.
- When $s^* = 0$, we must have $E\left[w^*(\mathbf{h})\right] \geq 1 - \epsilon$. $s^* = 0$ implies that either $g(\mathbf{h}, h_0^*) = 0$ or $w^*(\mathbf{h}) = 0$. Consequently, from Lemma 7(b) we have that $X_{w^*}(\mathbf{h})\mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0^*) = 0$. Therefore, in this case $\mathbf{P}^*(\mathbf{h}) = \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*)$. Since no supplemental power is allocated, $0 \leq v^*(\mathbf{h}) \leq 1$ can be any function that meets $E\left[w^*(\mathbf{h})\right] \geq 1 - \epsilon$. In

order to simplify the presentation and without loss of generality, we choose $v^*(\mathbf{h})$ so that $E[w^*(\mathbf{h})] = 1 - \epsilon$.[2]

The following theorem combines above results.

Theorem 4 *If problem (4.6) is feasible, an optimum power allocation is*

$$\mathbf{P}^*(\mathbf{h}) = \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*) + X_{w^*}(\mathbf{h}) \left[\mathbf{p}_{\mathrm{r}_0}(\mathbf{h}) - \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*)\right]^+ , \tag{4.45}$$

where

$$w^*(\mathbf{h}) = \begin{cases} 1 & g(\mathbf{h}, h_0^*) < s^* \\ v^*(\mathbf{h}) & g(\mathbf{h}, h_0^*) = s^* \\ 0 & g(\mathbf{h}, h_0^*) > s^* \end{cases} . \tag{4.46}$$

and h_0^*, s^*, *and* $0 \le v^*(\mathbf{h}) \le 1$ *are solutions to*

$$E[\langle \mathbf{P}^*(\mathbf{h}) \rangle] = p_{\mathrm{av}}, \quad E[w^*(\mathbf{h})] = 1 - \epsilon.$$

The optimum power allocation can be viewed as a two layer allocation: the first layer is the water-filling allocation, and the second layer is the supplemental allocation. The supplemental allocation provides the additional power and rate for the water-filling allocation to meet the basic rate requirement. If the channel states are so poor that the cost $g(\mathbf{h}, h_0^*)$ is above a threshold, the supplemental allocation is turned off and a service outage is declared. Thus, $g(\mathbf{h}, h_0^*)$ divides the channel space into a service set $g(\mathbf{h}, h_0^*) < s^*$ with rates $r \ge r_{\mathrm{o}}$, a boundary set $g(\mathbf{h}, h_0^*) = s^*$ with a probabilistic policy, and an outage set $g(\mathbf{h}, h_0^*) > s^*$ with rates $r < r_{\mathrm{o}}$. The service set can be further divided into a basic-rate set with rate $r = r_{\mathrm{o}}$, and an enhanced-rate set with rate $r > r_{\mathrm{o}}$.

4.3.4 Properties of the Optimum Solution

In this section, we study the properties of the optimum solution. By further examination of $\mathbf{P}^*(\mathbf{h})$ in Theorem 4, it can be seen that the optimum solution is a combination of basic-rate allocation and water-filling allocation in the non-boundary channel state set,

[2] In this case, the outage probability is less than or equal to $1 - E[w^*(\mathbf{h})] = \epsilon$.

and is randomized between these two at the boundary set. The optimum solution for the $M = 1$ fading channel can be found in Chapter 2. The optimum solution for $M = 2$ fading channels is depicted in Fig 4.2. The optimum solution can be classified into four types as a function of an increasing $p_{\rm av}$ for any given $(r_{\rm o}, \epsilon)$ as follows.

- $\mathbf{P}^*(\mathbf{h})$ is Type **I** when $p_{\rm av} = P_{\rm min}(r_{\rm o}, \epsilon)$. In this case, we have $\mathbf{p}_{\rm wf}(\mathbf{h}, h_0^*) = 0$ and $w^*(\mathbf{h}) = w'(\mathbf{h})$ from (4.21). The optimum solution is the same allocation as the outage capacity, that is

$$\mathbf{P}^*(\mathbf{h}) = \mathbf{P}_{\rm min}(\mathbf{h}) = X_{w'}(\mathbf{h})\mathbf{p}_{\rm r_0}(\mathbf{h}). \tag{4.47}$$

- $\mathbf{P}^*(\mathbf{h})$ is Type **II** when $\mathbf{p}_{\rm wf}(\mathbf{h}, h_0^*) = 0$ in the outage set. In this case, we have $w^*(\mathbf{h}) = w'(\mathbf{h})$ and the outage set is the same as for the outage capacity allocation. In this case, the cost function $g(\mathbf{h}, h_0^*)$ in the outage set reduces to $h_0^* \langle \mathbf{p}_{\rm r_0}(\mathbf{h}) \rangle - r_{\rm o}$. Therefore, the optimum outage set, defined as $g(\mathbf{h}, h_0^*) > s^*$, can be rewritten as $\langle \mathbf{p}_{\rm r_0}(\mathbf{h}) \rangle > s'$. Type II solution includes no transmission in the outage set, a probabilistic scheme in the boundary set, basic-rate allocation in the basic-rate set, and water-filling allocation with rate $r > r_{\rm o}$ in the enhanced-rate set.

- $\mathbf{P}^*(\mathbf{h})$ is Type **III** in the most general case. All other types can be considered as special cases of Type III. It includes water-filling allocation with rate $r < r_{\rm o}$ in the outage set, a probabilistic scheme in the boundary set, basic-rate allocation in the basic-rate set, and water-filling allocation with rate $r > r_{\rm o}$ in the enhanced-rate set.

- $\mathbf{P}^*(\mathbf{h})$ is Type **IV** when $\Pr\{r(\mathbf{h}, \mathbf{p}_{\rm wf}(\mathbf{h}, h_0^*)) \geq r_{\rm o}\} \geq 1 - \epsilon$ holds. In this case, we have $s^* = 0$ and $X_{w^*}(\mathbf{h})\mathbf{p}_{\rm s}(\mathbf{h}, h_0^*) = 0$. Thus, the optimum solution is $\mathbf{P}^*(\mathbf{h}) = \mathbf{p}_{\rm wf}(\mathbf{h}, h_0^*)$.

With increasing $p_{\rm av}$, $\mathbf{P}^*(\mathbf{h})$ gradually changes from Type I solution $\mathbf{P}_{\rm min}(\mathbf{h})$, the optimum solution for the outage capacity, to Type IV solution $\mathbf{p}_{\rm wf}(\mathbf{h}, h_0^*)$, the optimum solution for the ergodic capacity. The service outage achievable rate gradually changes from $r_{\rm o}(1-\epsilon)$ to the ergodic capacity. The outage probability is equal to ϵ for the types I-III solutions, and is less than ϵ for the type IV solution.

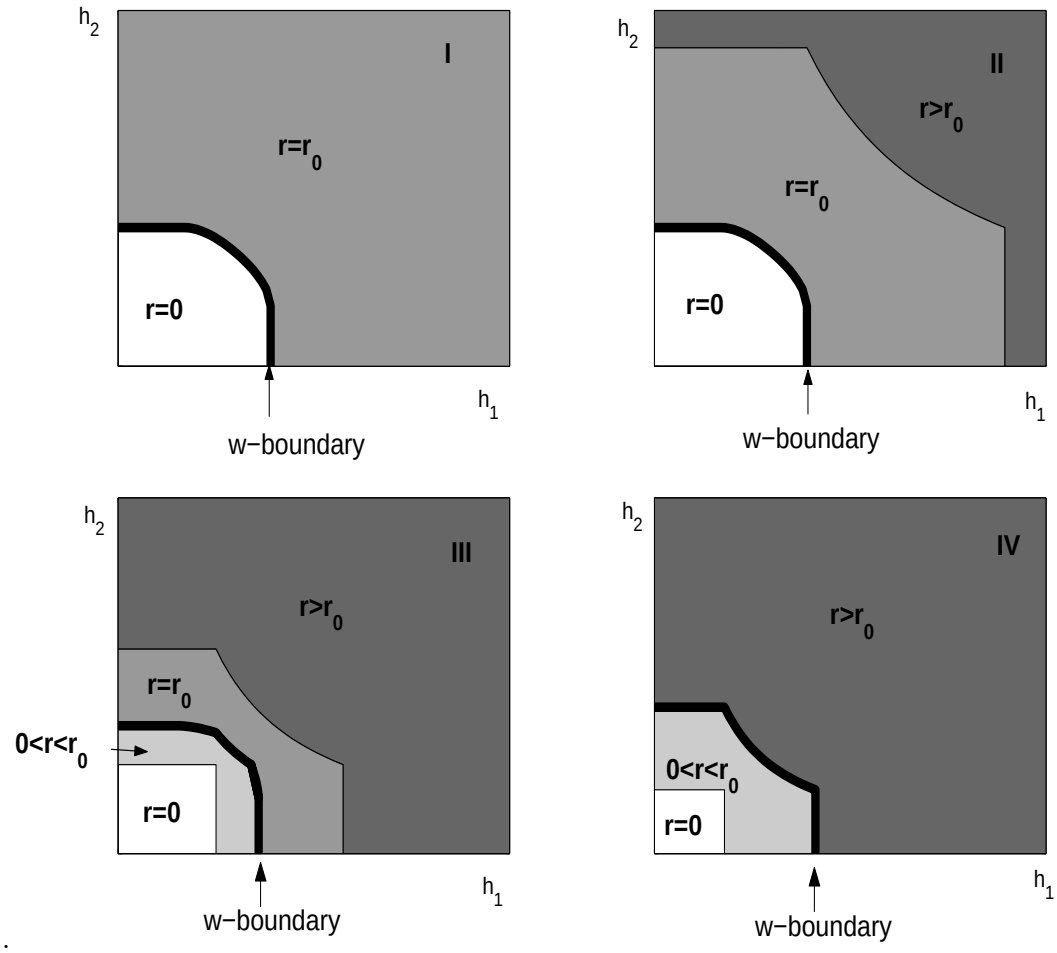


Figure 4.2: Optimum solution types I-IV in $M = 2$ parallel fading channels. The optimum solution is probabilistic only at the boundary set.

The optimum solution is probabilistic at the boundary set only when $s^* > 0$. For a continuous channel distribution, the boundary set has a probability measure zero when $s^* > 0$. Therefore, the optimum solution is deterministic for the continuous channel distribution. As stated before, the deterministic allocation problem for M-parallel fading channel is hard to solve directly. However, by considering the probabilistic allocation problem (4.6), we in fact obtain the optimum solution for the corresponding deterministic allocation problem for continuous channel distributions.

4.3.5 Computation of the Optimum Parameters

In this section, we study the algorithm that determines the parameters of the optimum scheme $\mathbf{P}^*(\mathbf{h})$. The optimum parameters h_0^*, s^*, and $v^*(\mathbf{h})$ are the solutions of the average power constraint and the outage probability constraint (necessary condition) as follows:

$$E\left[\langle \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)\rangle + \langle \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0)\rangle [1(g(\mathbf{h}, h_0) < s) + v(\mathbf{h}) 1(g(\mathbf{h}, h_0) = s)]\right] = p_{\mathrm{av}}, \quad (4.48)$$

$$\Pr\{g(\mathbf{h}, h_0) < s)\} + E\left[v(\mathbf{h}) 1(g(\mathbf{h}, h_0) = s)\right] = 1 - \epsilon, \quad (4.49)$$

$$0 \le v(\mathbf{h}) \le 1\,. \quad (4.50)$$

A solution of (4.48)-(4.50) must exist when Problem (4.6) is feasible. Moreover, Lemma 5 shows that any solution of these equations is the optimum parameter set (sufficiency). In the following, we discuss algorithms to solve (4.48)-(4.50) for continuous channel distributions and discrete channel distributions respectively.

In the case of a continuous channel distribution function $F(\mathbf{h})$, $\{g(\mathbf{h}, h_0) = s\}$ is a set of probability measure zero for $s > 0$, and thus (4.48)-(4.50) can be reduced to

$$E\left[\langle \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)\rangle + 1(g(\mathbf{h}, h_0) \le s)\mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0)\right] = p_{\mathrm{av}}, \quad (4.51)$$

$$\Pr\{g(\mathbf{h}, h_0) \le s\} = 1 - \epsilon\,. \quad (4.52)$$

The left sides of (4.51) and (4.52) are continuous functions of h_0 and s, and a variety of well known root finding algorithms can be used [69].

The case of a discrete channel distribution function is more complicated, since we have to determine the value of $v(\mathbf{h})$ for the boundary set. For given h_0 and s, (4.48)-(4.50) form a linear programming problem on $v(\mathbf{h})$. Standard linear programming approaches (such as employing artificial variables and simplex method) [57] can be used to determine whether there exist a feasible solution of $v(\mathbf{h})$. If there exists a solution of $v(\mathbf{h})$, the corresponding h_0 and s are the solutions we try to find. Therefore, the iterative algorithm is to search h_0 and s until a feasible solution of $v(\mathbf{h})$ is found. The two dimensional search for h_0 and s can be carried in a sequential manner as shown below. For a given h_0, since $0 \le v(\mathbf{h}) \le 1$, (4.49) implies that

$$\Pr\{g(\mathbf{h}, h_0) < s)\} \le 1 - \epsilon \le \Pr\{g(\mathbf{h}, h_0) \le s\}\,. \quad (4.53)$$

Thus, for a given h_0, s can be expressed as

$$s(h_0) = \sup\{x : \Pr\{g(\mathbf{h}, h_0) < x\} < 1 - \epsilon\} \ . \tag{4.54}$$

The linear programming approach for solving $v(\mathbf{h})$ is a numerical method, and usually requires a lot of computations. Moreover, it gives no insight on the structure of the solution. Therefore, in Appendix B we have also derived the exact feasibility condition and a closed form solution of $v(\mathbf{h})$ for any (h_0, s) by exploiting the structure of (4.48)-(4.50).

4.4 Near Optimum Allocation Schemes

In this section, we consider power allocations with the same two-layer structure

$$\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h})) = \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0) + X_w(\mathbf{h})\mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0)\,. \tag{4.55}$$

as the optimum solution $\mathbf{P}^*(\mathbf{h})$. As we can see, any $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ that satisfies $E\left[\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\right] = p_{\mathrm{av}}$ and $E\left[w(\mathbf{h})\right] \geq 1 - \epsilon$ is a feasible scheme for problem (4.6). In this section, by choosing some particular $(h_0, w(\mathbf{h}))$, we obtain two near optimum schemes in the form of (4.55). These schemes are simpler to implement than the optimum $\mathbf{P}^*(\mathbf{h})$ policy.

We first develop bounds for the average rate achieved by the $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ as follows.

Lemma 8 *The average rate achieved by* $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ *with* $E\left[w(\mathbf{h})\right] \geq 1-\epsilon$ *is bounded as*

$$R_{\mathrm{u}}(h_0) - r_o\epsilon \leq E\left[r(\mathbf{h}, \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h})))\right] \leq R_{\mathrm{u}}(h_0),$$

where $R_{\mathrm{u}}(h_0) = r_o + E\left[[r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)) - r_o]^+\right]$.

The upper bound is achieved when we have zero outage, and the lower bound is achieved when the rate during the outage is equal to zero.

It can be seen that for small ϵ the average rate performance is determined mainly by the value of $R_{\mathrm{u}}(h_0)$. Since $r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0))$ is a decreasing function of h_0, $R_{\mathrm{u}}(h_0)$ is

a decreasing function of h_0. Thus, in order to achieve a high average rate, the value of h_0 should be small.

For a given h_0, $w(\mathbf{h})$ that satisfies the average power and outage constraints is not unique. There could be an infinite number of choices of $w(\mathbf{h})$ for a given h_0 when $\mathbf{h}$ is a vector of continuous random variables. The bounds of Lemma 8 imply that the average rate performance is relatively insensitive to the value of $w(\mathbf{h})$ for a given h_0. When the outage probability is sufficiently small, there are many near optimum schemes with small h_0 and the exact shape of the outage set is not critical.

4.4.1 Near Optimum Power Allocation I

Consider a policy $\hat{\mathbf{P}}(\mathbf{h}) = \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, \hat{h}_0) + X_{\hat{w}}(\mathbf{h})\mathbf{p}_{\mathrm{s}}(\mathbf{h}, \hat{h}_0)$ with

$$\hat{w}(\mathbf{h}) = \begin{cases} 1 & \langle \mathbf{p}_{\mathrm{s}}(\mathbf{h}, \hat{h}_0)\rangle < \hat{s} \\ \hat{v}(\mathbf{h}) & \langle \mathbf{p}_{\mathrm{s}}(\mathbf{h}, \hat{h}_0)\rangle = \hat{s} \\ 0 & \langle \mathbf{p}_{\mathrm{s}}(\mathbf{h}, \hat{h}_0)\rangle > \hat{s} \end{cases}, \tag{4.56}$$

where $\hat{h}_0$, $\hat{s}$, and $0 \leq \hat{v}(\mathbf{h}) \leq 1$ are solutions to $E\left[\langle \hat{\mathbf{P}}(\mathbf{h})\rangle\right] = p_{\mathrm{av}}$ and $E\left[\hat{w}(\mathbf{h})\right] = 1 - \epsilon$. We will see that it is sufficient to choose $\hat{v}(\mathbf{h}) = \hat{v}$. For any solution $(\hat{h}_0, \hat{s}, \hat{v}(\mathbf{h}))$, it can be directly verified that $(\hat{h}_0, \hat{s}, \hat{v})$ with

$$\hat{v} = \frac{E\left[\hat{v}(\mathbf{h}) 1(\langle \mathbf{p}_{\mathrm{s}}(\mathbf{h}, \hat{h}_0)\rangle = \hat{s})\right]}{E\left[1(\langle \mathbf{p}_{\mathrm{s}}(\mathbf{h}, \hat{h}_0)\rangle = \hat{s})\right]}$$

is also a solution to $E\left[\langle \hat{\mathbf{P}}(\mathbf{h})\rangle\right] = p_{\mathrm{av}}$ and $E\left[\hat{w}(\mathbf{h})\right] = 1 - \epsilon$. This simplifies the computation of parameters relative to the computation of the optimum solution $\mathbf{P}^*(\mathbf{h})$. In this section, we show that $\hat{\mathbf{P}}(\mathbf{h})$ is a near optimum scheme for problem (4.6). Applying the generalized Kuhn-Tucker conditions, we have the following lemma characterizing $\hat{w}(\mathbf{h})$.

Lemma 9 *$\hat{w}(\mathbf{h})$ is the optimum solution of the following problem.*

$$\hat{w}(\mathbf{h}) = \arg \min_{0 \leq w(\mathbf{h}) \leq 1} E\left[w(\mathbf{h})\langle \mathbf{p}_{\mathrm{s}}(\mathbf{h}, \hat{h}_0)\rangle\right] \tag{4.57}$$

$$\textit{subject to} \quad E\left[w(\mathbf{h})\right] \geq 1 - \epsilon$$

The following lemma is a corollary of Lemma 9.

Lemma 10 *Policy* $\hat{\mathbf{P}}(\mathbf{h})$ *has the minimum water-filling parameter* h_0 *among all* $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ *that satisfy* $E\left[\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\right] = p_{\text{av}}$ *and* $E\left[w(\mathbf{h})\right] \geq 1 - \epsilon$.

Earlier in this section, it is shown that the average rate of $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ is mainly determined by $R_{\text{u}}(h_0)$, which is a decreasing function of h_0. This implies that $\hat{\mathbf{P}}(\mathbf{h})$ with the minimum h_0 should be a good scheme. Applying Lemma 8, we have

$$R^* - r_{\text{o}}\epsilon \overset{(a)}{\leq} R_{\text{u}}(h_0^*) - r_{\text{o}}\epsilon \overset{(b)}{\leq} R_{\text{u}}(\hat{h}_0) - r_{\text{o}}\epsilon \overset{(c)}{\leq} E\left[r(\mathbf{h}, \hat{\mathbf{P}}(\mathbf{h}))\right] \leq R^*. \tag{4.58}$$

Here R^* is the maximum average rate achieved by $\mathbf{P}^*(\mathbf{h})$. Inequalities (a) and (c) are direct results of the rate bounds in Lemma 8. Inequality (b) follows from the fact that $\hat{h}_0 \leq h_0^*$ by Lemma 10. Therefore, $\hat{\mathbf{P}}(\mathbf{h})$ achieves a rate between $R^* - r_{\text{o}}\epsilon$ and R^*, and is a near optimum solution for Problem (4.6) for small ϵ.

To determine $(\hat{h}_0, \hat{s}, \hat{v})$, we have to solve

$$E\left[\langle \mathbf{p}_{\text{wf}}(\mathbf{h}, h_0)\rangle + \langle \mathbf{p}_{\text{s}}(\mathbf{h}, h_0)\rangle 1(\langle \mathbf{p}_{\text{s}}(\mathbf{h}, h_0)\rangle < s)\right] + svE\left[1(\langle \mathbf{p}_{\text{s}}(\mathbf{h}, h_0)\rangle = s)\right] = p_{\text{av}}, \tag{4.59}$$

$$\Pr\{\langle \mathbf{p}_{\text{s}}(\mathbf{h}, h_0)\rangle < s)\} + vE\left[1(\langle \mathbf{p}_{\text{s}}(\mathbf{h}, h_0)\rangle = s)\right] = 1 - \epsilon, \tag{4.60}$$

$$0 \leq v \leq 1\,. \tag{4.61}$$

We search $(\hat{h}_0.\hat{s}, \hat{v})$ in an iterative way. For a given h_0, we first examine (4.60) and (4.61), and obtain $v(h_0)$ and $s(h_0)$ as a function of h_0 as

$$s(h_0) = \sup\left\{x : \Pr\{\langle \mathbf{p}_{\text{s}}(\mathbf{h}, h_0)\rangle < x\} < 1 - \epsilon\right\}, \tag{4.62}$$

$$v(h_0) = \frac{1 - \epsilon - \Pr\{\langle \mathbf{p}_{\text{s}}(\mathbf{h}, h_0)\rangle < s(h_0)\}}{\Pr\{\langle \mathbf{p}_{\text{s}}(\mathbf{h}, h_0)\rangle = s(h_0)\}}\,. \tag{4.63}$$

Then we adjust the value of h_0 until $E\left[\langle \hat{\mathbf{P}}(\mathbf{h})\rangle\right] = p_{\text{av}}$. Each time we adjust h_0, we have to compute $s(h_0)$ and $v(h_0)$ according to (4.62) and (4.63).

4.4.2 Near Optimum Power Allocation II

Consider a policy $\mathbf{P}'(\mathbf{h}) = \mathbf{p}_{\rm wf}(\mathbf{h}, h_0') + X_{w'}(\mathbf{h})\mathbf{p}_{\rm s}(\mathbf{h}, h_0')$ with $w'(\mathbf{h})$ given by (4.21) and h_0' satisfying $E\left[\langle \mathbf{P}'(\mathbf{h})\rangle\right] = p_{\rm av}$. Recall that

$$w'(\mathbf{h}) = \begin{cases} 1 & \langle \mathbf{p}_{\rm r_0}(\mathbf{h})\rangle < s' \\ v' & \langle \mathbf{p}_{\rm r_0}(\mathbf{h})\rangle = s' \\ 0 & \langle \mathbf{p}_{\rm r_0}(\mathbf{h})\rangle > s' \end{cases} . \tag{4.64}$$

Policy $\mathbf{P}'(\mathbf{h})$ allocates the supplemental power at channel states where $\langle \mathbf{p}_{\rm r_0}(\mathbf{h})\rangle$ is below a threshold. Since $w'(\mathbf{h})$ does not depend on h_0', the outage set of $\mathbf{P}'(\mathbf{h})$ is much simpler than $\mathbf{P}^*(\mathbf{h})$ and $\hat{\mathbf{P}}(\mathbf{h})$. Applying the equality

$$\mathbf{p}_{\rm wf}(\mathbf{h}, h_0) + \mathbf{p}_{\rm s}(\mathbf{h}, h_0) = \mathbf{p}_{\rm r_0}(\mathbf{h}) + [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0) - \mathbf{p}_{\rm r_0}(\mathbf{h})]^+,$$

$\mathbf{P}'(\mathbf{h})$ can be expressed equivalently as

$$\mathbf{P}'(\mathbf{h}) = \mathbf{P}_{\rm min}(\mathbf{h}) + [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0') - \mathbf{P}_{\rm min}(\mathbf{h})]^+ . \tag{4.65}$$

Recall that $\mathbf{P}_{\rm min}(\mathbf{h}) = X_{w'}(\mathbf{h})\mathbf{p}_{\rm r_0}(\mathbf{h})$ achieves the minimum sufficient power to meet the outage constraint. The physical meaning of $\mathbf{P}'(\mathbf{h})$ is that: we first assign $\mathbf{P}_{\rm min}(\mathbf{h})$ to meet the outage constraint with the minimum sufficient power, and then allocate the remaining power in an optimum way to maximize the excess rate. The $[\mathbf{p}_{\rm wf}(\mathbf{h}, h_0') - \mathbf{P}_{\rm min}(\mathbf{h})]^+$ is in fact a 'water-filling' allocation when $\mathbf{P}_{\rm min}(\mathbf{h})$ is viewed as an interference. In the following, we show that $\mathbf{P}'(\mathbf{h})$ is a near optimum scheme for Problem (4.6).

It is hard to show directly that $\mathbf{P}'(\mathbf{h})$ is a near optimum scheme. Our approach is to introduce a second scheme $\mathbf{P}''(\mathbf{h})$ as an intermediate step, as

$$\mathbf{P}''(\mathbf{h}) = \mathbf{P}_{\rm min}(\mathbf{h}) + [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0'') - \mathbf{p}_{\rm r_0}(\mathbf{h})]^+ , \tag{4.66}$$

where h_0'' is the solution to $E\left[\langle \mathbf{P}''(\mathbf{h})\rangle\right] = p_{\rm av}$. The following lemmas on $\mathbf{P}''(\mathbf{h})$ allow us to show that $\mathbf{P}''(\mathbf{h})$ is a near optimum scheme of problem (4.6).

Lemma 11 *The average rate achieved by* $\mathbf{P}''(\mathbf{h})$ *satisfies* $\mathrm{E}\left\{r(\mathbf{h}, \mathbf{P}''(\mathbf{h}))\right\} \geq R_{\rm u}(h_0'') - r_o\epsilon$.

Lemma 12 *We have* $h_0'' \leq \hat{h}_0$, *where the* $\hat{h}_0$ *is the water-filling parameter in* $\hat{\mathbf{P}}(\mathbf{h})$.

Applying Lemma 11 and Lemma 12, we have

$$R^* \overset{\text{(a)}}{\geq} \mathrm{E}\left\{r(\mathbf{h}, \mathbf{P}''(\mathbf{h}))\right\} \overset{\text{(b)}}{\geq} R_{\mathrm{u}}(h_0'') - r_{\mathrm{o}}\epsilon \overset{\text{(c)}}{\geq} R_{\mathrm{u}}(h_0^*) - r_{\mathrm{o}}\epsilon \overset{\text{(d)}}{\geq} R^* - r_{\mathrm{o}}\epsilon\,. \tag{4.67}$$

Inequality (a) holds since $\mathbf{P}''(\mathbf{h})$ is a feasible scheme for problem (4.6). Inequality (b) follows from Lemma 11. Inequality (c) holds since $R_{\mathrm{u}}(h_0)$ is a decreasing function of h_0 and $h_0'' < \hat{h}_0 \leq h_0^*$ by Lemma 10 and Lemma 12. Applying the rate bounds in Lemma 8 to the optimum allocation $\mathbf{P}^*(\mathbf{h})$, we have $R^* = E\left[r(\mathbf{h}, \mathbf{P}^*(\mathbf{h}))\right] \leq R_{\mathrm{u}}(h_0^*)$ and thus inequality (d) holds. Therefore, $\mathbf{P}''(\mathbf{h})$ is a near optimum scheme for problem (4.6) for small ϵ.

As we can see, both $\mathbf{P}'(\mathbf{h})$ and $\mathbf{P}''(\mathbf{h})$ first allocate $\mathbf{P}_{\min}(\mathbf{h})$ to meet the outage constraint, but the $\mathbf{P}'(\mathbf{h})$ allocates the remaining power in an optimum way to maximize the additional rate.

Lemma 13 *Scheme* $\mathbf{P}'(\mathbf{h})$ *achieves a higher average rate than* $\mathbf{P}''(\mathbf{h})$.

Therefore, $\mathbf{P}'(\mathbf{h})$ is a near optimum scheme for problem (4.6) for small ϵ.

The computation of $\mathbf{P}'(\mathbf{h})$ is much simpler than $\mathbf{P}^*(\mathbf{h})$ and $\hat{\mathbf{P}}(\mathbf{h})$, since its (s', v') do not depend on h_0'. The (s', v') can be determined by solving the outage probability constraint alone, which is given by (4.22) and (4.23). Then h_0' can be determined by solving $E\left[\langle \mathbf{P}'(\mathbf{h})\rangle\right] = p_{\mathrm{av}}$ using a line search technique. Therefore, in $\mathbf{P}'(\mathbf{h})$ the (s', v') are the same for different values of p_{av}, while in $\mathbf{P}^*(\mathbf{h})$ and $\hat{\mathbf{P}}(\mathbf{h})$ we have to compute (s, v) for each value of p_{av}.

The structure of $\mathbf{P}'(\mathbf{h})$ in (4.65) suggests a simple implementation of transmission of mixed real-time and non real-time services. The $\mathbf{P}_{\min}(\mathbf{h})$ can be used to transmit the real-time service with the basic-rate requirement, and $[\mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0') - \mathbf{P}_{\min}(\mathbf{h})]^+$ can be used to transmit the non real-time service. Two codebooks will be generated according to the corresponding power assignments to these two services, and transmitted simultaneously using superposition coding. The successive decoding is employed at receiver.

4.4.3 Discussion

We derived three allocation schemes $\mathbf{P}^*(\mathbf{h})$, $\hat{\mathbf{P}}(\mathbf{h})$, and $\mathbf{P}'(\mathbf{h})$. All of these three schemes have a similar two-layer structure: $\mathbf{p}_{\text{wf}}(\mathbf{h}, h_0) + X_w(\mathbf{h})\mathbf{p}_{\text{s}}(\mathbf{h}, h_0)$ but with different h_0 and $w(\mathbf{h})$. The $w(\mathbf{h})$ determines where to allocate the supplemental power according to a metric. The metric is $g(\mathbf{h}, h_0^*)$ for $\mathbf{P}^*(\mathbf{h})$, $\langle \mathbf{p}_{\text{s}}(\mathbf{h}, \hat{h}_0) \rangle$ for $\hat{\mathbf{P}}(\mathbf{h})$, and $\langle \mathbf{p}_{\text{r}_0}(\mathbf{h}) \rangle$ for $\mathbf{P}'(\mathbf{h})$. The service outage happens at the channel states where the metric is above a threshold. Since the metrics $g(\mathbf{h}, h_0^*)$, $\langle \mathbf{p}_{\text{s}}(\mathbf{h}, \hat{h}_0) \rangle$, and $\langle \mathbf{p}_{\text{r}_0}(\mathbf{h}) \rangle$ are all non-increasing functions of $\mathbf{h}$, outage occurs at poor channel states for a good scheme, which is consistent with the intuition.

In previous sections, we have shown that $\hat{\mathbf{P}}(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h})$ achieve a rate between $R^* - r_{\text{o}}\epsilon$ and R^*. It can be directly verified that for sufficiently small p_{av} such that $\mathbf{p}_{\text{wf}}(\mathbf{h}, h_0^*) = 0$ holds in the outage set, we have $\mathbf{P}^*(\mathbf{h}) = \hat{\mathbf{P}}(\mathbf{h}) = \mathbf{P}'(\mathbf{h})$. For sufficiently large average power, all three policies become water-filling allocation. Thus, for sufficiently small and large p_{av}, both $\hat{\mathbf{P}}(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h})$ are optimum. Moreover, as shown below, we have $\hat{\mathbf{P}}(\mathbf{h}) = \mathbf{P}'(\mathbf{h}) = \mathbf{P}^*(\mathbf{h})$ in an $M = 1$ fading channel for all parameters. When $M = 1$, since the metrics $g(h, h_0^*)$, $\langle \mathbf{p}_{\text{s}}(h, \hat{h}_0) \rangle$, and $\langle \mathbf{p}_{\text{r}_0}(h) \rangle$ are all non-increasing functions of h, all $w^*(h)$, $\hat{w}(h)$, and $w'(h)$ can be expressed in the same way as

$$w(h) = \begin{cases} 1 & h < h_b \\ v & h = h_b \\ 0 & h > h_b \end{cases}, \tag{4.68}$$

Hence, v and h_b are the same for all three schemes since they are the solutions of $E[w(h)] = 1 - \epsilon$. With the same $w(h)$, the average power constraints implies that $h_0^* = \hat{h}_0 = h_0'$. Thus, we have $\hat{\mathbf{P}}(\mathbf{h}) = \mathbf{P}'(\mathbf{h}) = \mathbf{P}^*(\mathbf{h})$ in an $M = 1$ fading channel.

4.5 Numerical Results

In most of our numerical results, the two near optimum schemes achieve an average rate almost equal to the maximum rate R^* achieved by $\mathbf{P}^*(\mathbf{h})$, and the lower bound $R^* - r_{\text{o}}\epsilon$ is loose, especially for large ϵ. To highlight the performance difference between these three schemes, we construct a particular two state model as follows: in an $M = 2$ channel, the

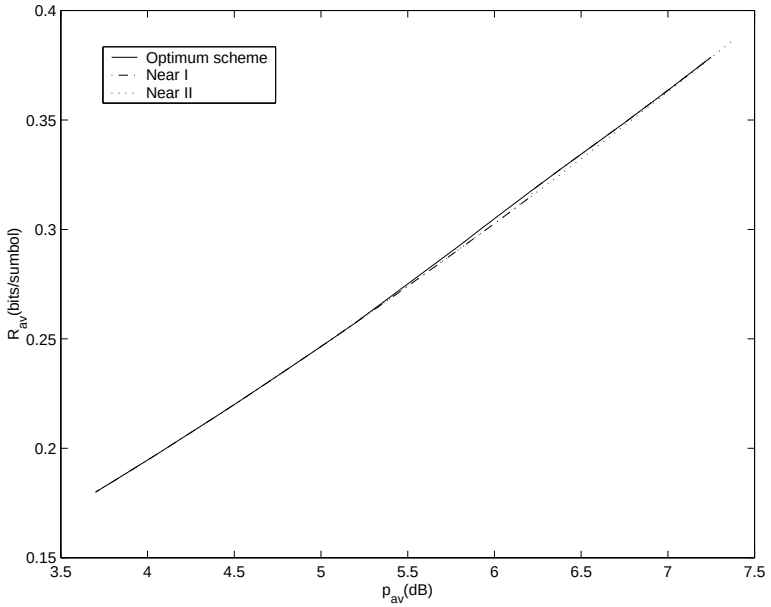


Figure 4.3: The average rate performance of the optimum scheme $\mathbf{P}^*(\mathbf{h})$ versus two near optimum schemes $\hat{\mathbf{P}}(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h})$ for a two state model with fixed $r_o = 0.36$ bits/symbol and $\epsilon = 1/2$.

channel state vector $\mathbf{h} = (h_1, h_2)$ is equal to $(0.1238, 0.1238)$ with probability $1/2$, and equal to $(0.1827, 0)$ with probability $1/2$. The average rate versus the average power performance for $\mathbf{P}^*(\mathbf{h})$, $\hat{\mathbf{P}}(\mathbf{h})$, and $\mathbf{P}'(\mathbf{h})$ in this model with fixed $r_o = 0.36$ bits/symbol and $\epsilon = 1/2$ is given by Figure 4.3. It can be seen that the optimum solution $\mathbf{P}^*(\mathbf{h})$ achieves a slightly higher average rate than the near optimum schemes $\hat{\mathbf{P}}(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h})$. For sufficiently small and high average power, both $\hat{\mathbf{P}}(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h})$ are equal to $\mathbf{P}^*(\mathbf{h})$. In this example, $\hat{\mathbf{P}}(\mathbf{h})$ is slightly better than $\mathbf{P}'(\mathbf{h})$ in a range of parameters. As shown in Fig 4.4 with $r_o = 0.5$ bits/symbol, $\mathbf{P}'(\mathbf{h})$ is almost as good as $\mathbf{P}^*(\mathbf{h})$, and both are better than $\hat{\mathbf{P}}(\mathbf{h})$. The relative performance difference is still much less than $\epsilon = 0.5$.

We apply the results to the $M = 2$ parallel Rayleigh fading channel model. To simplify the computations, we assume that the sub-channels are iid with the joint PDF:

$$f(h_1, h_2) = \begin{cases} e^{-(h_1+h_2)} & h_1 \geq 0, h_2 \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.69}$$

In Fig 4.5, the average rate versus the average power is plotted for $\mathbf{P}^*(\mathbf{h})$, $\hat{\mathbf{P}}(\mathbf{h})$, and

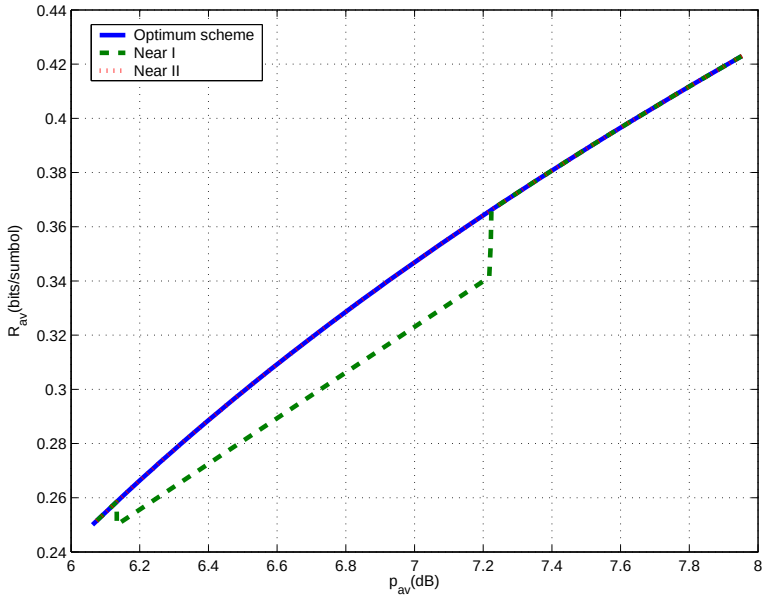


Figure 4.4: The average rate performance of the optimum scheme $\mathbf{P}^*(\mathbf{h})$ versus two near optimum schemes $\hat{\mathbf{P}}(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h})$ for a two state model with fixed $r_{\rm o} = 0.5$ bits/symbol and $\epsilon = 1/2$.

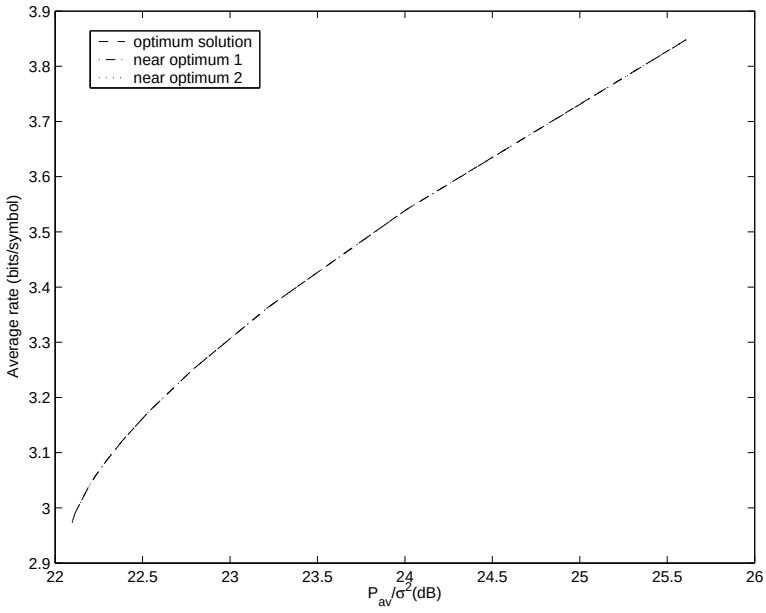


Figure 4.5: The average rate performance of the optimum scheme $\mathbf{P}^*(\mathbf{h})$ versus two near optimum schemes $\hat{\mathbf{P}}(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h})$ in Rayleigh fading channel with fixed $\epsilon = 0.01$ and $r_{\rm o} = 3$ bits/symbol.

$\mathbf{P}'(\mathbf{h})$ with fixed $\epsilon = 0.01$ and $r_{\rm o} = 3$ bits/symbol in Rayleigh fading channel. As we can see, the near optimum schemes are indistinguishable from the optimum solution $\mathbf{P}^*(\mathbf{h})$. In Fig 4.6, the service outage achievable rates with different $r_{\rm o}$ are plotted against the ergodic capacity and the outage capacity in $M = 2$ Rayleigh fading channel. As we can see, for the given outage probability $\epsilon = 0.01$, the outage capacity has a close to 2 dB loss in the average power compared to the ergodic capacity. A larger average power loss is expected when the outage probability is smaller. Between the outage capacity and the ergodic capacity, a number of service outage achievable rates with different r_0 exist. The service outage achievable rate is always between the outage capacity times $(1-\epsilon)$ and the ergodic capacity. Starting from $r_0(1-\epsilon)$, it approaches the ergodic capacity as the average power increases. The outage probability achieved by the water filling allocation with respect to different r_0 is also plotted against the service outage solution with a given $\epsilon = 0.01$ in Fig 4.7. It can be observed that, for a range of $P_{\rm av}$, the service outage solution achieves a rate very close to the ergodic capacity, and, at the same time, significantly reduces the outage probability. Hence, the service outage approach strikes good balance between average rate and outage probability.

Since the basic rate allocation $\mathbf{p}_{\rm r_0}(\mathbf{h})$ converges to $\mathbf{p}_{\rm wf}(\mathbf{h}, h_0)$, the water-filling allocation, as $M \to \infty$ [18], in the following, we compare the performance of $\mathbf{p}_{\rm wf}(\mathbf{h}, h_0)$ and $\mathbf{P}_{\rm min}(\mathbf{h})$ when M is large.

The water-filling allocation $\mathbf{p}_{\rm wf}(\mathbf{h}, h_0)$ achieves the same ergodic capacity per subchannel at all M. The instantaneous rate $r(\mathbf{h}, \mathbf{P}_{\rm min}(\mathbf{h}))$ achieved by $\mathbf{P}_{\rm min}(\mathbf{h})$ is either 0 during the outage set, or the outage capacity (that is the basic rate) during non-outage. The outage capacity converges to the ergodic capacity as $M \to \infty$. In fact, the outage capacity approaches the ergodic capacity very fast. In Fig 4.8, the outage capacity achieved by $\mathbf{P}_{\rm min}(\mathbf{h})$ with a fixed outage probability $\epsilon = 0.01$ is plotted against the ergodic capacity for $M = 1, 2, 5$ independent Rayleigh fading channels. As we can see, $\mathbf{P}_{\rm min}(\mathbf{h})$ with $\epsilon = 0.01$ achieves an outage capacity within 1 db of the ergodic capacity when $M = 5$.

To compare the instantaneous rate performance of $\mathbf{p}_{\rm wf}(\mathbf{h}, h_0)$ and $\mathbf{P}_{\rm min}(\mathbf{h})$, we define *the achievable outage rate* of a power allocation as the rate which can be supported

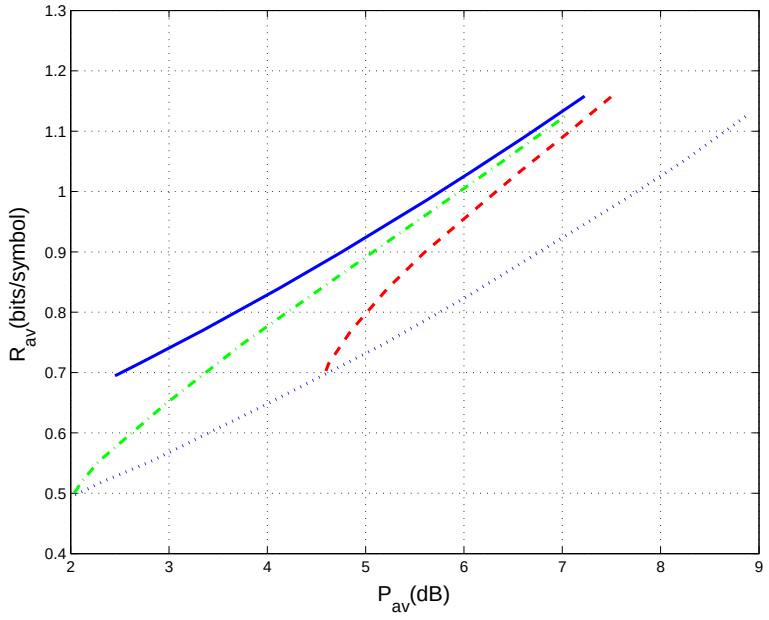


Figure 4.6: Comparison of the service outage achievable rate with other capacity notions in an $M = 2$ Rayleigh fading channel for a fixed $\epsilon = 0.01$.

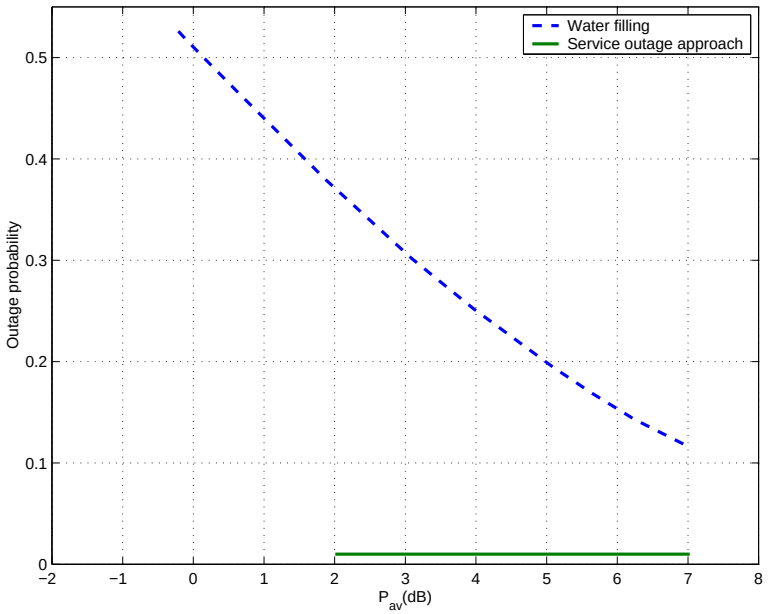


Figure 4.7: The outage probability of water-filling allocation with respect to $r_{\mathrm{o}} = 0.5$ is compared to the service outage solution in $M = 2$ Rayleigh fading channel.

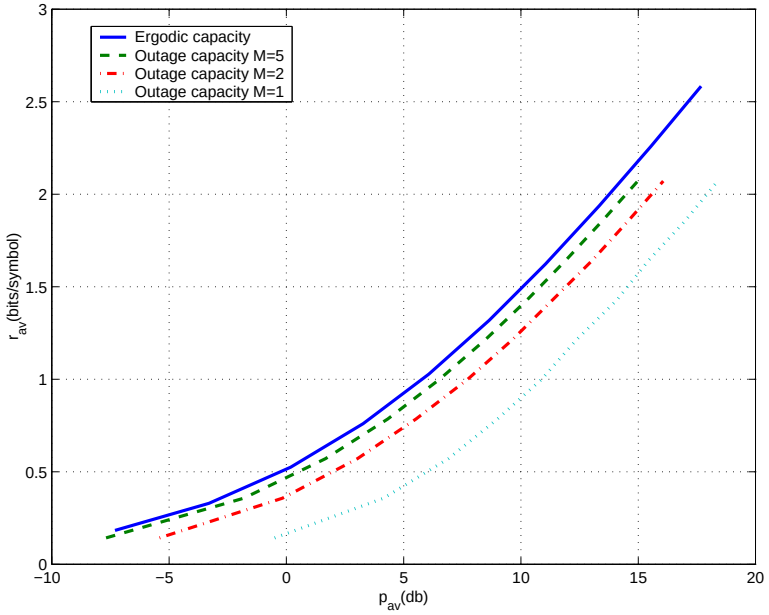


Figure 4.8: Comparison of the ergodic capacity per sub channel with the outage capacity for a fixed outage probability $\epsilon = 0.01$ in $M = 1, 2, 5$ independent Rayleigh fading channel.

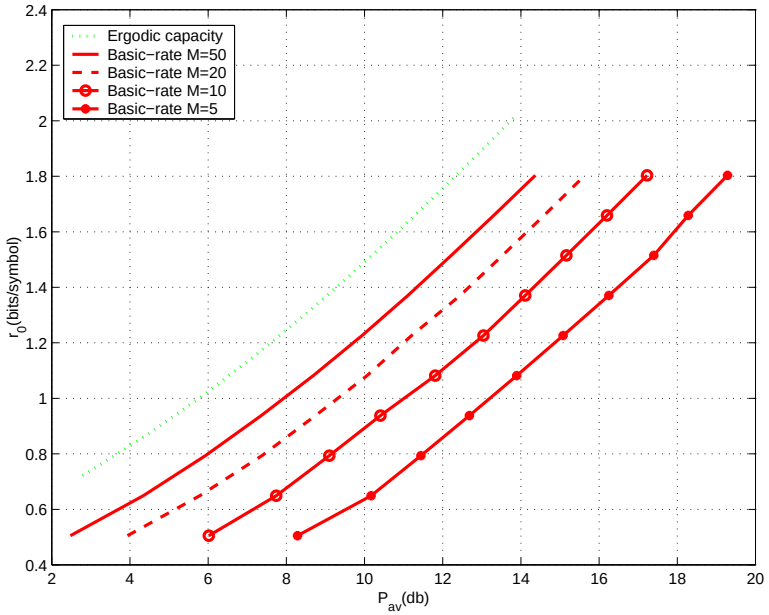


Figure 4.9: The achievable outage rate by the water-filling allocation $\mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)$ for a fixed outage probability $\epsilon = 0.01$ in $M = 5, 10, 20, 50$ independent Rayleigh fading channel.

throughout the fading process with an outage probability less than ϵ. The achievable outage rate is in fact the basic rate defined in the service outage based allocation problems. The achievable outage rate of $\mathbf{P}_{\min}(\mathbf{h})$ is just the outage capacity, which is close to the ergodic capacity for a moderate large M as shown in Fig 4.8. The achievable outage rate of $\mathbf{p}_{\text{wf}}(\mathbf{h}, h_0)$ for a given outage probability $\epsilon = 0.01$ in $M = 5, 10, 20, 50$ independent Rayleigh fading channel is plotted against the ergodic capacity in Fig 4.9. As we can see, the achievable outage rate of $\mathbf{p}_{\text{wf}}(\mathbf{h}, h_0)$ is still far from the ergodic capacity, which is almost equal to the outage capacity of $\mathbf{P}_{\min}(\mathbf{h})$ when $M = 50$. Therefore, when M is large, $\mathbf{P}_{\min}(\mathbf{h})$ outperforms $\mathbf{p}_{\text{wf}}(\mathbf{h}, h_0)$ in outage probability significantly, while at the same time achieves almost the same average rate as $\mathbf{p}_{\text{wf}}(\mathbf{h}, h_0)$. On the other hand, $\mathbf{P}_{\min}(\mathbf{h})$ is much more complicated than $\mathbf{p}_{\text{wf}}(\mathbf{h}, h_0)$ when M is large. In Fig 4.10, the ratio of the achievable outage rate and the ergodic capacity versus p_{av} is plotted for water-filling allocation at fixed outage probability $\epsilon = 0.01$ for $M = 1, 2, 5, 10, 20$ independent Rayleigh fading channel. As we can see, for a moderate large M, the water-filling allocation can support an achievable outage rate within a high percentage of ergodic capacity, which may be sufficient for some applications.

Since the instantaneous rate achieved by the water-filling allocation is a sum of iid random variables, the outage probability can be calculated using large deviation theory. See Appendix 4.C for a large deviation approximation formula. As shown in Fig 4.11, the large deviation approximation agrees with the simulation very well.

4.6 Conclusion

The service outage based allocation problem is to maximize the expected rate subject to the average power constraint and the outage probability constraint in the class of probabilistic power allocation schemes. The feasibility condition of this allocation problem can be obtained from the capacity versus outage probability problem [18]. The optimum power allocation is derived for an M-parallel fading channels model. The result can be applied to both discrete and continuous fading distributions.

The optimum power allocation is shown to be a combination of the water-filling

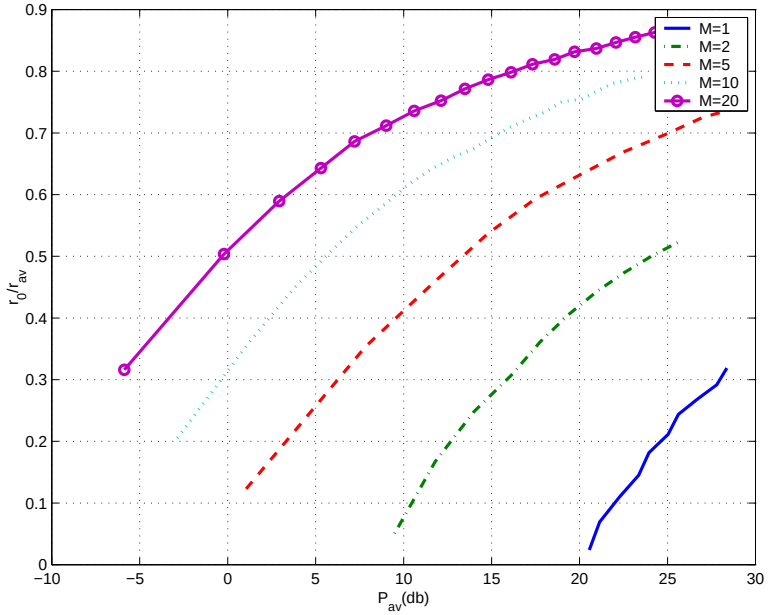


Figure 4.10: The ratio of the achievable outage rate and the ergodic capacity versus p_{av} for water-filling allocation at fixed outage probability $\epsilon = 0.01$ for $M = 1, 2, 5, 10, 20$ independent Rayleigh fading channel.

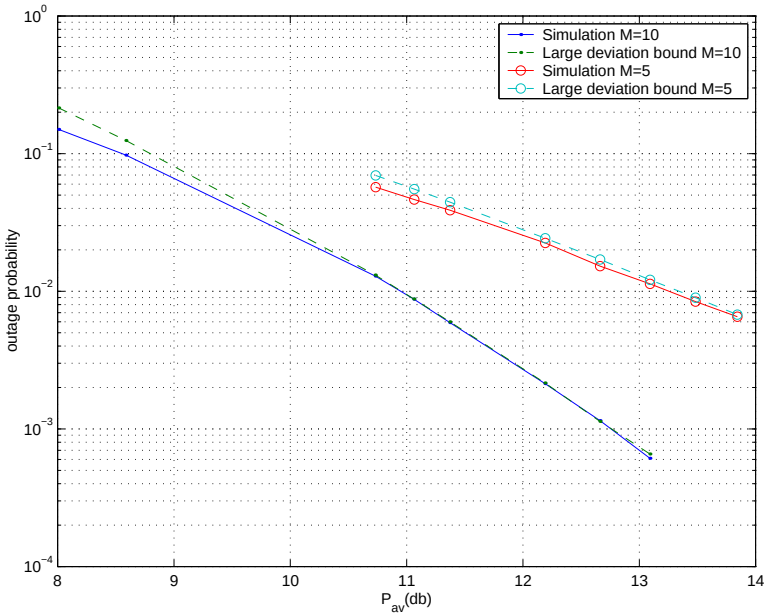


Figure 4.11: The outage probability of water-filling allocation for $r_0 = 1$ bits/symbol by simulation and large deviation bound for $M = 5$ and $M = 10$

allocation and the basic-rate allocation, and is deterministic except at the boundary set. It can be viewed as a two-layer allocation: the first layer is the water-filling allocation, and the second layer the supplemental power allocation. The supplemental power is only allocated at channel states where the supplemental cost $g(\mathbf{h}, h_0^*)$ is below a threshold. With increasing average power, the optimum power allocation gradually changes from $\mathbf{P}_{\min}(\mathbf{h})$, which is the optimum solution for the outage capacity, to $\mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0^*)$, which is the optimum solution for the ergodic capacity. The service outage based achievable rate R^* gradually changes from $r_{\mathrm{o}}(1-\epsilon)$ to the ergodic capacity. The service outage approach strikes a good balance between the outage probability and the average rate.

Two near optimum schemes are also derived: the $\hat{\mathbf{P}}(\mathbf{h})$ with the minimum h_0 and the $\mathbf{P}'(\mathbf{h})$ based on $\mathbf{P}_{\min}(\mathbf{h})$. Both $\hat{\mathbf{P}}(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h})$ have similar structures as the optimum solution, but the supplemental power is allocated according to different metric functions. We have $\hat{\mathbf{P}}(\mathbf{h}) = \mathbf{P}'(\mathbf{h}) = \mathbf{P}^*(\mathbf{h})$ in an $M = 1$ fading channel and for a range of parameters in $M \geq 2$ fading channels. Otherwise, $\hat{\mathbf{P}}(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h})$ achieve a rate between $R^* - r_{\mathrm{o}}\epsilon$ and R^*. The derivation of near optimum schemes shows that the exact shape of the outage set is not critical, a feasible scheme in the form of $\mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0) + X_w(\mathbf{h})\mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0)$ achieves a high average rate as long as the corresponding h_0 is small. The near optimum scheme $\mathbf{P}'(\mathbf{h})$ has an immediate application on transmission of mixed real-time and non real-time services. Its computation is also significantly simpler than the optimum solution $\mathbf{P}^*(\mathbf{h})$ and the near optimum scheme $\hat{\mathbf{P}}(\mathbf{h})$.

4.A Proofs

4.A.1 Lemma 4

Our approach is to show that for an arbitrary feasible probabilistic power allocation scheme $\mathbf{P}(\mathbf{h})$ with a conditional pdf $f_{\mathbf{P}|\mathbf{h}}(\mathbf{p}|\mathbf{h})$, we can always construct another feasible scheme $\mathbf{P}'(\mathbf{h})$ which is randomized between two deterministic schemes $\mathbf{p}_{\mathrm{a}}(\mathbf{h})$ and $\mathbf{p}_{\mathrm{b}}(\mathbf{h})$ with $r(\mathbf{h}, \mathbf{p}_{\mathrm{a}}(\mathbf{h})) \geq r_{\mathrm{o}}$ and the sharing factor $w(\mathbf{h})$ satisfying $E[w(\mathbf{h})] \geq 1 - \epsilon$. It can be shown that $\mathbf{P}'(\mathbf{h})$ achieves a higher average rate than $\mathbf{P}(\mathbf{h})$. This implies that there exists an optimum scheme which is randomized between two deterministic schemes, and

one of them has a rate higher or equal to $r_{\rm o}$ and $E\left[w(\mathbf{h})\right] \geq 1-\epsilon$.

The feasibility of $\mathbf{P}(\mathbf{h})$ implies that $E\left[\langle \mathbf{P}(\mathbf{h})\rangle\right] \leq p_{\rm av}$ and $\Pr\{r(\mathbf{h},\mathbf{P}(\mathbf{h})) < r_{\rm o}\} \leq \epsilon$. Deterministic schemes $\mathbf{p}_{\rm a}(\mathbf{h})$ and $\mathbf{p}_{\rm b}(\mathbf{h})$, and the weighting function $w(\mathbf{h})$ are constructed as follows:

$$\begin{aligned} w(\mathbf{h}) &= \Pr\{r(\mathbf{h},\mathbf{P}(\mathbf{h})) \geq r_{\rm o}|\mathbf{h}\} \\ \mathbf{p}_{\rm a}(\mathbf{h}) &= E\left[\mathbf{P}(\mathbf{h})|r(\mathbf{h},\mathbf{p}(\mathbf{h})) \geq r_{\rm o},\mathbf{h}\right] \\ \mathbf{p}_{\rm b}(\mathbf{h}) &= E\left[\mathbf{P}(\mathbf{h})|r(\mathbf{h},\mathbf{p}(\mathbf{h})) < r_{\rm o},\mathbf{h}\right] . \end{aligned} \tag{4.70}$$

Clearly, $1-w(\mathbf{h})$ is the outage probability of $\mathbf{P}(\mathbf{h})$ for a given $\mathbf{h}$. Since $\mathbf{P}(\mathbf{h})$ meets the outage probability constraint, we must have $E\left[1-w(\mathbf{h})\right] \leq \epsilon$. The $\mathbf{p}_{\rm a}(\mathbf{h})$ is the conditional average of $\mathbf{P}(\mathbf{h})$ whose rate is larger than or equal to $r_{\rm o}$, while the $\mathbf{p}_{\rm b}(\mathbf{h})$ is the conditional average of $\mathbf{P}(\mathbf{h})$ whose rate is smaller than $r_{\rm o}$. Since $r(\mathbf{h},\mathbf{p})$ is concave on $\mathbf{p}$ for a given $\mathbf{h}$, applying Jensen's inequality we have

$$\begin{aligned} r(\mathbf{h},\mathbf{p}_{\rm a}(\mathbf{h})) &= r(\mathbf{h},E\left[\mathbf{P}(\mathbf{h})|r(\mathbf{h},\mathbf{p}(\mathbf{h})) \geq r_{\rm o},\mathbf{h}\right]) \\ &\geq E\left[r(\mathbf{h},\mathbf{P}(\mathbf{h}))|r(\mathbf{h},\mathbf{p}(\mathbf{h})) \geq r_{\rm o},\mathbf{h}\right] \qquad (4.71) \\ r(\mathbf{h},\mathbf{p}_{\rm b}(\mathbf{h})) &= r(\mathbf{h},E\left[\mathbf{P}(\mathbf{h})|r(\mathbf{h},\mathbf{p}(\mathbf{h})) < r_{\rm o},\mathbf{h}\right]) \\ &\geq E\left[r(\mathbf{h},\mathbf{P}(\mathbf{h}))|r(\mathbf{h},\mathbf{p}(\mathbf{h})) < r_{\rm o},\mathbf{h}\right] . \qquad (4.72) \end{aligned}$$

Consider a new probabilistic scheme $\mathbf{P}'(\mathbf{h})$ such that $\mathbf{P}'(\mathbf{h}) = \mathbf{p}_{\rm a}(\mathbf{h})$ with probability $w(\mathbf{h})$ and $\mathbf{P}'(\mathbf{h}) = \mathbf{p}_{\rm b}(\mathbf{h})$ with probability $1-w(\mathbf{h})$. The average power of $\mathbf{P}'(\mathbf{h})$ is

$$E\left[\langle \mathbf{P}'(\mathbf{h})\rangle\right] = E\left[w(\mathbf{h})\langle \mathbf{p}_{\rm a}(\mathbf{h})\rangle + (1-w(\mathbf{h}))\langle \mathbf{p}_{\rm b}(\mathbf{h})\rangle\right] = E\left[\langle \mathbf{P}(\mathbf{h})\rangle\right] \leq p_{\rm av}\,. \tag{4.73}$$

Since $r(\mathbf{h},\mathbf{p}_{\rm a}(\mathbf{h})) \geq r_{\rm o}$ by (4.71) and $\Pr\{\mathbf{P}'(\mathbf{h}) = \mathbf{p}_{\rm a}(\mathbf{h})\} = w(\mathbf{h})$, we have

$$\Pr\{r(\mathbf{h},\mathbf{P}'(\mathbf{h})) \geq r_{\rm o}|\mathbf{h}\} \geq w(\mathbf{h})\,.$$

Thus, the outage probability of $\mathbf{P}'(\mathbf{h})$ satisfies

$$\Pr\{r(\mathbf{h},\mathbf{P}'(\mathbf{h})) < r_{\rm o}\} = 1 - E\left[\Pr\{r(\mathbf{h},\mathbf{P}'(\mathbf{h})) \geq r_{\rm o}|\mathbf{h}\}\right] \leq 1 - E\left[w(\mathbf{h})\right] \leq \epsilon\,. \tag{4.74}$$

From (4.73) and (4.74), $\mathbf{P}'(\mathbf{h})$ is also a feasible scheme for problem (4.6). Inequalities (4.71) and (4.72) imply that $\mathbf{P}'(\mathbf{h})$ achieves an average rate higher than or equal to

$\mathbf{P}(\mathbf{h})$, that is

$$E\left[r(\mathbf{h},\mathbf{P}'(\mathbf{h}))\right] = E\left[w(\mathbf{h})r(\mathbf{h},\mathbf{p}_{\mathrm{a}}(\mathbf{h})) + (1-w(\mathbf{h}))r(\mathbf{h},\mathbf{p}_{\mathrm{b}}(\mathbf{h}))\right] \geq E\left[r(\mathbf{h},\mathbf{P}(\mathbf{h}))\right] .$$

Thus, from any arbitrary feasible power allocation we can always construct a better feasible power allocation which is randomized between two deterministic power allocations. This implies that there must exist an optimum power allocation which is randomized between two deterministic power allocations. Furthermore, it is required that $r(\mathbf{h},\mathbf{p}_{\mathrm{a}}(\mathbf{h})) \geq r_{\mathrm{o}}$ and $E\left[w(\mathbf{h})\right] \geq 1-\epsilon$. Also it is easy to see that $E\left[\langle\mathbf{P}(\mathbf{h})\rangle\right] = p_{\mathrm{av}}$ should hold for the optimum solution; otherwise, a higher average rate can be achieved by increasing the power.

4.A.2 Proof of Lemma 5

In order to prove Lemma 5, we need the following propositions.

Proposition 5 *If $f(y)$ is a concave function over y, then function $l(x,y) = xf(y/x)$ is a concave function over non negative (x,y).*

Proof: Applying the fact that $f(y)$ is a concave function and $l(x,y) = xf(\frac{y}{x})$, we have

$$\begin{aligned}
&l(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \\
&= (\lambda x_1 + (1-\lambda)x_2) f\left(\frac{\lambda y_1 + (1-\lambda)y_2}{\lambda x_1 + (1-\lambda)x_2}\right) \\
&= (\lambda x_1 + (1-\lambda)x_2) f\left(\frac{\lambda x_1}{\lambda x_1 + (1-\lambda)x_2}\frac{y_1}{x_1} + \frac{(1-\lambda)x_2}{\lambda x_1 + (1-\lambda)x_2}\frac{y_2}{x_2}\right) \\
&\overset{\text{(a)}}{\geq} (\lambda x_1 + (1-\lambda)x_2)\left[\frac{\lambda x_1}{\lambda x_1 + (1-\lambda)x_2} f\left(\frac{y_1}{x_1}\right) + \frac{(1-\lambda)x_2}{\lambda x_1 + (1-\lambda)x_2} f\left(\frac{y_2}{x_2}\right)\right] \\
&= \lambda x_1 f\left(\frac{y_1}{x_1}\right) + (1-\lambda)x_2 f\left(\frac{y_2}{x_2}\right) \\
&= \lambda l(x_1,y_1) + (1-\lambda) l(x_2,y_2) .
\end{aligned} \tag{4.75}$$

Note that non-negativity of (x,y) is used in inequality (a). □

In the following proposition, we use $\nabla f(\mathbf{x})$ to denote the gradient of $f(\mathbf{x})$.

Proposition 6 *Let $\mathbf{x}$ and $\mathbf{y}$ be two vectors with equal lengths. Let $g(\mathbf{y}) = f(\mathbf{X}(\mathbf{y}))$ where $\mathbf{x} = \mathbf{X}(\mathbf{y})$ is a one to one transformation between $\mathbf{x}$ and $\mathbf{y}$. If $\hat{\mathbf{x}}$ is a solution to $\nabla f(\mathbf{x}) = 0$, then $\hat{\mathbf{y}} = \mathbf{X}^{-1}(\hat{\mathbf{x}})$ is also a solution to $\nabla g(\mathbf{y}) = 0$.*

Now we return to the proof of Lemma 5, and show that problem (4.25) can be transformed into a convex optimization problem. Define $\mathbf{q}_a(\mathbf{h}) = w(\mathbf{h})\mathbf{p}_{\mathrm{a}}(\mathbf{h})$ and $\mathbf{q}_b(\mathbf{h}) = (1 - w(\mathbf{h}))\mathbf{p}_{\mathrm{b}}(\mathbf{h})$. Problem (4.25) can be transformed as follows:

$$\max_{\mathbf{q}_a(\mathbf{h}),\mathbf{q}_b(\mathbf{h}),w(\mathbf{h})} \mathrm{E}\left\{ w(\mathbf{h})\, r\left(\mathbf{h}, \frac{\mathbf{q}_a(\mathbf{h})}{w(\mathbf{h})}\right) + (1 - w(\mathbf{h}))\, r\left(\mathbf{h}, \frac{\mathbf{q}_b(\mathbf{h})}{1 - w(\mathbf{h})}\right)\right\} \tag{4.76}$$

$$\text{subject to} \quad E\left[\langle \mathbf{q}_a(\mathbf{h})\rangle + \langle \mathbf{q}_b(\mathbf{h})\rangle\right] = p_{\mathrm{av}} \tag{4.76a}$$

$$E\left[w(\mathbf{h})\right] \geq 1 - \epsilon \tag{4.76b}$$

$$w(\mathbf{h})\, r\left(\mathbf{h}, \frac{\mathbf{q}_a(\mathbf{h})}{w(\mathbf{h})}\right) - w(\mathbf{h}) r_{\mathrm{o}} \geq 0 \tag{4.76c}$$

$$\mathbf{q}_a(\mathbf{h}) \geq 0 \quad \mathbf{q}_b(\mathbf{h}) \geq 0 \quad 0 \leq w(\mathbf{h}) \leq 1 \tag{4.76d}$$

Denoting $f_i(z) = \log(1 + h_i z)$, it follows from Proposition 5 that

$$w(\mathbf{h}) r(\mathbf{h}, w^{-1}(\mathbf{h})\mathbf{q}_a(\mathbf{h})) = \frac{1}{M}\sum_{i=1}^{M} w(\mathbf{h}) f_i\left(\frac{q_{\mathrm{a},i}(\mathbf{h})}{w(\mathbf{h})}\right) \tag{4.77}$$

is the sum of concave functions. Thus, $w(\mathbf{h})r(\mathbf{h}, w^{-1}(\mathbf{h})\mathbf{q}_a(\mathbf{h}))$ is a concave function over $(w(\mathbf{h}), \mathbf{q}_a(\mathbf{h}))$. Similarly $(1 - w(\mathbf{h}))r(\mathbf{h}, (1 - w(\mathbf{h}))^{-1}\mathbf{q}_b(\mathbf{h}))$ is a concave function over $(w(\mathbf{h}), \mathbf{q}_b(\mathbf{h}))$. Thus, the objective function is concave over $(w(\mathbf{h}), \mathbf{q}_a(\mathbf{h}), \mathbf{q}_b(\mathbf{h}))$. It can be seen that the equality constraint (4.76a) is a linear function over $(\mathbf{q}_a(\mathbf{h}), \mathbf{q}_b(\mathbf{h}))$, the constraint (4.76b) is a linear function over $w(\mathbf{h})$. Since the left side of constraint (4.76c) is a concave function over $(w(\mathbf{h}), \mathbf{q}_a(\mathbf{h}))$, constraint (4.76c) is a convex set. Thus, the constraints specify a convex feasible set. Therefore, according to the Kuhn-Tucker sufficient conditions theorem [43], the Kuhn-Tucker conditions are sufficient conditions for the transformed problem (4.76). Let $(\hat{\mathbf{p}_{\mathrm{a}}}(\mathbf{h}), \hat{\mathbf{p}_{\mathrm{b}}}(\mathbf{h}), \hat{w}(\mathbf{h}))$ be a solution of the Kuhn-Tucker conditions of the original problem (4.25). According to Proposition 6, it is easy to see that the corresponding transformed variables $(\hat{\mathbf{q}_a}(\mathbf{h}), \hat{\mathbf{q}_b}(\mathbf{h}), \hat{w}(\mathbf{h}))$ satisfies the Kuhn-Tucker conditions of the transformed problem (4.76). Therefore, $(\hat{\mathbf{p}_{\mathrm{a}}}(\mathbf{h}), \hat{\mathbf{p}_{\mathrm{b}}}(\mathbf{h}), \hat{w}(\mathbf{h}))$ is the optimum solution of the original problem (4.25), and, thus, the Kuhn-Tucker conditions of problem (4.25) are also sufficient.

4.A.3 Proof of Lemma 6

Condition (4.28) yields

$$p^*_{a,i}(\mathbf{h}) = \left(\frac{1+u^*(\mathbf{h})}{h^*_0} - \frac{1}{h_{\rm i}}\right)^+ \quad \text{for all} \quad i = 1, \ldots, M\,. \tag{4.78}$$

Condition (4.31) implies that:

1. When $u^*(\mathbf{h}) = 0$ we have $r(\mathbf{h}, \mathbf{p}^*_{\rm a}(\mathbf{h})) \geq r_{\rm o}$. Moreover, when $u^*(\mathbf{h}) = 0$, (4.78) implies $\mathbf{p}^*_{\rm a}(\mathbf{h}) = \mathbf{p}_{\rm wf}(\mathbf{h}, h^*_0)$. Thus, in this case, we have $r(\mathbf{h}, \mathbf{p}_{\rm wf}(\mathbf{h}, h^*_0)) \geq r_{\rm o}$.

2. When $u^*(\mathbf{h}) > 0$ we have $r(\mathbf{h}, \mathbf{p}^*_{\rm a}(\mathbf{h})) = r_{\rm o}$. Expression (4.78) and $r(\mathbf{h}, \mathbf{p}^*_{\rm a}(\mathbf{h})) = r_{\rm o}$ imply that $\mathbf{p}^*_{\rm a}(\mathbf{h}) = \mathbf{p}_{\rm r_0}(\mathbf{h})$ with $\frac{1+u^*(\mathbf{h})}{h^*_0} = \lambda(\mathbf{h})$. Since $u^*(\mathbf{h}) > 0$, we have $\lambda(\mathbf{h}) > 1/h^*_0$. According to Proposition 4(a), we have $\lambda(\mathbf{h}) > 1/h^*_0$ iff $r(\mathbf{h}, \mathbf{p}_{\rm wf}(\mathbf{h}, h^*_0)) < r_{\rm o}$. Therefore, when $r(\mathbf{h}, \mathbf{p}_{\rm wf}(\mathbf{h}, h^*_0)) < r_{\rm o}$, we have $\mathbf{p}^*_{\rm a}(\mathbf{h}) = \mathbf{p}_{\rm r_0}(\mathbf{h})$ with a rate equal to $r_{\rm o}$.

Therefore,

$$\mathbf{p}^*_{\rm a}(\mathbf{h}) = \begin{cases} \mathbf{p}_{\rm wf}(\mathbf{h}, h^*_0) & r(\mathbf{h}, \mathbf{p}_{\rm wf}(\mathbf{h}, h_0)) \geq r_{\rm o} \\ \mathbf{p}_{\rm r_0}(\mathbf{h}) & \text{otherwise} \end{cases}\,. \tag{4.79}$$

Lastly, condition (4.29) yields $\mathbf{p}^*_{\rm b}(\mathbf{h}) = \mathbf{p}_{\rm wf}(\mathbf{h}, h^*_0)$ directly.

4.A.4 Proof of Lemma 7

We need the following proposition to prove Lemma 7.

Proposition 7 *For $x \geq 0$, $t(x) = x - \log(1+x)$ is an increasing nonnegative function of x.*

Proof: When $x > 0$, the first derivative $t'(x) = 1 - 1/(1+x) > 0$. Thus, $t(x)$ is increasing in x when $x \geq 0$. Since $t(0) = 0$, $t(x) \geq 0$ for all $x \geq 0$. □

(a) To prove Lemma 7(a), we only need to show that function $g(\mathbf{h}, h_0)$ is a nonincreasing function of h_i for all $i = 1, 2, \ldots, M$.

When $\mathbf{p}_{\rm s}(\mathbf{h}, h_0) = 0$, we have $r(\mathbf{h}, \mathbf{p}_{\rm wf}(\mathbf{h}, h_0)) \geq r_{\rm o}$ and $r_{\rm s}(\mathbf{h}, h_0) = 0$. Thus, we have $g(\mathbf{h}, h_0) = 0$ when $\mathbf{p}_{\rm s}(\mathbf{h}, h_0) = 0$.

When $\mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) > 0$, we have $\mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) = \mathbf{p}_{\mathrm{r}_0}(\mathbf{h}) - \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)$ and $r_{\mathrm{s}}(\mathbf{h}, h_0) = r_{\mathrm{o}} - r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0))$. Thus, in this case, we have

$$g(\mathbf{h}, h_0) = h_0 \langle \mathbf{p}_{\mathrm{r}_0}(\mathbf{h}) \rangle - r_{\mathrm{o}} - q(\mathbf{h}, h_0), \tag{4.80}$$

where

$$q(\mathbf{h}, h_0) = h_0 \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0) - r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)). \tag{4.81}$$

We have

$$\frac{\partial q(\mathbf{h}, h_0)}{\partial h_i} = \begin{cases} \frac{1}{Mh_i}\left(\frac{h_0}{h_i} - 1\right) < 0 & h_i > h_0 \\ 0 & h_i \leq h_0 \end{cases} \tag{4.82}$$

Partial derivative of $\langle \mathbf{p}_{\mathrm{r}_0}(\mathbf{h}) \rangle$ can be computed according to the approach used in [18]. Without loss of generality, it is assumed that $h_1 \geq \ldots \geq h_M$. Reference [18] shows that

$$\frac{\partial \langle \mathbf{p}_{\mathrm{r}_0}(\mathbf{h}) \rangle}{\partial h_i} = \begin{cases} \frac{1}{Mh_i}\left(\frac{1}{h_i} - \lambda(\mathbf{h})\right) & i = 1, \ldots, \mu \\ 0 & i = u+1, \ldots, M \end{cases}, \tag{4.83}$$

where μ is an integer employed in $\lambda(\mathbf{h})$. Parameter μ has a property such that $\lambda(\mathbf{h}) \geq h_i^{-1}$ for $i \leq \mu$ and $\lambda(\mathbf{h}) < h_i^{-1}$ for $i > \mu$ [18]. Thus, we have $(\partial/\partial h_i)\langle \mathbf{p}_{\mathrm{r}_0}(\mathbf{h}) \rangle \leq 0$.

From (4.82) and (4.83), it follows that

- when $h_i \leq h_0$, we have $(\partial/\partial h_i) g(\mathbf{h}, h_0) = h_0 (\partial/\partial h_i) \langle \mathbf{p}_{\mathrm{r}_0}(\mathbf{h}) \rangle \leq 0$.
- when $h_i > h_0$, we have

$$(\partial/\partial h_i) g(\mathbf{h}, h_0) = h_0 (\partial/\partial h_i) \langle \mathbf{p}_{\mathrm{r}_0}(\mathbf{h}) \rangle - (\partial/\partial h_i) q(\mathbf{h}, h_0).$$

  In the case of $\mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) > 0$, by Proposition 4 we have $h_0^{-1} < \lambda(\mathbf{h})$, and thus $h_i^{-1} < h_0^{-1} < \lambda(\mathbf{h})$. By the definition of $\lambda(\mathbf{h})$, we have $\lambda(\mathbf{h}) \geq h_i^{-1}$ iff $i \leq \mu$. Thus, in this case we must have $i < \mu$. Therefore, we have

$$(\partial/\partial h_i) g(\mathbf{h}, h_0) = \frac{1}{Mh_i}(1 - h_0 \lambda(\mathbf{h})) \leq 0.$$

Thus, $g(\mathbf{h}, h_0)$ is a non increasing function of h_i for all $i = 1, 2, \ldots, M$.

(b) As shown in (a), we have $g(\mathbf{h}, h_0) = 0$ when $\mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) = 0$.

When $\mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) > 0$, we want to show that $g(\mathbf{h}, h_0)$ is strictly positive.

In this case, we have

$$\begin{aligned} r_{\mathrm{s}}(\mathbf{h}, h_0) &= r_{\mathrm{o}} - r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)) \\ &= r(\mathbf{h}, \mathbf{p}_{\mathrm{r_0}}(\mathbf{h}, h_0)) - r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)) \\ &= 1/M \sum_{i=1}^{M} \log\left(1 + \frac{h_i p_{\mathrm{s,i}}(\mathbf{h}, h_0)}{1 + h_i p_{\mathrm{wf},i}(h_i, h_0)}\right) \end{aligned} \tag{4.84}$$

Let

$$g_i(\mathbf{h}, h_0) = h_0 p_{\mathrm{s,i}}(\mathbf{h}, h_0) - \log\left(1 + \frac{h_i p_{\mathrm{s,i}}(\mathbf{h}, h_0)}{1 + h_i p_{\mathrm{wf},i}(h_i, h_0)}\right),$$

then we have $g(\mathbf{h}, h_0) = 1/M \sum_{i=1}^{M} g_i(\mathbf{h}, h_0)$.

1. When $p_{\mathrm{wf},i}(h_i, h_0) = 0$, we have $h_i \leq h_0$ and $p_{\mathrm{s,i}}(\mathbf{h}, h_0) = p_{\mathrm{r_0},i}(\mathbf{h})$. Then

$$\begin{aligned} g_i(\mathbf{h}, h_0) &= h_0 p_{\mathrm{r_0},i}(\mathbf{h}) - \log(1 + h_i p_{\mathrm{r_0},i}(\mathbf{h})) \\ &= \frac{h_0}{h_i} h_i p_{\mathrm{r_0},i}(\mathbf{h}) - \log(1 + h_i p_{\mathrm{r_0},i}(\mathbf{h})) \geq t(h_i p_{\mathrm{r_0},i}(\mathbf{h}))\,. \end{aligned} \tag{4.85}$$

2. When $p_{\mathrm{wf},i}(h_i, h_0) > 0$, we have $1 + h_i p_{\mathrm{wf},i}(h_i, h_0) = h_i/h_0$, and thus

$$g_i(\mathbf{h}, h_0) = h_0 p_{\mathrm{s,i}}(\mathbf{h}, h_0) - \log(1 + h_0 p_{\mathrm{s,i}}(\mathbf{h}, h_0)) = t(h_0 p_{\mathrm{s,i}}(\mathbf{h}, h_0))\,.$$

Here, function $t(x)$ is an increasing nonnegative function of x when $x \geq 0$ by Proposition 7. Therefore, we have $g_i(\mathbf{h}, h_0) > 0$ for all i, and thus $g(\mathbf{h}, h_0) > 0$ when $\mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) > 0$.

4.A.5 Proof of Lemma 8

The average rate achieved by $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ is

$$\begin{aligned} &E\left[r(\mathbf{h}, \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h})))\right] \\ &= E\left[r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)) + w(\mathbf{h}) r_{\mathrm{s}}(\mathbf{h}, h_0)\right] \\ &= E\left[r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)) + r_{\mathrm{s}}(\mathbf{h}, h_0)\right] - E\left[(1 - w(\mathbf{h})) r_{\mathrm{s}}(\mathbf{h}, h_0)\right]\,. \end{aligned} \tag{4.86}$$

The first term in (4.86) is

$$\begin{aligned} & E\left[r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)) + r_{\mathrm{s}}(\mathbf{h}, h_0)\right] \\ & = E\left[r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)) + [r_{\mathrm{o}} - r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0))]^+\right] \\ & \overset{(a)}{=} E\left[r_{\mathrm{o}} + [r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)) - r_{\mathrm{o}}]^+\right] \\ & = R_{\mathrm{u}}(h_0)\,. \end{aligned} \tag{4.87}$$

Equality (a) follows from $a + (b-a)^+ = b + (a-b)^+$ for any a, b, and c.

The second term in (4.86) is bounded between 0 and $r_{\mathrm{o}}\epsilon$, since

$$0 \le r_{\mathrm{s}}(\mathbf{h}, h_0) = [r_{\mathrm{o}} - r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0))]^+ \le r_{\mathrm{o}} \tag{4.88}$$

$$0 \le E\left[1 - w(\mathbf{h})\right] \le \epsilon\,. \tag{4.89}$$

Thus,

$$R_{\mathrm{u}}(h_0) - r_{\mathrm{o}}\epsilon \le E\left[r(\mathbf{h}, \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h})))\right] \le R_{\mathrm{u}}(h_0)\,. \tag{4.90}$$

4.A.6 Proof of Lemma 10

We need the following proposition to prove Lemma 10.

Proposition 8 *The average power* $E\left[\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\right]$ *is decreasing in* h_0 *for a given* $w(\mathbf{h})$.

Proof: Proposition 8 The average power achieved by $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ can be expressed as follows:

$$\begin{aligned} & E\left[\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\right] \\ & = E\left[\langle \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)\rangle + w(\mathbf{h})\langle \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0)\rangle\right] \\ & = E\left[(1 - w(\mathbf{h}))\langle \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)\rangle + w(\mathbf{h})(\langle \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)\rangle + \langle \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0)\rangle)\right] \\ & = E\left[(1 - w(\mathbf{h}))\langle \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)\rangle + w(\mathbf{h})\left(\langle \mathbf{p}_{\mathrm{r}_0}(\mathbf{h})\rangle + \langle [\mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0) - \mathbf{p}_{\mathrm{r}_0}(\mathbf{h})]^+\rangle\right)\right]\,. \end{aligned} \tag{4.91}$$

Since $\langle \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)\rangle$ is decreasing in h_0, the above expression implies that $E\left[\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\right]$ is a decreasing function of h_0 for a given $w(\mathbf{h})$. □

Now we return to the proof of Lemma 10.

For any $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ with $E\left[\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\right] = P_{\text{av}}$ and $E\left[w(\mathbf{h})\right] \geq 1-\epsilon$, consider a scheme $\mathbf{P}(\mathbf{h}, \hat{h}_0, w(\mathbf{h}))$. By Lemma 9, we have $E\left[\hat{w}(\mathbf{h})\langle \mathbf{p}_{\text{s}}(\mathbf{h}, \hat{h}_0)\rangle\right] \leq E\left[w(\mathbf{h})\langle \mathbf{p}_{\text{s}}(\mathbf{h}, \hat{h}_0)\rangle\right]$ for any $w(\mathbf{h})$ that satisfies $E\left[w(\mathbf{h})\right] \geq 1-\epsilon$. Then we have

$$\begin{aligned} & E\left[\langle \mathbf{P}(\mathbf{h}, \hat{h}_0, w(\mathbf{h}))\rangle\right] \\ = \; & E\left[\langle \mathbf{p}_{\text{wf}}(\mathbf{h}, \hat{h}_0)\rangle + w(\mathbf{h})\langle \mathbf{p}_{\text{s}}(\mathbf{h}, \hat{h}_0)\rangle\right] \\ \geq \; & E\left[\langle \mathbf{p}_{\text{wf}}(\mathbf{h}, \hat{h}_0)\rangle + \hat{w}(\mathbf{h})\langle \mathbf{p}_{\text{s}}(\mathbf{h}, \hat{h}_0)\rangle\right] \\ = \; & E\left[\langle \mathbf{P}(\mathbf{h}, \hat{h}_0, \hat{w}(\mathbf{h}))\rangle\right] = p_{\text{av}} \\ = \; & E\left[\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\right] \end{aligned} \tag{4.92}$$

Since $E\left[\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\right]$ is a decreasing function of h_0 for a given $w(\mathbf{h})$ by Proposition 8, we have $\hat{h}_0 \leq h_0$. Hence, $\hat{\mathbf{P}}(\mathbf{h})$ has the minimum water-filling parameter among all $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ that satisfies $E\left[\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\right] = P_{\text{av}}$ and $E\left[w(\mathbf{h})\right] \geq 1-\epsilon$.

4.A.7 Proof of Lemma 11

In this section, we show that $E\left[r(\mathbf{h}, \mathbf{P}''(\mathbf{h}))\right] \geq R_{\text{u}}(h_0'') - r_{\text{o}}\epsilon$.

For any $x_1 \geq 0$ and $x_2 \geq 0$, we have

$$\log(1 + x_1 + x_2) \leq \log(1 + x_1) + \log(1 + x_2),$$

and thus

$$\log(1 + x_1) \geq \log(1 + x_1 + x_2) - \log(1 + x_2).$$

Therefore, for any two power allocations $\mathbf{p}_{\text{a}}(\mathbf{h})$ and $\mathbf{p}_{\text{b}}(\mathbf{h})$ we have

$$E\left[r(\mathbf{h}, \mathbf{p}_{\text{a}}(\mathbf{h}))\right] \geq E\left[r(\mathbf{h}, \mathbf{p}_{\text{a}}(\mathbf{h}) + \mathbf{p}_{\text{b}}(\mathbf{h}))\right] - E\left[r(\mathbf{h}, \mathbf{p}_{\text{b}}(\mathbf{h}))\right]. \tag{4.93}$$

Let $\mathbf{p}_{\text{u}}(\mathbf{h}, h_0) = \mathbf{p}_{\text{r}_0}(\mathbf{h}) + [\mathbf{p}_{\text{wf}}(\mathbf{h}, h_0) - \mathbf{p}_{\text{r}_0}(\mathbf{h})]^+$. Expanding $\mathbf{P}''(\mathbf{h})$ and applying $\mathbf{P}_{\min}(\mathbf{h}) = X_{w'}(\mathbf{h})\mathbf{p}_{\text{r}_0}(\mathbf{h})$, we have

$$\begin{aligned} \mathbf{P}''(\mathbf{h}) &= \mathbf{P}_{\min}(\mathbf{h}) + [\mathbf{p}_{\text{wf}}(\mathbf{h}, h_0'') - \mathbf{p}_{\text{r}_0}(\mathbf{h})]^+ \\ &= \mathbf{P}_{\min}(\mathbf{h}) - \mathbf{p}_{\text{r}_0}(\mathbf{h}) + \mathbf{p}_{\text{r}_0}(\mathbf{h}) + [\mathbf{p}_{\text{wf}}(\mathbf{h}, h_0'') - \mathbf{p}_{\text{r}_0}(\mathbf{h})]^+ \\ &= \mathbf{p}_{\text{u}}(\mathbf{h}, h_0'') - (\mathbf{p}_{\text{r}_0}(\mathbf{h}) - \mathbf{P}_{\min}(\mathbf{h})) \\ &= \mathbf{p}_{\text{u}}(\mathbf{h}, h_0'') - X_{1-w'}(\mathbf{h})\mathbf{p}_{\text{r}_0}(\mathbf{h}). \end{aligned} \tag{4.94}$$

Applying Proposition 4 (b) and (c), it is easy to show that

$$r(\mathbf{h}, \mathbf{p}_{\mathrm{u}}(\mathbf{h}, h_0'')) = r_{\mathrm{o}} + [r(\mathbf{h}, \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0'')) - r_{\mathrm{o}}]^+ . \tag{4.95}$$

Thus, $E\left[r(\mathbf{h}, \mathbf{p}_{\mathrm{u}}(\mathbf{h}, h_0''))\right] = R_{\mathrm{u}}(h_0'')$.

Since $E\left[w'(\mathbf{h})\right] = 1 - \epsilon$, we have

$$E\left[r(\mathbf{h}, X_{1-w'}(\mathbf{h})\mathbf{p}_{\mathrm{r}_0}(\mathbf{h}))\right] = E\left[1 - w'(\mathbf{h})\right] r_{\mathrm{o}} = r_{\mathrm{o}}\epsilon . \tag{4.96}$$

Thus, applying (4.93) we have

$$\begin{aligned} E\left[r(\mathbf{h}, \mathbf{P}''(\mathbf{h}))\right] &\geq E\left[r(\mathbf{h}, \mathbf{p}_{\mathrm{u}}(\mathbf{h}, h_0''))\right] - E\left[r(\mathbf{h}, X_{1-w'}(\mathbf{h})\mathbf{p}_{\mathrm{r}_0}(\mathbf{h}))\right] \\ &= R_{\mathrm{u}}(h_0'') - r_{\mathrm{o}}\epsilon . \end{aligned} \tag{4.97}$$

4.A.8 Proof of Lemma 12

To show $h_0'' \leq \hat{h}_0$, we only need to show that $h_0'' \leq h_0$ for any $\mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))$ that satisfies $E\left[\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\right] = p_{\mathrm{av}}$ and $E\left[w(\mathbf{h})\right] \geq 1 - \epsilon$.

We have

$$\begin{aligned} & E\left[\langle \mathbf{P}''(\mathbf{h})\rangle\right] \\ &= E\left[\langle \mathbf{P}_{\min}(\mathbf{h}) + [\mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0'') - \mathbf{p}_{\mathrm{r}_0}(\mathbf{h})]^+\rangle\right] = p_{\mathrm{av}} \\ &= E\left[\langle \mathbf{P}(\mathbf{h}, h_0, w(\mathbf{h}))\rangle\right] \\ &= E\left[\langle \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0) + w(\mathbf{h})\mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0)\rangle\right] \\ &\overset{(a)}{\geq} E\left[\langle \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0) + [w(\mathbf{h})\mathbf{p}_{\mathrm{r}_0}(\mathbf{h}) - \mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0)]^+\rangle\right] \\ &\overset{(b)}{=} E\left[\langle w(\mathbf{h})\mathbf{p}_{\mathrm{r}_0}(\mathbf{h}) + [\mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0) - w(\mathbf{h})\mathbf{p}_{\mathrm{r}_0}(\mathbf{h})]^+\rangle\right] \\ &\overset{(c)}{\geq} E\left[\langle w(\mathbf{h})\mathbf{p}_{\mathrm{r}_0}(\mathbf{h}) + [\mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0) - \mathbf{p}_{\mathrm{r}_0}(\mathbf{h})]^+\rangle\right] \\ &\overset{(d)}{\geq} E\left[\langle \mathbf{P}_{\min}(\mathbf{h}) + [\mathbf{p}_{\mathrm{wf}}(\mathbf{h}, h_0) - \mathbf{p}_{\mathrm{r}_0}(\mathbf{h})]^+\rangle\right] . \end{aligned} \tag{4.98}$$

Inequality (a) and (c) follows from $0 \leq w(\mathbf{h}) \leq 1$. Equality (b) follows from $a + (b-a)^+ = b + (a-b)^+$. Inequality (d) holds since $\mathbf{P}_{\min}(\mathbf{h})$ achieves the minimum power that needed to support r_{o} with probability $1 - \epsilon$, that is $E\left[\langle \mathbf{P}_{\min}(\mathbf{h})\rangle\right] \leq E\left[\langle w(\mathbf{h})\mathbf{p}_{\mathrm{r}_0}(\mathbf{h})\rangle\right]$ for any $w(\mathbf{h})$ that satisfies $E\left[w(\mathbf{h})\right] \geq 1 - \epsilon$.

Thus, it follows that

$$E\left[\langle[\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0'')-\mathbf{p}_{\mathrm{r}_0}(\mathbf{h})]^+\rangle\right]\geq E\left[\langle[\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0)-\mathbf{p}_{\mathrm{r}_0}(\mathbf{h})]^+\rangle\right]\,. \tag{4.99}$$

Since $\langle\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0)\rangle$ is a decreasing function of h_0, we have $h_0''\leq h_0$, and thus $h_0''\leq\hat{h}_0$.

4.A.9 Proof of Lemma 13

In this section, we show that $\mathbf{P}'(\mathbf{h})$ achieves a higher average rate than $\mathbf{P}''(\mathbf{h})$.

For any nonnegative $\mathbf{p}_{\mathrm{a}}(\mathbf{h})$ and $\mathbf{p}_{\mathrm{b}}(\mathbf{h})$, let Π denote a set of probabilistic schemes with average power p_{av} such that $\mathbf{P}(\mathbf{h})=\mathbf{p}_{\mathrm{r}_0}(\mathbf{h})+\mathbf{p}_{\mathrm{a}}(\mathbf{h})$ with probability $w'(\mathbf{h})$ and $\mathbf{P}(\mathbf{h})=\mathbf{p}_{\mathrm{b}}(\mathbf{h})$ with probability $1-w'(\mathbf{h})$. Here $w'(\mathbf{h})$ is given by (4.21). It is easy to show that $\mathbf{P}'(\mathbf{h})\in\Pi$ with $\mathbf{p}_{\mathrm{a}}(\mathbf{h})=[\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0')-\mathbf{p}_{\mathrm{r}_0}(\mathbf{h})]^+$ and $\mathbf{p}_{\mathrm{b}}(\mathbf{h})=\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0')$, and $\mathbf{P}''(\mathbf{h})\in\Pi$ with $\mathbf{p}_{\mathrm{a}}(\mathbf{h})=\mathbf{p}_{\mathrm{b}}(\mathbf{h})=[\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0'')-\mathbf{p}_{\mathrm{r}_0}(\mathbf{h})]^+$.

Consider the following optimization problem

$$\max_{\mathbf{p}_{\mathrm{a}}(\mathbf{h}),\mathbf{p}_{\mathrm{b}}(\mathbf{h})} E\left[w'(\mathbf{h})r(\mathbf{h},\mathbf{p}_{\mathrm{r}_0}(\mathbf{h})+\mathbf{p}_{\mathrm{a}}(\mathbf{h}))+(1-w'(\mathbf{h}))r(\mathbf{h},\mathbf{p}_{\mathrm{b}}(\mathbf{h}))\right] \tag{4.100}$$

$$\text{subject to}\quad E\left[w'(\mathbf{h})\langle\mathbf{p}_{\mathrm{r}_0}(\mathbf{h})+\mathbf{p}_{\mathrm{a}}(\mathbf{h})\rangle+(1-w'(\mathbf{h}))\langle\mathbf{p}_{\mathrm{b}}(\mathbf{h})\rangle\right]\leq p_{\mathrm{av}} \tag{4.100a}$$

$$\mathbf{p}_{\mathrm{a}}(\mathbf{h})\geq 0,\quad \mathbf{p}_{\mathrm{b}}(\mathbf{h})\geq 0\,.$$

Applying the generalized Kuhn-Tucker conditions [56], the optimum solution of (4.100) is

$$\mathbf{p}_{\mathrm{a}}^*(\mathbf{h})=[\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0')-\mathbf{p}_{\mathrm{r}_0}(\mathbf{h})]^+,\quad \mathbf{p}_{\mathrm{b}}^*(\mathbf{h})=\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0')\,,$$

Thus, $\mathbf{P}'(\mathbf{h})$ is the optimum power allocation that maximizes the average rate in set Π. Therefore, $\mathbf{P}'(\mathbf{h})$ achieves a higher average rate than $\mathbf{P}''(\mathbf{h})$.

4.B Closed Form Solution for Sub-problem

As shown in section 4.3.5, to determine $(h_0^*,s^*,v^*(\mathbf{h}))$ for $\mathbf{P}^*(\mathbf{h})$, it requires solving a linear programming problem on $v(\mathbf{h})$ for given h_0 and s.

For a given h_0 and the corresponding $s(h_0)$ in (4.54), let

$$p=p_{\mathrm{av}}-E\left[\langle\mathbf{p}_{\mathrm{wf}}(\mathbf{h},h_0)\rangle-\langle\mathbf{p}_{\mathrm{s}}(\mathbf{h},h_0)\rangle[1(g(\mathbf{h},h_0)<s(h_0))]\right], \tag{4.101}$$

$$\delta=1-\epsilon-\Pr\{g(\mathbf{h},h_0)<s(h_0)\}\,. \tag{4.102}$$

Define event B as $g(\mathbf{h}, h_0) = s(h_0)$. Equations (4.48)-(4.50) become

$$E\left[v(\mathbf{h})\langle \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0)\rangle | B\right] = p, \tag{4.103}$$

$$E\left[v(\mathbf{h})|B\right] = \delta \tag{4.104}$$

$$0 \le v(\mathbf{h}) \le 1\,. \tag{4.105}$$

To determine whether (4.103)-(4.105) is feasible, we first solve the following two optimization problems.

$$p_{\min} = \min_{0\le v(\mathbf{h})\le 1} E\left[v(\mathbf{h})\langle \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0)\rangle | B\right] \tag{4.106}$$

$$\text{subject to} \quad E\left[v(\mathbf{h})|B\right] = \delta, \tag{4.106a}$$

and

$$p_{\max} = \max_{0\le v(\mathbf{h})\le 1} E\left[v(\mathbf{h})\langle \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0)\rangle | B\right] \tag{4.107}$$

$$\text{subject to} \quad E\left[v(\mathbf{h})|B\right] = \delta, \tag{4.107a}$$

Let $v_1(\mathbf{h})$ and $v_2(\mathbf{h})$ denote the corresponding optimum solution for $p_{\min}$ and $p_{\max}$ respectively.

We have the following lemma for the sub-problem.

Lemma 14 *For a given h_0, problem (4.103)-(4.105) is feasible iff $p_{\min} \le p \le p_{\max}$. When it is feasible, one solution is $v(\mathbf{h}) = \lambda v_1(\mathbf{h}) + (1-\lambda)v_2(\mathbf{h})$, where λ is the solution to $p = \lambda p_{\min} + (1-\lambda)p_{\max}$.*

The $v_1(\mathbf{h})$ can be obtained by solving problem (4.106) using the Kuhn-Tucker conditions. We have

$$v_1(\mathbf{h}) = \begin{cases} 1 & \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) < t_1 \\ v_1 & \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) = t_1 \\ 0 & \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) > t_1 \end{cases}, \tag{4.108}$$

where parameters t_1 and $0 \le v_1 \le 1$ are solutions of (4.106a) as

$$t_1 = \sup\left\{x : \Pr\{B, \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) < x\} < \delta\right\} \tag{4.109}$$

$$v_1 = \frac{\delta - \Pr\{B, \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) < t_1\}}{\Pr\{B, \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) = t_1\}}\,. \tag{4.110}$$

Similarly, solving problem (4.107), we have

$$v_2(\mathbf{h}) = \begin{cases} 1 & \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) > t_2 \\ v_2 & \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) = t_2 \\ 0 & \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) < t_2 \end{cases}, \tag{4.111}$$

where t_2 and v_2 are

$$t_2 = \inf\{x : \Pr\{B, \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) > x\} < \delta(h_0)\} \tag{4.112}$$

$$v_2 = \frac{\delta - \Pr\{B, \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) > t_2\}}{\Pr\{B, \mathbf{p}_{\mathrm{s}}(\mathbf{h}, h_0) = t_2\}}. \tag{4.113}$$

In summary, the algorithm to find out the optimum parameters is as follows:

1. choose h_0.
2. compute s according to (4.54).
3. compute p and δ according to (4.101) and (4.102).
4. compute $v_1(\mathbf{h}, h_0)$ and $v_2(\mathbf{h}, h_0)$ according to (4.108) and (4.111). Compute the corresponding $p_{\min}$ and $p_{\max}$.
5. If $p_{\min} \leq p \leq p_{\max}$, we have $v^*(\mathbf{h}) = \lambda v_1(\mathbf{h}) + (1-\lambda)v_2(\mathbf{h})$, where λ is the solution to $p = \lambda p_{\min} + (1-\lambda)p_{\max}$. The corresponding h_0 and s are the optimum solutions.
6. Otherwise, adjust h_0 go to step 2.

4.C Large Deviation Approximation

In this section, we introduce techniques to compute the tail probability

$$\Pr\{\frac{1}{N}\sum_{n=1}^{N} z_n < A\},$$

where z_n are independent, identically distributed random variables with pdf $f_Z(z)$ and $A < \mathrm{E}\{z\}$. When the mean and variance of Z are known, the tail probability may be computed by using a Gaussian approximation. The Gaussian approximation, though can be evaluated with extreme ease, is not particular accurate for tail probablity. The

calculation of the tail probability falls into the area of large deviation theory [32, 34], which is concerned with the sum of a large number of random variables.

First, we examine the Chernoff bound. Define $u(s)$ as

$$u(s) = -\log(\mathrm{E}\left\{e^{-sz}\right\}). \tag{4.114}$$

Applying the Chernoff bound, we have

$$\Pr\{\frac{1}{N}\sum_{n=1}^{N} z_n < A\} \leq \min_{s>0} \mathrm{E}\left\{e^{s(NA-\sum_{n=1}^{N} z_n)}\right\} = \min_{s>0} e^{-N(u(s)-sA)}. \tag{4.115}$$

The optimum s^* is the solution to $u'(s) = A$. Since $u'(s)$ is decreasing and $A < \mathrm{E}\{z\} = u'(0)$, the parameter s^* is positive.

The above bound can be sharpened by the theory of large deviations. See [34] for the detail. Here we only present the results. The first two derivatives of $u(s)$ are given by

$$u'(s) = \frac{\mathrm{E}\left\{ze^{-sz}\right\}}{\mathrm{E}\left\{e^{-sz}\right\}} \tag{4.116}$$

$$u''(s) = -\frac{\mathrm{E}\left\{z^2e^{-sz}\right\}}{\mathrm{E}\left\{e^{-sz}\right\}} + [u'(s)]^2. \tag{4.117}$$

Notice that $u'(0)$ and $-u''(0)$ are, respectively, the mean and variance of z. From [34], the tail probability is given by

$$\Pr\{\frac{1}{N}\sum_{n=1}^{N} z_n < A\} = e^{-N(u(s^*)-s^*A)}\left(\frac{1}{s^*\sqrt{2\pi N(-u''(s^*))}} + o\left(\frac{1}{\sqrt{N}}\right)\right). \tag{4.118}$$

Chapter 5

Adaptive Transmission for Mixed Services over Fading Channels

In this chapter, we explore adaptive transmission for systems with mixed real-time and non real-time services over fading channels.

Wireless systems are expected to provide a variety of services including voice, file transfer, email, Internet access, video, and audio. These services can be divided into two basic classes: real-time service and non real-time service. Without loss of generality, we examine the systems with mixed voice and data services where the voice represents the real-time service and the data represents non real-time service. A lot of work has been done in the area of integration of voice and data in multiuser scenarios to serve both voice and data users [39, 85, 86, 88]. Typically, a static environment where the channel conditions are fixed is assumed and the focus is on the media access control (MAC) protocols. In this work, we concentrate on transmission of mixed services in a single wireless link, where the channel condition changes with time. This is motivated by the increasing popularity of multimedia applications on the world wide web that may contain both non real-time information such as text and real-time information such as voice or video.

The adaptive transmission of mixed voice and data services in a fading channel was previously discussed in [1]. In [1] the inphase (I) channel was dedicated to voice and the quadrature (Q) channel was dedicated to data. It is assumed in [1] that the power for both services is constant throughout the fading. Since the voice service requires a constant rate, the power allocation for voice is the channel inversion. The remaining power is allocated to the data service, and the code rate for data varies according to the channel state and the availabe power. Reference [1] didn't address the problem of

finding optimum transmission scheme in a system with mixed services.

In this chapter, we first determine the optimum transmission scheme for a system with mixed services from an information theoretic point of view. We show that the optimum transmission scheme is the service outage based allocation described in Chapter 4. The service outage based allocation and its implementations with joint transmission are discussed in detail in Section 5.1. In Section 5.2, it is shown that the service outage based allocation can also be implemented by an adaptive channel partition scheme. The optimum fixed channel partition scheme is derived and compared to the adaptive scheme in Section 5.3. In addition, a suboptimum fixed partition scheme called equal power density partition is studied and observed to be close to the optimum fixed scheme in $M = 1$ Rayleigh fading channel in Section 5.4

5.1 Service Outage Based Capacity and Joint Transmission Strategies

Consider a system with mixed voice and data services in BF-AWGN channels. The voice service is a real time service and let $r_{\rm o}$ denote its information source rate. When the transmission rate is less than $r_{\rm o}$, the information bits are useless and are thrown away, and this event is termed an outage. In order to guarantee good voice quality, the service outage probability must be less than a specified value ϵ. Therefore, the transmission rate must be at least the voice target rate $r_{\rm o}$ with probability $1-\epsilon$, nevertheless, it is not restricted to be just equal to $r_{\rm o}$. We allow for a higher transmission rate than $r_{\rm o}$, so that any excess rate can be used to transmit data. As we can see, the maximum average rate achieved by this system is just the service outage based achievable rate discussed in previous chapters.

In the system with mixed voice and data services, the service outage based achievable rate can be achieved by joint encoding and decoding of both services. Particular, near optimum allocation II discussed in Chapter 4 has a structure appealing to transmission of mixed services. Recall that near optimum allocation II can be expressed as

$$\mathbf{P}'(\mathbf{h}) = \mathbf{P}_{\min}(\mathbf{h}) + [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0') - \mathbf{P}_{\min}(\mathbf{h})]^+ . \tag{5.1}$$

Here $\mathbf{P}_{\min}(\mathbf{h}) = X_{w'}(\mathbf{h})\mathbf{p}_{\rm r_0}(\mathbf{h})$ achieves the minimum sufficient power to maintain the

basic rate r_0 with outage probability ϵ. Let $\mathbf{p}_{\rm ex}(\mathbf{h}) = [\mathbf{p}_{\rm wf}(\mathbf{h}, h_0') - \mathbf{P}_{\rm min}(\mathbf{h})]^+$ denote the second term in (5.1), called *excessive power allocation*. It maximizes the excessive rate in the presence of $\mathbf{P}_{\rm min}(\mathbf{h})$.

Near optimum allocation II can be implemented in the mixed services system using superposition coding and successive decoding with different orders as shown below. To describe successive decoding schemes, for two arbitrary power allocations $\mathbf{p}_{\rm a}(\mathbf{h})$ and $\mathbf{p}_{\rm b}(\mathbf{h})$, we define

$$r(\mathbf{h}, \mathbf{p}_{\rm b}(\mathbf{h})|\mathbf{p}_{\rm a}(\mathbf{h})) = \frac{1}{M}\sum_{i=1} M \log\left(1 + \frac{h_i p_{b,i}(\mathbf{h})}{1 + h_i p_{a,i}(\mathbf{h})}\right). \tag{5.2}$$

Rate $r(\mathbf{h}, \mathbf{p}_{\rm b}(\mathbf{h})|\mathbf{p}_{\rm a}(\mathbf{h}))$ is the rate achieved by $\mathbf{p}_{\rm b}(\mathbf{h})$ when $\mathbf{p}_{\rm a}(\mathbf{h})$ is an additive interference. Using chain rule, we have

$$\begin{aligned} r(\mathbf{h}, \mathbf{p}_{\rm a}(\mathbf{h}) + \mathbf{p}_{\rm b}(\mathbf{h})) &= r(\mathbf{h}, \mathbf{p}_{\rm a}(\mathbf{h})) + r(\mathbf{h}, \mathbf{p}_{\rm b}(\mathbf{h})|\mathbf{p}_{\rm a}(\mathbf{h})) \\ &= r(\mathbf{h}, \mathbf{p}_{\rm b}(\mathbf{h})) + r(\mathbf{h}, \mathbf{p}_{\rm a}(\mathbf{h})|\mathbf{p}_{\rm b}(\mathbf{h}))\,. \end{aligned} \tag{5.3}$$

Two joint transmission strategies can be employed in the system as follows:

- Strategy I: first decode data and then voice, that is

$$r(\mathbf{h}, \mathbf{P}'(\mathbf{h})) = r(\mathbf{h}, \mathbf{P}_{\rm min}(\mathbf{h})) + r(\mathbf{h}, \mathbf{p}_{\rm ex}(\mathbf{h})|\mathbf{P}_{\rm min}(\mathbf{h}))\,. \tag{5.4}$$

  In this approach, voice is allocated with the minimum power $\mathbf{P}_{\rm min}(\mathbf{h})$, and data is allocated with the excessive power $\mathbf{p}_{\rm ex}(\mathbf{h})$. At the receiver, data is decoded first and is subtracted from the received signal, then voice is decoded. In a BF-AWGN channel model, if the data service uses variable-rate multiple-codebooks, the decoding delay associated with the voice service is the time to decode one short codeword for data, which may be acceptable for many cases. If instead, the data service uses a constant-rate single-codebook averaged out all fading states, the decoding delay associated with the voice service is likely to be unacceptable. Fortunately, in many wireless systems, variable-rate short codes are usually used for non realtime data service.

- Stategy II: first decode voice and then data. In the outage set, only data is transmitted. In the service set, we have $\mathbf{P}_{\rm min}(\mathbf{h}) = \mathbf{p}_{\rm r_0}(\mathbf{h})$. Using the idea in [45],

in the service set we have

$$\begin{aligned} r(\mathbf{h}, \mathbf{P}'(\mathbf{h})) &= r(\mathbf{h}, \mathbf{p}_{\rm ex}(\mathbf{h})e^{-r_0} + \mathbf{p}_{\rm r_0}(\mathbf{h}) + \mathbf{p}_{\rm ex}(\mathbf{h})(1-e^{-r_0})) \\ &= r(\mathbf{h}, \mathbf{p}_{\rm ex}(\mathbf{h})e^{-r_0}) + r(\mathbf{h}, \mathbf{p}_{\rm r_0}(\mathbf{h}) + \mathbf{p}_{\rm ex}(\mathbf{h})(1-e^{-r_0})|\mathbf{p}_{\rm ex}(\mathbf{h})e^{-r_0})\,. \end{aligned}$$

Using the fact that $r(\mathbf{h}, \mathbf{p}_{\rm r_0}(\mathbf{h})) = r_0$, it is easy to show that

$$r(\mathbf{h}, \mathbf{p}_{\rm r_0}(\mathbf{h}) + \mathbf{p}_{\rm ex}(\mathbf{h})(1-e^{-r_0})|\mathbf{p}_{\rm ex}(\mathbf{h})e^{-r_0}) = r_0\,. \tag{5.5}$$

Therefore, in this approach, in the service set, data service is allocated with power $\mathbf{p}_{\rm ex}(\mathbf{h})e^{-r_0}$, and voice service is allocated with power $\mathbf{p}_{\rm r_0}(\mathbf{h}) + \mathbf{p}_{\rm ex}(\mathbf{h})(1-e^{-r_0})$. At the receiver, voice is decoded first and is subtracted from the received signal, then the data is decoded. This approach allows the voice service to be decoded with a delay independent of the data service and the rate of channel variation.

5.2 Adaptive Channel Partition

In this section, we show that, without using joint coding, the service outage based capacity can in fact be achieved by an adaptive channel partition scheme, where services are transmitted in separate channels. By using the adaptive channel partition, the system complexity is significantly reduced compared to the joint transmission schemes, and also it brings more flexibility in the design of systems with mixed services.

In an channel partition scheme, the channel is partitioned into voice and data subchannels. Similar to frequency division (FD), time division (TD), and orthogonal code division (CD) in multiuser communications [8, 81], the partition between different services can be done in either frequency, time, or code space. For an adaptive channel partition scheme, the fraction of channel (time, bandwidth, or code) allocated to each service is adapted according to the channel state. Orthogonal partition schemes in different spaces are equivalent to each other, in the sense that they can achieve the same rate [8, 81]. Without loss of generality, we use partitioning in time to illustrate the basic ideas and all the results can be applied to the partition schemes in other spaces. For partition in time, the time is divided into a sequence of frames with the same time duration. Within each frame the time is divided into the voice slot and the data slot.

The channel state is assumed to be constant within each frame, but may vary from frame to frame.

Let $\alpha(\mathbf{h})$ denote the fraction of time allocated to the voice slot at channel state $\mathbf{h}$. Then the remaining $1-\alpha(\mathbf{h})$ of time is allocated to the data slot at channel state $\mathbf{h}$. If the power allocations in both the voice slot and the data slot are the service outage based allocation, denoted by $\mathbf{p}(\mathbf{h}, h_0)$, then the voice rate achieved in the frame is $\alpha(\mathbf{h})r(\mathbf{h}, \mathbf{p}(\mathbf{h}, h_0))$ and the data rate is $(1-\alpha(\mathbf{h}))r(\mathbf{h}, \mathbf{p}(\mathbf{h}, h_0))$. The total rate $r(\mathbf{h}, \mathbf{p}(\mathbf{h}, h_0))$ can be achieved in the frame with $\mathbf{h}$. Let $\mathcal{H}_{\mathrm{o}}$ denote the outage set for scheme $\mathbf{p}(\mathbf{h}, h_0)$. Choose the partition ratio as

$$\alpha(\mathbf{h}) = \begin{cases} \frac{r_{\mathrm{o}}}{r(\mathbf{h}, \mathbf{p}(\mathbf{h}, h_0))} & \mathbf{h} \notin \mathcal{H}_{\mathrm{o}} \\ 0 & \mathbf{h} \in \mathcal{H}_{\mathrm{o}} \end{cases} . \tag{5.6}$$

Then, the voice rate r_{o} can be provided with probability $1-\epsilon$. The proposed adaptive channel partition scheme employs the following three modes of operations to achieve the service outage based capacity:

1. In the outage set, the partition ratio $\alpha(\mathbf{h}) = 0$. In this case, the whole channel is dedicated to the data service with rate $r(\mathbf{h}, \mathbf{p}(\mathbf{h}, h_0))$.

2. In the basic-rate set with $r(\mathbf{h}, \mathbf{p}(\mathbf{h}, h_0)) = r_{\mathrm{o}}$, the partition ratio $\alpha(\mathbf{h}) = 1$. In this case, the whole channel is dedicated to the voice service.

3. In the enhanced-rate set with $r(\mathbf{h}, \mathbf{p}(\mathbf{h}, h_0)) > r_{\mathrm{o}}$, the partition ratio $\alpha(\mathbf{h}) = r_{\mathrm{o}}/r(\mathbf{h}, \mathbf{p}(\mathbf{h}, h_0)) < 1$. In this case, the channel is partitioned into voice subchannel and data subchannel. The partition ratio $\alpha(\mathbf{h})$ decreases with channel state vector $\mathbf{h}$. Therefore, at a better channel state vector a larger faction of the channel is allocated to data.

5.3 Optimum Fixed Channel Partition Scheme

In Section 5.1, it is shown that the service outage based capacity can be achieved by a adaptive channel partition scheme. In this section, we study fixed channel partition schemes, whose partition ratio is fixed throughout the whole communication session.

Let α denote the fraction of channel allocated to the voice service. Then the remaining $1-\alpha$ fraction of channel is allocated to the data service. Let $\mathbf{p}_{\rm v}(\mathbf{h})$ and $\mathbf{p}_{\rm d}(\mathbf{h})$ denote the power allocations in the voice and data subchannels respectively. In the following allocation problem, we would like to find out $(\mathbf{p}_{\rm v}(\mathbf{h}), \mathbf{p}_{\rm d}(\mathbf{h}), \alpha)$ that maximize the excessive rate for the data service, while at the same time maintain a basic rate r_0 for the voice service with an outage probability ϵ.

$$\begin{aligned} \max_{\mathbf{p}_{\rm v}(\mathbf{h}),\mathbf{p}_{\rm d}(\mathbf{h}),\alpha} & (1-\alpha)E\left[r(\mathbf{h},\mathbf{p}_{\rm d}(\mathbf{h}))\right] \\ \text{subject to} \quad & \alpha E\left[\langle \mathbf{p}_{\rm v}(\mathbf{h})\rangle\right] + (1-\alpha)E\left[\langle \mathbf{p}_{\rm d}(\mathbf{h})\rangle\right] \le p_{\rm av} \\ & \Pr\{\alpha r(\mathbf{h},\mathbf{p}_{\rm v}(\mathbf{h})) < r_0\} \le \epsilon \\ & 0 \le \alpha \le 1, \quad \mathbf{p}_{\rm v}(\mathbf{h}) \ge 0, \quad \mathbf{p}_{\rm d}(\mathbf{h}) \ge 0\,. \end{aligned} \tag{5.7}$$

Clearly, the excessive rate for the data service increase as the average power allocated for the data services increases. Therefore, allocation problem (5.7) can be solved in the following two steps:

1. Find $\mathbf{p}_{\rm v}(\mathbf{h})$ that minimizes the average power needed for voice for given α. The optimum solution is denoted by $\mathbf{p}_{\rm v}(\mathbf{h},\alpha)$.

2. Find $\mathbf{p}_{\rm d}(\mathbf{h})$ and α that maximize the average excessive rate for data, when the power allocation for voice is $\mathbf{p}_{\rm v}(\mathbf{h},\alpha)$. The optimum solutions are denoted as $\mathbf{p}^*_{\rm d}(\mathbf{h})$ and α^*.

Clearly, the optimum solution $\mathbf{p}_{\rm v}(\mathbf{h},\alpha)$ in step one is equal to $\mathbf{P}_{\min}(\mathbf{h})$ with an effective basic rate r_0/α in the voice subchannel. That is

$$\mathbf{p}_{\rm v}(\mathbf{h},\alpha) = \begin{cases} \mathbf{p}_{r_0/\alpha}(\mathbf{h}) & \langle \mathbf{p}_{r_0/\alpha}(\mathbf{h})\rangle \le s \\ 0 & \text{otherwise} \end{cases}, \tag{5.8}$$

where s to the solution to $\Pr\{\langle \mathbf{p}_{r_0/\alpha}(\mathbf{h})\rangle > s\} = \epsilon$. Define

$$P_{\rm v}(\alpha) = E\left[\mathbf{p}_{\rm v}(\mathbf{h},\alpha)\right] = \int_{\langle \mathbf{p}_{r_0/\alpha}(\mathbf{h})\rangle \le s} \langle \mathbf{p}_{r_0/\alpha}(\mathbf{h})\rangle dF(\mathbf{h}) \tag{5.9}$$

as the power density in the voise subchannel. Then the minimum average power for voice is $\alpha P_{\rm v}(\alpha)$ and the maximum average power for data is $P_{\rm av} - \alpha P_{\rm v}(\alpha)$.

Let $\alpha_{\min}$ to be the solution to $\alpha P_{\rm v}(\alpha) = P_{\rm av}$. Problem (5.7) is feasible iff $\alpha \geq \alpha_{\min}$. Allocation problem in step two can be written as

$$\begin{aligned} &\max_{\mathbf{p}_{\rm d}(\mathbf{h}),\alpha} (1-\alpha) E\left[r(\mathbf{h}, \mathbf{p}_{\rm d}(\mathbf{h}))\right] \\ &\text{subject to} \quad \alpha P_{\rm v}(\alpha) + (1-\alpha) E\left[\langle \mathbf{p}_{\rm d}(\mathbf{h})\rangle\right] \leq P_{\rm av} \\ &\qquad\qquad \mathbf{p}_{\rm d}(\mathbf{h}) \geq 0, \quad \alpha_{\min} \leq \alpha \leq 1\,. \end{aligned} \tag{5.10}$$

Using Lagrange multiplie h_0, define

$$l(\mathbf{p}_{\rm d}(\mathbf{h}), \alpha, h_0) = (1-\alpha) r(\mathbf{h}, \mathbf{p}_{\rm d}(\mathbf{h})) - h_0(\alpha P_{\rm v}(\alpha) + (1-\alpha)\mathbf{p}_{\rm d}(\mathbf{h}))\,. \tag{5.11}$$

The average rate for data service is equal to zero when $\alpha = \alpha_{\min}$ or $\alpha = 1$. Therefore, when $\alpha_{\min} < 1$, the optimum α^* must be a non-boundary point, that is, we must have $\alpha_{\min} < \alpha^* < 1$. According to the Kuhn-Tucker necessary conditions theorem, the optimum solution $(\mathbf{p}_{\rm d}^*(\mathbf{h}), \alpha^*)$ should satisfy the following conditions:

$$\left.\frac{\partial l(\mathbf{p}_{\rm d}(\mathbf{h}), \alpha, h_0)}{\partial \mathbf{p}_{\rm d}(\mathbf{h})}\right|_{\mathbf{p}_{\rm d}(\mathbf{h})=\mathbf{p}_{\rm d}^*(\mathbf{h})} \begin{cases} = 0 & \mathbf{p}_{\rm d}^*(\mathbf{h}) > 0 \\ \leq 0 & \mathbf{p}_{\rm d}^*(\mathbf{h}) = 0 \end{cases} \tag{5.12}$$

$$\left.\frac{\partial l(\mathbf{p}_{\rm d}(\mathbf{h}), \alpha, h_0)}{\partial \alpha}\right|_{\alpha=\alpha^*} = 0, \qquad \alpha_{\min} < \alpha^* < 1 \tag{5.13}$$

$$\alpha^* P_{\rm v}(\alpha^*) + (1-\alpha^*) E\left[\mathbf{p}_{\rm d}^*(\mathbf{h})\right] = P_{\rm av} \tag{5.14}$$

Condition (5.12) yields

$$\mathbf{p}_{\rm d}^*(\mathbf{h}) = \mathbf{p}_{\rm wf}(\mathbf{h}, h_0)\,. \tag{5.15}$$

The optimum power allocation in the data subchannel is the water-filling allocation. Define the power density in the data subchannel as $P_{\rm d}(h_0) = E\left[\mathbf{p}_{\rm d}^*(\mathbf{h})\right]$, and the rate density in the data subchannel as $R_{\rm d}(h_0) = E\left[r(\mathbf{h}, \mathbf{p}_{\rm d}^*(\mathbf{h}))\right]$. Then it follows that h_0 and α^* must satisfy

$$h_0[P_{\rm d}(h_0) - P_{\rm v}(\alpha^*) - \alpha^* P_{\rm v}'(\alpha^*)] - R_{\rm d}(h_0) = 0 \tag{5.16}$$

$$\alpha^* P_{\rm v}(\alpha^*) + (1-\alpha^*) P_{\rm d}(h_0) = P_{\rm av}\,. \tag{5.17}$$

$P_{\rm v}'(\alpha)$ is the derivative of $P_{\rm v}(\alpha)$ over α.

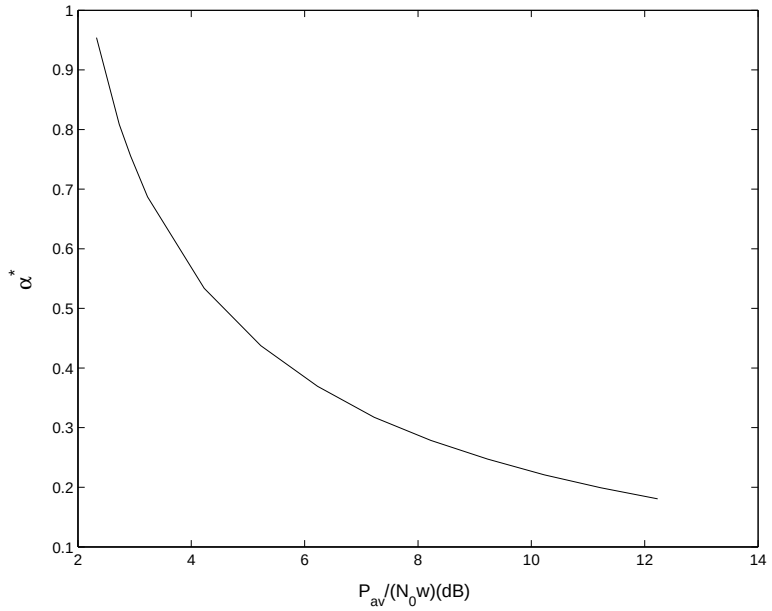


Figure 5.1: Optimum fixed partition ratio α^* versus $P_{\rm av}$ in $M = 1$ Rayleigh fading channel for $r_{\rm o} = 0.5$ bits/s/Hz and $\epsilon = 0.01$.

Calculation of $P_{\rm v}(\alpha)$ and $P'_{\rm v}(\alpha)$ is complicated in an $M \geq 2$ parallel fading channel model. Here we only present the results for $M = 1$ fading channel. When $M = 1$, we have

$$p_{\rm v}(h, \alpha) = \begin{cases} 0 & h < h_\epsilon \\ \frac{e^{r_0/\alpha}-1}{h} & h \geq h_\epsilon \end{cases} . \tag{5.18}$$

It follows that

$$P_{\rm v}(\alpha) = \int_{h_\epsilon}^{\infty} \frac{e^{r_0/\alpha} - 1}{h} dF(h) \, . \tag{5.19}$$

Function $\alpha P_{\rm v}(\alpha)$ is convex in α when $M = 1$. By replacing $\mathbf{p}_{\rm d}(\mathbf{h})$ with a new variable $\mathbf{q}_{\rm d}(\mathbf{h}) = (1 - \alpha)\mathbf{p}_{\rm d}(\mathbf{h})$, allocation problem (5.10) can be transformed into a convex optimization problem when $M = 1$. Thus, the Kuhn-Tucker conditions (5.12)-(5.14) are also sufficient for the optimum solution when $M = 1$, which implies that any solutions of (5.16) and (5.17) are the optimum solution for problem (5.10). The h_0 and α^* can be obtained numerically through two dimensional search.

Consider a Rayleigh fading channel with a probability density function $f(h) = e^{-h}$ for $h > 0$. In Figure 5.1, the optimum partition ratio α^* versus $P_{\rm av}$ is presented for $M = 1$ Rayleigh fading channel with fixed $r_{\rm o} = 0.5$ bits/s/Hz and $\epsilon = 0.01$. The optimum partition ratio decreases with average power. In Figure 5.2 the total average

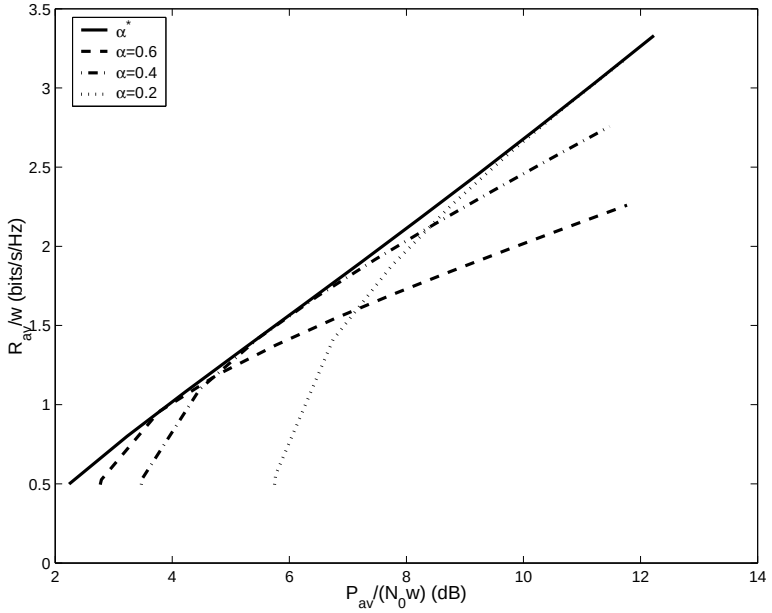


Figure 5.2: Comparison of the optimum fixed partition ratio α^* with arbitrary partition ratio in $M = 1$ Rayleigh fading channel for $r_{\rm o} = 0.5$ bits/s/Hz, and $\epsilon = 0.01$.

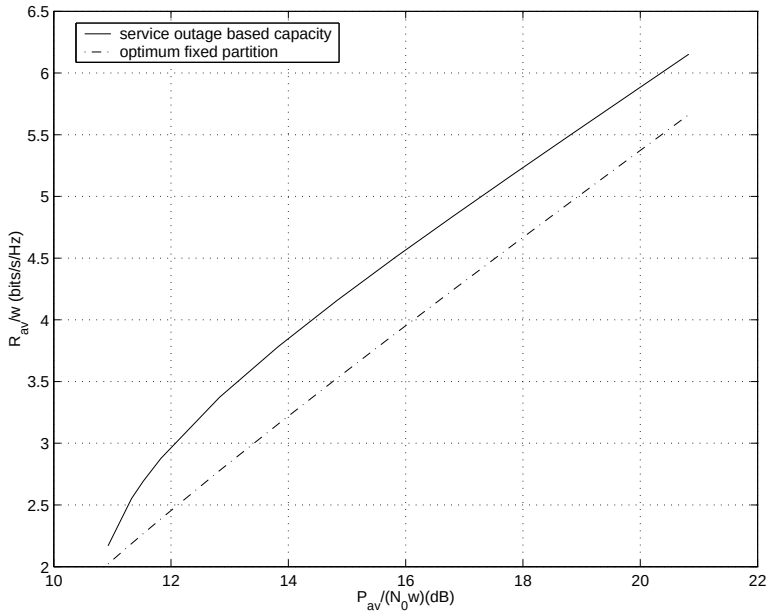


Figure 5.3: Comparison of service outage approach with optimum fixed partition scheme in $M = 1$ Rayleigh fading channel for $r_{\rm o} = 2$ bits/s/Hz, and $\epsilon = 0.01$.

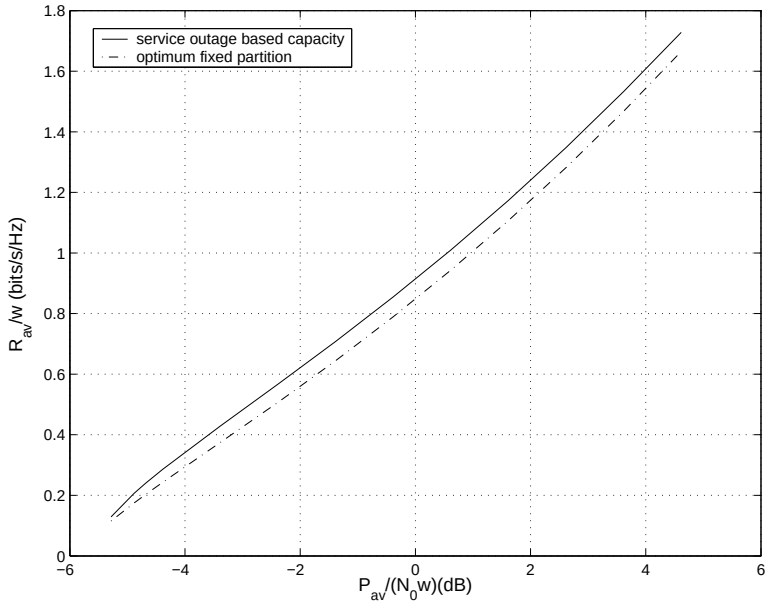


Figure 5.4: Comparison of service outage approach with optimum fixed partition scheme in $M = 1$ Rayleigh fading channel for $r_{\rm o} = 0.1$ bits/s/Hz, and $\epsilon = 0.01$.

rate achieved by the optimum partition ratio is plotted against the performance by arbitrary partition ratios. It can be seen that a significant performance degradation can be incurred if an arbitrary partition ratio is used.

We compare the service outage approach with the optimum fixed partition scheme in Figures 5.3 and 5.4 for $M = 1$ Rayleigh fading channel. Figure 5.3 demonstrates that in the case of high voice spectral efficiency $r_{\rm o} = 2$ bits/s/Hz, the fixed partition scheme incurs almost 2 dB loss in average power. In the case of low voice spectral efficiency $r_{\rm o} = 0.1$ bits/s/Hz as shown in Figure 5.4, the fixed partition scheme has less than 0.5 dB loss in the average power. From the numerical results, we observe that the optimum fixed partition scheme can be used without significant loss in the system performance when the voice spectral efficiency is $r_{\rm o} \ll 1$.

5.4 Equal Power Density Partition in $M = 1$ fading channel

Since the optimum fixed partition ratio does not has a closed form solution even in $M = 1$ fading channel, we propose an equal power density partition scheme, which has a simple close form solution in $M = 1$ fading channel. It is observed that the equal

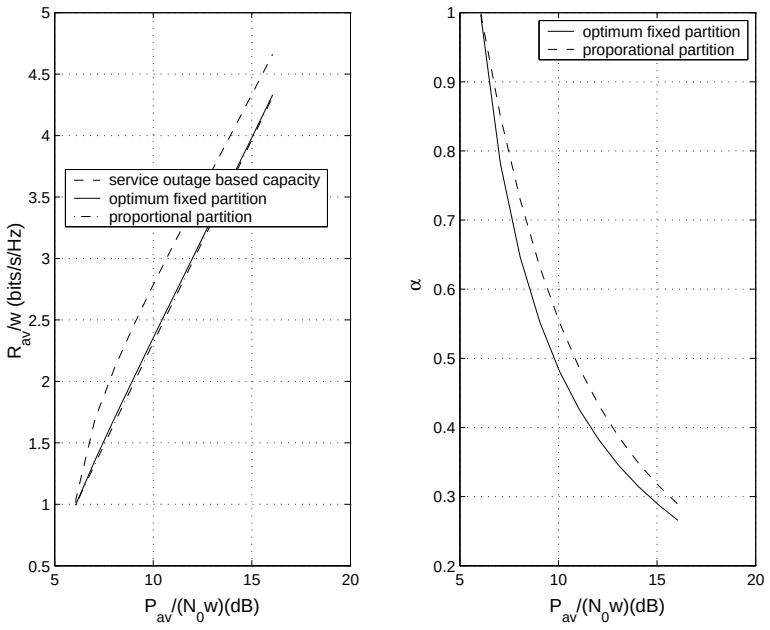


Figure 5.5: Comparison of equal power density partition scheme with the optimum fixed partition scheme in $M = 1$ Rayleigh fading channel with $r_{\rm o} = 1$ bits/s/Hz and $\epsilon = 0.01$. The average rate versus SNR is on the left, and the corresponding partition ratio versus SNR is on the right.

power density partition scheme achieves an average rate close to the optimum fixed partition scheme in $M = 1$ fading channel.

In the equal power density partition scheme, the partition ratio $\hat{\alpha}$ is chosen so that

$$P_{\rm v}(\hat{\alpha}) = P_{\rm d}(h_0, \hat{\alpha}) = P_{\rm av}.$$

In $M = 1$ fading channel, $\hat{\alpha}$ has a simple closed form solution as

$$\hat{\alpha} = \frac{r_{\rm o}}{\log\left(1 + \frac{P_{\rm av}}{\int_{h_\epsilon}^{\infty} 1/hf(h)dh}\right)}. \tag{5.20}$$

$\hat{\alpha}$ decreases with $P_{\rm av}$. It has been observed that the equal power density partition achieves an average rate close to the optimum fixed partition in $M = 1$ Rayleigh fading channel for a wide range of parameters. In Figure 5.5 and 5.6, the equal power density partition scheme is compared to the optimum fixed partition scheme for $r_{\rm o} = 1$ bits/s/Hz and $r_{\rm o} = 0.05$ bits/s/Hz respectively. It is observed that $R(\hat{\alpha}, P_{\rm av})$ is closer to $R(\alpha^*, P_{\rm av})$ when the voice spectral efficiency $r_{\rm o}$ is higher. In Figure 5.5 with $r_{\rm o} = 1$ bits/s/Hz, the ratio $R(\hat{\alpha}, P_{\rm av})/R(\alpha^*, P_{\rm av}) \approx 0.98$. Even with a very low voice spectral

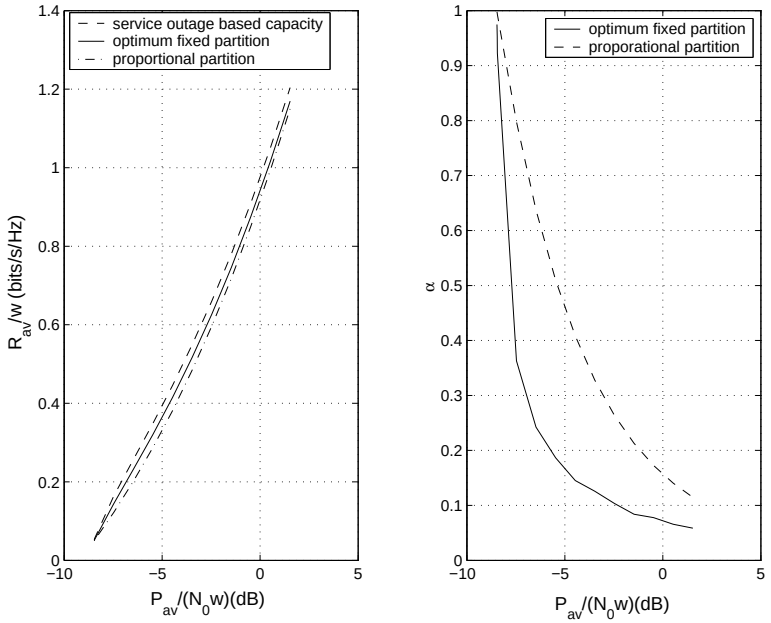


Figure 5.6: Comparison of equal power density partition scheme with the optimum fixed partition scheme in $M = 1$ Rayleigh fading channel with $r_{\rm o} = 0.05$ bits/s/Hz and $\epsilon = 0.01$. The average rate versus SNR is on the left, and the corresponding partition ratio versus SNR is on the right.

efficiency $r_{\rm o} = 0.05$ bits/s/Hz as shown in Fig 5.6, the difference between $R(\hat{\alpha}, P_{\rm av})$ and $R(\alpha^*, P_{\rm av})$ is negligible. The same phenomena is also observed for different ϵ.

Notice that although the resulting average rates $R(\hat{\alpha}, P_{\rm av})$ and $R(\alpha^*, P_{\rm av})$ are close to each other, the corresponding partition ratios $\hat{\alpha}$ and α^* can be quite different. In Figure 5.6 with $r_{\rm o} = 0.05$ bits/s/Hz, the corresponding $\hat{\alpha} = 0.46$ is almost three times of $\alpha^* = 0.17$. Thus, the equal power density partition scheme is a near optimum scheme, but it is in general not an approximation of the optimum fixed partition scheme.

Chapter 6

Variable-rate Turbo Bit-Interleaved Coded Modulation

In this chapter, we study the performance of variable-rate turbo bit-interleaved coded modulation (Turbo-BICM). A continuously varying rate can be obtained by changing the binary code rate through both random puncturing and modulation constellation size. The closed form approximation for the rate vs SNR performance of the variable-rate Turbo-BICM scheme is obtained and is applied to the design of adaptive transmission schemes in slow fading channels.

6.1 Introduction

Turbo codes proposed by Berrou et al. represent a recent breakthrough in coding theory [9, 10], which has stimulated a large amount of new research. These codes are parallel concantenated convolutional codes (PCCC) whose encoder is formed by two or more [28] constituent systematic encoders joined through one (or more) interleavers. Other types of turbo-like code concatenated with interleavers, such as serial concatenation codes [5] and repeat accumulate (RA) [26] codes were proposed. Some of the early investigations on turbo code design can be found, for example, in [27]- [29], and the iterative decoding algorithms in [41]. In depth research on understanding the excellent performance of turbo codes were explored in [6,7] and in [48]. Some tight upper bounds on the error probability of ML decoding for turbo codes were reported in [25], [75]. Coding theorems for turbo code ensembles were recently derived in [44] for a single channel and in [54, 55, 76] for parallel channels.

Successful attempts have also been undertaken to combine the binary turbo codes with higher order modulations (e.g. 8-PSK, 16-QAM). Several bandwidth efficient turbo code schemes, such as turbo bit-interleaved coded modulation (Turbo-BICM) [35],

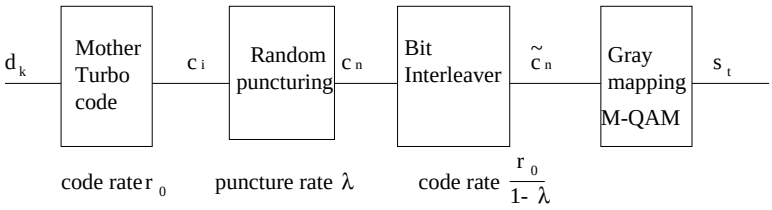


Figure 6.1: The turbo bit interleaved coded modulation transmitter.

turbo trellis coded modulation (Turbo-TCM) [74], parallel concatenated trellis coded modulation (PCTCM) [5], and bit-interleaved coded modulation with iterative decoding (BICM-ID) [53], were proposed. Compared to other schemes, Turbo-BICM has the advantage of simplicity. Although simple, Turbo-BICM has been shown to achieve bit error rate (BER) performance close to the capacity limit for a range of spectral efficiency values, for both additive white Gaussian noise (AWGN) channels [2, 35] and Rayleigh fading channels [36].

Turbo-BICM employs the capacity achieving binary turbo code [10] in the general bit-interleaved coded modulation (BICM) structure. BICM was formally described by Zehavi [87]. Unlike Ungerboecks trellis-coded modulation [80], the BICM separates modulation and coding as two independent entities. It is a bandwidth-efficient coding technique based on serial concatenation of binary error-correcting code, bit-by-bit channel interleaver, and a high order modulation (e.g. PSK, M-QAM) [19, 87]. At its receiver, appropriate soft-decision bit metrics are generated from the received signals and input into the binary decoder. The BICM achieves a better diversity order than the symbol interleaved coded modulation in a fading channel [19, 87]. Reference [19] presented, in a comprehensive fashion, the theory underlying BICM and provided a general information-theoretical framework for this concept.

In this work, a continuously variable-rate Turbo-BICM is obtained by changing the binary code rate through both random puncturing and the modulation constellation size. Random puncturing as a tool for analysis of variable rate binary turbo-codes has been studied in [54, 55, 76]. In the first part of this chapter, we apply recent results on

parallel channel performance of turbo code ensembles [54, 55, 76], to obtain a *union-Bhattacharyya rate threshold* for variable-rate Turbo-BICM based on a reliable channel region for turbo codes transmitted over parallel channels. A closed form approximation of this rate threshold is determined for an AWGN channel. This rate threshold is shown to predict well the Turbo-BICM rate performance [61].

In the second part of this chapter, we design the adaptive Turbo-BICM in a slow fading channel, by applying the closed form approximation of the union-Bhattacharyya rate threshold. The performance of a fixed-rate Turbo-BICM in a fast fading environment, where one fixed-rate codeword experience many independent fading states through channel interleaver, has been studied in [36] and was shown to achieve a rate within 3 dB of ergodic capacity. However, this fixed-rate approach is not applicable in a slow fading environment due to the delay constraint. To achieve a high average rate in a slow fading environment, adaptive transmission needs to be employed. Several information theoretic capacity notions describing adaptive transmission system performance and the corresponding optimum power allocations have been developed in [18,38,45,58]. In [21,37], the performance of adaptive transmission for both uncoded and trellis coded modulation with MQAM modulation was studied and was shown to be far away from the ergodic capacity. In this work, under the assumption that one codeword only spans one fading state, we design variable-rate Turbo-BICM schemes by adapting the power, puncture rate, modulation constellation size according to the current channel state. Proposed approaches can achieve an average rate within $2-3$ dB of the ergodic capacity [60].

6.2 System Model

In this section, we introduce the system model for variable-rate Turbo-BICM. As shown in Fig 6.1, the system consists of a mother turbo code with rate r_0, a random puncturing device with puncturing rate λ, a bit-by-bit channel interleaver, and an M-QAM modulation with Gray mapping. The mother turbo code results from parallel concatenation of two recursive systematic convolutional (RSC) codes. Fig 6.2 shows an example of parallel concatenated turbo code. It is assumed that the systematic bits of

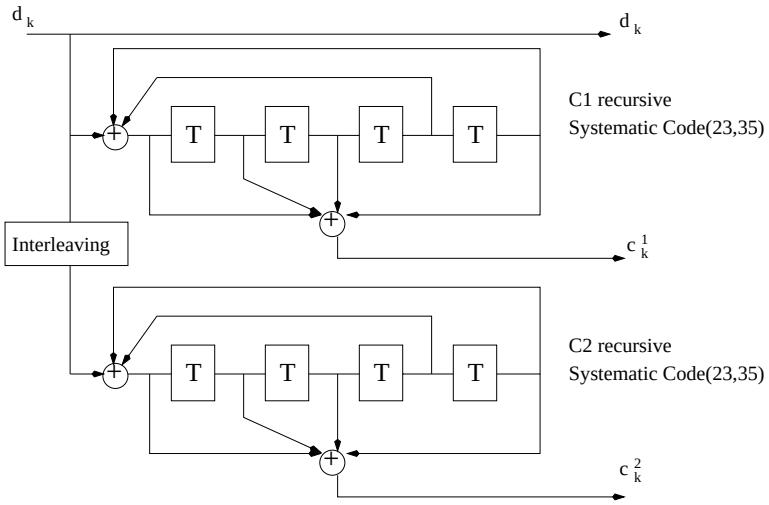


Figure 6.2: Rate $r_0 = 1/3$ standard turbo-encoder

the second encoder are not transmitted. The random puncturing device provides a large code family with various rates by varying the puncturing rate λ. In order to obtain an uncorrelated noise for adjacent coded bits after deinterleaving at the turbo-decoder input, a bit-by-bit channel interleaver is employed between the puncturing device and the modulator. After the interleaver, $2m$ coded bits are associated with a complex signal point in an $M = 2^{2m}$ QAM constellation using Gray mapping. The Gray mapping is asymptotically optimum in conjunction with BICM [19]. Usually, a periodic puncturing device with certain puncturing pattern is used for generating a family of rate compatible codes [16, 41]. Random puncturing as a tool for analysis of variable rate binary turbo-codes was introduced in [55]. In the case of random puncturing, each coded bit is independently punctured with probability λ. For a mother turbo code with rate r_0, let K denote the turbo code interleaving length and N the codeword length, then we have $K = r_0 N$. After random puncturing with puncturing rate λ, the resulting expected code rate, as well as the asymptotic rate when $N \to \infty$, equals to $r_0/(1-\lambda)$. After channel interleaving and M-QAM modulation, the rate is

$$r = m\frac{r_0}{1-\lambda}\,. \tag{6.1}$$

The unit of rate r is bits/s/dimension.

As shown in Fig 6.3, the Turbo-BICM receiver consists of bit metric generator,

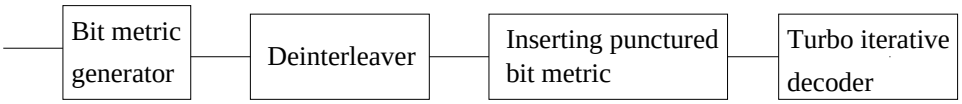


Figure 6.3: The turbo bit interleaved coded modulation receiver

bit metric deinterleaver, inserting bit metrics for punctured bits, and turbo iterative decoding. For each received complex signal, $2m$ bit metrics are computed according to the parallel binary random modulation model introduced in Section 6.3.1. The resulting bit metrics for the whole codeword are then deinterleaved. The bit metrics corresponding to the punctured code bit are inserted. Then the iterative decoding algorithm for binary turbo code is employed by applying these bit metrics.

6.3 Union-Bhattacharyya Rate Threshold

In this section, we first review some basic results on BICM in Section 6.3.1 and the parallel-channel coding theorem in Section 6.3.2. Then, a rate threshold for Turbo-BICM and its approximation in an AWGN channel are derived in Section 6.3.3 and Section 6.3.4. The simulation results on Turbo-BICM code performance in AWGN channel is presented in 6.3.5, and is shown to agree well with the prediction by the rate threshold.

Although $M = 2^{2m}$ QAM constellation is used in the system, all the analysis in this paper is based on 2^m-ary modulation in one dimension. This is because the Gray mapping in two dimensions employed in the system can be separated into two Gray mappings in each dimension, and also with a coherent receiver the in-phase and quadrature demodulator outputs are independent of each other. Therefore, it is sufficient to study the performance of BICM based on the 2^m-ary constellation .

6.3.1 Bit Interleaved Coded Modulation (BICM)

As shown in [19,87], BICM with an ideal interleaver can be analyzed based on a parallel binary input random modulation channel model. For a 2^m-ary constellation $\mathcal{X}$, each

signal point is labeled with m bits. Similar to [19, 87], the constellation $\mathcal{X}$ can be partitioned into two disjoint subsets $\mathcal{X}_0^i, \mathcal{X}_1^i$ for each bit position $i = 1, 2, \ldots, m$. Subset $\mathcal{X}_0^i$ includes all the signals $x \in \mathcal{X}$ whose label has the value 0 in position i. The size of set $\mathcal{X}_0^i$ is $2^{(m-1)}$. Similar definition applies to subset $\mathcal{X}_1^i$. An example of a 4-ary constellation partition is depicted in Fig 6.4. As shown in Fig 6.5, the equivalent channel

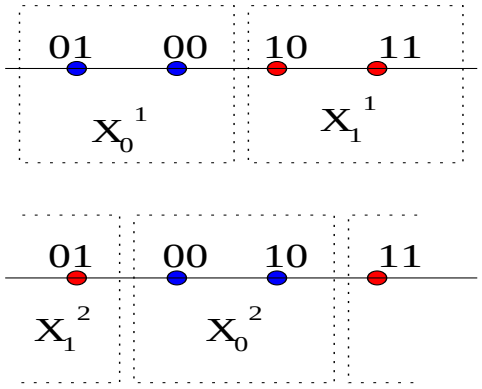


Figure 6.4: 4-ary constellation partition.

model consists of a set of m parallel independent and memoryless binary input channels, which are connected to the encoder output by a random switch. Each sub-channel corresponds to a position in the binary label of the signals in $\mathcal{X}$ and are associated with signal subsets $\mathcal{X}_0^i, \mathcal{X}_1^i$. Due to the random interleaver between the binary code and modulation, a coded bit b assigned to the i-th channel randomly chooses a signal $x \in \mathcal{X}_b^i$ with probability $2^{-(m-1)}$. Let $p(y|x)$ denote the transition probability of the memoryless physical channel. Then the transition probability of the i-th channel in the equivalent channel model can be expressed as

$$p_i(y|b) = \frac{1}{2^{(m-1)}} \sum_{x \in \mathcal{X}_b^i} p(y|x) \quad i = 1, 2, \ldots, m\,. \tag{6.2}$$

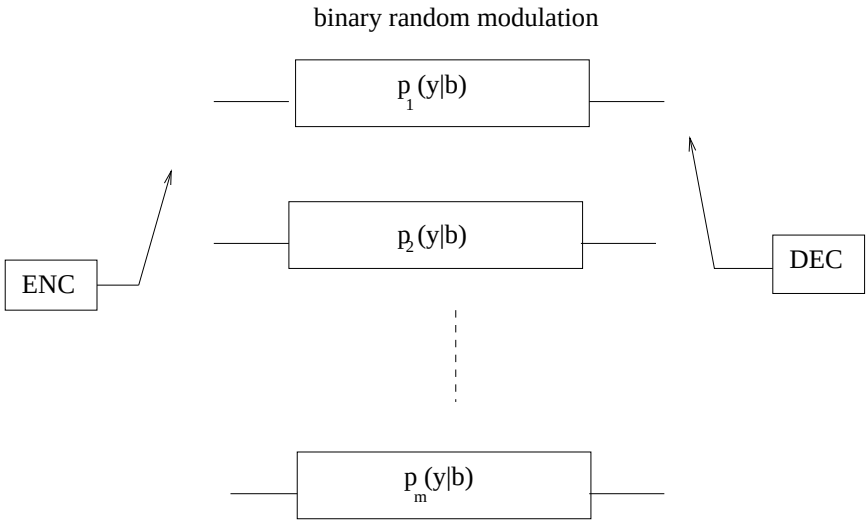


Figure 6.5: Equivalent parallel channel model for BICM with ideal random interleaving

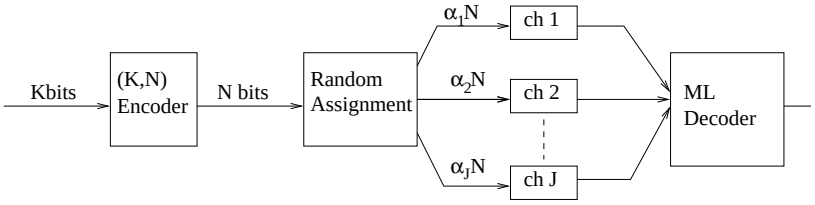


Figure 6.6: System model for parallel-channel coding theorem with random assignment

6.3.2 Union Bhattacharyya Reliable Channel Region

Coding theorems for binary Turbo code Ensembles were presented in [44] for a single channel. More recently, references [54,55,76] extend the coding theorem to parallel channels with random assignment, whose system model is shown in Fig 6.6. A (K, N) block code $\mathcal{C}$ with rate $r_0 = K/N$ is transmitted through J independent discrete memoryless channels (DMCs) in parallel. Each channel is modeled by its transition probability $p_j(y|b)$, with a binary input $b \in \{0, 1\}$ and an arbitrary output y. For a given codeword, it is assumed that each coded bit is randomly assigned to J parallel channels independently with probability α_j, for $1 \leq j \leq J$. The α_j is called the assignment rate for Channel j. Then, asymptotically, the code rate for the j-th channel is $r_j = r_0/\alpha_j$. The coding theorem is characterized by the Bhattacharyya noise parameters γ_j for parallel

channels and the mother turbo code ensemble threshold $c_0^{[\mathcal{C}]}$.

1. The *Bhattacharyya noise parameter* for the j-th channel is defined as

$$\gamma_j = \int \sqrt{p_j(y|0)p_j(y|1)}dy\,, \tag{6.3}$$

where $p_j(y|0)$ and $p_j(y|1)$ are channel transition probabilities. $\gamma_j \in [0,1]$ characterizes the 'noisiness' of the channel. The channel is more noisy when γ_j is closer to 1.

2. For a code ensemble $[\mathcal{C}]$, let $\overline{A}_h^{[\mathcal{C}](N)}$ denote the average number of codewords with weight h, termed *average weight enumerator*(AWE). Let D_N be a sequence of numbers such that

$$D_N \to \infty \quad \text{and} \quad \frac{D_N}{N^\epsilon} \to 0 \quad \forall \epsilon > 0\,. \tag{6.4}$$

The turbo code *ensemble threshold* $c_0^{[\mathcal{C}]}$is defined as

$$c_0^{[\mathcal{C}]} = \limsup_{N\to\infty} \max_{D_N < h \le N} \frac{\ln \overline{A}_h^{[\mathcal{C}](N)}}{h}\,. \tag{6.5}$$

It is observed [44, 54] that most asymptotically 'good' codes such as turbo codes, RA codes, and LDPC codes satisfy the following Assumption 1.

Assumption 1

$$\lim_{N\to\infty} \sum_{h=1}^{D_N} \overline{A}_h^{[\mathcal{C}](N)} = 0. \tag{6.6}$$

Theorem 5 *A code ensemble $[\mathcal{C}]$, which satisfies Assumption 1, is transmitted through J binary-input arbitrary-output channels in parallel with a set of Bhattacharyya noise parameters $\{\gamma_j\}$ and assignment rates $\{\alpha_j\}$, for $j = 1, \ldots, J$. If*

$$\sum_{j=1}^{J} \gamma_j \alpha_j < \exp(-c_0^{[\mathcal{C}]}). \tag{6.7}$$

then, the average ML decoding word error probability

$$\lim_{N\to\infty} \overline{P}_W^{[\mathcal{C}](N)} = 0\,. \tag{6.8}$$

Experimental evidence has shown that turbo code exhibits a threshold behavior, that is, the error probability is approaching zero asymptotically when the noise parameter is less than a threshold, and is bounded away from zero otherwise. Theorem 5 is only an achievable condition, and provides a lower bound to the actual "threshold" of the turbo code.

6.3.3 Equivalent Parallel Channel Model and the Union-Bhattacharyya Rate Threshold

Similar to BICM, the Turbo-BICM with random puncturing can be modeled by an equivalent parallel channel model with random assignments. The puncturing device can be regarded as the $(m+1)$-th sub-channel with $\gamma_{m+1} = 1$. As shown in Fig 6.7, the

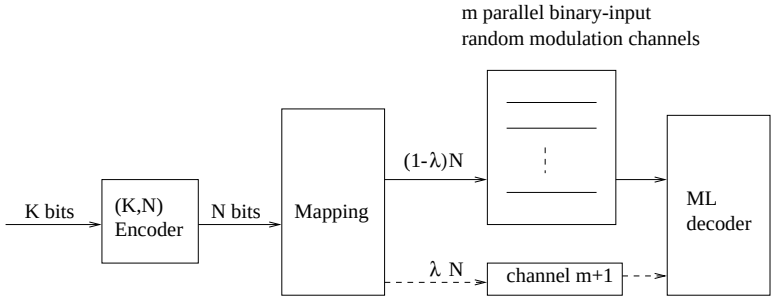


Figure 6.7: Equivalent parallel-channel model for BICM with random puncturing

mother code is transmitted through $m+1$ parallel channels with random assignment: the punctured bits are sent to a bad channel with $\gamma_{m+1} = 1$ with an assignment rate $\alpha_{m+1} = \lambda$, while the remaining bits are transmitted through m-parallel channels with γ_i with an assignment rate $\alpha_i = (1-\lambda)\frac{1}{m}$ for $i = 1, 2, \ldots, m$.

Following (6.2), the Bhattacharyya noise parameter for the i-th channel is

$$\gamma_i \;=\; \int \sqrt{\frac{1}{2^{m-1}} \sum_{x \in \mathcal{X}_0^i} p(y|x) \cdot \frac{1}{2^{m-1}} \sum_{z \in \mathcal{X}_1^i} p(y|z)}\, dy \quad i = 1, 2, \ldots, m\,. \tag{6.9}$$

Following the approaches in [54, 76] and applying Theorem 5, the average ML decoding word error probability approaches zero asymptotically if

$$\lambda \cdot 1 + (1-\lambda) \cdot \frac{1}{m} \sum_{i=1}^{m} \gamma_i < \exp(-c_0^{[\mathcal{C}]})\,. \tag{6.10}$$

Define the average Bhattacharyya noise parameter over m-parallel channels as

$$\overline{\gamma} = \frac{1}{m}\sum_{i=1}^{m}\gamma_i\,. \tag{6.11}$$

It follows that the word error probability goes to zero if the puncturing rate satisfies

$$\lambda < \frac{\exp(-c_0^{[\mathcal{C}]}) - \overline{\gamma}}{1-\overline{\gamma}} \triangleq \lambda_0(\overline{\gamma}),. \tag{6.12}$$

Since the puncturing rate must be nonnegative, we have that

$$\overline{\gamma} \le \exp(-c_0^{[\mathcal{C}]})\,. \tag{6.13}$$

Define a *union-Bhattacharyya rate threshold* as

$$\begin{aligned} r_{\mathrm{UB}}(\overline{\gamma}) &= \frac{mr_0}{1-\lambda_0(\overline{\gamma})} \qquad \text{for} \quad \lambda_0(\overline{\gamma}) \ge 0 \\ &= \begin{cases} 0 & \overline{\gamma} > \exp(-c_0^{[\mathcal{C}]}) \\ \frac{mr_0}{1-\exp(-c_0^{[\mathcal{C}]})}(1-\overline{\gamma}) & \overline{\gamma} \le \exp(-c_0^{[\mathcal{C}]}) \end{cases}. \end{aligned} \tag{6.14}$$

Then following Theorem 5, for a given average noise parameter $\overline{\gamma}$ any rate less than $r_{\mathrm{UB}}(\overline{\gamma})$ is achievable. $r_{\mathrm{UB}}(\overline{\gamma})$ is a decreasing function of $\overline{\gamma}$. When the channel is too noisy such that $\overline{\gamma} > \exp(-c_0^{[\mathcal{C}]})$, the rate threshold is zero, and a higher rate threshold can be achieved when the channel is better.

6.3.4 Approximate Rate Threshold for an AWGN Channel

In an AWGN channel with one-side noise spectrum density N_0, the noise variance per dimension is $N_0/2$. The channel transition probability per dimension is

$$p(y|x) = \frac{1}{\sqrt{\pi N_0}}\exp\frac{(y-x)^2}{N_0}\,.$$

In an AWGN channel, the Bhattacharyya noise parameter and the corresponding achievable rate only depend on signal to noise ratio $\frac{P}{N_0}$ and the modulation parameter m. Here P is the transmission energy per complex signal, and thus $P/2$ is the transmission energy per dimension. Let $\overline{\gamma}\left(\frac{P}{N_0}, m\right)$ and $r_{\mathrm{UB}}\left(\frac{P}{N_0}, m\right)$ denote the average Bhattacharyya noise parameter and the union-Bhattacharyya rate threshold in an AWGN channel respectively. In this section, we derive a closed form approximation for $r_{\mathrm{UB}}\left(\frac{P}{N_0}, m\right)$.

$\eta_1(m)$, the solution of

$$\overline{\gamma}\left(\eta_1(m), m\right) = \exp\left(-c_0^{[\mathcal{C}]}\right), \tag{6.15}$$

exists, since $\overline{\gamma}\left(\frac{P}{N_0}, m\right)$ is continuous and monotonic in $\frac{P}{N_0}$. It follows that

$$r_{\mathrm{UB}}\left(\frac{P}{N_0}, m\right) = 0 \quad \text{if} \quad \frac{P}{N_0} < \eta_1(m), \tag{6.16}$$

i.e., $\eta_1(m)$ is the minimum SNR for a nonzero r_{UB}.

In the following, we derive a closed form approximation of $r_{\mathrm{UB}}(\frac{P}{N_0}, m)$ when $\frac{P}{N_0} \geq \eta_1(m)$. We first obtain a closed form approximation of $\overline{\gamma}(\frac{P}{N_0}, m)$. It is easy to show that in an AWGN channel with $m = 1$ (QPSK modulation), we have

$$\overline{\gamma}(\frac{P}{N_0}, m = 1) = \exp\left\{-\frac{P}{2N_0}\right\}. \tag{6.17}$$

Though the expression for $\overline{\gamma}$ is simple at $m = 1$, it can only be calculated numerically for $m \geq 2$.

For any two signals $x \in \mathcal{X}$ and $z \in \mathcal{X}$, let $d(x, z)$ denote the Euclidean distance between x and z. In an AWGN channel, we have

$$\int \sqrt{p(y|x)p(y|z)}dy = \exp\left(-\frac{d^2(x,z)}{4N_0}\right). \tag{6.18}$$

Proposition 9 *For any $a_i \geq 0$ with $i = 1, 2, \ldots, N$, we have*

$$\sqrt{\sum_{i=1}^{N} a_i} \leq \sum_{i=1}^{N} \sqrt{a_i}.$$

The bound is tight when there exists one dominant term, that is, there exists an $a_j \gg a_i$ for all $i \neq j$.

From Proposition 9, and (6.18), we obtain an upper bound on $\overline{\gamma}\left(\frac{P}{N_0}, m\right)$ as follows:

$$\begin{aligned}
\overline{\gamma}\left(\frac{P}{N_0}, m\right) &= \frac{1}{m}\sum_{i=1}^{m}\int\sqrt{\frac{1}{2^{m-1}}\sum_{x\in\mathcal{X}_0^i} p(y|x)\cdot\frac{1}{2^{m-1}}\sum_{z\in\mathcal{X}_1^i} p(y|z)}dy \\
&\leq \frac{1}{m2^{m-1}}\sum_{i=1}^{m}\sum_{x\in\mathcal{X}_0^i}\sum_{z\in\mathcal{X}_1^i}\int\sqrt{p(y|x)p(y|z)}dy \\
&= \frac{1}{m2^{m-1}}\sum_{i=1}^{m}\sum_{x\in\mathcal{X}_0^i}\sum_{z\in\mathcal{X}_1^i}\exp\left(-\frac{d^2(x,z)}{4N_0}\right) \\
&\triangleq \overline{\gamma}_{\mathrm{u}}(\frac{P}{N_0}, m)
\end{aligned} \tag{6.19}$$

Let $d_{\min}$ denote the minimum Euclidean distance between two signal points in a 2^m-ary constellation $\mathcal{X}$. Then, from [70] we have

$$\frac{d_{\min}^2}{N_0} = \frac{6}{2^{2m}-1}\frac{P}{N_0}\,. \tag{6.20}$$

It can be shown that the upper bound (6.19) is tight for a high SNR such that $\frac{d_{\min}^2}{N_0} \gg 1$. Under this condition, signal points are separated far away from each other compared to the noise variance. Then for any y the summation $\sum_{x\in\mathcal{X}_0^i}\sum_{x\in\mathcal{X}_1^i} p(y|x)p(y|z)$ is dominated by one term with $\hat{x} \in \mathcal{X}_0^i$ and $\hat{z} \in \mathcal{X}_1^i$, which are closest signals to y. Thus, by Proposition 9 the inequality (a) is tight at high SNR.

Definition 4 In an 2^m-ary constellation, define $N_{\rm d}(m,x)$ as the number of signals within the complement set $\mathcal{X}_{\overline{b}}^i$, whose distance from $x \in \mathcal{X}_b^i$ is the minimum distance $d_{\min}$. Define

$$a_2(m) = \frac{1}{m2^m}\sum_{i=1}^{m}\sum_{b=0}^{1}\sum_{x\in\mathcal{X}_b^i} N_{\rm d}(m,x)\,. \tag{6.21}$$

By combining the terms with the minimum distance, the upper bound $\overline{\gamma}_{\rm u}$ can be written as

$$\begin{aligned}
\overline{\gamma}_{\rm u}(\frac{P}{N_0},m) &= a_2(m)\exp\left(-\frac{d_{\min}^2}{4N_0}\right) + o\left(\frac{d_{\min}^2}{N_0}\right) \\
&\approx a_2(m)\exp\left(-\frac{3}{2(2^{2m}-1)}\frac{P}{N_0}\right) \qquad \text{for} \quad \frac{d_{\min}^2}{N_0} \gg 1 \\
&\triangleq \overline{\gamma}_{\rm appr}(\frac{P}{N_0},m)\,.
\end{aligned} \tag{6.22}$$

Therefore, combining (6.19) and (6.22), we have $\overline{\gamma} \approx \overline{\gamma}_{\rm u} \approx \overline{\gamma}_{\rm appr}$ when $d_{\min}^2/N_0 \gg 1$. As shown from the numerical results in Fig 6.8 and Fig 6.9, the approximation is tight when $d_{\min}^2/N_0 \geq 2$. By translating the aprroximation condition in terms of SNR, we summerize the above result in the following lemma.

Lemma 15 *The average Bhattacharyya noise parameter is approximated as*

$$\overline{\gamma}\left(\frac{P}{N_0},m\right) \approx a_2(m)\exp\left(-\frac{3}{2(2^{2m}-1)}\frac{P}{N_0}\right)\,, \tag{6.23}$$

when

$$\frac{P}{N_0} \geq \frac{2^{2m}-1}{3} \triangleq \eta_2(m)\,. \tag{6.24}$$

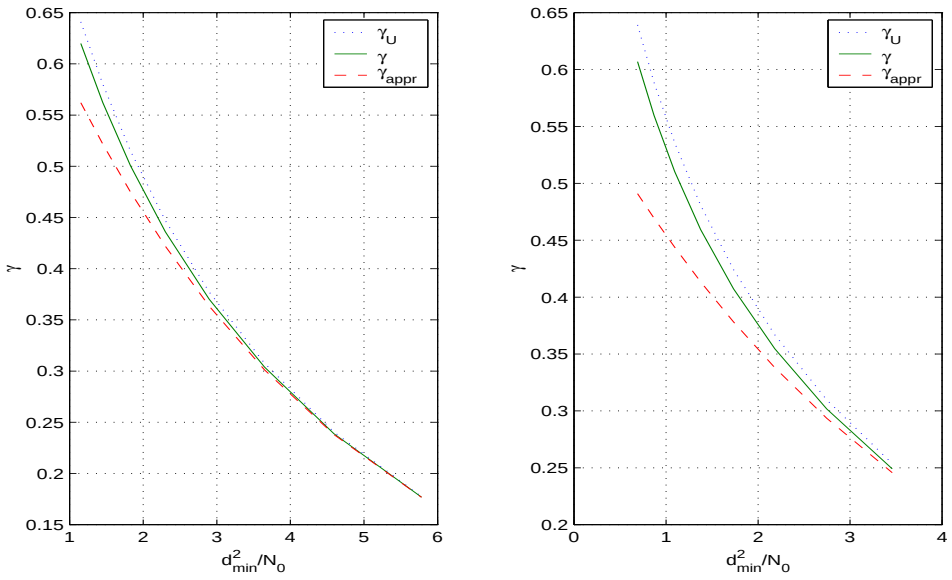


Figure 6.8: The average Bhattacharyya noise parameter $\overline{\gamma}$, the upper bound $\overline{\gamma}_{\mathrm{u}}$, the approximated noise parameter $\overline{\gamma}_{\mathrm{appr}}$ versus $d^2_{\min}/N_0$ for 16-QAM, 64-QAM in AWGN channel.

It is also observed from the numerical results that $r_{\mathrm{UB}}(\frac{P}{N_0}, m)$ is parallel to the capacity of the AWGN channel when $\frac{P}{N_0} \leq \eta_2(m)$. Then, we have

$$r_{\mathrm{UB}}\left(\frac{P}{N_0}, m\right) \approx 1/2 \log_2\left(1 + \frac{P}{N_0}\frac{1}{\Gamma(m)}\right) \quad \text{for} \quad \eta_1(m) \leq \frac{P}{N_0} \leq \eta_2(m)\,.$$

Here $\Gamma(m)$ denotes the gap between the rate threshold and the capacity.

In summary, the union-Bhattacharyya rate threshold $r_{\mathrm{UB}}\left(\frac{P}{N_0}, m\right)$ for Turbo-BICM can be approximated with a function

$$\tilde{r}_{\mathrm{UB}}\left(\frac{P}{N_0}, m\right) =$$

$$\begin{cases} 0 & \frac{P}{N_0} < \eta_1(m) \\ 1/2 \log_2\left(1 + \frac{P}{N_0}\frac{1}{\Gamma(m)}\right) & \eta_1(m) \leq \frac{P}{N_0} \leq \eta_2(m) \\ a_1(m)\left(1 - a_2(m)\exp\left(-a_3(m)\frac{P}{N_0}\right)\right) & \frac{P}{N_0} \geq \eta_2(m) \end{cases}, \tag{6.25}$$

where parameters $a_1(m)$ and $a_3(m)$ are

$$a_1(m) = \frac{mr_0}{1 - \exp(-c_0^{[\mathcal{C}]})}, \quad a_3(m) = \frac{3}{2(2^{2m} - 1)}\,. \tag{6.26}$$

The $\Gamma(m)$ can be determined such that $\tilde{r}_{\mathrm{UB}}\left(\frac{P}{N_0}, m\right)$ is a continuous function of $\frac{P}{N_0}$.

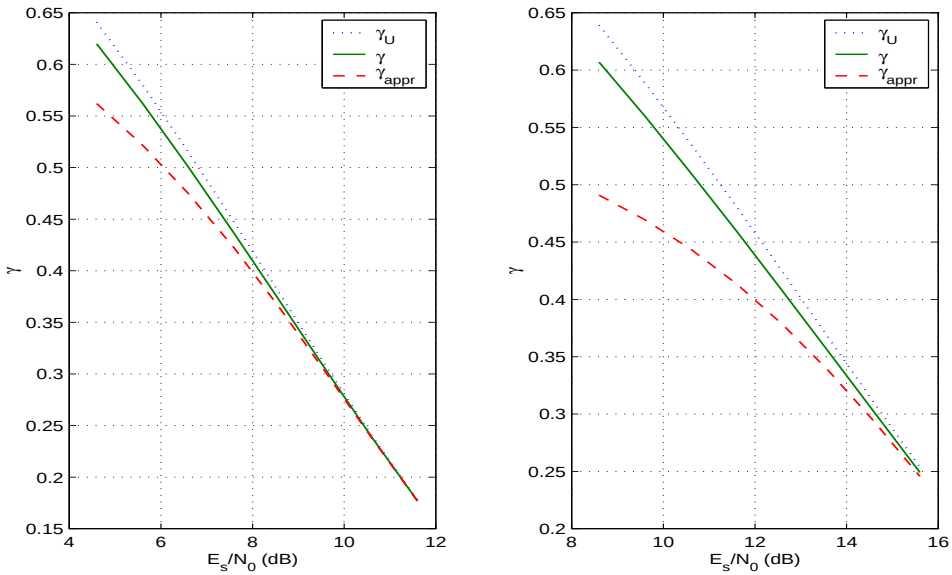


Figure 6.9: The average Bhattacharyya noise parameter $\overline{\gamma}$, the upper bound $\overline{\gamma}_{\mathrm{u}}$, the approximated noise parameter $\overline{\gamma}_{\mathrm{appr}}$ versus E_{av}/N_0 for 16-QAM, 64-QAM in AWGN channel

	$\eta_1(m)$(dB)	$\eta_2(m)$(dB)	$\alpha(m)$	$\Gamma(m)(dB)$
$m=1$	-0.06	0	1	2.3107
$m=2$	4.649	6.9897	3/4	2.8434
$m=3$	8.41	13.222	7/12	3.8049

Table 6.1: Parameters employed in $\tilde{r}_{\mathrm{UB}}\left(\frac{P}{N_0}, m\right)$.

To illustrate the above results, we use a turbo code with a rate $r_0 = 1/3$, where the same polynomial generators $(23, 35)$ is used for two elementary encoders C_1 and C_2 as shown in Fig 6.2. For this turbo code, we have $r_0 = 1/3$ and $c_0^{[\mathcal{C}]} = 0.51$. The parameters employed in the rate threshold approximation are given in Table 6.1. The union-Bhattacharyya rate threshold and its approximation for QPSK, 16QAM, and 64 QAM are presented in Fig 6.10. We can see that $\tilde{r}_{\mathrm{UB}}(\frac{P}{N_0}, m)$ is a good approximation of $r_{\mathrm{UB}}(\frac{P}{N_0}, m)$.

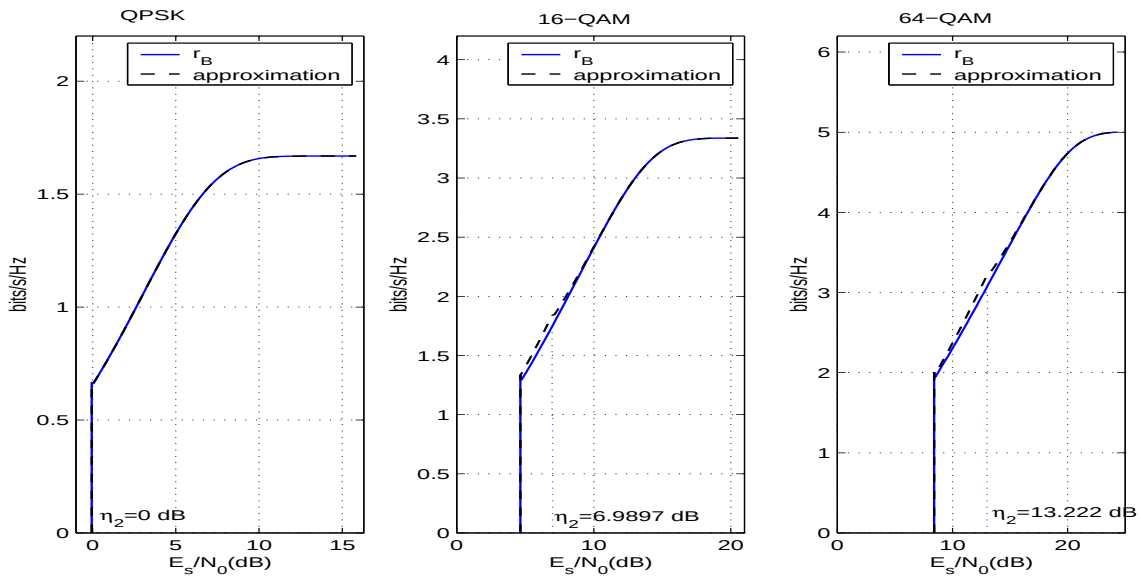


Figure 6.10: Comparison of the rate threshold $r_{\mathrm{UB}}\left(\frac{P}{N_0}, m\right)$ and its approximation $\tilde{r}_{\mathrm{UB}}\left(\frac{P}{N_0}, m\right)$ for QPSK, 16-QAM, and 64-QAM modulation.

6.3.5 Simulations and Discussions

In this section, we present the simulation results and show that the union-Bhattacharyya rate threshold $r_{\mathrm{UB}}(\frac{P}{N_0}, m)$ can help predict the rate performance of Turbo-BICM very well for sufficiently large K and a range of p_{e}.

For a typical turbo code, its codeword error probability as a function of SNR can be described by the following three distinct regions:

- *High error rate region:* Below a certain critical SNR, high error probability is observed, with almost no improvement until a critical SNR (the threshold) is reached.

- *Waterfall region:* As the SNR is increased past the threshold, the error probability drops rapidly, and the curve resembles a waterfall in this region.

- *Error-floor region:* After the waterfall region, the slope of the error probability curve becomes very low, reaching the so called error floor. The error floor is determined by the minimum free distance of the code.

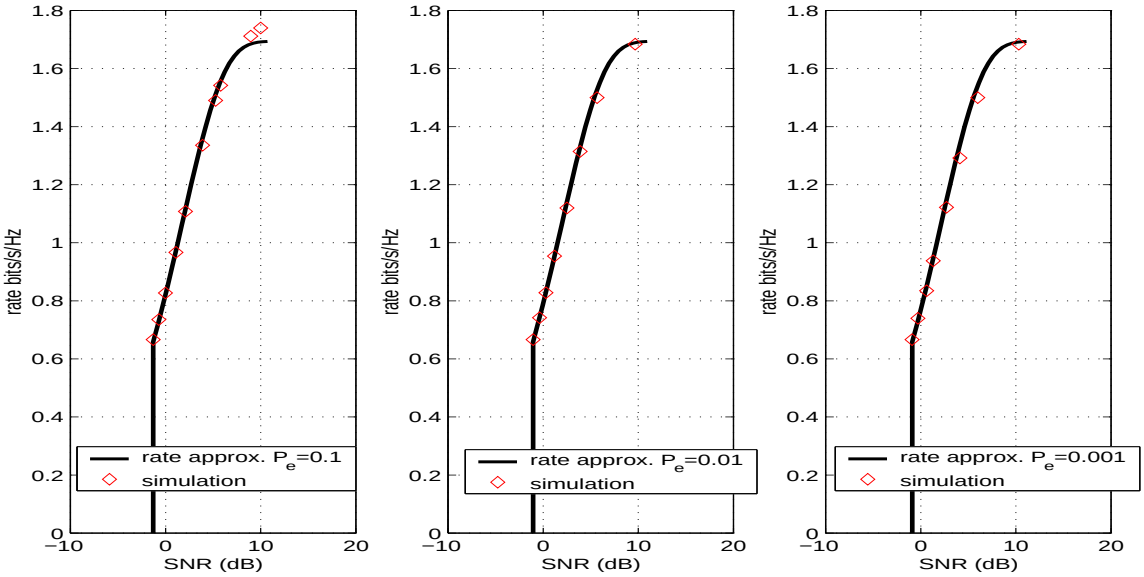


Figure 6.11: The approximated rate $\tilde{r}_{\mathrm{UB}}(\frac{P}{N_0}\xi(p_{\mathrm{e}},K),m)$ and the rate obtained from simulation for Turbo-BICM with QPSK modulation ($m=1$) with $K=1000$, and $p_{\mathrm{e}}=0.1$, $p_{\mathrm{e}}=0.01$, and $p_{\mathrm{e}}=0.001$ respectively.

As codeword length N increases, the 'water-fall' slope is steeper and the error floor is lower. This 'threshold' behavior is also observed in turbo BICM schemes with M-QAM modulation.

Let $r(\frac{P}{N_0},m,p_{\mathrm{e}},K)$ denote the rate that achieves a word error probability p_{e} at an SNR $\frac{P}{N_0}$ by a Turbo-BICM with modulation parameter m and an information sequence length K using the iterative decoding algorithm. It is observed that $r(\frac{P}{N_0},m,p_{\mathrm{e}},K)$ is almost parallel to $\tilde{r}_{\mathrm{UB}}(\frac{P}{N_0},m)$. The gap depends on the error probability p_{e} and the information sequence length K, denoted by $\xi(p_{\mathrm{e}},K)$. The rate performance of Turbo-BICM for sufficiently large K and a range of p_{e} can be approximated as

$$r\left(\frac{P}{N_0},m,p_{\mathrm{e}},K\right)\approx\tilde{r}_{\mathrm{UB}}\left(\frac{P}{N_0}\xi(p_{\mathrm{e}},K),m\right). \tag{6.27}$$

In Fig 6.11, the approximated rate is plotted against the rate obtained from simulation for Turbo-BICM with QPSK modulation ($m=1$) with $K=1000$, and $p_{\mathrm{e}}=0.1$, $p_{\mathrm{e}}=0.01$, and $p_{\mathrm{e}}=0.001$ respectively. The turbo code used in simulation is shown in Fig 6.2. It is observed that for $K=1000$, we have $\xi(p_{\mathrm{e}}=0.1,K)\approx 1.23dB$ dB, $\xi(p_{\mathrm{e}}=0.01,K)\approx 1dB$ dB, and $\xi(p_{\mathrm{e}}=0.001,K)\approx 0.83$ dB. In Fig 6.12, the approximate

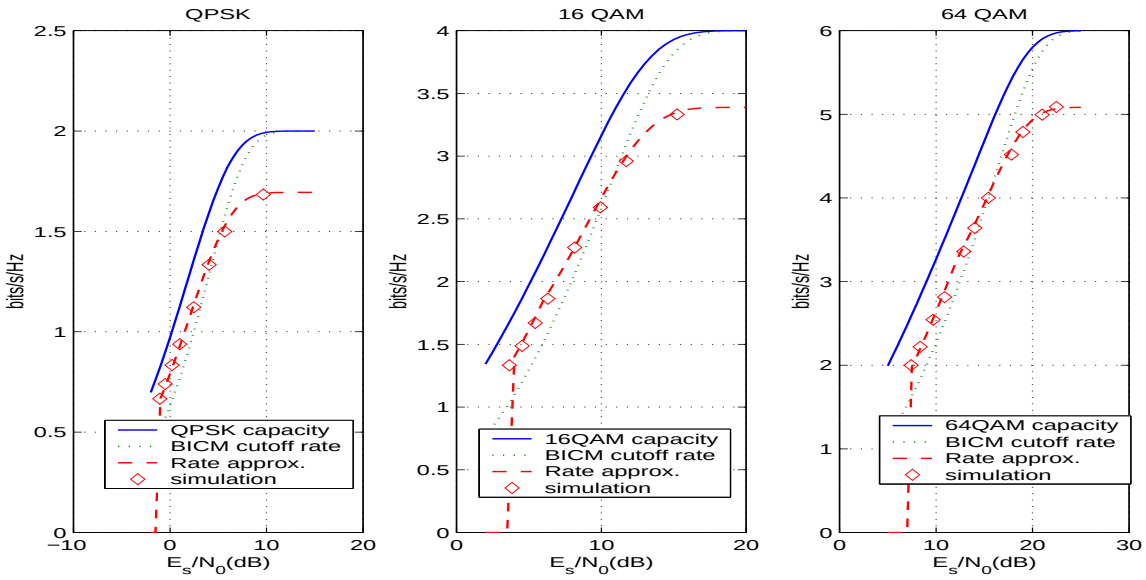


Figure 6.12: MQAM capacity, BICM cutoff rate, rate approximation, and simulation with $p_e = 0.01$.

rate is plotted against the rate obtained from the simulation for QPSK, 16-QAM and 64-QPSK with $p_e = 0.01$. It can be seen that the approximated rate predicts the rate performance very well. In Fig 6.12, the M-QAM capacity and BICM cutoff rate are also presented for comparison. It can be seen that at low SNR where the puncturing rate is low, the rate is above the cutoff rate, while at high SNR where the puncturing rate is too high, the rate is below the cutoff rate.

6.4 Adaptive Turbo-BICM for Slow Fading Channels

In this section, based on the closed form expression of the rate threshold [59] developed in Section 6.3, we describe the optimum power and modulation allocation for Turbo-BICM for a slow fading channel. The computational complexity of finding the optimum allocations is exponential in the size of channel state space. A dual problem solution, whose computational complexity increases linearly with the size of channel state space, is described. It achieves rates close to the optimum solution. Two simple schemes: water-filling with optimum modulation and equal power allocation with optimum modulation are shown to have a good performance. Proposed approaches can achieve an

average rate within $2-3$ dB of the ergodic capacity.

6.4.1 The Allocation Problem for a BF-AWGN Channel

Under the assumption that the phase can be completely compensated at the receiver, the slow fading channel model is as follows:

$$y = \sqrt{g}x + w\,, \tag{6.28}$$

where g is the fading gain, x is the transmitted complex signal with average energy P, and noise w is a complex zero-mean Gaussian iid random variable with covariance $E[ww^\dagger] = \frac{N_0}{2}I_2$. Let the effective noise density be $n = N_0/g$. By incorporating the fading gain into the noise term, we get an equivalent channel model as $y = x + z$, with $E[zz^\dagger] = \frac{n}{2}I_2$. The effective noise density n is a random variable which now models the effects of the fading. We assume that the fading is slow relative to the codeword length, that is, the effective noise density n stays the same within one codeword but may change from one codeword to another (block fading model). We also assume that perfect channel state information is available at both transmitter and receiver.

The puncture rate, power, and modulation constellation can be adapted according to the channel state information to achieve a better performance than the non-adaptive scheme. Let $P(n)$ denote the power allocation, and $m(n)$ denote the constellation parameter assigned to the channel state described by the noise density realization n. Optimal allocations $P(n)$ and $m(n)$ maximize the expected rate $E\left[r(\frac{P(n)}{n}, m(n), p_{\mathrm{e}}, K)\right]$ for a given target error probability p_{e}. Since $r(\frac{P(n)}{n}, m(n), p_{\mathrm{e}}, K)$ is related to the rate threshold $\tilde{r}_{\mathrm{UB}}(\frac{P(n)}{n}, m(n))$ through a constant gap $\xi(p_{\mathrm{e}}, K)$, equivalently, we solve the following problem

$$R^* = \max_{P(n), m(n)} E\left[\tilde{r}_{\mathrm{UB}}\left(\frac{P(n)}{n}, m(n)\right)\right] \tag{6.29}$$

$$\text{subject to} \quad E\left[P(n)\right] \le P_{\mathrm{av}} \tag{6.29a}$$

$$P(n) \ge 0, \quad m(n) \in \{0, 1, 2, 3\}\,. \tag{6.29b}$$

Once the optimum solution of Problem (6.29) is available, it is easy to obtain the corresponding allocation that maximizes the rate performance $E\left[r(\frac{P(n)}{n}, m, p_{\mathrm{e}}, K)\right]$.

In this paper, it is assumed that n is a discrete random variable with an alphabet size N. Problem (5) is a mixed-integer nonlinear optimization problem with continuous and discrete variables and nonlinearities in the objective function and constraints [5]. Due to its combinatorial nature, it belongs to the class of NP-complete problems [5] (see also [9]). In the following sections, we consider several approaches to solving Problem (5).

6.4.2 Optimum Allocation

Although there are more efficient algorithms than the naive exhaustive search, they are usually more complicated. These algorithms commonly involve a combination of basic nonlinear optimization subproblems and a cutting plane mixed-integer linear programming problem (see, e.g., [40]). In this subsection, we study an exhaustive search approach.

The exhaustive-search algorithm consists of two steps:

1. For any given sequence $m(n) \in \{(0,1,2,3)^N\}$, find the power allocation that maximizes the average rate as follows:

$$\begin{aligned} R(m(n)) = \max_{P(n)} & E\left[\tilde{r}_{\mathrm{UB}}\left(\frac{P(n)}{n}, m(n)\right)\right] \\ \text{subject to} \quad & E\left[P(n)\right] \leq P_{\mathrm{av}} \\ & P(n) \geq 0\,. \end{aligned} \tag{6.30}$$

   For a given value of n, the rate threshold $\tilde{r}_{\mathrm{UB}}\left(\frac{P(n)}{n}, m(n)\right)$ is described in three regions and its first derivative is not continuous at the boundary points of these regions. Therefore, subproblem (6.30) is not a convex optimization problem and the Kuhn-Tucker conditions theorem can not be applied directly. A naive algorithm for solving this non-convex optimization problem is given in Appendix 6.A.

2. Find the optimum sequence $m^*(n)$ that maximizes $R(m(n))$ by enumerating all 4^N possible $m(n)$ sequences.

This algorithm is guaranteed to find an optimum solution, but with a price of high computational complexity. The computational complexity increases exponentially with

the size of N.

6.4.3 Lagrange Dual Problem Solution

In this section, we study a dual of problem (6.29). The computational complexity of the dual problem increases linearly with N. It provides lower and upper bounds to problem (6.29). For some ranges of P_{av} where the duality gap is zero, the dual problem solution is also the optimum solution of the primal problem (6.29).

Let λ be the Lagrange multiplier associated with the constraint (6.29a). Define a set of dual functions at each channel state n as

$$\begin{aligned} g(\lambda, n) = & \max_{m,P} \tilde{r}_{\text{UB}}\left(\frac{P}{n}, m\right) - \lambda(P - P_{\text{av}}) \\ & \text{subject to} \quad P \geq 0, \quad m \in \{0, 1, 2, 3\}. \end{aligned} \tag{6.31}$$

The primal problem (6.29) is decomposed into N dual problems. Since each dual problem only has two variables: m with an alphabet size 4 and P having three continuous descriptions in three distinct regions, the computational complexity of each dual problem is small. Thus, the overall complexity of the dual approach increases linearly with the alphabet size N.

Let $m(n, \lambda)$ and $P(n, \lambda)$ denote the optimum solution of problem (6.31). For any $\lambda \geq 0$ whose corresponding $m(n, \lambda)$ and $P(n, \lambda)$ are feasible solutions of Problem (2.11), we have

$$\begin{aligned} & g(\lambda, n) \\ & = E\left[\tilde{r}_{\text{UB}}\left(\frac{P(n,\lambda)}{n}, m(n,\lambda)\right)\right] + \lambda(P_{\text{av}} - E\left[P(n,\lambda)\right]) \\ & \overset{(a)}{\geq} R^* \overset{(b)}{\geq} E\left[\tilde{r}_{\text{UB}}\left(\frac{P(n,\lambda)}{n}, m(n,\lambda)\right)\right]. \end{aligned} \tag{6.32}$$

Inequality (a) follows from the weak duality bound [14]. Inequality (b) follows from the feasibility condition. Inequality (6.32) provides upper and lower bounds for the optimum solution. In the case where the optimum solution is hard to obtain, the bounds can be used to estimate the optimum solution.

To obtain a feasible solution close to the optimum solution, we chose λ' to minimize

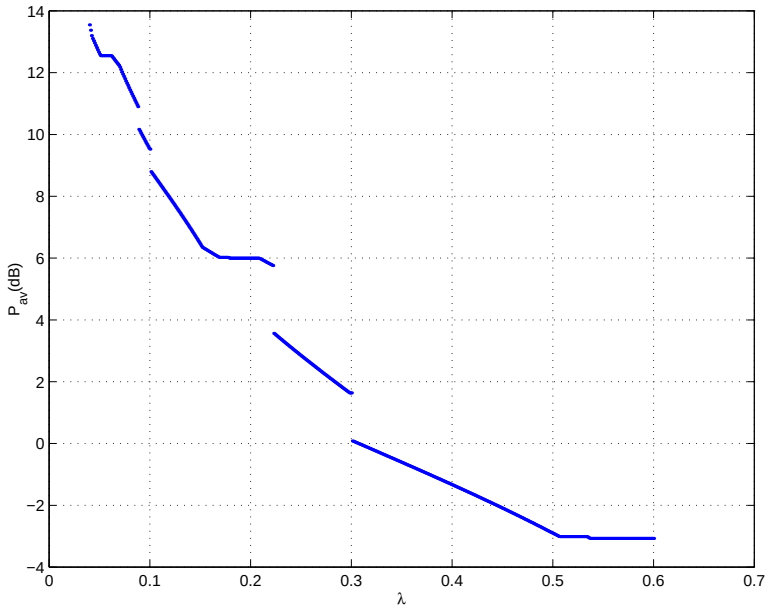


Figure 6.13: $E\left[P(\lambda, n)\right]$ versus λ in a two-state model: $n_1 = 1$ and $n_2 = 3$ with equal probability.

the gap between the bounds as[1]

$$\begin{aligned} \lambda' &= \arg\min_{\lambda \geq 0} \lambda(P_{\text{av}} - E\left[P(\lambda, n)\right]) \\ &\quad \text{subject to} \quad P_{\text{av}} - E\left[P(\lambda, n)\right] \geq 0\,. \end{aligned}$$

The $P(\lambda', n)$ and $m(\lambda', n)$ are feasible solutions of the primal problem (2.11). In Fig 6.13, $E\left[P(\lambda, n)\right]$ versus λ is plotted for a two-state model: $n_1 = 1$ and $n_2 = 3$ with equal probability. As we can see, since $P(\lambda, n)$ is the optimum solution to the mixed-integer optimization problem (6.31), its expected value $E\left[P(n, \lambda)\right]$ is not a continuous function of λ. If there exists a λ' such that $E\left[P(n, \lambda')\right] = P_{\text{av}}$ for a given P_{av}, the duality gap is zero and the corresponding $P(\lambda', n)$ and $m(\lambda', n)$ are the optimum solution for the primal problem (2.11). If no solution of $E\left[P(n, \lambda)\right] = P_{\text{av}}$ exists for a given P_{av}, the dual solution is sub-optimum. A probabilistic scheme may be employed for the range of P_{av} where the dual solution is sub-optimum. The probabilistic approach is out of the scope of this paper. Interested readers may refer to [59] for an example of

[1]The tightest upper bound is achieved by $\lambda^* = \min_{\lambda \geq 0} g(\lambda, n)$, but the corresponding $P(\lambda^*, n)$ and $m(\lambda^*, n)$ may not be a feasible solution to problem (2.11).

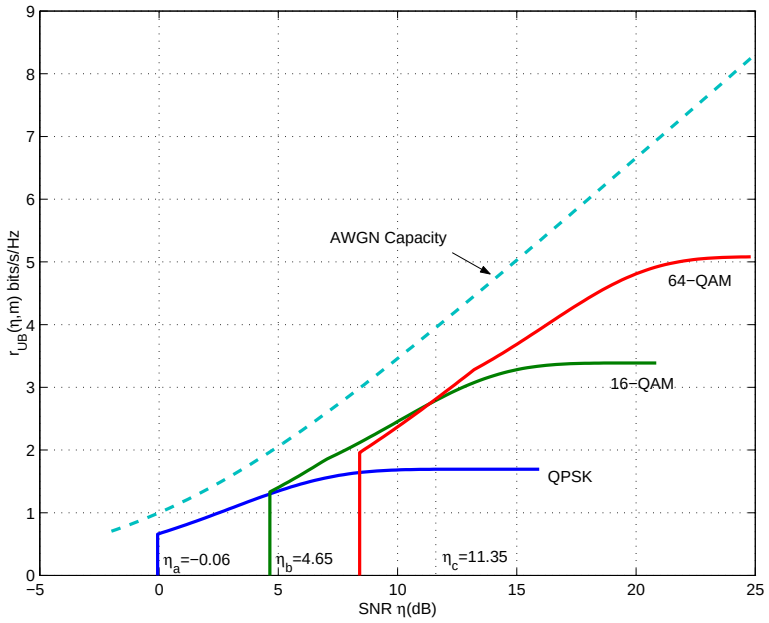


Figure 6.14: AWGN capacity and rate threshold approximation $\tilde{r}_{\mathrm{UB}}(\eta, m)$ for QPSK, 16-QAM, 64-QAM.

probabilistic schemes.

6.4.4 Simple Power Allocations with Optimum Modulation

In previous sections, we have discussed several approaches to solving Problem (6.29). The optimum solution as well as the dual problem solution, can be computationally intensive. In this section, we propose two simple allocation schemes: water-filling allocation with optimum modulation, and equal power allocation with optimum modulation. These two simple schemes, as shown in Section V, have a performance close to the optimum solution.

We first determine the optimum modulation mode for a given $P(n)$ as

$$m(n, P(n)) = \arg\max_{m\in\{0,1,2,3\}} \tilde{r}_{\mathrm{UB}}\left(\frac{P(n)}{n}, m\right). \tag{6.33}$$

The rate threshold $\tilde{r}_{\mathrm{UB}}(\eta, m)$ as a function of received SNR η for all $m = 0, 1, 2, 3$ is plotted in Fig 6.14. As can be observed, for any two modulation modes m_1 and m_2, the corresponding rate functions $\tilde{r}_{\mathrm{UB}}(\frac{P(n)}{n}, m_1)$ and $\tilde{r}_{\mathrm{UB}}(\frac{P(n)}{n}, m_2)$ intersect only once.

Let $\eta_{\rm a}$, $\eta_{\rm b}$, and $\eta_{\rm c}$ denote the SNR at the intersections as

$$\tilde{r}_{\rm UB}(\eta_{\rm a},1)=0,\quad \tilde{r}_{\rm UB}(\eta_{\rm b},1)=\tilde{r}_{\rm UB}(\eta_{\rm b},2),$$
$$\tilde{r}_{\rm UB}(\eta_{\rm c},2)=\tilde{r}_{\rm UB}(\eta_{\rm c},3)\,. \tag{6.34}$$

We have $\eta_{\rm a}=0$ dB, $\eta_{\rm b}=4.65$ dB, and $\eta_{\rm c}=11.35$ dB.

The optimum modulation mode for given $P(n)$ is

$$m(n,P(n))=\begin{cases}0 & \frac{P(n)}{n}\le\eta_{\rm a}\\ 1 & \eta_{\rm a}\le\frac{P(n)}{n}\le\eta_{\rm b}\\ 2 & \eta_{\rm b}\le\frac{P(n)}{n}\le\eta_{\rm c}\\ 3 & \frac{P(n)}{n}\ge\eta_{\rm c}\end{cases}. \tag{6.35}$$

We propose two power allocation schemes, the water-filling allocation and the equal power allocation. The water-filling allocation is $P_{\rm wf}(n,\lambda)=(\lambda-n)^+$, where λ is the solution to $E\left[P_{\rm wf}(n,\lambda)\right]=P_{\rm av}$. The equal power allocation is $P_{\rm eq}(n)=P_{\rm av}$. The motivation of choosing these two power allocations is as follows. The rate threshold $\tilde{r}_{\rm UB}(\frac{P(n)}{n},m)$ is almost parallel to AWGN capacity with a constant gap when $\frac{P(n)}{n}\le\eta_2(m)$, and goes away from the AWGN capacity when $\frac{P(n)}{n}\ge\eta_2(m)$. The optimum modulation mode allocation implies switching to a higher order modulation after the rate vs. SNR relationship diverges from the AWGN capacity. As shown in Fig 6.14, the rate threshold achieved by the power allocation with the optimum modulation allocation follows closely the trend of the AWGN capacity curve. Since the water-filling allocation is the optimum power allocation and the equal power allocation is near optimum at high SNR for a BF-AWGN channel, these two schemes should provide performance close to the optimum solution.

6.4.5 Numerical Results

In this section, we compare different approaches for a simple two-state channel model, and a N-state discrete channel model approximating the Rayleigh fading channel model.

Due to high computational complexity, the exact optimum solution of problem (2.11) is only available for small N. To compare the optimum solution with other approaches,

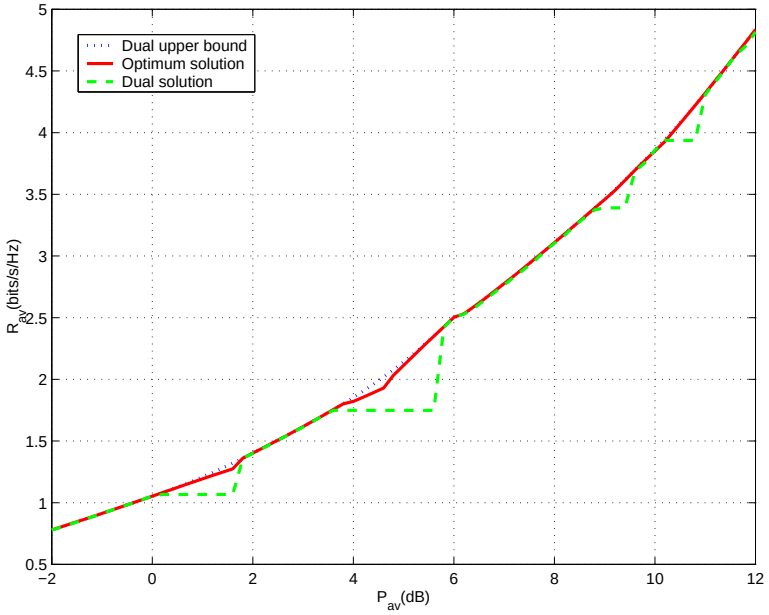


Figure 6.15: In a two-state channel: $n_1 = 1$ and $n_2 = 3$ with equal probability, compare the optimum solution with the dual solution and the dual upper bound.

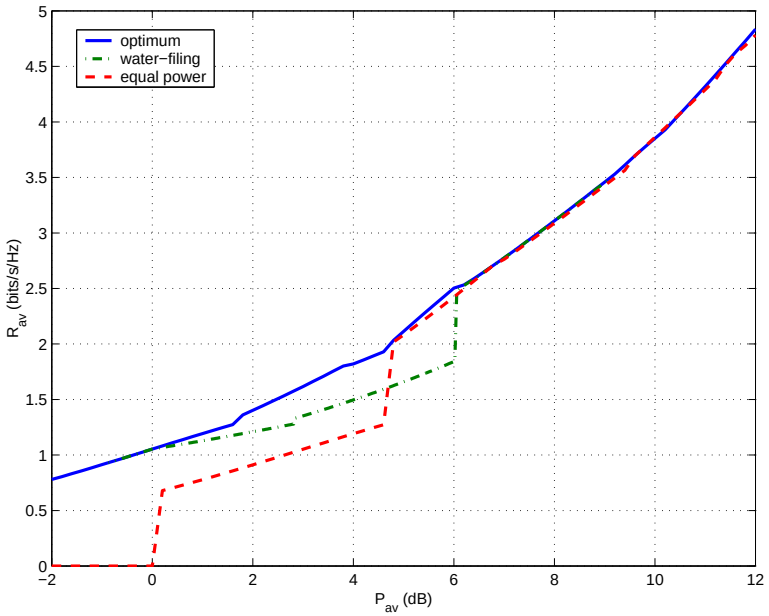


Figure 6.16: In a two-state channel: $n_1 = 1$ and $n_2 = 3$ with equal probability, compare the optimum solution with water-filling with optimum modulation and equal power allocation with optimum modulation.

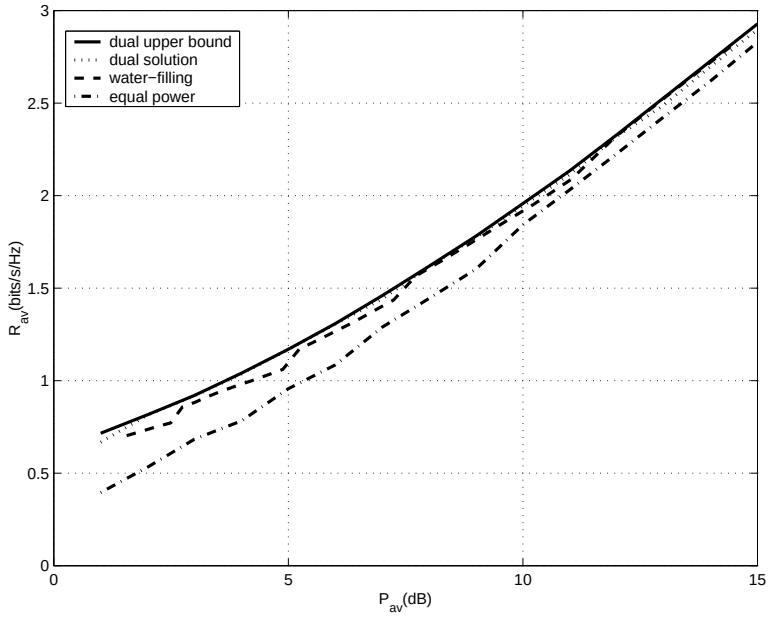


Figure 6.17: In a $N = 10$ state channel model obtained from Rayleigh fading, compare the optimum solution with the dual solution, the dual upper bound, water-filling with optimum modulation and equal power allocation with optimum modulation .

we study a simple two-state channel: $n_1 = 1$ and $n_2 = 3$ with equal probability 1/2. In Fig 6.15, the optimum solution is plotted against the dual solution and the corresponding upper bound for the two-state channel model. As can be seen, the optimum solution is between the dual solution and its upper bound. In Fig 6.16, the optimum solution is plotted against the water-filling allocation with the optimum modulation and the equal power allocation with the optimum modulation for the two-state channel model. Compared to the optimum solution, at low SNR these two simple sub-optimum schemes have some performance loss, while at high SNR they achieves a performance almost the same as the optimum solution.

We also study the performance in Rayleigh fading channel. In a Rayleigh fading channel, we assume that the channel gain has a pdf as $f(h) = exp(-h)$ for $h \geq 0$. The channel state space is divided into N regions with equal probability. Then a N-state discrete channel model is obtained by using N boundary points $0 = h_1 < h_2 < \ldots < h_N$ to represent N regions. The equivalent noise variance at state i is $n_i = 1/h_i$. Since the optimum solution is hard to obtained for large N, we use the dual solution and its

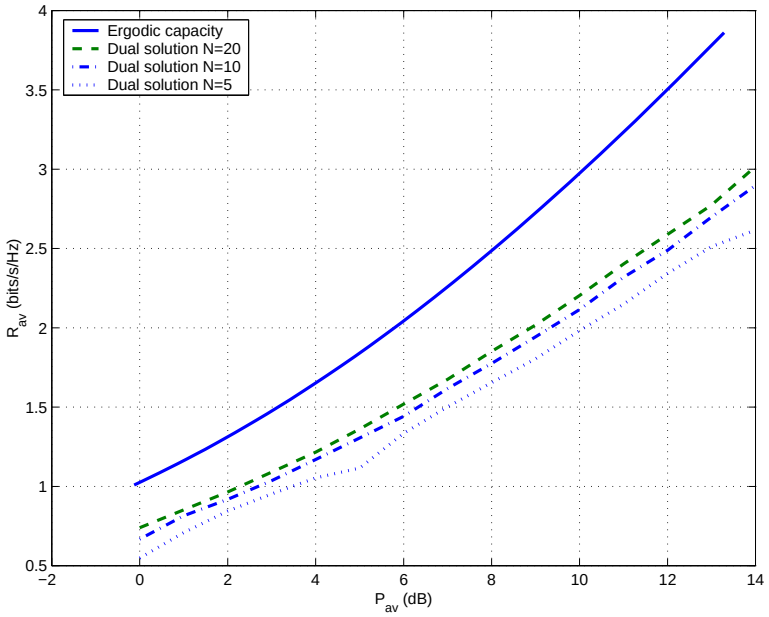


Figure 6.18: Compare the ergodic capacity of Rayleigh fading channel with the real rate performance of Turbo-BICM using dual solutions in a $N = 5, 10, 20$ state channel model obtained from Rayleigh fading.

upper bound to estimate the optimum solution. As shown in Fig 6.17, in a $N = 10$ state fading channel model the dual solution almost achieve the same rate as its upper bound. Therefore, the dual solution is almost as good as the optimum solution. The water-filling with optimum modulation and equal power allocation with optimum modulation are also presented in Fig 6.17. As we can see, these two simple schemes achieve a performance close to the dual solution.

As shown in Section 6.3.5, the rate performance of a Turbo-BICM can be estimated by shifting the rate threshold a constant gap. For a Turbo-BICM with an information sequence $K = 1000$ and code word error probability $p_e = 0.01$, the gap is almost 1 dB. In Fig 6.18, the rate performance of the adaptive Turbo-BICM is plotted by shifting the rate threshold 1 dB in $N = 5, 10, 20$ state fading channel models and compared to the ergodic capacity. As can be seen, the adaptive Turbo-BICM can achieve an average rate with $2 - 3$ dB of ergodic capacity.

6.A Algorithm for Solving Problem (6.30)

Let n_i with $i = 1, \ldots, N$ denote N possible values of the random variable n. Let f_i, m_i, and P_i denote the corresponding probability, modulation constellation size, and power assignment at n_i respectively. Problem (6.30) can be rewritten as

$$R(m_1, \ldots m_N) = \max_{P_1,\ldots,P_N} \sum_{i=1}^{N} \tilde{r}_{\mathrm{UB}} \left(\frac{P_i}{n_i}, m_i \right) f_i \tag{6.36}$$

$$\text{subject to} \quad \sum_{i=1}^{N} P_i f_i \leq P_{\mathrm{av}}$$

$$P_i \geq 0 \quad \text{for} \quad i = 1, \ldots, N\,.$$

Although rate $\tilde{r}_{\mathrm{UB}} \left(\frac{P_i}{n_i}, m_i \right)$ is not continuous in the first derivation for $P_i \geq 0$, it is concave and continuous in all derivatives within each of three regions of P_i: $0 \leq \frac{P_i}{n_i} < \eta_1(m_i)$, $\eta_1(m_i) \leq \frac{P_i}{n_i} < \eta_2(m_i)$, and $\frac{P_i}{n_i} \geq \eta_2(m_i)$. The space of $(P_1, \ldots, P_N)$ can be divided into 3^N subspace, where in each subspace P_i for all $i = 1, \ldots, N$ belongs to one of its three regions. Problem (6.36) can be decomposed into 3^N subproblem, where in each sub-problem the whole space of $(P_1, \ldots, P_N)$ is replaced by one subspace. Since each subproblem is a convex optimization problem, it can be solved using Kuhn-Tucker conditions theorem. The solution of each subproblem is a candidate solution for problem (6.36). Compare the average rates achieved by 3^N candidate solutions of subproblems, and the one who achieves the highest average rate is the optimum solution of problem (6.36).

Chapter 7

Conclusion and Future Work

7.1 Thesis Summary

In this thesis, we have studied service outage based allocation problems and adaptive turbo bit-interleaved coded modulation in fading channels.

The service outage based allocation problem explores variable rate transmission schemes and combines the concepts of ergodic capacity and outage capacity for fading channels. The ergodic capacity determines the maximum achievable rate for non real-time applications, and the outage capacity is developed for constant rate real-time applications. None of them is appropriate for variable rate multimedia applications for the next generation wireless networks. In order to address this problem, the service outage based allocation problem is proposed. A service outage occurs when the instantaneous transmission rate is smaller than the basic rate specified by the application. The service outage allocation problem is to find the optimum power allocation that maximizes the average rate subject to an outage probability constraint and an average power constraint. In this work, it is assumed that perfect channel state information is available at both transmitter and receiver.

We first derived the optimum power allocation in the class of deterministic allocation schemes for a single fading channel with a continuous channel distribution. The feasibility of this problem is related to the capacity versus outage problem. When the problem is feasible, the optimum power policy is shown to be a combination of water filling and channel inversion allocation, where the outage occurs at a set of channel states below a certain threshold. For a given pair (r_o, ϵ), as P_{av} increases, and the optimum scheme gradually changes from the on-off channel inversion allocation to the water filling allocation. We also studied the service outage based allocation problem

in terms of the energy efficiency, which is the ratio of average rate and average power. The optimum operating average power that maximizes the energy efficiency subject to the service outage constraint is obtained.

We extended the service outage based allocation problem to the general M-parallel fading channel. The results can be applied to both discrete and continuous channel distribution. The M-parallel flat fading channel model can characterize a number of diversity system, including an OFDM system with frequency selective fading and the multiple antenna signal model when the perfect channel state information is available at transmitter and singular value decomposition is employed. The allocation problem is formulated in the general class of probabilistic schemes. In a probabilistic scheme, multiple power vectors can be assigned to the same channel state vector with a conditional pdf. The class of probabilistic schemes is more general than the class of deterministic schemes. The optimum power allocation is shown to be a combination of the water-filling allocation and the basic-rate allocation, and is deterministic except at the boundary set. It can be viewed as a two-layer allocation: the first layer is the water-filling allocation, and the second layer the supplemental power allocation. The supplemental power is only allocated at channel states where the corresponding supplemental cost is below a threshold. With increasing average power, the optimum power allocation gradually changes from the minimum outage power allocation to the water-filling allocation. The resulting service outage based average rate gradually changes from the basic rate times $1 - \epsilon$ to the ergodic capacity. The service outage approach strikes a good balance between the outage probability and the average rate.

We also derived two near optimum scheme for M-parallel fading channel. The two near optimum schemes have the similar two layer structure as the optimum scheme, except that they have different outage sets. In near optimum solution I, outage occurs when the sum of supplemental power allocation is above a threshold. In near optimum solution II, outage occurs when the sum of the basic-rate power allocation is above a threshold. Near optimum allocation II also significantly reduces the computational complexity of the optimum scheme and near optimum solution I.

We have also investigated the transmission of both realtime and non realtime services

in fading channels. The optimum service outage based power allocation maximizes the average rate in such system, and it can be implemented using an adaptive channel partition scheme, where different services are transmitted in separated channels and the partition ratio varies with fading states. We also examined the fixed channel partition scheme. In a fixed channel partition scheme, the optimum power allocation in the subchannel for the realtime service is the minimum outage power allocation, where the optimum power allocation in the subchannel for the non realtime service is the water-filling allocation. The optimum fixed channel partition ratio can be computed numerically. We also examined another fixed channel partition scheme: the equal power density partition scheme, where the partition ratio can be easily obtained in a single fading channel. It is observed from the numerical results that the equal power density partition scheme is almost as good as the fixed optimum channel partition scheme, and both of them have a negligible loss compared to the optimum service outage solution when the basic rate is small.

We also studied the performance of a continuously variable-rate Turbo-BICM by changing the rate through both random puncturing and the modulation constellation size. We apply recent results on parallel channel performance of turbo code ensembles, to obtain a union-Bhattacharyya rate threshold for variable-rate Turbo-BICM based on a reliable channel region for turbo codes transmitted over parallel channels. A closed form approximation of this rate threshold is determined for an AWGN channel. This rate threshold is shown to predict well the Turbo-BICM rate performance. By applying the closed form approximation of the union-Bhattacharyya rate threshold, we designed the adaptive Turbo-BICM in a slow fading channel. The optimum power and modulation allocations are described. A dual problem solution which achieves a rate close to the optimum solution with significantly reduced computational complexity is described. Two simple schemes: water-filling with optimum modulation and equal power allocation with optimum modulation are also presented and shown to achieve a good performance. Proposed adaptive schemes are shown to achieve a rate within $2-3$ db of the ergodic capacity of a Rayleigh fading channel.

7.2 Future Directions

In this thesis, we studied the service outage based allocation problem for a single user. Similar allocation problems can be formulated and studied for multiple users. Each user may have its own basic rate and outage probability requirements. The capacity region may be identified subject to the outage constraints and average power constraints for each user. The minimum rate capacity region for multiple users, which corresponds to the service outage based problem with zero outage, was recently derived in [45]. The multi-user outage capacity region were characterized in [52]. The approaches and techniques used in these works may be helpful in deriving the multi-user service outage based capacity region.

In this work, perfect channel state information is assumed at both transmitter and receiver. The allocation problems with imperfect channel state information can also be studied. The channel state information can be estimated at the receiver and fed back to the transmitter. In general, the channel state estimates at the transmitter can be different from the channel state estimates at the receiver, and they can be modeled by their joint distribution with the true channel state information. The channel capacity with imperfect channel state information has been addressed in several works, including [17], [64] and [84].

In this work, it is assumed that there is an infinite number of packets available at the source and the transmission delay is the main concern. However, in a packet-radio network, data arrivals in bursts, which results in queuing delay. How to allocation power and rate according to the channel states, buffer length, and arrival statistics is an interesting research topic.

In our adaptive Turbo-BICM, the random puncturing is carried on both systematic bits and non-systematic bits. For sufficiently high SNR such that the coded scheme can work well, it is better to only puncture the non-systematic bits to obtain a higher rate. For a given target error probability, the uniform random puncturing scheme saturates in a lower rate than the uncoded system at high SNR, while the non-systematic random puncturing scheme approaches the uncoded scheme at high SNR. We may investigate

the performance of the non-systematic puncturing scheme, and/or a hybrid system with uniform random puncturing at low SNR and non-systematic random puncturing at high SNR.

The adaptive Turbo-BICM can also be extended to fast fading, where one codeword spans multiple fading blocks, and compared to non-adaptive scheme. As number of fading blocks increases within one codeword, one may expect that the benefit of an adaptive scheme over a non-adaptive scheme is negligible due to the sufficient diversity order within one codeword. It will be interesting to examine when it is sufficient to apply a non-adaptive scheme in a fast fading environment.

References

[1] M. Alouini and A. J. Goldsmith. An adaptive modulation scheme for simultaneous voice and data transmission over fading channels. *IEEE Journal on Selected Areas in Communications*, May 1999.

[2] A.S. Barbulescu, W. Farell, P. Gray, and M. Rice. Bandwidth-efficient turbo coding for high-speed mobile satellite communication. In *Proc. Int. Symp. Turbo Codes and Related Topics, Brest, France*, pages 119–126, September 1997.

[3] M. S. Bazaraa and C. M. Shetty. *Nonlinear programming theory and algorithms.* John Wiley and Sons, New York, 1979.

[4] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi. CDMA/HDR: A Bandwidth-Efficient High-Speed Wireless Data Service for Nomadic Users. *IEEE Communications Magazine*, 38(7), July 2000.

[5] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara. Bandwidth efficient parallel concatenated coding schemes. *Electronic lett.*, (24):2067–2069, 1995.

[6] S. Benedetto and G. Montorsi. Unveiling turbo codes: some results on parallel concatenated coding schemes. *IEEE Trans. Inform. Theory*, pages 409–428, Mar. 1996.

[7] S. Benedetto and G. Montorsi. Design of parallel cancatenated convolutional cpdes. *IEEE Trans. Commun.*, pages 591–600, May 1996.

[8] P. P. Bergmans and T. M. Cover. Cooperative broadcasting. *IEEE Transactions on Information Theory*, IT-20(3), May 1974.

[9] C. Berrou and A. Glavieux. Near optimum error correcting coding and decoding: Turbo codes. *IEEE Trans. Commun.*, (10):1261–1271, 1996.

[10] C. Berrou, A. Glavieux, and P. Thitimajshima. Near shannon limit error-correcting coding and decoding: Turbo-codes. In *Proc. IEEE International conference on communications 1993,Geneva, Switzerland*, May 1993.

[11] R. A. Berry. *Power and delay trade-offs in Fading channels.* PhD thesis, Massachusetts Institute of Technology, June 2000.

[12] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, 1995.

[13] E. Biglieri, J. Proakis, and S. Shamai. Fading channels: Information-theoretic and communications aspects. *IEEE Transactions on Information Theory*, 44(6):2619–2692, Oct. 1998.

[14] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, Feb. 2004.

[15] D. M. Burley. *Studies in optimization*. Wiley, New York, 1974.

[16] J. B. Cain, G. C. Clark, and J. M. Geist. Punctured convolutional codes of rate (n-1)/n and simplified maximum likelihood decoding. *IEEE Trans. Information Theory*.

[17] G. Caire and S. Shamai. On the capacity of sime channels with channel side information. *IEEE Transactions on Information Theory*, 45:2007–2019, Sep. 1999.

[18] G. Caire, G. Taricco, and E. Biglieri. Optimum power control over fading channels. *IEEE Transactions on Information Theory*, 45(5):1468–1489, July 1999.

[19] G. Caire, G. Taricco, and E. Biglieri. Bit-interleaved coded modulation. *IEEE Trans. Information Theory*, (3):927–946, May 1998.

[20] A. Chockalingam and Michele Zorzi. Energy efficiency of media access protocols for media data network. *IEEE Transactions on Communication*, (11), November 1998.

[21] S. G. Chua and A. J. Goldsmith. Adaptive coded modulation for fading channels. *IEEE Transactions on Communication*, May 1998.

[22] J. C. Clegg. *Calculus of Variations*. John Wiley and Sons, New York, 1967.

[23] B. E. Collins and R. L. Cruz. Transmission policies for time varying channels with average delay constraints. In *1999 Allerton Conference on Communication, Control, and Computing, Monticello, IL*, 1999.

[24] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley and Sons, New York, 1991.

[25] D. Divsalar. A simple tight bound on error probability of block codes with application to turbo codes. *TMO Progress Report 42-139*, November, 1999.

[26] D. Divsalar, H. Jin, and R. J. McEliece. Coding theorems for 'turbo-like' codes. In *1998 Allerton conference*, Spet. 1998.

[27] D. Divsalar and F. Pollara. Turbo codes for deep-space communications. *TDA Progress Report 42-120*, Feb., 1995.

[28] D. Divsalar and F. Pollara. Multiple turbo codes for deep-space communications. *TDA Progress Report 42-121*, May, 1995.

[29] D. Divsalar and F. Pollara. On the design of turbo codes. *TDA Progress Report 42-123*, Nov., 1995.

[30] Sudhir Dixit, Yile Guo, and Zoe Antoniou. Resource management and quality of service in third-generation wireless networks. *IEEE Communications Magazine*, 39(2), Feb. 2001.

[31] ETSI. General Packet Radio Service(GPRS); Mobile Station (MS)-Base Station System(BSS) Interface; Radio Link Control/Medium Access Control (RLC/MAC) Protocol (GSM 04.60). *Tech. Spec.*, 8(1), Nov. 1999.

[32] W. Feller. *An Introduction to Probability Theory and its Applications, Vol. II.* New York: Wiley, 1987.

[33] Furuskar. EDGE: enhanced dara rates for GSM and TDMA/136 evolution. *IEEE Personal Communications*, (3):56–66, June 1999.

[34] R. Gallager. *Information Theory and Reliable Communication.* John Wiley and Sons, New York, 1968.

[35] S. L. Goff, A. Glavieux, and C. Berrou. Turbo-codes and high spectral efficiency modulation. In *International Conference on Communications (ICC'94)*, pages 645–649, May 1994.

[36] S. Le Goff. Performance of bit-interleaved turbo-coded modulation on Rayleigh faidng channels. *Electronic Letters*, pages 731–733, 2000.

[37] A. J. Goldsmith and S. Chua. Variable-rate variable-power MQAM for fading channels. *IEEE Transactions on Communications*, 45(10), Oct. 1997.

[38] A. J. Goldsmith and P. Varaiya. Capacity of fading channels with channel side information. *IEEE Transactions on Information Theory*, 43(6):pp. 1986–1992, Nov. 1997.

[39] D. J. Goodman, R. A. Valenzuela, K. T. Gayliard, and B. Ramamurthi. Packet reservation multiple access for local wireless communications. *IEEE Transactions on Communications*, August 1989.

[40] I. E. Grossmann. Review of nonlinear mixed-interger and disjunctive programming techniques. *submitted June 2001/Rev. April 2002.*

[41] J. Hagenauer. Rate-compatible punctured convolutional codes (RCPC codes) and their applications. *IEEE Transactions on Communications*, (4):389–400, April 1988.

[42] S. V. Hanly and D. N. C. Tse. Multiaccess fading channels: part II: Delay-limited capacities. *IEEE Transactions on Information Theory*, 44(7):2816–2831, Nov. 1997.

[43] A. D. Ioffe and V. M. Tihomirov. *Theory of extremal problems.* North-Holland, New York, 1979.

[44] H. Jin and R.J. McEliece. Coding theorems for turbo code ensembles. *IEEE Trans. on Information Theory*, pages 1451–1461, June 2002.

[45] N. Jindal and A. Goldsmith. Capacity and optimum power allocation for fading broadcast channels with minimum rates. *IEEE Transaction on Information Theory*, pages 2895–2909, Nov. 2003.

[46] N. Jindal and A. J. Goldsmith. Capacity and optimum power allocation for fading broadcast channels with minimum rates. In *IEEE Globecom, San Antonio, TX*, volume 2, pages 25–29, Nov., 2001.

[47] G. Kuo, Amitabh Mishra, and Ramjee Prasad. Guest editorial: QoS and resource allocation in 3rd-generation wireless networks. *IEEE Communications Magazine*, 39(2), Feb. 2001.

[48] L.C.Perez, J. Seghers, and D.J.Costello. A distance spectrum interpretation of turbo codes. *IEEE Transactions on Information Theory*, pages 1698–1709, November 1996.

[49] William C. Y. Lee. *Mobile Cellular Telecommunications analog and Digital systems*. Mcgraw-hill, 1995.

[50] L. Li and A. J. Goldsmith. Minimum outage probability and optimal power allocation for fading multiple-access channels. In *Proc. International Symposium on Information Theory*, June 2000.

[51] L. Li and A. J. Goldsmith. Capacity and optimum resource allocation for fading broadcast channels: Part I: Ergodic capacity. *IEEE Transaction on Information Theory*, 47:1083–1102, March 2001.

[52] L. Li and A. J. Goldsmith. Capacity and optimum resource allocation for fading broadcast channels: Part II: Outage capacity. *IEEE Transaction on Information Theory*, 47:1103–1127, March 2001.

[53] X. Li, A. Chindapol, and J. A. Ritcey. Bit-interleaved coded modulation with iterative decoding and 8-PSK signaling. *IEEE Trans. on Communications*, (8):1250–1257, August 2002.

[54] R. Liu, P. Spasojević, and E. Soljanin. Reliable channel regions for good codes transmitted over parallel channels. *to be submitted to IEEE Trans. Inform. Theory.*

[55] R. Liu, P. Spasojević, and E. Soljanin. Punctured turbo code ensembles. In *Proc. IEEE Information Theory Workshop 2003, Paris, France*, March 2003.

[56] D. G. Luenberger. *Optimization by vector space methods*. John Wiley and Sons, New York, 1968.

[57] D. G. Luenberger. *Linear and nonlinear programming*. Addison-Wesley, Mass., 1984.

[58] J. Luo, L. Lin, R. Yates, and P. Spasojević. Service outage based power and rate allocation. *in IEEE Transaction on Information Theory*, 49(1):323–330, January 2003.

[59] J. Luo, L. Lin, R. Yates, and P. Spasojević. Service outage based power and rate allocation. In *Proceeding of Conference on Information Science and Systems, Johns Hopkins Univ., MD*, March 2001.

[60] J. Luo, R. Liu, and P. Spasojević. Adaptive turbo bit-interleaved coded modulation in slow fading. In *Proceeding of Conference on Information Science and Systems*, 2004.

[61] J. Luo, R. Liu, and P. Spasojević. Variable-rate turbo bit-interleaved coded modulation. In *submitted to IEEE International Symposium on Information Theory*, 2004.

[62] J. Luo, R. Yates, and P. Spasojević. Service outage based capacity and power allocation in parallel fading channel. In *IEEE International Symposium on Information Theory*, page 108, June 2002.

[63] J. Luo, R. Yates, and P. Spasojević. Energy efficient power allocation based on service outage. In *Proceeding of Allerton Conference on Communication, Control, and Computing, Monticello, IL*, October 2001.

[64] M. Medard. The effect upon channel capacity in wireless communications of perfect and imperfect knowlege of the channel. *IEEE Transactions on Information Theory*, 46:933–946, May 2000.

[65] S. Nanda, K. Balachandran, and S. Kumar. Adaptation techniques in wireless packet data services. *IEEE Communications Magazine*, 38(1), Jan. 2000.

[66] R. Negi, M. Charikar, and J. Cioffi. Minimum outage transmission over fading channel with delay constraint. In *Proc. of the IEEE International Communication Conference(ICC'99)*, pages 282–286, June 2000.

[67] G. O. Okikiolu. *Aspects of the theory of bounded integral operators in L^P space.* Academic Press, New York, 1971.

[68] L. H. Ozarow, S. Shamai, and A. D. Wyner. Information theoretic considerations for cellular mobile radio. *IEEE Transactions on Vehicular Technology*, 43(2):359–378, May 1994.

[69] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C.* Camberidge University Press, New York, 1992.

[70] J. G. Proakis. *Digital Communications.* Mc Graw Hill, 1995.

[71] Xiaoxin Qiu and Kapil Chawla. On the performance of adaptive modulation in cellular systems. *IEEE Transactions on Communication*, (6), June 1999.

[72] D. Rajan, A. Sabharwal, and B. Aazhang. Delay-bounded packet scheduling of bursty traffic over wireless channels. *IEEE Transactions on Information Theory*, submitted in November 2001.

[73] Theodore S. Rappaport. *Wireless Communications Principles and Practice.* Prentice Hall, 1997.

[74] P. Roberton and T. Woerz. Bandwidth-efficient turbo trellis-coded modulation using punctured component codes. *IEEE Journal on Selected Areas in Communications*, (2):206–218, Feb. 1998.

[75] S. Shamai and I. Sason. Variations on the gallager bounds, connections and applications. *IEEE Trans. on Information Theory*, (12):3029 – 3051, December 2002.

[76] E. Soljanin, R. Liu, and P. Spasojević. Hybrid ARQ with random transmission assignments. In *DIMACS Workshop on Network Information Theory*, Piscataway, NJ, March 17–19, 2003.

[77] I. E. Telatar and R. G. Gallager. Combining queueing theory with information theory for multiaccess. *IEEE Journal on Selected Areas in Communications*, August 1995.

[78] TIA/EIA/IS-95-A. *Mobile station-base station compatibility standard for dual-mode wideband spread spectrum cellular system.* 1993.

[79] D. Tse and S. Hanly. Multi-access fading channels: Part i: Polymatroid structure, optimal resource allocation and throughput capacities. *IEEE Transactions on Information Theory*, (7):2796–2815, Nov. 1998.

[80] G. Ungerboeck. Channel coding with multilevel/phase signals. *IEEE Trans. On Information Theory, vol=IT-28, pages=56-67, year=Jan. 1982.*

[81] S. Verdú. *Multiuser Detection.* Cambridge University Press, New York, 1998.

[82] S. Verdú. A general formula for channel capacity. *IEEE Transactions on Information Theory*, 40(4):1147–1157, July 1994.

[83] S. Verdú. On channel capacity per unit cost. *IEEE Transactions on Information Theory*, 36(5):1019–1030, Sept. 1990.

[84] H. Viswananthan. Capacity of markov channels with receiver csi and delayed feedback. *IEEE Transactions on Information Theory*, 45:761–771, March 1999.

[85] N. Wilson, R. Genesh, K. Joseph, and D. Raychaudhuri. Packet CDMA versus dynamic TDMA for multiple access in an integrated voice/data PCN. *IEEE Journal on Selected Areas in Communications*, pages 870–884, August 1993.

[86] G. Wu, K. Mukumoto, and A. Fukud. Analysis of an integrated voice and data transmission system using packet reservation multiple access. *IEEE Transaction on Vehicular Technology*, pages 289–297, May 1994.

[87] E. Zehavi. 8-PSK trellis codes for a Rayleigh channel. *IEEE Trans. on Commun.*, (5):873–884, May 1992.

[88] K. Zhang and K. Pahlavan. An integrated voice/data system for mobile indoor radio networks. *IEEE Transactions on Vehicular Technology*, Feb. 1990.

[89] M. Zorzi and R. R. Rao. Energy constrained error control for wireless channels. *IEEE Personal Communication Magzine*, (12), December 1997.