ADAPTIVE TRANSMISSION IN FADING ENVIRONMENT

BY LANG LIN

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

Professor Roy D. Yates and Professor Predrag Spasojević

and approved by

New Brunswick, New Jersey

January, 2004

ABSTRACT OF THE DISSERTATION

Adaptive Transmission in Fading Environment

by Lang Lin

Dissertation Director: Professor Roy D. Yates and Professor Predrag Spasojević

Adaptive transmission that varies the parameters at the transmitter using channel state information is a promising technique for improving bandwidth and/or power efficiency in time-varying wireless environments. To understand the full potential of adaptive transmission, the achievable throughput of adaptive transmission systems with practical constraints needs to be determined.

In this thesis, we formulate and analyze power constrained average reliable throughput maximization with a finite number of code rates/power levels and with channel uncertainty. The focus is on communication over slow fading channels where the channel does not change over the duration of a codeword. Average reliable throughput is based on the concept of information outage, which can be related to the error performance of advanced coding designs.

For an adaptive transmission system limited by encoding/decoding only with a finite choices of code rates, it is necessary to determine the finite set of code rates that enables the system to achieve maximum throughput. If the transmitted power also varies within a finite set of values, optimum designs are closely approximated by designs found by an iterative algorithm that maximizes the throughput by alternatively varying the sets of power levels and code rates. Numerical evaluations show that systems with ten

power levels and ten code rates can achieve a performance within one dB of the ergodic capacity for several fading channels.

If the transmitted power can vary continuously, throughput maximizing designs can be found by a one-parameter line search algorithm. In this case, the required number of code rates can be reduced substantially. In addition, throughput maximizing designs for an arbitrary code rate set can be obtained analytically.

When perfect channel state information is not available at the transmitter, both the transmitted power and the code rate must be adjusted according to channel estimates. If the transmitted power and the code rate vary continuously, throughput maximizing designs must simultaneously satisfy two non-linear criteria and, thus, have an ordered structure. We propose an algorithm which finds either good or optimum designs. However, for a finite number of code rates, the throughput maximization can be solved in the case of a constant outage constraint.

Numerical results show that carefully designed adaptive transmission systems with a small number of code rates (power levels) or with reasonably good channel state information can achieve throughput values close to ergodic capacity.

Acknowledgements

At the end of this six-year journey of doctorate degree study, there are many people that I am indebted to and thank for.

First, I would like to express my gratitude to Prof. Roy D. Yates and Prof. Predrag Spasojević. Without their support, encouragement, and guidance, this thesis would not be possible. As their student, I am really appreciative of their insight and elegant style in research.

I would like to express my thanks to my PhD committee members, Prof. Dipankar Raychaudhuri and Prof. Christopher Rose. I would like to thank Prof. Stuart C. Schwartz of Princeton University for being the external examiner in the committee. I would also like to thank Prof. David Goodman of Polytechnic University and Dr. Jack Holtzman of Qualcomm for letting me join the WINLAB in 1997. The past six years are truly invaluable in my life.

I would like to thank all my friends and colleagues in Rutgers for making my stay colorful and enjoyable.

Dedication

To My Parents

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

For wireless data communication systems [2, 4, 5, 29], adaptive transmission is a communication strategy that varies the parameters at the transmitter according to channel state information to improve both bandwidth and power efficiency in communicating over time-varying fading channels [28, 34, 35]. In practice, a number of parameters can be adapted, including transmitted power level [1, 25], symbol transmission rate [14], modulation formats [21], channel coding strategies [20], or any combination of the above [3, 23, 27, 33, 49, 58, 59].

In this thesis, the primary goal is to explore the maximum throughput that adaptive transmission can achieve in a fading environment. When considering the maximum achievable throughput within a given bandwidth, all of the adaptive transmission techniques can be reduced to the case of adapting only two primary parameters: the code rate and the transmission power level. Thus, the discussion of the ultimate gain of adaptive transmission falls into the regime of information theory.

One of the earliest works on improving spectrum efficiency of communications over time varying channels by employing adaptive transmission is [53]. In [53], Shannon demonstrated an example where adaptive transmission with channel state information can achieve higher throughput than non-adaptive transmission. In addition, for time-varying channels with a finite number of channel states and input/output symbols, Shannon obtained the ergodic capacity for the case where the perfect CSI is available at the transmitter (CSIT) but no CSI at the receiver (CSIR). Moreover, in [53], Shannon stated without proof an explicit ergodic capacity formula corresponding to the scenario with both perfect CSIT and CSIR.

Specifically, let the channel state be a stationary and ergodic stochastic process represented by $S(t)$, which assumes values over a finite set $\mathcal{S} = \{s_i, i = 0, 1, \ldots, N_s - 1\}$, where N_s is the cardinality of $\mathcal{S}$. Let P_i be the probability mass function of a particular channel state s_i and C_i be the corresponding capacity. Then, the overall capacity is

$$C = \sum_{i=0}^{N_s-1} C_i P_i. \tag{1.1}$$

A proof can be found in [60].

Consider a unit bandwidth fading channel modeled by

$$Y = \sqrt{S}X + W, \tag{1.2}$$

where S, X, Y, and W are the channel state (channel fading gain), the transmitted signal, the received signal, and a circularly symmetric additive white Gaussian noise (AWGN) with zero mean and variance N_0. The channel state S is a real random variable of unit mean with probability density function (PDF) $f(s)$, cumulative distribution function (CDF) $F(s)$, and domain $\mathcal{S} = \{s | s \geq 0\}$. Given a policy with transmitted power $p(s)$ corresponding to a particular channel state s, the maximum mutual information is

$$R(sp(s)) = \log\left(1 + \frac{sp(s)}{N_0}\right), \tag{1.3}$$

and (1.1) implies

$$C(p(s)) = \int_{s\in\mathcal{S}} R(sp(s)) f(s)\, ds, \tag{1.4}$$

where $C(p(s))$ is the capacity corresponding to a particular power allocation scheme. For systems with average power $\overline{p}$, the ergodic capacity is

$$C = \max_{p(s):\ \int_{s\in\mathcal{S}} p(s)f(s)\,ds \leq \overline{p}} C(p(s)). \tag{1.5}$$

In [22], it was shown that the ergodic capacity can be achieved by a multiplexing strategy where messages are coded with a variable code rate and codewords at rate $r(s)$ are transmitted with a transmitted power $p(s)$ when the channel state is s. In order to achieve the ergodic capacity, the multiplexing strategy implies that the transmitted code rate $r(s)$ must be the same as $R(sp(s))$. Therefore, it can be shown that the optimum $p(s)$ is in the following form [22],

$$p(s) = N_0 \left(\frac{1}{s_0} - \frac{1}{s} \right)^+, \tag{1.6}$$

where $(\cdot)^+$ denotes the operator $\max(\cdot, 0)$ and s_0 is a positive *cutoff* value such that

$$\int_{s_0}^{\infty} p(s) f(s)\, ds = \overline{p}. \tag{1.7}$$

For time varying systems, the optimum power allocation in (1.6) has the interpretation of water-filling in time. In addition, it requires that the system allows for adapting the transmitted power and code rate to an infinite number of channel states.

In [10], it is shown that varying the transmitted power is sufficient to achieve the ergodic capacity when the codeword is longer than the time required for the channel to experience ergodicity. In this case, messages are encoded with a code rate equal to the ergodic capacity and codeword symbols are transmitted with variable power. Even though such an approach simplifies the overall design by dropping the need for code rate adaptation, it may incur a severe decoding delay and, thus, unacceptable quality of service. Therefore, within the scope of this thesis, we allow for varying both the transmitted power and code rate.

In [22], the authors demonstrated through numerical evaluation that there is a very large throughput improvement if adaptive transmission is used. However, in [22], an optimum adaptive transmission system that achieves the ergodic capacity requires knowledge of the current channel state at the transmitter. In addition, the transmission parameters, e.g., the transmitted power and/or code rate assignments must vary continuously. Both of these requirements have been widely adopted in further work on information theoretic aspects of communication over fading channels [10, 57]. Unfortunately, these requirements are very stringent for practical system implementations.

When the parameter space of the transmitted power and/or code rate is a finite set, it is possible that the transmitted code rate $r(s)$ does not match the mutual information $R(sp(s))$. For instance, given a fixed power assignment $p(s) = \overline{p}$, since s can be any non-negative real number, in order to have $r(s) = R(sp(s))$ for all $s \in \mathcal{S}$, arbitrarily high code rates $r(s)$ must be employed. When $r(s) > R(sp(s))$, an information outage event occurs.

The information outage is an intrinsic characteristic of communication over fading channels with a decoding delay constraint [24, 45] or, alternatively, with codewords not long enough to experience ergodic fading. For delay-limited cases, the strict sense Shannon capacity is zero [45]. The fundamental reason is that, given a codeword with any positive code rate and transmitted at any positive power level, and a channel state which can be arbitrarily close to zero, the probability of the event that $r(s) > R(sp(s))$ is non-zero. Thus, the maximum achievable rate for reliable communication is zero.

During an outage, a transmission is not considered reliable and, thus, it is frequently convenient to assume that the transmitted data can be ignored [8]. This assumption leads to the *capacity versus outage* problem which focuses on the tradeoff between the outage probability and the supportable rate; see, for example, [10, 26, 51, 52, 55]. The practice of ignoring data received during an outage is supported by the fact that the outage probability matches well the error probability of actual codes [30, 42]. In fact, there is a large class of channel coding schemes achieving the performance predicted by the channel capacity [6, 18]. Consequently, we characterize the performance of a system design based on the concept of *average reliable throughput* (ART), defined as the average data rate **assuming zero rate when the channel is in outage**.

The capacity-versus-outage approach or the capacity distribution approach can be generalized as follows. For a compound channel characterized by θ, the capacity is denoted by C_θ. Given any rate r, we can define a set Θ_r that is the largest of all sets $\left\{\theta \middle| C_\theta \geq r\right\}$. Then, the outage probability is $P_{\text{out}} = \Pr[\theta \notin \Theta_r]$.

Consider the simple case of a flat Rayleigh fading channel $f(s) = e^{-s}$, with the perfect CSIR only. For this compound channel, if $p(s) = \overline{p}$, the instantaneous channel capacity with channel state s is given by

$$C(s) = R(sp(s)) = R(s\overline{p}). \tag{1.8}$$

Given a rate r, the outage probability is

$$P_{\text{out}} = \Pr[C(s) < r] = \Pr\left[\log\left(1 + \frac{s\overline{p}}{N_0}\right) < r\right] = 1 - \exp\left(-\frac{N_0}{s\overline{p}}(e^r - 1)\right). \tag{1.9}$$

Thus, $P_{\text{out}} = 0$ only if $r = 0$. Therefore, the capacity in the Shannon's sense is zero.

Channel estimation is another important factor limiting the performance of adaptive transmission systems. In practice, the channel state can be estimated by a wide range of estimation techniques including pilot-aided [13], joint pilot-aided and data detection [50], and blind [61]. A survey of the channel estimation techniques can be found in [17]. Due to complexity and timing constraints, in adaptive transmission systems, it may not always be possible to have perfect CSIT.

Historically, it has been a challenging problem to determine the capacity without perfect CSI. In [9, 48], the ergodic capacity for systems with a more general model of CSI has been studied. Reference [48] generalizes Shannon's original result [53] to the case with both uncertain CSIT and CSIR. In [9], a general ergodic capacity formula for the case when the channel state process has memory and a generalization of several results for the additive white Gaussian noise (AWGN) channel with fading are provided. In [44], the loss in the mutual information due to the imperfect CSIR is studied for the general fading channel models with inter-symbol interference, time-variation, multiple access, and no feedback (thus, no adaptive transmission). Other relevant references can be found in the survey [8].

As we noted earlier, the ergodic capacity may impose an undesirable delay. When there is a delay requirement, since the code rate and the transmitted power level are based on channel estimates, it is possible that the instantaneous mutual information corresponding to the channel state is less than the assigned code rate and an information outage occurs. Therefore, we can again characterize the performance of a system design by ART.

1.2 Road Map

In this thesis, we try to apply the concept of information outage to provide some preliminary quantitative answers for the following questions in slow fading channels.

1. What is the performance of an adaptive transmission system with a finite number of code rates and/or power levels?

2. What is the performance of an adaptive transmission system with channel state uncertainty?

In Chapter 2, the throughput maximization of an adaptive transmission system with a finite number of transmitted power levels and code rates for communication over slow fading channels is analyzed based on the concept of information outage. Properties of throughput maximizing policies lead to an iterative algorithm that yields good system designs. Numerical results show that carefully designed discrete adaptive transmission systems with a small number of power levels and code rates can achieve throughput values close to ergodic capacity for several fading channel models.

In Chapter 3, an adaptive transmission system that supports a discrete set of code rates and continuously variable transmit power is examined. We find that for a large class of fading channels, the maximum throughput can be obtained by a line search over a single parameter. Numerical results show that in a Rayleigh fading channel, there is only a gap of 1 dB between the ergodic capacity and the throughput of a 2-rate adaptive transmission system when the throughput is less than 4 bits/sec/Hz.

In Chapter 4, for adaptive systems with a finite set of code rates and imperfect CSIT, we formulate the throughput maximization with both an average power constraint and an information outage constraint. It is verified that, for the optimal transmission policy, the transmission only needs to adapt to a sufficient statistic for the channel state. For a Rayleigh fading channel with a simple training scheme, numerical results show that, with a reasonable amount of training and a small set of code rates, the adaptive transmission can achieve a performance close to the ergodic capacity.

In Chapter 5, based on the concept of information outage, the throughput optimization of adaptive transmission with channel state uncertainty is formulated for slow fading channels. Even without a constraint on the channel state distribution and the approach to channel estimation, the optimum solution of the problem exhibits a well ordered structure. We propose an algorithm to find either good or optimum designs. For a Rayleigh fading channel with a simple training scheme, numerical results show that with a reasonable amount of training, the adaptive transmission can achieve a performance very close to the ergodic capacity.

Chapter 2

Adaptive Transmission with a Finite Set of Code Rates and Power Levels

2.1 Introduction

In [22], an optimum adaptive transmission system achieving ergodic capacity requires knowledge of the current channel state at the transmitter. In addition, both transmitted power and code rate assignments must adapt continuously to changes in the channel state. Unfortunately, these requirements are hard to satisfy in practice.

To accommodate these practical constraints, in [21], the spectral efficiency of an adaptive transmission with a variable-rate and variable-power M-ary quadrature amplitude modulation (MQAM) scheme is studied. It is shown that when perfect CSIT and CSIR is available, for Rayleigh and lognormal fading channels [54], a simple suboptimum adaptive MQAM design with only five or six different QAM constellations can achieve a performance within 1-2 dB of the maximum spectral efficiency that requires unlimited constellation sets. In [20], an adaptive trellis coded modulation (TCM) based MQAM constellations is proposed and it has an effective coding gain of 3 dB relative to the uncoded MQAM. More recently, an adaptive transmission design based on outdated channel information and either MQAM or TCM is proposed in [19].

Instead of directly involving any modulation or coding design, this chapter examines achievable throughput of a discrete adaptive system with a finite number of power levels and code rates. In this problem, the challenge is twofold. First, the throughput maximization problem for such a system is hard to formulate, especially when channel coding is involved. Second, the problem requires joint optimization of multiple parameters.

We assume a slow multiplicative fading environment where the channel is constant during the transmission of a codeword. Although the receiver is assumed to have knowledge of the exact channel state, the channel state space is partitioned into a finite number of quantization intervals and the transmitter is assumed to know the current quantization interval only. Based on the current interval, the transmitter selects a corresponding pair from a finite set of code rate and power level pairs (rate-power pairs) to encode and transmit the information message, respectively. It is assumed that there are multiple codebooks and each is used for its corresponding quantization interval. Since each codeword experiences an additive white Gaussian noise (AWGN) channel, random Gaussian codes are employed. With an infinite number of intervals, the discrete model coincides with the continuous adaptation of [22] for the slow fading channel model assumed in this chapter.

For the proposed discrete adaptive system, it is possible that the instantaneous mutual information corresponding to a channel state is less than the assigned code rate. Consequently, we characterize the performance of a system design based on the concept of *average reliable throughput* (ART), defined as the average data rate **assuming zero rate when the channel is in outage**. Henceforth, the central topic is maximizing ART of a discrete adaptive system communicating over a slow fading channel.

The reader may question the necessity of introducing information outage in a discrete adaptive transmission system. Since the throughput during an outage is not counted towards ART, any transmission (with non-zero power) during an outage is a waste and, intuitively, should be avoided. It will be shown that zero outage can be achieved if the zero-rate zero-power pair is included in the set of rate-power pairs.

One may observe that if the zero-rate zero-power pair is not included, it can be added to the original set of rate-power pairs by introducing little additional complexity in the circuitry of the codec and the power amplifier. Therefore, such a change seems advantageous. However, introducing a new rate-power pair increases the number of quantization intervals and, consequently, may incur significant changes in the system design.

One example of a system change is in the feedback channel design required to communicate a different number of quantization interval indices in frequency division duplex (FDD) systems. A second example of a system design change is in the required quality of channel estimation. Indeed, we assumed perfect channel state information (CSI) at the receiver and current interval information at the transmitter. For FDD systems, one explanation for such an assumption is that the estimator at the receiver obtains perfect CSI instantly and forwards the current interval information to the transmitter through a noiseless feedback channel instantly. This model is also appropriate when there are two different estimators at the receiver. One fast estimator uses less received data and is, thus, less accurate. It can only determine the current interval instantly. The other estimator uses significantly more data and can determine the exact channel state reliably for decoding. In such a system, varying the number of intervals changes the estimation resolution requirement for the fast estimator. Thus, systems with different numbers of intervals cannot be compared with complete fairness. Similar arguments apply to time division duplex (TDD) designs.

Given these system design considerations, we only optimize over designs with the same number of intervals and, furthermore, we do not assume automatic inclusion of the zero-rate zero-power pair in a design. In this scenario, we will see that designs that include the zero-rate zero-power pair to achieve zero outage generally do not achieve the maximum ART. In fact, zero outage represents an additional constraint instead of a property of the optimal solution. For instance, the zero-outage design for a fixed power and rate transmission, which is a discrete adaptive transmission design only with one rate/power interval, simply does not transmit at all. In this context, our work studies the necessity of information outage within the framework of an adaptive system.

Following the formulation of the discrete adaptive system design problem, we focus on exploring optimum (ART-maximizing) policies where a *policy* is defined as an ensemble of the channel state space partition and corresponding power and rate allocations. In this work, we will show that, for an optimum policy, an outage can occur only for a set of channel states within the first quantization interval. In other quantization intervals, the assigned rate supports the worst channel state of that interval. An optimum policy can be uniquely characterized by the channel state space partition, first interval rate assignment, and the power level assignments. The optimum power level allocation has a water-filling character. When the number of levels approaches infinity, reasonably designed discrete adaptive transmission schemes with a water-filling power assignment and an equal probability partition of the channel state space can achieve the ergodic capacity.

We emphasize that finding an optimum policy is still a challenging non-convex optimization problem. Since brute force search over a space of policies at a high resolution has complexity that increases exponentially with the number of quantization levels, we present an iterative algorithm that numerically evaluates the lower bounds of the maximum ART and determines corresponding sub-optimum policies. The computational complexity of the algorithm is linear in the number of levels and the achieved rates can be very close to the true maximum ART.

2.2 System Model and Problem Formulation

We consider a multiplicative flat fading channel model similar to that in [22]. The complex received signal

$$Y = \sqrt{S}X + W, \tag{2.1}$$

where S is the channel (fading) state, X is the complex transmitted signal, and W is a circularly symmetric additive white Gaussian noise (AWGN) with variance N_0. The channel state S is a real random variable of unit mean with a probability density function (PDF) $f(s)$, a cumulative distribution function (CDF) $F(s)$, and a domain $\mathcal{S} = \{s | s \geq 0\}$. It is also assumed that the fading is sufficiently *slow* that the channel state is constant during the transmission of a codeword.

The proposed adaptive transmission system quantizes any channel state s to one of L levels $v_0 < v_1 < \ldots < v_{L-1}$, where $v_0 = 0$. The L channel state quantization intervals are denoted by $\mathcal{U}_l = [v_l, v_{l+1})$ for $l = 0, \ldots, L-1$, where $v_L = \infty$. Note that the set $\{\mathcal{U}_0, \ldots, \mathcal{U}_{L-1}\}$ partitions $\mathcal{S}$. When the channel state $s \in \mathcal{U}_l$, the encoder at the transmitter generates codewords of a code rate r_l and the codewords are transmitted at a power level $p_l = E\left\{|X|^2 \,|s \in \mathcal{U}_l\right\}$, where $E\{\cdot\}$ denotes expectation.

Since (2.1) describes an AWGN channel for any given $s \in \mathcal{S}$, the corresponding maximum mutual information is given by $\log(1 + p_l s/N_0)$. Let $R(\phi) \stackrel{\triangle}{=} \log(1 + \phi/N_0)$, the maximum mutual information associated with any state $s \in \mathcal{U}_l$ is $R(p_l s)$.

An *adaptive transmission policy* is defined by a set of quantization levels $\{v_l\}$ and the corresponding set of power and rate assignment pairs $\{(p_l, r_l)\}$, or equivalently, by the triple of L by 1 vectors

$$(\mathbf{p}, \mathbf{v}, \mathbf{r}) = ([p_0, \ldots, p_{L-1}]^\top, [v_0, \ldots, v_{L-1}]^\top, [r_0, \ldots, r_{L-1}]^\top). \tag{2.2}$$

Such a policy is also referred to as an L-level policy when policies with different L's are addressed. Throughout this chapter, unless otherwise noted, $(\mathbf{p}, \mathbf{v}, \mathbf{r})$ indicates an L-level policy even though it seems more appealing to put parameter L together with $(\mathbf{p}, \mathbf{v}, \mathbf{r})$ to specify an L-level policy.

An illustration of the power assignment of an arbitrary policy is in Fig. 2.1, where the heights of boxes indicate the power levels assigned to their respective quantization intervals. Within each interval $\mathcal{U}_l$, the shaded region indicates the *outage interval* where the channel state does not support reliable communication with the assigned rate r_l given implicitly through q_l, as explained in the next section [1].

For any $s \in \mathcal{U}_l$, given a power level and code rate assignment pair (p_l, r_l), it is guaranteed that the information will be successfully received only if $R(p_l s) \geq r_l$. Hence, following the established outage probability definition [55], we define the conditional outage probability as

$$P_{\text{out}}(r_l, p_l|l) = \Pr\left[R(p_l s) < r_l | s \in \mathcal{U}_l\right]. \tag{2.3}$$

Given a policy $(\mathbf{p}, \mathbf{v}, \mathbf{r})$, the *average reliable throughput* (ART) is

$$R_L(\mathbf{p}, \mathbf{v}, \mathbf{r}) = \sum_{l=0}^{L-1} \Pr\left[s \in \mathcal{U}_l\right]\left[1 - P_{\text{out}}(r_l, p_l|l)\right] r_l. \tag{2.4}$$

Note that this definition hinges on the assumption that no information is successfully received during an outage. Here, we adopt the following convention: $p_l = 0$ implies that

[1] $\mathbf{q} = [q_0, \ldots, q_{L-1}]^\top$ is defined in the next section, where it is shown that $(\mathbf{p}, \mathbf{v}, \mathbf{q})$ is sufficient to specify a policy of interest. So, the illustration shows the full detail of a policy.

$r_l = 0$ and, consequently, there is no decoding error and no outage when no transmission is attempted. Let $F(s_1, s_2) \stackrel{\triangle}{=} F(s_2) - F(s_1) = \Pr[s_1 \leq S < s_2]$, the *average transmitted power* of the policy $(\mathbf{p}, \mathbf{v}, \mathbf{r})$ is

$$\rho(\mathbf{p}, \mathbf{v}, \mathbf{r}) = \sum_{l=0}^{L-1} F(v_l, v_{l+1}) p_l. \tag{2.5}$$

If the average transmitted power is upper bounded by $\overline{p}$, the *set of feasible L-level transmission policies* is

$$\pi_L(\overline{p}) = \{(\mathbf{p}, \mathbf{v}, \mathbf{r}) |\, \rho(\mathbf{p}, \mathbf{v}, \mathbf{r}) \leq \overline{p}\}. \tag{2.6}$$

We define the maximum average reliable throughput over all L-level policies to be

$$C_L = \max_{(\mathbf{p}, \mathbf{v}, \mathbf{r}) \in \pi_L(\overline{p})} R_L(\mathbf{p}, \mathbf{v}, \mathbf{r}). \tag{2.7}$$

A corresponding ART-maximizing policy is referred to as an *optimum policy* $(\mathbf{p}^*, \mathbf{v}^*, \mathbf{r}^*)$.

2.3 Properties of Optimum Policies

In this section, we present sketches of power allocations as a function of the channel state and illustrate a number of useful properties of optimum policies. These properties are helpful in simplifying the optimization problem (2.7).

It is intuitive that any outage interval (shaded interval in Fig. 2.1) is contiguous within its respective quantization interval $\mathcal{U}_l$ and would include the left end point v_l of $\mathcal{U}_l$. Specifically, if a channel state $s_1 \in \mathcal{U}_l$ does not support reliable communication at rate r_l, no channel state $s < s_1$ in $\mathcal{U}_l$ will support reliable communication at rate r_l. Thus, there can be at most L outage intervals for an L-level policy. In the following, we develop an implicit characterization of the assigned rates for an optimal policy.

Lemma 1 *For an optimum policy $(\mathbf{p}^*, \mathbf{v}^*, \mathbf{r}^*)$, we have that*

$$r_l^* \in \left[\, R(p_l^* v_l^*), R(p_l^* v_{l+1}^*) \,\right). \tag{2.8}$$

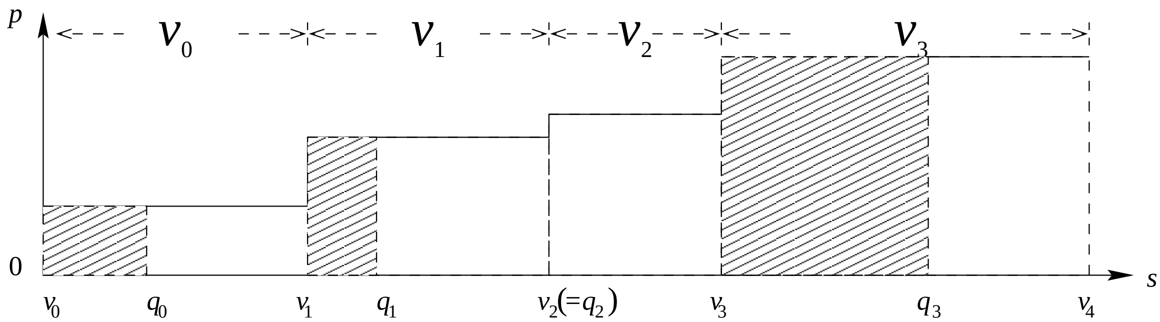

Figure 2.1: Illustration of an arbitrary power assignment. The rate assignment is implicitly shown through q_l. The shaded regions correspond to channel outage intervals.

This proof, as well as proofs of other lemmas and theorems, can be found in Appendix.

Without loss of generality, we assume that any policy of interest $(\mathbf{p}, \mathbf{v}, \mathbf{r})$ also satisfies the condition (2.8). Consequently, for any policy $(\mathbf{p}, \mathbf{v}, \mathbf{r})$ for which (2.8) holds and for any quantization interval $\mathcal{U}_l$, strict monotonicity of $R(\cdot)$ implies that we can find a unique channel state $q_l \in \mathcal{U}_l$ such that $r_l = R(p_l q_l)$. This defines a one to one mapping between channel states $\{q_0, \ldots, q_{L-1}\}$ and the respective rate assignments $\{r_0, \ldots, r_{L-1}\}$ for a given power policy. In particular, q_l is the worst channel state in $\mathcal{U}_l$ that still allows for reliable communication at a rate r_l. Therefore, it will be more convenient to redefine transmission policies of interest as vector triples $(\mathbf{p}, \mathbf{v}, \mathbf{q})$. Accordingly, such a policy assignment can be illustrated by using a plot such as Fig. 2.1, where all vector parameters are depicted.

With the introduction of $\mathbf{q}$, the conditional outage probability in (2.3) can be interpreted as the probability that $s \in \mathcal{U}_l$ is worse than the worst reliable channel q_l, i.e.,

$$P_{\text{out}}(r_l, p_l|l) = \Pr\left[R(p_l s) < R(p_l q_l) | s \in \mathcal{U}_l\right] \tag{2.9}$$

$$= F(v_l, q_l)/F(v_l, v_{l+1}). \tag{2.10}$$

With (2.4) and (2.10), the corresponding ART can be expressed in terms of the vector $\mathbf{q}$ as

$$R_L(\mathbf{p}, \mathbf{v}, \mathbf{q}) = \sum_{l=0}^{L-1} F(q_l, v_{l+1}) R(p_l q_l). \tag{2.11}$$

The next lemma, which follows from the monotonicity of $R(p_l q_l)$ with respect to p_l, says that an optimum policy meets the average transmitted power constraint with equality.

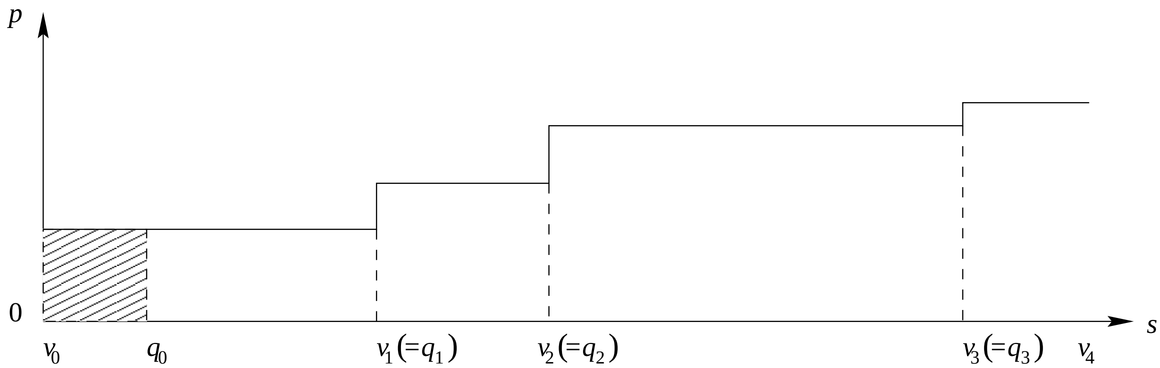

Figure 2.2: The policy which achieves C_L will have $q_1 = v_1$ and $q_3 = v_3$.

Lemma 2 *For an optimum policy* $(\mathbf{p}^*, \mathbf{v}^*, \mathbf{q}^*)$*, we have that*

$$\rho(\mathbf{p}^*, \mathbf{v}^*, \mathbf{q}^*) = \sum_{l=0}^{L} F(v_l^*, v_{l+1}^*) p_l^* = \overline{p}. \tag{2.12}$$

As clear from (2.11) and Fig. 2.1, an outage interval does not contribute to the overall average rate. Thus, one could intuitively assume that an optimum policy should minimize such intervals. An extreme case is $q_l = v_l$, which implies that there is no outage interval.

Theorem 1 *Given an arbitrary policy* $(\mathbf{p}, \mathbf{v}, \mathbf{q})$*, there exists a policy* $(\mathbf{p}', \mathbf{v}', \mathbf{q})$ *such that* $v_l' = q_l$ *for all* $l > 0$ *and* $R_L(\mathbf{p}', \mathbf{v}', \mathbf{q}) \geq R_L(\mathbf{p}, \mathbf{v}, \mathbf{q})$.

In the general case of arbitrary power levels, Theorem 1 is due to the concavity and monotonicity of $R(\cdot)$. In the special case when a policy has an increasing power allocation as depicted in Fig. 2.2, it is relatively simple to demonstrate Theorem 1. As noted earlier from (2.11) and Fig. 2.1, if $v_{l+1} < q_{l+1}$, then channel states s in the outage interval $[v_{l+1}, q_{l+1})$ do not contribute towards ART. Thus, we can increase v_{l+1} to eliminate the outage interval without changing the value of p_l. In this case, ART increases, while the average transmitted power does not increase. Since this procedure can be applied to any interval $l > 0$, for an optimum policy, an outage can occur only in the $\mathcal{U}_0$ interval.

Theorem 1 shows that policies of interest are fully determined by vectors $\mathbf{p}$ and $\mathbf{q}$

with the quantization intervals given by

$$\mathcal{U}_l = \begin{cases} [0, q_1) & l = 0 \\ [q_l, q_{l+1}) & 1 \leq l \leq L-1 \end{cases}, \tag{2.13}$$

where we define $q_L = \infty$.

In the following, we assume that the CDF $F(s)$ is a strictly increasing function of s and, instead of optimizing the policy $(\mathbf{p}, \mathbf{q})$, we equivalently optimize a policy $(\mathbf{p}, \mathbf{a})$, where $a_l = F(q_l)$ for all l. Using the shorthand $q(a) = F^{-1}(a)$, ART in terms of the pair $(\mathbf{p}, \mathbf{a})$ has the form:

$$R_L(\mathbf{p}, \mathbf{a}) = \sum_{l=0}^{L-1} (a_{l+1} - a_l) R(p_l q(a_l)). \tag{2.14}$$

The average transmitted power constraint is

$$\rho(\mathbf{p}, \mathbf{a}) = a_0 p_0 + \sum_{l=0}^{L-1} (a_{l+1} - a_l) p_l \leq \overline{p}. \tag{2.15}$$

Among the set of L-level policies,

$$\pi_L(\overline{p}) = \{(\mathbf{p}, \mathbf{a}) | \rho(\mathbf{p}, \mathbf{a}) = \overline{p}\}, \tag{2.16}$$

and a feasible policy from (2.16) is an optimum policy if it achieves the maximum ART

$$C_L = \max_{(\mathbf{p}, \mathbf{a}) \in \pi_L(\overline{p})} R_L(\mathbf{p}, \mathbf{a}). \tag{2.17}$$

Hence, Theorem 1 now implies the following corollary.

Corollary 1 *For an optimum policy $(\mathbf{p}^*, \mathbf{a}^*)$, either only $\mathcal{U}_0^*$ includes an outage interval $[0, q(a_0^*))$ or $p_0^* = 0$.*

With the previous characterization, the form of the functions becomes similar to those in [15, 32], where the optimization is under bit-error rate constraints for real constellations.

Although the new optimization problem (2.17) is somewhat simpler than (2.7), it is still non-convex and difficult to solve in the general case. The following theorem provides a further characterization of the optimum policies. It shows that, given any adaptation partition now defined by $\mathbf{q}$, the optimum power assignment is a water-filling assignment [16].

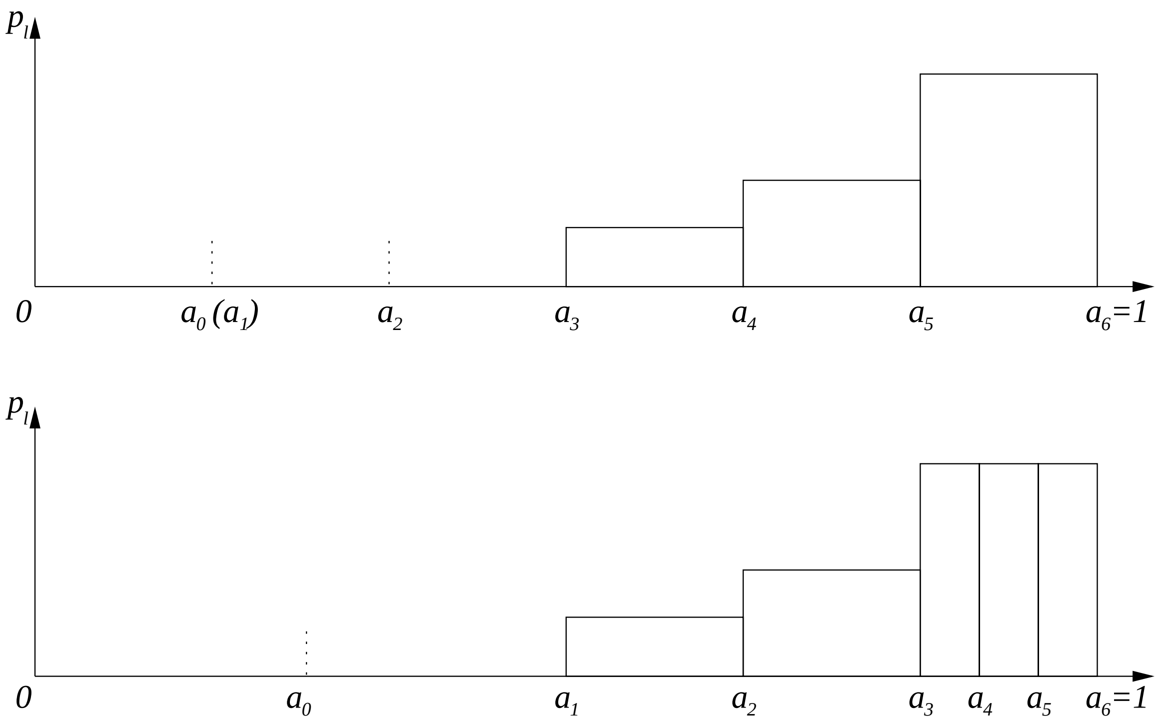

Figure 2.3: Illustration of the repartitioning step for $L = 6$ and $m = 3$.

Theorem 2 *Given a partition* $\mathbf{a}$*, the optimum power allocation is water-filling,*

$$p_l = \begin{cases} \left(\frac{a_1 - a_0}{\lambda a_1} - \frac{N_0}{q(a_0)}\right)^+ & l = 0 \\ \left(\frac{1}{\lambda} - \frac{N_0}{q(a_l)}\right)^+ & l > 0 \end{cases}, \tag{2.18}$$

where the water-filling level $1/\lambda$ *is chosen to satisfy the average power constraint* $\rho(\mathbf{p}, \mathbf{a}) = \overline{p}$.

Note that the powers p_l allocated according to Theorem 2 are non-decreasing in l. Also, given an arbitrary partition $\mathbf{a}$, water-filling may result in a collection of $l' > 1$ intervals $\{\mathcal{U}_l | l \leq l'\}$ with power $p_l = 0$. However, there is no benefit in terms of ART to design policies with more than one zero power interval. The following lemma shows that ART can be increased by subdividing intervals with non-zero power.

Lemma 3 *A quantization interval* $[a_l, a_{l+1})$ *with* $p_l > 0$ *can be split into two new quantization intervals,* $[a_l, x)$ *and* $[x, a_{l+1})$*, such that its contribution towards ART strictly increases while the average transmitted power does not change.*

As illustrated in Fig. 2.3, greater efficiency can result from repartitioning an arbitrary policy by merging all intervals with zero power into a single interval and subdividing a non-zero power interval. Thus Lemma 3 yields the following corollary.

Corollary 2 *For an optimum policy* $(\mathbf{p}^*, \mathbf{a}^*)$ *only* p_0^* *can be zero.*

An important question is whether p_0^* is always zero for an optimal policy $(\mathbf{p}^*, \mathbf{a}^*)$. Since the throughput during an outage does not contribute anything towards ART, transmission during an outage is wasting power. From the policy characterization in this section, outage happens only if p_0 is not zero. So, an intuitively good policy should set p_0 to be zero. However, such a conjecture is not true for $L = 1$ where $p_0 = 0$ means no transmission for any time. In fact, the intuitively satisfying zero-outage policy requires at least $L = 2$. Nevertheless, only for relatively a small $\overline{p}$ with respect to N_0, the optimal policies have zero outage and *outage can not be avoided for ART-maximizing policies with a relatively large* $\overline{p}$, even for $L = 2$. Such an example will be shown in Section 2.6. For zero-outage policies with $L = 2$, increasing p_1 (and, therefore, r_1) with a fixed $\overline{p}$ requires to increase the fraction of the time that the transmitter is off. The best zero-outage policy suffers from this trade-off. Consequently, allowing non-zero outage offers greater flexibility to trade off the possibility of outage against the additional efficiency of supporting more than one rate.

2.4 An Asymptotically Optimum Policy

The ergodic capacity of a fading channel [22] can be written as

$$C = \int_{s_0}^{\infty} \log\left(\frac{s}{s_0}\right) f(s) ds, \tag{2.19}$$

where s_0 is a cut-off value which is strictly positive for a finite average transmitted power constraint

$$\int_{s_0}^{\infty} p(s) f(s) ds = \overline{p}, \tag{2.20}$$

where $p(s)$ is a continuous water-filling power assignment given by

$$p(s) = N_0 \left(\frac{1}{s_0} - \frac{1}{s}\right)^+. \tag{2.21}$$

Intuitively, for the discrete adaptive system, when L increases to infinity, the corresponding maximum ART C_L should converge to the ergodic capacity. For a reasonably good policy $(\mathbf{p}, \mathbf{a})$, it is expected that the corresponding ART $R_L(\mathbf{p}, \mathbf{a})$ should also converge to the ergodic capacity. This property is referred to as *asymptotic optimality.*

Although we gained some insight on construction of good policies in the previous section, the design of an optimum partition $\mathbf{a}^*$ remains unknown. Here, we verify the asymptotic optimality of a policy based on a channel state partition which is uniform in probability. This design is a building block and the starting point of the iterative algorithm in the next section. We compare the ergodic capacity with ART of an L-level policy $(\mathbf{p}^\dagger, \mathbf{a}^\dagger)$ with the rate/quantization level assignment

$$a_l^\dagger = \begin{cases} 1/L, & l = 0 \\ l/L, & l = 1, 2, \ldots, L-1 \end{cases}, \tag{2.22}$$

and the water-filling power assignment (2.18). Here, we deliberately set $a_0^\dagger = a_1^\dagger$ which leads to $p_0^\dagger = 0$ after water-filling. Although suboptimal, this choice will simplify subsequent arguments. Based on (2.18), we have that $p_l^\dagger$ for $l \neq 0$ can be expressed as follows.

$$p_l^\dagger = N_0 \left(\frac{1}{\lambda^\dagger N_0} - \frac{1}{q(a_l^\dagger)} \right)^+. \tag{2.23}$$

Since $(\mathbf{p}^\dagger, \mathbf{a}^\dagger)$ is sub-optimum, $R_L^\dagger = R_L(\mathbf{p}^\dagger, \mathbf{a}^\dagger)$ satisfies

$$R_L^\dagger < C_L < C. \tag{2.24}$$

The similarity between the power allocation functions $p(s)$ in (2.21) and $p_l^\dagger$ in (2.23) is helpful in demonstrating that $R_L^\dagger$ is asymptotically optimal, as verified on in the proof of the following theorem.

Theorem 3 $\lim_{L\to\infty} R_L^\dagger = C.$

The implication of Theorem 3 is that, for sufficiently large L, the optimization over the partition $\mathbf{q}$ is of less importance. Allocating the power in a water-filling manner and assigning the corresponding code rates allow for an average throughput which is close to the ergodic capacity for a given "reasonable" partition $\mathbf{a}$. On the other

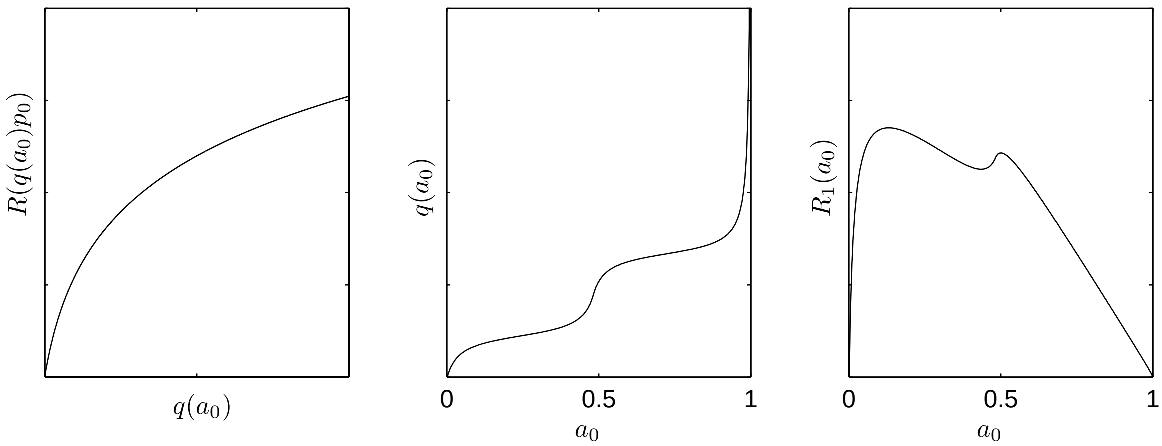

Figure 2.4: Local optima ($q_0 = q(a_0)$).

hand, an arbitrarily chosen partition $\mathbf{a}$ may not lead to an asymptotically optimum policy. Moreover, arguments given here imply that the joint partition and power-rate optimization will offer more significant improvement when L is small.

2.5 Iterative Policy Improvement

Finding an optimum policy $(\mathbf{p}^*, \mathbf{a}^*)$ that achieves the maximum ART is a challenging optimization problem. The following simple example illustrates that the problem could be non-convex and that there can be multiple local maxima. In this example, we assume $L = 1$ so that the equality in the average transmitted power constraint yields $p_0 = \bar{p}$. The objective function, ART, is

$$R_1(a_0) \quad = \quad (1 - a_0)R(q(a_0)p_0), \tag{2.25}$$

where $q_0 = q(a_0)$ and $R(q_0p_0)$ is a monotonically increasing function shown in Fig. 2.4(a). Let q_0 be as shown in Fig. 2.4(b), then $R_1(a_0)$ has two local maxima and is not concave as shown in Fig. 2.4(c).

One way of finding the optimum policy is brute force maximization of (2.17) over a quantized space of all possible pairs of power and partition assignments. This approach entails quantization of continuous policy variables $\mathbf{p}$ and $\mathbf{a}$ and can only be taken for a very small number of quantization levels L since its complexity increases exponentially in L. In Fig. 2.5, we present an iterative algorithm that finds a good policy $(\mathbf{p}^\ddagger, \mathbf{a}^\ddagger)$ with

1. $k = 0$. Choose interval boundaries $a_l^{(k)} = l/L$ for $l = 1, \ldots, L$ and set $a_0^{(k)} = 1/L$.

2. *water-filling:* Given $\mathbf{a}^{(k)}$, find $\mathbf{p}^{(k)}$ from the water-filling assignment (2.18).

3. *repartitioning:* If $p_1^{(k)} = 0$, let $n = \min\{m | p_m^{(k)} > 0\}$. Define the new policy $(\mathbf{p}, \mathbf{a})$ as

$$
\begin{aligned}
a_l &= \begin{cases} a_n^{(k)}/2, & l = 0 \\ a_{l+n-1}^{(k)}, & l = 1, \ldots, L-n \\ a_{L-1}^{(k)} + (1 - a_{L-1}^{(k)})\frac{l-n}{L-n}, & \text{otherwise} \end{cases} \\
p_l &= \begin{cases} 0, & l = 0 \\ p_{l+n-1}^{(k)}, & l = 1, \ldots, L-n \\ p_{L-1}^{(k)}, & \text{o.w.} \end{cases}
\end{aligned}
$$

Otherwise, $\mathbf{p} = \mathbf{p}^{(k)}$ and $\mathbf{a} = \mathbf{a}^{(k)}$.

4. *water-spilling:* For $l = L-1, L-2, \ldots, 1$:

- Using (2.27), let $(p'_{l-1}, a'_l) = \arg\max_{(p_{l-1}(a_l), a_l)} R_L(\mathbf{p}, \mathbf{a})$.
- Set

$$
\begin{aligned}
\mathbf{p} &= [p_0, \ldots, p_{l-2}, p'_{l-1}, p'_l, \ldots, p'_{L-2}, p_{L-1}]^\top \\
\mathbf{a} &= [a_0, \ldots, a_{l-1}, a'_l, a'_{l+1}, \ldots, a'_{L-1}]^\top
\end{aligned}
$$

Let $a'_0 = \arg\max_{a_0} R_L(\mathbf{p}, \mathbf{a})$. Set $\mathbf{a}^{(k)} = [a'_0, a'_1, \ldots, a'_{L-1}]^\top$ and $\mathbf{p}^{(k)} = \mathbf{p}$.

5. If $R_L(\mathbf{p}^{(k)}, \mathbf{a}^{(k)}) - R_L(\mathbf{p}^{(k-1)}, \mathbf{a}^{(k-1)}) < \epsilon$ let $\mathbf{p}^\ddagger = \mathbf{p}^{(k)}$ and $\mathbf{a}^\ddagger = \mathbf{a}^{(k)}$. Define $R_L^\ddagger = R_L(\mathbf{p}^\ddagger, \mathbf{a}^\ddagger)$ and stop. Otherwise, set $\mathbf{a}^{(k+1)} = \mathbf{a}^{(k)}$, $k = k+1$, and go back to step 2.

Figure 2.5: The iterative algorithm consisting of water-filling, repartitioning, and water-spilling.

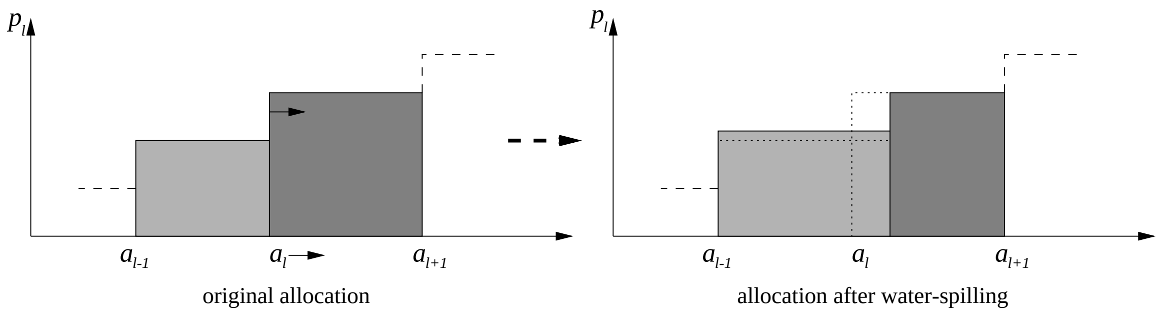

Figure 2.6: Illustration of the water-spilling step. As the boundary a_l moves to the right, power from interval $\mathcal{U}_l = [a_l, a_{l+1})$ spills over the boundary to fill the interval $\mathcal{U}_{l-1}$. The figure on the right depicts the increase in p_{l-1} as the boundary a_l is increased.

$R_L^{\ddagger} = R_L(\mathbf{p}^{\ddagger}, \mathbf{a}^{\ddagger})$. The first two steps initialize the algorithm with the asymptotically optimum policy $(\mathbf{p}^{\dagger}, \mathbf{a}^{\dagger})$ with $R_L^{\dagger}$. The rest of the algorithm consists of the following three local optimization techniques:

- Water-filling (Theorem 2) to optimize the power allocation given a partition;
- Water-spilling (described below) to optimize the intervals while satisfying the power constraint (2.16);
- Repartitioning (Lemma 3) to re-allocate zero power intervals.

The convergence of the algorithm follows from the fact that ART will be nondecreasing at each step of the algorithm and is upper-bounded by C [22]. Thus,

$$R_L^{\dagger} \leq R_L^{\ddagger} \leq C, \tag{2.26}$$

Even though the algorithm may not converge to the optimal solution, the difference between $R_L^{\ddagger}$ and C_L typically is very small.

We note that a water-filling step may result in several zero power assignments ($p_l = 0$), which is suboptimal according to Corollary 2. We will see that water-spilling may not be able to remedy this. Thus we employ repartitioning, as illustrated in Fig. 2.3, as an intermediate step. Under repartitioning, intervals with zero power are merged to satisfy Corollary 2 and $\mathcal{U}_{L-1}$ is partitioned following Lemma 3. These two operations ensure that ART will increase.

For iterative adjustment of the partitions, we employ a local optimization technique called *water-spilling* that varies boundaries a_l one at a time. Fig. 2.6 depicts a policy before and after such local optimization. The water-spilling technique is designed to satisfy the average transmitted power constraint (2.15). Since the quantization intervals in Fig. 2.6 are represented in terms of $\mathbf{a}$, the average transmitted power assigned to a given interval is equal to the area of its respective rectangle. To ensure that the power constraint (2.15) is satisfied with equality regardless of the change in a_l, the sum of the areas of the shaded rectangles must remain the same. Consequently, when we shift a_l to the right (or, equivalently, increase the rate assignment r_l), power spills from the interval l to raise the power p_{l-1}. Let

$$p_{l-1}(a_l) = \begin{cases} \left[\bar{p}_l - p_l(a_{l+1} - a_l)\right] / \left(a_l - a_{l-1}\right), & l > 1 \\ \left[\bar{p}_l - p_l(a_{l+1} - a_l)\right] / a_l, & l = 1 \end{cases}, \tag{2.27}$$

where $\bar{p}_l$ is the average transmitted power over $\mathcal{U}_{l-1}$ and $\mathcal{U}_l$. The slight difference in (2.27) for $p_0(a_1)$ is due to the possibility of outage. Consequently, water-spilling is

$$\max_{a_l:\, a_{l-1} \le a_l \le a_{l+1}} \quad (a_l - a_{l-1}) \log(1 + p_{l-1}(a_l)a_{l-1}) + (a_{l+1} - a_l) \log(1 + p_l a_l) \tag{2.28}$$

Determining the family of twice differentiable distributions $F(s)$ for which the water-spilling objective function in (2.28) is concave in a_l is straightforward. For such distributions any convex searching algorithm can be used in this step. Nevertheless, since water-spilling solves a single parameter optimization problem, any line-search algorithm can be employed over an interval (a_{l-1}, a_{l+1}) for determining a_l[2].

2.6 Numerical Results

In this section, we present numerical results for the maximum ART C_L and lower bounds of C_L obtained by brute force maximization over a finely quantized space of all possible policies and by using the proposed iterative algorithm, respectively. We

[2]It can be shown that there are impulsive PDFs for which the objective function has multiple local maxima, and for such distribution functions local maximization algorithms would fail to determine the optimum solution.

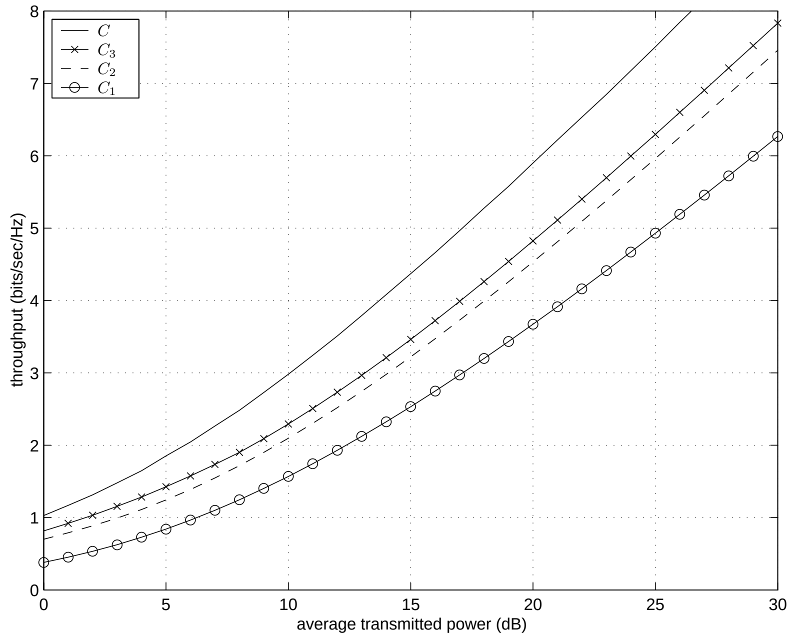

Figure 2.7: C_L with $L = 1, 2, 3$ for Rayleigh fading.

compare C_L and lower bounds with the ergodic capacity C for two different fading models: Rayleigh and log-normal fading.

Following [10], within this section, we assume that $N_0 = 1$ for (2.1) without loss of generality. Therefore, the system performance is captured by throughput versus average transmitted power ($\overline{p}$) curves.

2.6.1 Rayleigh Fading Channel

For a Rayleigh fading channel, the fading CDF is

$$F_{\mathrm{R}}(s) = 1 - e^{-s}, \qquad s \in \mathcal{S}. \tag{2.29}$$

In this case, it is easily verified that the water-spilling objective function in (2.28) is concave. Fig. 2.7 shows there is a 6 to 7 dB gap between the curves of C and C_1 for a value around 1 to 2 bits/sec/Hz. As C_1 is the maximum ART for the constant-rate constant-power transmission, the gap indicates potential gains of the adaptive transmission system. By applying an $L = 2$ levels adaptive transmission policy, the required $\overline{p}$ can be reduced by approximately 3 dB in comparison with that by using the constant-power and constant-rate policy. In other words, an adaptive system with

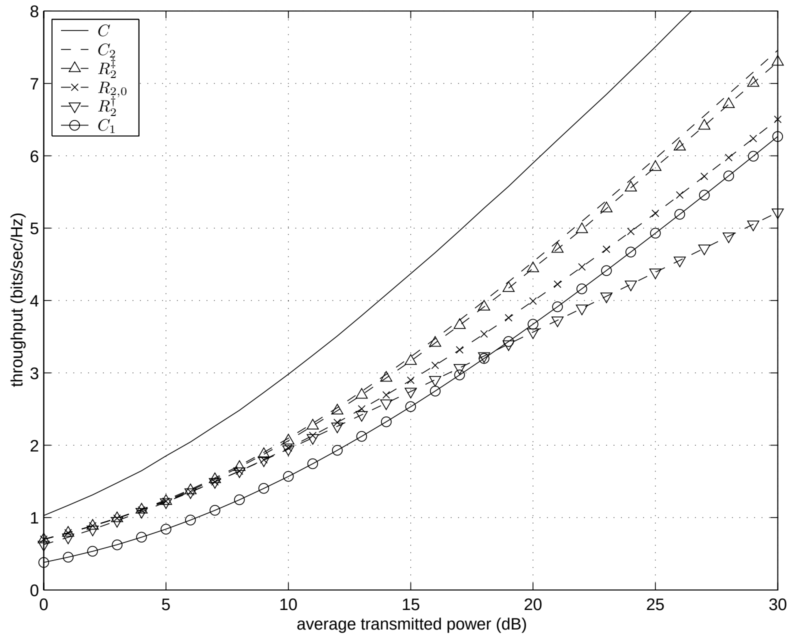

Figure 2.8: A comparison of C, C_L, $R_L^{\dagger}$, $R_L^{\ddagger}$, and the maximum ART of zero-outage policies ($R_{L,0}$) for Rayleigh fading.

$L = 2$ can eliminate about half of the power requirement gap between the curves of the ergodic and the non-adaptive maximum ART. Furthermore, increasing L from 2 to 3 yields another 1 dB reduction in the average transmitted power requirement. Note that C_2 and C_3 were obtained by brute force maximization (searching).

In Fig. 2.8, for $L = 2$, we compare C_L, $R_L^{\ddagger}$, $R_L^{\dagger}$, and the maximum ART of zero-outage policies (indicated by $R_{L,0}$). For $R_2^{\dagger}$, the corresponding policy $(p^{\dagger}, q^{\dagger})$ has $p_0^{\dagger} = 0$ owing to $a_0^{\dagger} = a_1^{\dagger}$. In addition, since we force $a_1^{\dagger} = 0.5$, the transmitter is off half of the time. With these artificial constraints, $(p^{\dagger}, q^{\dagger})$ is not a good policy for improving ART and $R_2^{\dagger}$ is even worse than C_1 for high $\overline{p}$'s.

Note that $(p^{\dagger}, q^{\dagger})$ is also a zero-outage policy. Therefore, $R_{2,0}$ is always higher than $R_2^{\dagger}$. Moreover, since the transmitter is turned off when channel is bad, $R_{2,0}$ is always better than C_1 as well. Even though $R_{2,0}$ is an intuitively good design, it only overlaps with C_2 for low $\overline{p}$'s and for high $\overline{p}$'s, the artificial constraint of zero-outage leads to a big performance penalty. For instance, at an average transmitted power of 20 dB, 10% of ART is lost due to this zero-outage constraint. On the other hand, $R_{2,0}$ is relatively easy to obtain since only a one parameter optimization is needed.

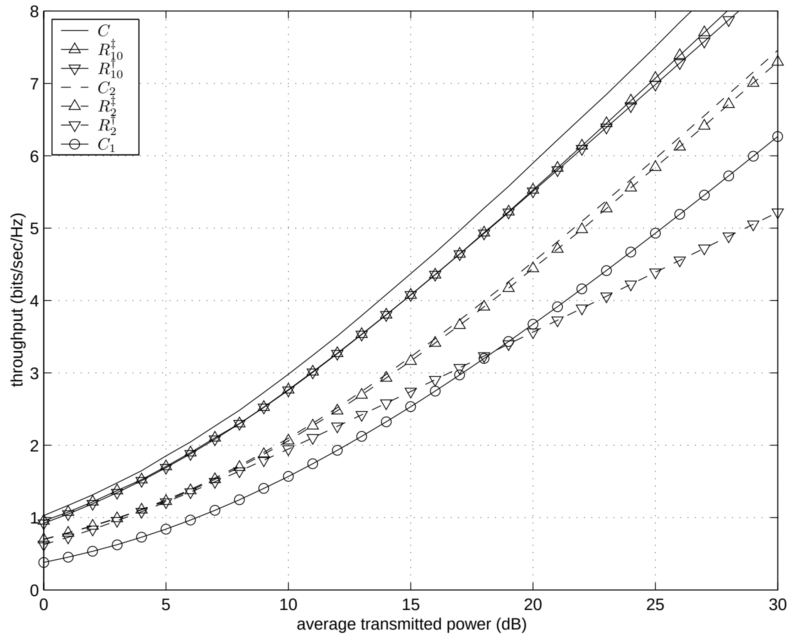

Figure 2.9: A comparison of C, C_L, $R^{\dagger}_{L}$, and $R^{\ddagger}_{L}$ for Rayleigh fading.

With the proposed iterative algorithm, we can find policies achieves $R^{\ddagger}_{2}$ which is within a fraction of a dB from C_2 regardless of the fact that the iterative algorithm starts from $(p^{\dagger}, q^{\dagger})$. Thus, the effectiveness of the algorithm is demonstrated.

In Fig. 2.9, results with $L = 10$ are shown. $R^{\ddagger}_{10}$ is within a dB from C when the throughput is less than 5 bits/second/Hz. Since $R^{\ddagger}_{L}$ is a lower bound of C_L, we are sure that for $L > 10$, C_L is very close to C. In addition, the difference between $R^{\ddagger}_{10}$ and $R^{\dagger}_{10}$ is very small and this shows the asymptotic optimality of $(p^{\dagger}, q^{\dagger})$ for relatively large L's.

2.6.2 Nakagami Fading Channel

Generally, a PDF of a Nakagami distribution [46] is

$$f_{\mathrm{N}}(t) = \begin{cases} \frac{2}{\Gamma(m)} \left(\frac{m}{\Omega}\right)^m t^{2m-1} e^{-mt^2/\Omega}, & t \geq 0 \\ 0, & \text{otherwise} \end{cases} . \tag{2.30}$$

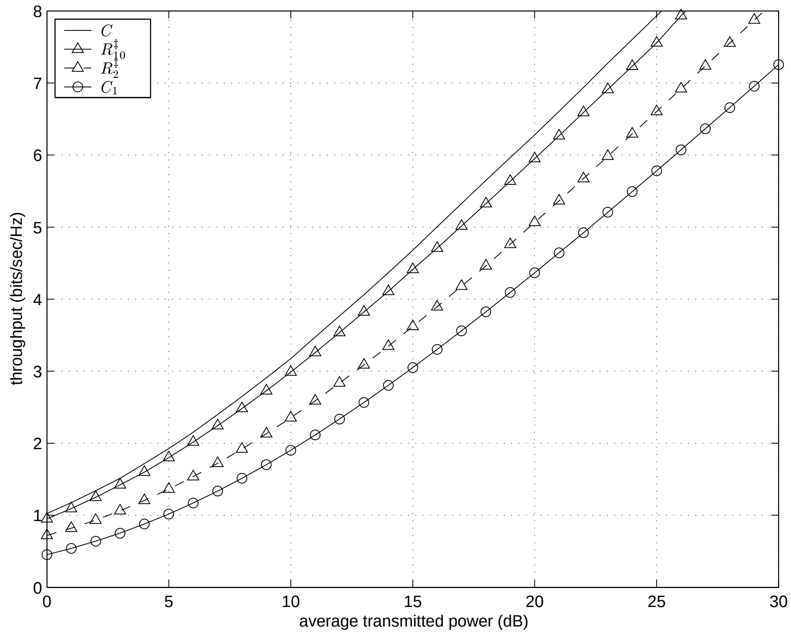

Figure 2.10: $R_L^{\ddagger}$ with $L \leq 10$ for Nakagami $(m = 2)$ fading channel.

where m is a parameter, $\Omega = E\{t^2\}$ and $t = \sqrt{s\Omega}$. Similarly to [22], we choose the Nakagami fading channel with $m = 2$ for which the CDF is

$$f_{\mathrm{N}}(t) = \begin{cases} 2\left(\frac{2}{\Omega}\right)^2 t^3 e^{-2t^2/\Omega}, & t \geq 0 \\ 0, & \text{otherwise} \end{cases}. \tag{2.31}$$

Hence,

$$f_{\mathrm{N}}(s) = 4se^{-2s}, \tag{2.32}$$

and

$$F_{\mathrm{N}}(s) = 1 - (1 + 2s)e^{-2s}. \tag{2.33}$$

In this case, (2.28) is also a concave maximization since

$$\frac{e^{2s}}{1+2s}\left[2 + \frac{(1-2s)(1+2s)}{4s^2}\right] = \frac{e^{2s}}{1+2s}\left[1 + \frac{1}{4s^2}\right] \geq 0. \tag{2.34}$$

In Fig. 2.10, we use the same iterative algorithm to evaluate $R_L^{\ddagger}$ for a Nakagami fading channel with $m = 2$. These curves are very similar to those of the Rayleigh fading channel. In fact, this is not surprising since the Rayleigh fading channel is the Nakagami fading channel with $m = 1$.

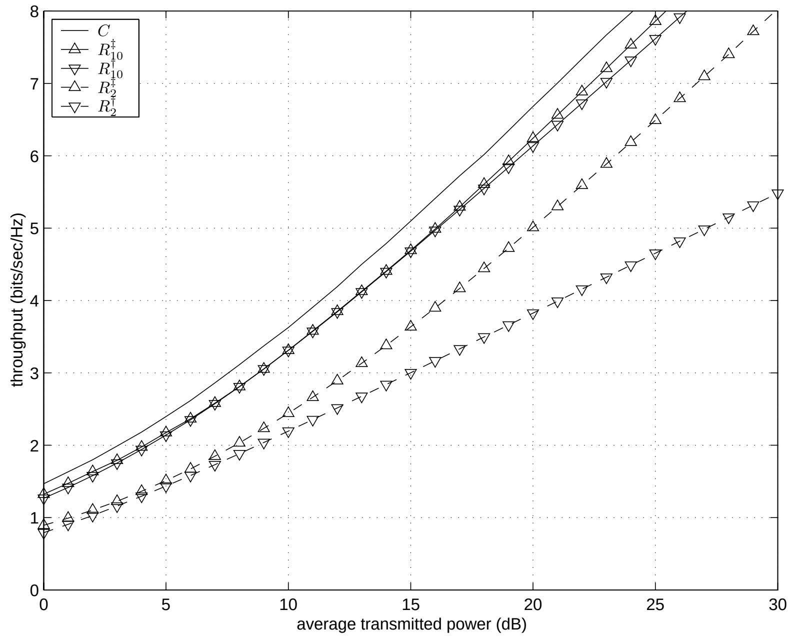

Figure 2.11: A comparison of C, $R_L^{\dagger}$, and $R_L^{\ddagger}$ for log-normal fading ($\sigma = 6$).

2.6.3 Log-Normal Fading Channel

With a mean μ and a variance σ, a PDF of a log-normal random variable s [47] is

$$f_{\mathrm{LN}}(s) = \frac{1}{\sqrt{2\pi}\sigma}\frac{10}{\log 10}\frac{1}{s}e^{-\frac{(10\log_{10} s-\mu)^2}{2\sigma^2}}. \tag{2.35}$$

Fig. 2.11, relates C, $R_L^{\dagger}$, and $R_L^{\ddagger}$ for the log-normal channel with $\sigma = 6$ is shown. Here, we have similar observations to those of the Rayleigh fading channel. However, it seems that in comparison to the Rayleigh fading, the log-normal fading channel requires relatively larger L for an adaptive system in order to achieve a performance close to C.

2.7 Conclusion

Following the pioneering work [22] and the recently introduced concepts of information outage and outage probability [8, 45, 55], we define the throughput maximization problem for discrete adaptive transmission systems. Problem solution is an optimum policy consisting of a channel state space partition and a power and rate allocation. Properties of optimum policies aid in understanding and simplifying the throughput maximization problem.

No closed form optimum policies have been found, nevertheless, we suggest three approaches to obtaining good policies. The first approach is to exhaustively search for good policies over the feasible policy set with a high resolution. Unfortunately, this approach is only effective for cases where L is relatively small since its computational complexity increases exponentially in L. Another approach is to design a reasonably good policy, as shown in Section 2.4, which provides a very good approximation to an optimum policy when L is large. The third approach is an iterative local search algorithm which is particularly useful when the first approach is impractical and the second one is inadequate.

Finally, we have demonstrated that a carefully designed discrete adaptive system with a small number of power levels and code rates can achieve average throughput close to the one obtained by optimum continuous adaptive transmission systems in several slow fading environments.

2.A Proofs

Proof: Lemma 1

When p_l^* is zero, (2.8) is obvious. Otherwise, suppose there is an optimum policy $(\mathbf{p}, \mathbf{v}, \mathbf{r})$ for which $p_l > 0$ and $r_l \notin [R(p_l v_{l+1}), R(p_l v_{l+1}))$ for some l. We construct the policy $(\mathbf{p}, \mathbf{v}, \mathbf{r}')$ with $\mathbf{r}' = [r_0, \ldots, r_{l-1}, r_l', r_{l+1}, \ldots, r_{L-1}]^\top$ with $r_l' = R(p_l v_l)$. It is straightforward to show that $(\mathbf{p}, \mathbf{v}, \mathbf{r}') \in \pi_L(\overline{p})$. We consider two cases. First, if $r_l > R(p_l v_{l+1})$, then

$$R_L(\mathbf{p}, \mathbf{v}, \mathbf{r}') - R_L(\mathbf{p}, \mathbf{v}, \mathbf{r}) = F(v_l, v_{l+1}) r_l' > 0, \qquad (2.36)$$

which is a contradiction. Second, if $r_l < R(p_l v_l)$,

$$R_L(\mathbf{p}, \mathbf{v}, \mathbf{r}') - R_L(\mathbf{p}, \mathbf{v}, \mathbf{r}) = (r_l' - r_l) F(v_l, v_{l+1}) > 0, \qquad (2.37)$$

which is also a contradiction.

□

Proof: Lemma 2

Given a policy $(\mathbf{p},\mathbf{v},\mathbf{q})$, with average transmitted power $\sum_{l=0}^{L-1} F(v_l, v_{l+1})p_l = \overline{p}-\epsilon$, we can construct a new policy $(\mathbf{p}',\mathbf{v},\mathbf{q}) \in \pi_L(\overline{p})$ where $\mathbf{p}' = [p_0+\epsilon,\ldots,p_{L-1}+\epsilon]^\top$. This policy achieves

$$\begin{aligned} R_L(\mathbf{p}',\mathbf{v},\mathbf{q}) &= \sum_{l=0}^{L-1} F(q_l, v_{l+1})R((p_l+\epsilon)q_l) && (2.38)\\ &> R_L(\mathbf{p},\mathbf{v},\mathbf{q}). && (2.39)\end{aligned}$$

$\square$

Proof: Theorem 1

Given an arbitrary policy $(\mathbf{p},\mathbf{v},\mathbf{q}) \in \pi_L(\overline{p})$, suppose that there is an l such that $v_{l+1} < q_{l+1}$ and $0 \le l < L-1$. We will construct a new policy $(\mathbf{p}',\mathbf{v}',\mathbf{q}) \in \pi_L(\overline{p})$ such that $v'_{l+1} = q_{l+1}$ and $R_L(\mathbf{p}',\mathbf{v}',\mathbf{q}) \ge R_L(\mathbf{p},\mathbf{v},\mathbf{q})$. If such scenarios appear more than once, we can repeat the same construction for each such l.

Let

$$\hat{p} = F(q_{l+1}, v_{l+2})p_{l+1} + \sum_{i\neq\{l,l+1\}} F(v_i, v_{i+1})p_i, \tag{2.40}$$

denote the portion of the average transmitted power $\rho(\mathbf{p},\mathbf{v},\mathbf{q})$ associated with channel states $s \notin [v_l, q_{l+1})$. Therefore, we can write

$$\begin{aligned} \rho(\mathbf{p},\mathbf{v},\mathbf{q}) &= \hat{p} + F(v_l, v_{l+1})p_l + F(v_{l+1}, q_{l+1})p_{l+1} && (2.41)\\ &= \hat{p} + F(v_l, q_{l+1})\tilde{p}, && (2.42)\end{aligned}$$

where

$$\tilde{p} = \left\{F(v_l, v_{l+1})p_l + F(v_{l+1}, q_{l+1})p_{l+1}\right\}/F(v_l, q_{l+1}). \tag{2.43}$$

Based on $\tilde{p}$, we construct a new policy $(\mathbf{p}',\mathbf{v}',\mathbf{q})$ with $\mathbf{v}' = [v_0,\ldots,v_l,q_{l+1},v_{l+2},\ldots,v_{L-1}]^\top$ and $\mathbf{p}' = [p_0,\ldots,p_{l-1},\tilde{p},p_{l+1},\ldots,p_{L-1}]^\top$. In the policy $(\mathbf{p}',\mathbf{v}',\mathbf{q})$, transmitted power $\tilde{p}$ is used for all states $s \in [v_l, q_{l+1})$. Hence,

$$\rho(\mathbf{p}',\mathbf{v}',\mathbf{q}) = \hat{p} + F(v_l, q_{l+1})\tilde{p}. \tag{2.44}$$

It follows from (2.42) that $\rho(\mathbf{p}',\mathbf{v}',\mathbf{q}) = \rho(\mathbf{p},\mathbf{v},\mathbf{q})$. The corresponding ART is

$$R_L(\mathbf{p}',\mathbf{v}',\mathbf{q}) = \hat{R} + F(q_l, q_{l+1})R(q_l\tilde{p}), \tag{2.45}$$

where $\hat{R} = \sum_{i \neq l} F(q_i, v_{i+1}) R(q_i p_i)$ denotes the contributions to $R_L(\mathbf{p}, \mathbf{v}, \mathbf{q})$ from channel states $s \notin \mathcal{U}_l$. Defining $\alpha = F(q_l, v_{l+1})/F(q_l, q_{l+1})$, we observe from (2.43) that

$$\tilde{p} = \frac{p_l F(v_l, q_l) + \alpha p_l F(q_l, q_{l+1}) + p_{l+1} F(v_{l+1}, q_{l+1})}{F(v_l, q_{l+1})} \tag{2.46}$$

$$\geq \frac{p_l F(v_l, q_l) + \alpha p_l F(q_l, q_{l+1})}{F(v_l, q_{l+1})}. \tag{2.47}$$

Since $\alpha \leq 1$,

$$\tilde{p} \geq \frac{\alpha p_l (F(v_l, q_l) + F(q_l, q_{l+1}))}{F(v_l, q_{l+1})} = \alpha p_l. \tag{2.48}$$

From (2.45), (2.48) and the fact that $R(\cdot)$ is monotonic increasing, we observe that

$$R_L(\mathbf{p}', \mathbf{v}', \mathbf{q}) \geq \hat{R} + F(q_l, q_{l+1}) R\left[q_l(\alpha p_l + 0)\right] \tag{2.49}$$

$$\overset{(a)}{\geq} \hat{R} + F(q_l, q_{l+1}) \left[\alpha R(q_l p_l) + (1 - \alpha) R(0)\right] \tag{2.50}$$

$$\overset{(b)}{=} \hat{R} + F(q_l, q_{l+1}) \alpha R(q_l p_l) \tag{2.51}$$

$$= \hat{R} + F(q_l, v_{l+1}) R(q_l p_l) = R_L(\mathbf{p}, \mathbf{v}, \mathbf{q}). \tag{2.52}$$

Note that inequality (a) is due to the concavity of $R(\cdot)$ while (b) holds because $R(0) = 0$.

□

Proof: Theorem 2

Following the standard Lagrange procedure, we have

$$J = \lambda \overline{p} + \sum_{l=0}^{L-1} (a_{l+1} - a_l) \left[R(p_l q(a_l)) - \lambda p_l\right] - \lambda a_0 p_0. \tag{2.53}$$

For $l > 0$,

$$\frac{\partial J}{\partial p_l} = (a_{l+1} - a_l) \left(\frac{q(a_l)/N_0}{1 + p_l q(a_l)/N_0} - \lambda \right) = 0. \tag{2.54}$$

This leads to

$$p_l = \left(\frac{1}{\lambda} - \frac{N_0}{q(a_l)} \right)^+. \tag{2.55}$$

Similarly, if $a_0 > 0$, we have

$$\frac{\partial J}{\partial p_0} = (a_1 - a_0) \frac{q(a_0)/N_0}{1 + p_0 q(a_0)/N_0} - \lambda a_1 = 0, \tag{2.56}$$

and, thus,

$$p_0 = \left(\frac{a_1 - a_0}{\lambda a_1} - \frac{N_0}{q(a_0)} \right)^+. \tag{2.57}$$

If $a_0 = 0$, then $p_0 = 0$. Based on (2.16), $1/\lambda$ can be solved and (2.18) follows immediately.

$\square$

Proof: Lemma 3

$$
\begin{aligned}
(a_l - a_{l-1}) R\left(p_l q(a_l)\right) &= (a_l - x) R\left(p_l q(a_l)\right) + (x - a_{l-1}) R\left(p_l q(a_l)\right) \\
&< (a_l - x) R\left(p_l q(x)\right) + (x - a_{l-1}) R\left(p_l q(a_l)\right).
\end{aligned}
$$

$\square$

Proof: Theorem 3

For a sufficiently large L, there exists an $l_0(L) > 0$ such that $s_0 \in \mathcal{U}^\dagger_{l_0(L)}$. We use notation $l_0 \triangleq l_0(L)$ and $q^\dagger_l \triangleq q(a^\dagger_l)$ in order to simplify the following derivations. In this case, $q^\dagger_{l_0} \le s_0 \le q^\dagger_{l_0+1}$. It follows from (2.21) that

$$
p(s) \ge N_0 \left(\frac{1}{q^\dagger_{l_0+1}} - \frac{1}{s} \right)^+ . \tag{2.58}
$$

For $s \in \mathcal{U}^\dagger_l$, $s \ge q^\dagger_l$, so that

$$
p(s) \ge N_0 \left(\frac{1}{q^\dagger_{l_0+1}} - \frac{1}{q^\dagger_l} \right)^+ \qquad s \in \mathcal{U}^\dagger_l. \tag{2.59}
$$

With the same average transmitted power $\overline{p}$ for both the discrete power assignment (2.23) and the continuous policy (2.21), we have

$$
\begin{aligned}
\frac{1}{L} \sum_{l=1}^{L-1} N_0 \left(\frac{1}{\lambda^\dagger N_0} - \frac{1}{q^\dagger_l} \right)^+ &= \sum_{l=0}^{L-1} \int_{\mathcal{U}^\dagger_l} p(s) f(s)\, ds & (2.60) \\
&\ge \frac{1}{L} \sum_{l=1}^{L-1} N_0 \left(\frac{1}{q^\dagger_{l_0+1}} - \frac{1}{q^\dagger_l} \right)^+ . & (2.61)
\end{aligned}
$$

We observe that the inequality (2.61) implies $q^\dagger_{l_0+1} > \lambda^\dagger N_0$. It follows from (2.23) that

$$
p^\dagger_l \ge N_0 \left(\frac{1}{q^\dagger_{l_0+1}} - \frac{1}{q^\dagger_l} \right)^+ . \tag{2.62}
$$

Now we derive a lower bound to $R^\dagger_L$.

$$
R^\dagger_L = \sum_{l=1}^{L-1} \frac{1}{L} R(p^\dagger_l q^\dagger_l) \ge \sum_{l=l_0+1}^{L-1} \frac{1}{L} R(p^\dagger_l q^\dagger_l). \tag{2.63}
$$

Applying the lower bound (2.62), we obtain

$$R_L^\dagger \geq R_{\text{lo}}(L) = \frac{1}{L}\sum_{l=l_0+1}^{L-1} \log\left(\frac{q_l^\dagger}{q_{l_0+1}^\dagger}\right). \tag{2.64}$$

Next, we upperbound C by $R_{\text{lo}}(L)$ plus terms that will go to zero with increasing L. Since $s_0 \geq q_{l_0}^\dagger$, we see from (2.19) that

$$C \leq \int_{q_{l_0}^\dagger}^{\infty} \log\left(\frac{s}{q_{l_0}^\dagger}\right) f(s)\,ds \tag{2.65}$$

$$= \sum_{l=l_0+1}^{L-1} \int_{\mathcal{U}_{l-1}^\dagger} \log\left(\frac{s}{q_{l_0}^\dagger}\right) f(s)\,ds + I_L^\dagger, \tag{2.66}$$

where $I_L^\dagger$ denotes the integral

$$I_L^\dagger = \int_{\mathcal{U}_{L-1}^\dagger} \log\left(\frac{s}{q_{l_0}^\dagger}\right) f(s)\,ds. \tag{2.67}$$

Below, we will take some additional care to upperbound $I_L^\dagger$. Since $s \in \mathcal{U}_{l-1}^\dagger$ implies $s \leq q_l^\dagger$, we have from (2.66) that

$$C \leq \frac{1}{L}\sum_{l=l_0+1}^{L-1} \log\left(\frac{q_l^\dagger}{q_{l_0}^\dagger}\right) + I_L^\dagger \tag{2.68}$$

$$= \frac{1}{L}\sum_{l=l_0+1}^{L-1} \log\left(\frac{q_l^\dagger}{q_{l_0+1}^\dagger}\frac{q_{l_0+1}^\dagger}{q_{l_0}^\dagger}\right) + I_L^\dagger \tag{2.69}$$

$$\leq R_{\text{lo}}(L) + \log\left(\frac{q_{l_0+1}^\dagger}{q_{l_0}^\dagger}\right) + I_L^\dagger. \tag{2.70}$$

Using the shorthand $\mathcal{U}_{L-1}^\dagger$ for the event $S \in \mathcal{U}_{L-1}^\dagger$, we employ (2.67) to write

$$I_L^\dagger = \Pr[\mathcal{U}_{L-1}^\dagger]\left(E\{\log S|\mathcal{U}_{L-1}^\dagger\} - \log q_{l_0}^\dagger\right). \tag{2.71}$$

Since $\Pr[\mathcal{U}_{L-1}^\dagger] = 1/L$, and since the log function is concave,

$$I_L^\dagger \leq \frac{1}{L}\log E\{S|\mathcal{U}_{L-1}^\dagger\} - \frac{\log q_{l_0}^\dagger}{L}. \tag{2.72}$$

The conditional expectation can also be upperbounded as

$$E\{S|\mathcal{U}_{L-1}^\dagger\} = \frac{1}{\Pr[\mathcal{U}_{L-1}^\dagger]}\int_{\mathcal{U}_{L-1}^\dagger} s f(s)\,ds \tag{2.73}$$

$$\leq L\int_0^\infty s f(s)\,ds = LE\{S\}. \tag{2.74}$$

From (2.72) and (2.74), we have that

$$I_L^\dagger \leq \frac{1}{L} \log\left(\frac{LE\{S\}}{q_{l_0}^\dagger}\right). \tag{2.75}$$

Finally, we observe that $q_{l_0}^\dagger \leq s_0 \leq q_{l_0+1}^\dagger$ implies that

$$F(s_0) - F(q_{l_0}^\dagger) \leq F(q_{l_0+1}^\dagger) - F(q_{l_0}^\dagger) = \frac{1}{L}. \tag{2.76}$$

Continuity of the distribution function $F(\cdot)$ implies that $\lim_{L\to\infty} q_{l_0}^\dagger = \lim_{L\to\infty} q_{l_0+1}^\dagger = s_0$. This permits us to conclude that

$$\lim_{L\to\infty} \log\left(\frac{q_{l_0+1}^\dagger}{q_{l_0}^\dagger}\right) = 0. \tag{2.77}$$

Similarly, (2.75) implies

$$\lim_{L\to\infty} I_L^\dagger = 0. \tag{2.78}$$

Applying (2.77) and (2.78) to (2.70), we see that $C \leq \lim_{L\to\infty} R_{\mathrm{lo}}(L)$. Since $C \geq R_L^\dagger$ and (2.64) implies $\lim_{L\to\infty} R_L^\dagger \geq \lim_{L\to\infty} R_{\mathrm{lo}}(L)$, the theorem follows.

$\square$

Chapter 3

Adaptive Transmission with a Finite Set of Code Rates

3.1 Introduction

In [22], two adaptive transmission policies are compared: (1) the truncated inversion policy with continuously variable transmitted power but a fixed code rate and (2) fixed transmitted power with continuously variable code rates. It is shown that the latter is superior in terms of the maximum average throughput. Both policies represent limiting cases of practical policies that support a finite set of code rates and power levels.

In Chapter 2, the optimization of a discrete adaptive transmission design based on information outage has been studied. In this case, the problem formulation was limited to the scenario where the numbers of code rates and transmitted power levels are the same. However, this restriction may not necessarily reflect realistic design constraints; it is common to have significantly fewer code rates than power levels. For instance, in an IS-95 system [1] which has been available commercially for more than 10 years, the transmission power adapts on a grid of 1-dB steps over a dynamic range of 60 dB or more. On the other hand, even for the most recent adaptive system designs [4, 5], the number of code rates is only around ten.

In this chapter, we examine an extreme case of system design with a finite number of code rates and a continuously variable transmitted power. Though similar problems for adaptive MQAM systems have been studied in [21, 32], this chapter emphasizes the optimum system design with the physical constraints of average power and a finite set of codebooks. The primary design problem will be the selection of code rates and the corresponding assignments of rate and power based on the channel state.

We assume additive white Gaussian noise (AWGN) and a slow multiplicative fading environment with a channel state which is constant during the transmission of a codeword. It is assumed that the exact current channel state information is known at both the transmitter and the receiver. The channel state space is partitioned into a countable number of intervals. Upon each transmission, a message is encoded at a rate corresponding to the current channel state interval and a power level corresponding to the current channel state. Since each codeword experiences an additive white Gaussian noise (AWGN) channel, random Gaussian codes organized in multiple codebooks are employed.

For the proposed adaptive system, similar to the scenario in Chapter 2, it is also possible that the instantaneous mutual information corresponding to a channel state is less than the assigned code rate. In this case, an information outage event occurs. Consequently, we can still characterize the performance of a system design based on the concept of *average reliable throughput* (ART), defined as the average data rate **assuming zero rate when the channel is in outage** Chapter 2.

Following the formulation of the finite code rate set problem, we explore optimum (ART-maximizing) policies where a *policy* is defined by a channel state space partition together with the corresponding transmitted power and rate allocation. In this work, we will show that, for an optimum policy, there is no outage in the sense that the transmitted power is nonzero only if the transmitter employs a nonzero code rate that can be successfully decoded at the receiver. Furthermore, the transmitter employs zero transmitted power only for a subset of worst channel states. We show that an optimum policy employing L codebooks can be uniquely characterized by a partition of the channel state space with $L+1$ intervals: the zero rate/power interval in addition to L intervals corresponding to L nonzero code rates. In particular, the optimum power allocation has a water-filling character which is uniquely determined by the channel state space partition.

Different from the situation in Chapter 2, here we can reduce the throughput maximization problem, which is a joint optimization over the space of channel state space partitions and the code rate/power assignments, to a one-parameter search. This outcome is quite surprising since the throughput maximization problem is generally not a convex optimization problem and can have multiple local maxima for an arbitrary distribution of the channel state. Within this chapter, the channel state distribution is only constrained to be continuous and differentiable.

In addition, we have obtained the optimum partition for a given set of rates to be assigned and an average power constraint. Such a solution is of particular interest to heuristic practical designs where codes can be selected only from a limited set of good channel codes before one chooses an optimum channel state partition. Though not surprisingly, our derivation also shows that a partition of $L+1$ channel state intervals is indeed the optimum choice. As a byproduct, for the MQAM spectral efficiency maximization problem introduced in [21] and addressed in [19, 32], we provide the optimum solution.

Finally, numerical results show that, for a Rayleigh fading channel, there is a gap of only 1 dB between the ergodic capacity and the throughput of a 2-rate adaptive transmission system when the throughput is less than 6 bits/sec/Hz. Thus, in comparison with the results in Chapter 2,we find that power adaptation can indeed be very helpful for adaptive transmission with discrete code rates. This agrees with the conclusion for power adaptation in MQAM systems in [32].

3.2 System Model and Problem Formulation

We consider a multiplicative flat fading channel model similar to that in [22]. The complex received signal

$$Y = \sqrt{S}X + W, \tag{3.1}$$

where S is the channel (fading) state, X is the complex transmitted signal, and W is a circularly symmetric additive white Gaussian noise (AWGN) with variance N_0. The channel state S is a real random variable of unit mean with a probability density function (PDF) $f(s)$, a cumulative distribution function (CDF) $F(s)$, and a domain $\mathcal{S} = \{s|s \geq 0\}$. Only in Section 3.4 is $F(s)$ assumed to be continuous, differentiable, and strictly increasing in s. In this chapter, fading is assumed to be sufficiently *slow* so that the channel state is constant during the transmission of a codeword.

A generalized adaptive transmission system can be modeled as follows. At any channel state s, the transmitter transmits codewords coded at a rate $r(s)$ with a power level $p(s) = E\left\{|X|^2 \, |s\right\}$, where $E\{\cdot\}$ denotes expectation. The code rate $r(s)$ is chosen from a set, $\mathcal{R}_0 = \{r_0 = 0\} \cup \mathcal{R}$ where $\mathcal{R} = \{r_1, \ldots, r_L\}$. Without loss of generality, we assume that $r_{l-1} < r_l$ for $l = 1, \ldots, L$. An adaptive *transmission policy* can uniquely be specified by the assigned code rate $r(s)$ and the corresponding allocated power level $p(s)$. Such a policy is denoted by the tuple $(\mathcal{R}, r(s), p(s))$.

Let

$$\mathcal{V}_l = \{s | r(s) = r_l\}, \; l = 0, \ldots, L, \tag{3.2}$$

denote the set of channel states in which rate r_l is employed. Each policy specifies a partition of the set of channel states $\mathcal{S} = \cup_{l=0}^{L} \mathcal{V}_l$. In general, $\mathcal{V}_l$ can be a countable union of intervals in $\mathcal{S}$ or simply a measurable set of channel states. Since there are L non-zero code rates, we call $(\mathcal{R}, r(s), p(s))$ an L-level policy.

Since (3.1) is an AWGN channel for any given $s \in \mathcal{S}$, the corresponding maximum mutual information is given by $\log(1 + p(s)s/N_0)$. Adopting the notation

$$R(\phi) = \log(1 + \phi/N_0), \qquad \phi \geq 0, \tag{3.3}$$

the maximum mutual information associated with any state s is $R(p(s)s)$. For any $s \in \mathcal{V}_l$, given a code rate r_l of a capacity achieving Gaussian codebook, and a power allocation $p(s)$, the information is guaranteed to be successfully received iff $R(p(s)s) \geq r_l$. We define the binary outage indicator function

$$I_{\text{out}}(r, s, p(s)) = \begin{cases} 1, & R(p(s)s) < r, \\ 0, & \text{otherwise.} \end{cases} \tag{3.4}$$

The expected value of $I_{\text{out}}(r, s, p(s))$ over the channel states S is the information outage introduced in [55].

Given a policy $(\mathcal{R}, r(s), p(s))$, the *average reliable throughput* (ART) is

$$R_L(\mathcal{R}, r(s), p(s)) = \sum_{l=0}^{L} r_l \int_{\mathcal{V}_l} \left[1 - I_{\text{out}}(r_l, s, p(s))\right] f(s) \, ds. \tag{3.5}$$

Let

$$f_l(s) = \begin{cases} f(s)/\Pr\left[s \in \mathcal{V}_l\right], & s \in \mathcal{V}_l, \\ 0, & \text{otherwise}, \end{cases} \tag{3.6}$$

denote the conditional PDF of the channel state given that it belongs to $\mathcal{V}_l$. The conditional average power given $s \in \mathcal{V}_l$ is

$$P_l = \int_{\mathcal{V}_l} p(s) f_l(s)\, ds, \tag{3.7}$$

Also, the *average transmitted power* for the policy $(\mathcal{R}, r(s), p(s))$ is

$$\rho_L(\mathcal{R}, r(s), p(s)) = \sum_{l=0}^{L} P_l \Pr\left[s \in \mathcal{V}_l\right]. \tag{3.8}$$

Throughout this chapter, we only consider the policies with $r(s)$ and $p(s)$ such that both (3.5) and (3.7) are meaningful, i.e., the integrals are either Riemann or Lebesgue integrable. For convenience, given a rate set $\mathcal{R}$, we define

$$\mathcal{R}_0 = \{r_0 = 0\} \cup \mathcal{R}. \tag{3.9}$$

The objective will be to maximize ART $R_L(\mathcal{R}, r(s), p(s))$ subject to an average power constraint and a constraint that there are L codebooks of L distinct nonzero rates:

$$\max_{\mathcal{R}, r(s), p(s)} \quad R_L(\mathcal{R}, r(s), p(s)), \tag{3.10}$$

$$\text{subject to} \quad \rho_L(\mathcal{R}, r(s), p(s)) \le \overline{p}, \tag{3.10a}$$

$$|\mathcal{R}| = L, \tag{3.10b}$$

$$r(s) \in \mathcal{R}_0, \tag{3.10c}$$

$$p(s) \ge 0, \tag{3.10d}$$

$$r(s) \ge 0. \tag{3.10e}$$

In (3.5), for any $\mathcal{V}_l$ with either $r_l = 0$ or $P_l = 0$, there is zero contribution towards ART and, consequently, it is optimum to assign $p(s) = 0$ over all such $\mathcal{V}_l$. For a $\mathcal{V}_l$ with nonzero r_l and P_l, we observe that an optimum policy must be locally optimum over

$\mathcal{V}_l$. Local optimality requires that given $\mathcal{V}_l$, P_l, and r_l, $p(s)$ must be the solution of

$$\max_{p(s)} \quad r_l \int_{\mathcal{V}_l} \left[1 - I_{\text{out}}(r_l, s, p(s))\right] f_l(s)\, ds, \tag{3.11}$$

$$\text{subject to} \quad \int_{\mathcal{V}_l} p(s) f_l(s)\, ds \le P_l, \tag{3.11a}$$

$$p(s) \ge 0. \tag{3.11b}$$

Given any non-negative r_l, (3.11) is equivalent to the following local outage minimization problem

$$\min_{p(s)} \quad \int_{\mathcal{V}_l} I_{\text{out}}(r_l, s, p(s)) f_l(s)\, ds, \tag{3.12}$$

$$\text{subject to} \quad \int_{\mathcal{V}_l} p(s) f_l(s)\, ds \le P_l, \tag{3.12a}$$

$$p(s) \ge 0. \tag{3.12b}$$

The solution of (3.12) is presented in [10, Proposition 4] and will be summarized here. We define a channel inversion power allocation

$$\psi(s, r) \quad = \quad \begin{cases} \frac{N_0}{s}\left(e^r - 1\right), & s > 0, r > 0 \\ 0, & r = 0 \end{cases}. \tag{3.13}$$

which represents the minimum power required to communicate reliably at the rate r for the channel state s. Let

$$r_l^{\dagger} = \max\left\{ r \middle| \int_{\mathcal{V}_l} \psi(s, r) f_l(s)\, ds \le P_l \right\}, \tag{3.14}$$

denote the largest possible assigned rate over $\mathcal{V}_l$ without outage given the average power allocation P_l over $\mathcal{V}_l$. If $r_l \le r_l^{\dagger}$, the solution of (3.12) is trivial; we allocate power $\psi(s, r_l)$ and achieve zero outage over $\mathcal{V}_l$. On the other hand, if $r_l > r_l^{\dagger}$, outage within $\mathcal{V}_l$ is inevitable and the corresponding optimum power allocation is the truncated channel inversion

$$p(s) = \begin{cases} 0, & s \in \mathcal{V}_l \backslash \bar{\mathcal{V}}_l \\ \psi(s, r_l), & s \in \bar{\mathcal{V}}_l \end{cases}, \tag{3.15}$$

where $\bar{\mathcal{V}}_l \subset \mathcal{V}_l$ is a subset of better channel states in $\mathcal{V}_l$, i.e., $s_1 \in \mathcal{V}_l \backslash \bar{\mathcal{V}}_l$ and $s_2 \in \bar{\mathcal{V}}_l$ implies $s_1 < s_2$. Moreover, the set $\bar{\mathcal{V}}_l$ is chosen to satisfy the average power constraint

$$P_l = \int_{\bar{\mathcal{V}}_l} \psi(s, r_l) f_l(s)\, ds. \tag{3.16}$$

Note that $\psi(s, r)$ is not defined for $s = 0$ and $r > 0$. This is due to the fact that, in the vicinity of $s = 0$, the channel is too poor to support any positive rate. Thus, the following must hold: $r(0) = 0$ and $p(0) = 0$.

In the optimization problem (3.11), we assume that r_l is known. However, for any $\mathcal{V}_l$, given the conditional average power P_l, we can also choose r_l to maximize the conditional ART. Specifically, local optimality implies that the optimum rate/power allocation must solve

$$\max_{r_l} \max_{p(s)} \quad r_l \int_{\mathcal{V}_l} [1 - I_{\text{out}}(r_l, s, p(s))]\, f_l(s)\, ds, \tag{3.17}$$

$$\text{subject to} \quad \int_{\mathcal{V}_l} p(s) f_l(s)\, ds \le P_l. \tag{3.17a}$$

$$p(s) \ge 0. \tag{3.17b}$$

Since rate $r_l^\dagger$ defined by (3.14) is achievable under (3.17), any rate $r_l < r_l^\dagger$ is suboptimal for (3.17). For any $r_l' > r_l^\dagger$, the outage probability within $\mathcal{V}_l$ is nonzero. Thus, for any optimum rate r_l, the optimum power allocation in (3.17) has the form (3.15). Since there is no transmission for $s \in \mathcal{V}_l \backslash \bar{\mathcal{V}}_l$, we can incorporate this set of channel states into the set of zero-power zero-rate channel states $\mathcal{V}_0 = \{s | r(s) = 0\}$ and redefine $\mathcal{V}_l = \bar{\mathcal{V}}_l$. For the new policy, there is no outage for states $s \in \mathcal{V}_0$; we reliably and trivially achieve zero rate by using zero power. Moreover, the assigned rate $r(s)$ and the corresponding power $p(s)$ satisfy

$$p(s) = \psi(s, r_l), \qquad s \in \mathcal{V}_l, \tag{3.18}$$

$$r(s) = \log(1 + p(s)s). \tag{3.19}$$

Note that (3.18) and (3.19) are consistent with the definition of $\mathcal{V}_l$ which implies that $s \in \mathcal{V}_l$ whenever $r(s) = r_l$. Consequently, it is clearly sufficient to denote any optimum policy by $(\mathcal{R}, r(s))$. Moreover, (3.19) and (3.10d) imply (3.10e).

From (3.5), (3.7), and (3.8), the corresponding overall average power and ART are

$$\rho_L(\mathcal{R}, r(s)) = \sum_{l=1}^{L} \int_{\mathcal{V}_l} \psi(s, r_l) f(s)\, ds, \tag{3.20}$$

$$R_L(\mathcal{R}, r(s)) = \sum_{l=1}^{L} \int_{\mathcal{V}_l} r_l f(s)\, ds. \tag{3.21}$$

Our objective is to maximize $R_L(\mathcal{R}, r(s))$ subject to both the average power constraint $\rho_L(\mathcal{R}, r(s)) \le \overline{p}$ and the constraint that the rate set $\{r(s)\}$ has a cardinality $L+1$ and includes the zero rate $r_0 = 0$.

Even though (3.21) is much simpler than (3.5), a simple solution is not available. In the next two sections, we will obtain necessary and sufficient conditions for optimum policies by applying the Lagrange multiplier method and the Karush-Kuhn-Tucker conditions [7]. Given these conditions, the optimum policies can be found by a relatively simple search method.

3.3 Partition Optimization

In this section, we address the subproblem of finding the optimum $r(s)$ given a specific rate set $\mathcal{R}_0$. This problem is of interest since a valid strategy for designing adaptive transmission systems is to choose a subset of good error control codes before deciding $r(s)$, i.e., the channel state partition $\{\mathcal{V}_l\}$. Furthermore, since this subproblem and the MQAM spectral efficiency maximization problem in [21] (solved only in a suboptimal manner in [21]) are closely related, the optimum solution presented here is also the optimum solution of the problem in [21].

For a given rate set $\mathcal{R}$ including L distinct positive rates, the throughput maximization problem (3.10) becomes

$$\max_{r(s)} \quad R_L(\mathcal{R}, r(s)), \tag{3.22}$$

$$\text{subject to} \quad \rho_L(\mathcal{R}, r(s)) \le \overline{p}, \tag{3.22a}$$

$$r(s) \in \mathcal{R}_0. \tag{3.22b}$$

In comparison with the problem (3.10), (3.22) has fewer constraints. Specifically, given $\mathcal{R}$, with (3.22b), (3.10b) and (3.10e) are redundant. In addition, with (3.22a), we implicitly take $p(s)$ in the form specified in (3.18), which further implies that (3.10d) is automatically satisfied.

The maximization problem (3.22) is a variation of the bit-loading problem in [11].

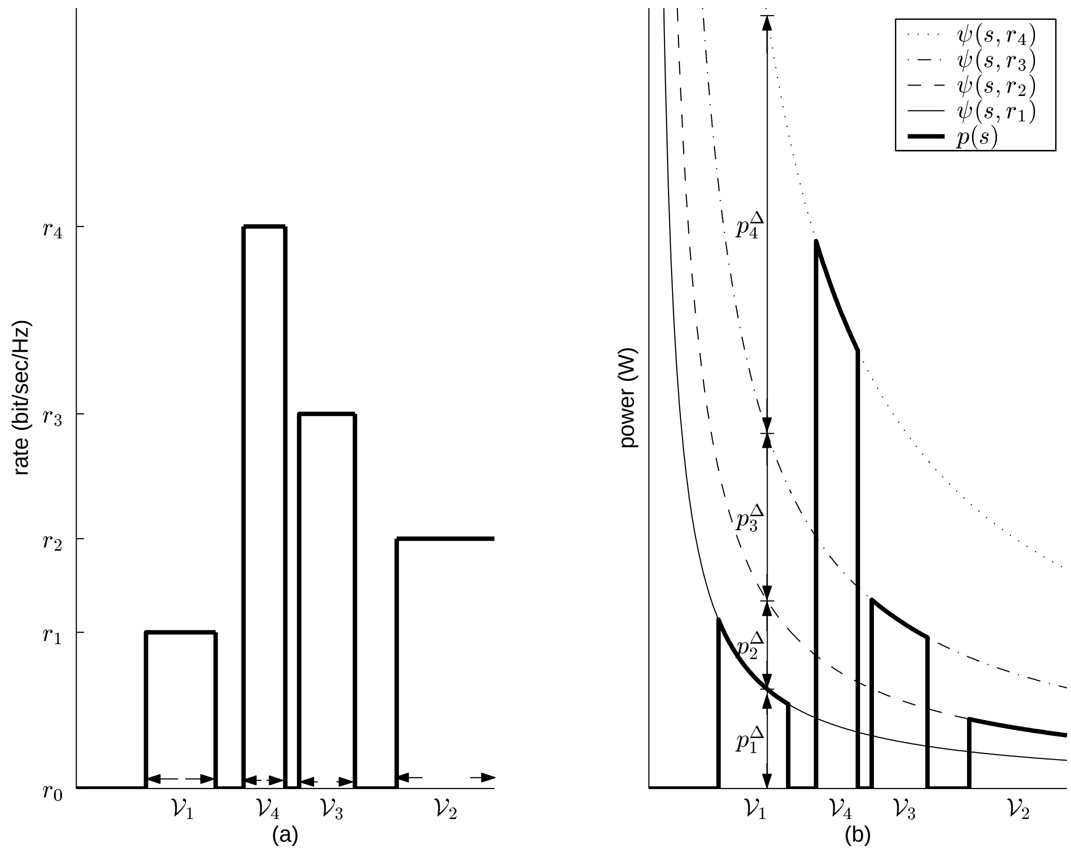

Figure 3.1: Illustration of an arbitrary policy satisfying (3.18) and (3.19). Note that $\mathcal{V}_0 = \mathcal{S} \backslash \cup_{l=1}^{4} \mathcal{V}_l$ is not shown. (a) rate assignment, (b) power allocation.

Both problems belong to the general class of Knapsack Problems (KPs) [31]. The traditional bit-loading problem is to optimize the rate/power allocation over a finite number of parallel channels where any rate assigned to a channel can only be an integer. In this work, the rate/power allocation is over $\mathcal{S}$, which is an uncountable set. Furthermore, elements of $\mathcal{R}$ are not necessarily integers.

In [11] (see also [12]), it is proved that a greedy rate/power allocation is the optimum solution for the bit-loading problem in parallel channels. We will show that the same allocation is indeed optimum for (3.22). Moreover, such a solution implies that quantization is the optimum partition. The optimum quantization partition boundaries are also derived.

Fig. 3.1 depicts a policy with an arbitrary, and not particularly intelligent, rate and partition assignment. The wave-like power assignment in Fig. 3.1(b) is the result of the local optimization in (3.17). The average rate $R_L(\mathcal{R}, r(s))$ in (3.21) is the integral of the rate pulses in Fig. 3.1(a) weighted by the channel state PDF $f(s)$. Similarly, the average power $\rho_L(\mathcal{R}, r(s))$ in (3.20) is the integral of wave crests of Fig. 3.1(b) weighted by $f(s)$. To express these integrals in a more useful form, we introduce the incremental rate and power functions

$$\begin{aligned} r_l^{\Delta} &= r_l - r_{l-1}, && (3.23) \\ p_l^{\Delta}(s) &= \psi(s, r_l) - \psi(s, r_{l-1}), && (3.24) \end{aligned}$$

for $l = 1, \ldots, L$. Since elements in $\mathcal{R}$ are a strictly increasing sequence, the incremental rates r_l^{Δ} and power $p_l^{\Delta}(s)$ are both positive.

At any channel state s, the transmitted code rate and the corresponding required transmitted power assignments can be expressed as sums of r_l^{Δ} and $p_l^{\Delta}(s)$,

$$r(s) = \sum_{l=1}^{L} I_l(s) r_l^{\Delta}, \tag{3.25}$$

$$p(s) = \sum_{l=1}^{L} I_l(s) p_l^{\Delta}(s), \tag{3.26}$$

where the coefficients, $\{I_l(s) | \, l = 1, \ldots, L\}$, in (3.25) and (3.26) is a set of binary 0/1-value functions. Since $r_l \geq r_{l'}$ for $l' = 1, \ldots, l$, (3.25) implies that

$$\mathcal{V}_l = \left\{ s \middle| I_{l'}(s) = 1, l' \leq l, \text{ and } I_{l'}(s) = 0, l' > l \right\}. \tag{3.27}$$

A given rate set $\mathcal{R}$ specifies the incremental rates $\{r_l^{\Delta}\}$ and the incremental powers $\{p_l^{\Delta}(s)\}$. Thus, given $\mathcal{R}$, $\mathcal{I} = \{I_l(s) | \, l = 1, \ldots, L\}$ describes a policy of interest. The corresponding ART and average power are

$$\begin{aligned} R_L(\mathcal{I}) &= \sum_{l=1}^{L} \int_0^{\infty} I_l(s) r_l^{\Delta} f(s) \, ds, && (3.28) \\ \rho_L(\mathcal{I}) &= \sum_{l=1}^{L} \int_0^{\infty} I_l(s) p_l^{\Delta}(s) f(s) \, ds. && (3.29) \end{aligned}$$

And the maximization problem (3.22) becomes a problem of searching for the optimum $\mathcal{I}$. Note that (3.27) requires a valid policy to satisfy the precedence constraint

$$I_{l'}(s) \geq I_l(s), \qquad l' \leq l. \tag{3.30}$$

The precedence constraint simply says that if $I_l(s) = 1$, then $I_{l'}(s) = 1$ for all $l' \leq l$. In addition, if $I_l(s) = 0$, then $I_{l'}(s) = 0$ for all $l' > l$.

In order to identify a policy that provides an optimum solution to the throughput maximization problem (3.22), we introduce the *incremental efficiency* (or, simply, *efficiency*)

$$\eta_l(s) = \frac{r_l^{\Delta} f(s)}{p_l^{\Delta}(s) f(s)} = \frac{s}{N_0} \frac{r_l - r_{l-1}}{e^{r_l} - e^{r_{l-1}}}, \tag{3.31}$$

which is a ratio between an increment in the throughput from $r_{l-1} f(s)$ to $r_l f(s)$ at state s and the corresponding power expenditure $p_l^{\Delta}(s) f(s)$. For optimization with integer rates, the efficiency concept may not be necessary [12]. Nevertheless, it is a key for solving the problem with non-integer code rates.

Lemma 4 *The incremental efficiency $\eta_l(s)$ has the following properties:*

(a) $\eta_l(0) = 0$ *for all* l.

(b) For fixed l, $\eta_l(s)$ *increases in* s.

(c) For fixed s, $\eta_l(s)$ *decreases in* l.

ART can now be expressed as

$$R_L(\mathcal{I}) = \sum_{l=1}^{L} \int_0^{\infty} I_l(s) \eta_l(s) p_l^{\Delta}(s) f(s)\, ds. \tag{3.32}$$

Together with the constraint (3.29), the maximization of $R_L(\mathcal{I})$ forms a Knapsack Problem [31], which is solved by the following policy.

Definition 1 The *most power efficient quantization* (MPEQ) is a policy $\mathcal{I}^* = \{I_l^*(s)\}$ where

$$I_l^*(s) = \begin{cases} 1, & \eta_l(s) \geq \lambda_{\mathrm{M}}(\mathbf{r}), \\ 0, & otherwise, \end{cases} \qquad l = 1, \ldots, L, \tag{3.33}$$

and the positive constant $\lambda_{\mathrm{M}}(\mathbf{r})$ is determined by the average power constraint $\rho_L(\mathcal{I}^*) = \overline{p}$.

Note that the uniqueness of the efficiency lower bound $\lambda_{\mathrm{M}}(\mathbf{r})$ and MPEQ is an issue discussed later in the subsection.

For any MPEQ, $s \in \mathcal{V}_{l-1}$ and $s' \in \mathcal{V}_l$ imply that for some $\lambda_{\mathrm{M}}(\mathbf{r}) > 0$, $I_l^*(s) = 0$ and $I_l^*(s') = 1$, i.e., $\eta_l(s) < \eta_l(s')$ and, consequently, $s < s'$. Thus, MPEQ leads to a quantization in the sense that $s \in \mathcal{V}_{l-1}$ and $s' \in \mathcal{V}_l$ imply that $s < s'$. In addition, Lemma 4(c) guarantees that MPEQ satisfies the precedence constraint (3.30).

Theorem 4 *For any policy $\mathcal{I} = \{I_l(s)\}$ satisfying the average power constraint $\rho_L(\mathcal{I}) \leq \overline{p}$,*

$$R_L(\mathcal{I}) \leq R_L(\mathcal{I}^*) \tag{3.34}$$

We emphasize that Theorem 4 holds regardless of whether the policy $I_l(s)$ satisfies the precedence constraint (3.30) or not.

The MPEQ solution is easy to obtain analytically. An MPEQ policy is fully determined by the efficiency threshold $\lambda_{\mathrm{M}}(\mathbf{r})$ where $r(s) \geq r_l$ iff $\eta_l(s) \geq \lambda_{\mathrm{M}}(\mathbf{r})$. From Lemma 4 and the continuity of the domain $\mathcal{S}$, there exists q_l such that $\eta_l(q_l) = \lambda_{\mathrm{M}}(\mathbf{r})$ and $\eta_l(s) \geq \lambda_{\mathrm{M}}(\mathbf{r})$ iff $s \geq q_l$.

Thus, q_l is the boundary separating $\mathcal{V}_{l-1}$ and $\mathcal{V}_l$. It follows from (3.31) and the equality $\eta_l(q_l) = \lambda_{\mathrm{M}}(\mathbf{r})$ that

$$q_l = \frac{\lambda_{\mathrm{M}}(\mathbf{r}) N_0 \left(e^{r_l} - e^{r_{l-1}}\right)}{r_l - r_{l-1}}, \qquad l = 1, \ldots, L. \tag{3.35}$$

Since $\eta_l(s)$ decreases in l for fixed s, (3.35) is an increasing sequence of L boundaries $q_1, \ldots, q_L$. Moreover, these boundaries indicate that for a given set of rates $\mathcal{R}$, an optimum policy is given by a quantization of the channel state set $\mathcal{S}$ into exactly $L+1$ intervals corresponding to the set of rates $\mathcal{R}$. Thus, an optimum policy for a given $\mathcal{R}$ can be represented by the vector

$$\mathbf{q} = [q_1, \ldots, q_L]^\top. \tag{3.36}$$

An MPEQ policy is illustrated in Fig. 3.2, where $q_0 = 0$.

To study the uniqueness of MPEQ, we assume $\mathcal{I}^* = \{I_l^*(s)\}$ and $\mathcal{I}^{*\prime} = \left\{I_l^{*\prime}(s)\right\}$ be two MPEQs corresponding to $\lambda_{\mathrm{M}}(\mathbf{r})$ and $\lambda'_{\mathrm{M}}(\mathbf{r})$, respectively. The corresponding boundaries of the partition intervals are $\mathbf{q}$ and $\mathbf{q}' = [q'_1, \ldots, q'_L]^\top$.

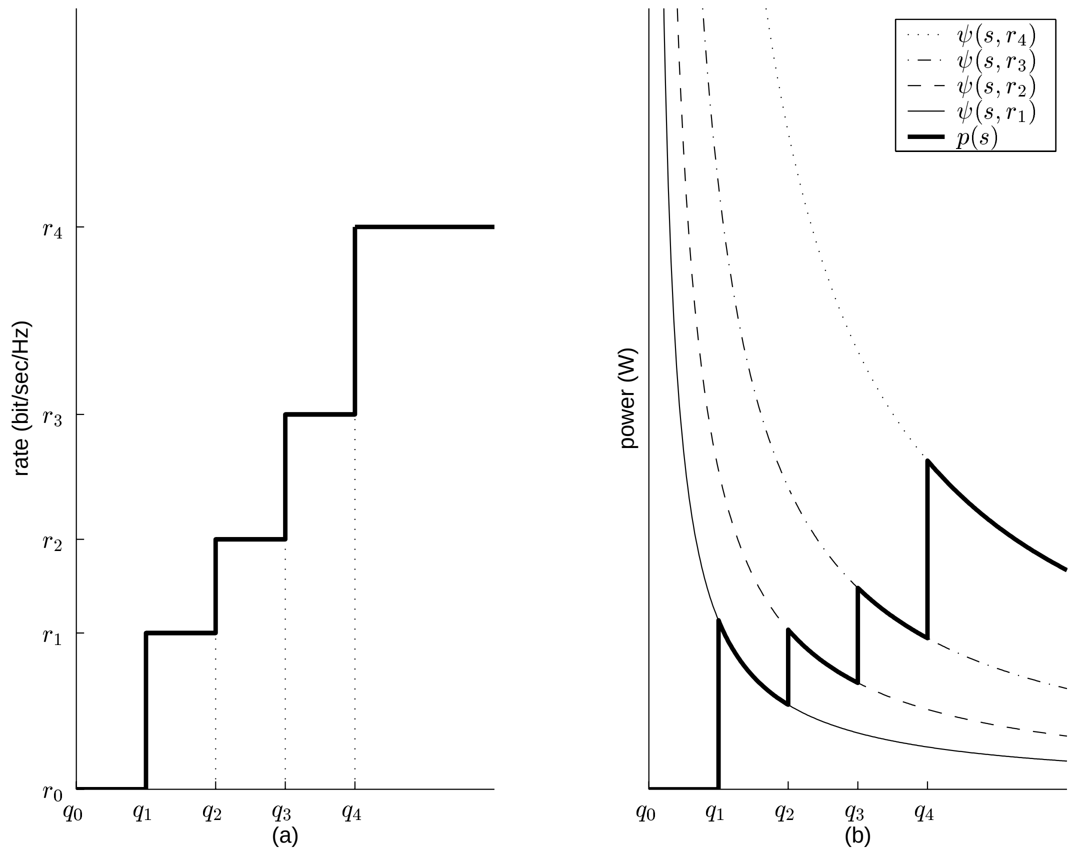

Figure 3.2: Illustration of an MPEQ policy. (a) rate staircase, (b) power staircase.

Proposition 1 *If* $\lambda_{\mathrm{M}}(\mathbf{r}) > \lambda'_{\mathrm{M}}(\mathbf{r})$*, we have*

1) if $I_l^*(s) = 1$ *then* $I_l^{*'}(s) = 1$,

2) $\rho_L(\mathcal{I}^*) \leq \rho_L(\mathcal{I}^{*\prime})$,

3) $q_l' < q_l$ *for all* $l = 1, \ldots, L$,

4) $\rho_L(\mathcal{I}^*) = \rho'_L(\mathcal{I}^{*\prime})$ *if* $f(s) = 0$ *for all* $s \in [q_l', q_l), l = 1, \ldots, L$.

The claim 4) in Proposition 1 implies the following theorem.

Theorem 5 *MPEQ is unique if* $f(s) > 0$ *for* $s \in \mathcal{S} \setminus \{0\}$.

3.3.1 MQAM Spectral Efficiency Maximization

Problem (3.22) is closely related to a special case of the spectral efficiency maximization problem for an adaptive MQAM system [21], which suggests employing a set of $L+1$ predetermined MQAM constellations of sizes in $\mathcal{M} = \{M_0 = 1, M_1, \ldots, M_L\}$. The

$M_0 = 1$ constellation corresponds to turning off the transmitter. The transmission scheme requires that given a channel state s in $\mathcal{V}_l$, a transmitter transmits a QAM symbol constellation of size $M(s) = M_l,\ M_l \in \mathcal{M}$, with power $p(s)$.

Without loss of generality, we assume $M_{l-1} < M_l$ for $l = 1, \ldots, L$. To guarantee a specified bit error rate (BER) P_b for all channel states, the transmitted power is

$$p(s) = \frac{M(s) - 1}{Ks}, \tag{3.37}$$

where $K = -1.5/\log(5P_\mathrm{b})$ [21].

Let $r_l = \log(M_l),\ l = 0, \ldots, L$. Consequently, $r(s) = \log\left(1 + p(s)sK\right)$. The spectral efficiency maximization problem for $L+1$ QAM signaling and a continuously varying power is as follows

$$\max_{r(s)} \quad R_L(\mathcal{R}, r(s)), \tag{3.38}$$

$$\text{subject to} \quad \int_{\mathcal{S}} p(s) f(s)\, ds \le \overline{p}, \tag{3.38a}$$

$$r(s) \in \mathcal{R}_0, \tag{3.38b}$$

$$p(s) = \frac{e^{r(s)} - 1}{Ks} = \frac{\psi(s, r(s))}{KN_0}, \tag{3.38c}$$

where $\mathcal{R}_0 = \{r_0\} \cup \mathcal{R}$ and $\mathcal{R} = \{r_1, \ldots, r_L\}$. Following the procedure derived in this section, we obtain the optimum partition as a quantization with boundaries

$$q_l = \frac{\lambda\left(e^{r_l} - e^{r_{l-1}}\right)}{K(r_l - r_{l-1})} \tag{3.39}$$

$$= \frac{M_l - M_{l-1}}{K(\log(M_l) - \log(M_{l-1}))}, \qquad l = 1, \ldots, L, \tag{3.40}$$

where λ is determined by the average power constraint. Note that q_l in (3.40) are typically not the same as those suboptimal partition boundaries provided in [21] for any adaptive MQAM systems. An evaluation of this policy in Rayleigh fading is given in Section 3.5.3.

3.4 Optimum Policies

In the previous sections, we have demonstrated that an ART-maximizing policy, defined in Section 3.2, with a rate set $\mathcal{R}_0$ must be an MPEQ corresponding to $\mathcal{R}_0$. More specifically, Theorem 4 shows that any candidate ART-maximizing policy with a rate set $\mathcal{R}_0$ must have channel states partitioned into $|\mathcal{R}_0|$ intervals. Furthermore, we found that the optimum solution employs rate $r(s) = r_l \in \mathcal{R}_0$ in the interval $\mathcal{Q}_l = [q_l, q_{l+1})$, where q_l is defined in (3.35) with $q_0 = 0$ and $q_{|\mathcal{R}_0|} = \infty$. In addition,

$$p(s) = \psi(s, r_l), \qquad s \in \mathcal{Q}_l,\ l = 0, 1, \ldots, |\mathcal{R}|. \tag{3.41}$$

Consequently, any policy of interest can be specified by $(\mathbf{r}, \mathbf{q})$ where the vectors

$$\mathbf{r} = [r_1, \ldots, r_{|\mathcal{R}|}]^\top, \tag{3.42}$$

$$\mathbf{q} = [q_1, \ldots, q_{|\mathcal{R}|}]^\top, \tag{3.43}$$

with the corresponding $p(s)$ given by (3.41). An optimum policy is denoted by $(\mathbf{r}^*, \mathbf{q}^*)$.

Clearly, for a policy specified by $(\mathbf{r}, \mathbf{q})$, if the q_l are not distinct, the policy degrades to a policy with a number of distinct rates smaller than $|\mathcal{R}_0|$. Consequently, without loss of generality, we concentrate on $\mathbf{q}$ with $|\mathcal{R}|$ distinct elements.

For an $L = |\mathcal{R}|$ level policy $(\mathbf{r}, \mathbf{q})$, the corresponding ART is

$$R_L(\mathbf{r}, \mathbf{q}) = \sum_{l=1}^{L} r_l F_l, \tag{3.44}$$

where

$$F_l = \Pr[s \in \mathcal{Q}_l]. \tag{3.45}$$

The conditional average power in the interval $\mathcal{Q}_l$ is

$$P_l(\mathbf{r}, \mathbf{q}) = \int_{\mathcal{Q}_l} \psi(s, r_l) f_l(s)\, ds, \tag{3.46}$$

where the conditional PDF of channel state s is

$$f_l(s) = \begin{cases} f(s)/F_l, & s \in \mathcal{Q}_l \\ 0, & \text{otherwise} \end{cases}. \tag{3.47}$$

For an interval $\mathcal{Q}_l$, the conditional average channel quality ω_l is defined by

$$\frac{1}{\omega_l} = \int_{\mathcal{Q}_l} \frac{N_0}{s} f_l(s)\, ds, \qquad l = 1, \ldots, L. \tag{3.48}$$

Here, ω_l is normalized by N_0 in order to simplify the latter derivations and $0 < w_1 < \ldots < w_L$. Equations (3.13), (3.46), and (3.48) imply

$$P_l(\mathbf{r}, \mathbf{q}) \;=\; \frac{(e^{r_l} - 1)}{\omega_l}. \tag{3.49}$$

The original ART maximization problem (3.10) now becomes

$$\max_{\mathbf{r},\mathbf{q}} \quad R_L(\mathbf{r}, \mathbf{q}) \tag{3.50}$$

$$\text{subject to} \quad \rho_L(\mathbf{r}, \mathbf{q}) \leq \overline{p}, \tag{3.50a}$$

where

$$R_L(\mathbf{r}, \mathbf{q}) \;=\; \sum_{l=1}^{L} F_l r_l = \sum_{l=1}^{L} r_l \int_{\mathcal{Q}_l} f(s)\, ds, \tag{3.51}$$

$$\rho_L(\mathbf{r}, \mathbf{q}) \;=\; \sum_{l=1}^{L} F_l P_l(\mathbf{r}, \mathbf{q}) = \sum_{l=1}^{L} (e^{r_l} - 1) \int_{\mathcal{Q}_l} \frac{N_0}{s} f(s)\, ds. \tag{3.52}$$

We note that problem (3.50) can have multiple local maxima. For example, when $L = 1$,

$$\rho_L(\mathbf{r}, \mathbf{q}) \to \rho_1(r_1, q_1) = (e^{r_1} - 1) \int_{q_1}^{\infty} \frac{N_0}{s} f(s)\, ds. \tag{3.53}$$

Now, since $\rho_1(r_1, q_1) = \overline{p}$ is necessary to achieve the maximum ART, r_1 must be a function of q_1 defined by

$$r_1 = \log\left(1 + \frac{\overline{p}}{\int_{q_1}^{\infty} \frac{N_0}{s} f(s)\, ds}\right). \tag{3.54}$$

Consequently,

$$R_L(\mathbf{r}, \mathbf{q}) \to R_1(r_1, q_1) \to R_1(q_1) = \log\left(1 + \frac{\overline{p}}{\int_{q_1}^{\infty} \frac{N_0}{s} f(s)\, ds}\right) \int_{q_1}^{\infty} f(s)\, ds. \tag{3.55}$$

An illustration of $R_1(q_1)$ with multiple local maxima is shown in Fig. 3.3. Note that in Fig. 3.3, the $f(s)$ violates the continuity assumption in Section 3.2. However, it is not very hard to imagine that a continuous $f(s)$ similar to that in Fig. 3.3 will lead to a similar scenario with multiple local maxima.

3.4.1 Water-Filling Power Allocation

According to (3.49), given a channel state partition $\mathbf{q}$ that specifies the conditional average channel quality ω_l, the rate r_l is uniquely expressed in terms of the conditional average power, P_l, in the interval $\mathcal{Q}_l$:

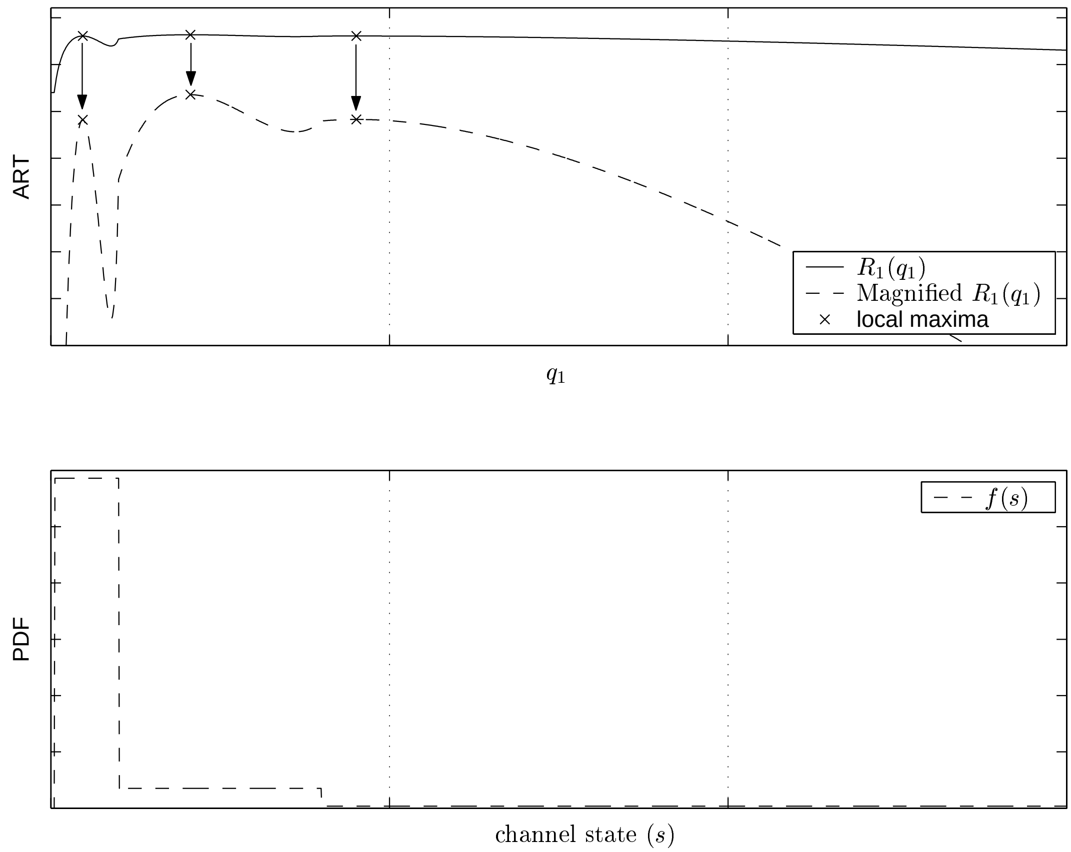

Figure 3.3: A $R_1(q_1)$ with a simple piecewise linear $f(s)$ has three local maxima. Arrows show the corresponding local maxima before and after magnification.

$$r_l = \log\left(1 + P_l \omega_l\right). \tag{3.56}$$

Given a fixed $\mathbf{q}$, the ART maximization problem (3.50) can be written as

$$\max_{P_1,\ldots,P_L} \quad \sum_{l=1}^{L} \log\left(1 + P_l \omega_l\right) F_l \tag{3.57}$$

$$\text{subject to} \quad \sum_{l=1}^{L} P_l F_l \leq \overline{p}, \tag{3.57a}$$

$$P_l \geq 0, \; l = 1, \ldots, L. \tag{3.57b}$$

It is straightforward to show using the Karush-Kuhn-Tucker (KKT) conditions that the optimum solution of (3.57) for given $\mathbf{q}$ is

$$P_l'(\mathbf{q}) = \left(\frac{1}{\lambda_{\mathrm{W}}(\mathbf{q})} - \frac{1}{\omega_l}\right)^+, \tag{3.58}$$

where $1/\lambda_{\mathrm{W}}(\mathbf{q})$ is the water-filling level which satisfies

$$\overline{p} = \sum_{l=1}^{L} F_l P_l'(\mathbf{q}). \tag{3.59}$$

We have shown that an optimum policy $(\mathbf{r}^*, \mathbf{q}^*)$ must have $|\mathcal{R}| = L$. Hence, (3.56) implies that $P_l^* = P_l'(\mathbf{q}^*) > 0$ for $l = 1, \dots, L$. That is, for optimum policies, the operator $(\cdot)^+$ in (3.58) should not have any impact. This in turn enforces the implicit requirement on $\mathbf{q}^*$ in the form of $\lambda_{\mathrm{W}}(\mathbf{q}^*) < \omega_l^*$ for $l = 1, \dots, L$, for policies of interest.

Consequently, (3.59) implies

$$\frac{1}{\lambda_{\mathrm{W}}(\mathbf{q})} = \frac{\overline{p} + \int_{q_1}^{\infty} \frac{N_0}{s} f(s)\, ds}{1 - F(q_1)}. \tag{3.60}$$

and (3.56) implies that the optimum rates corresponding to $\mathbf{q}$ are given by

$$r_l' = \log\left(\frac{\omega_l}{\lambda_{\mathrm{W}}(\mathbf{q})}\right), \tag{3.61}$$

Since ω_l characterizes the channel states of interval $\mathcal{Q}_l$, the water-filling result is analogous to those in both the original continuous adaptive transmission problem [22] and the parallel Gaussian channel problem [16].

Note that the water-filling power allocation (3.58) is optimum within the set of policies $(\mathbf{r}, \mathbf{q})$ that explicitly require $p(s)$ in the form of (3.41). For a poorly chosen partition $\mathbf{q}$, $(\mathbf{r}, \mathbf{q})$ policies with $p(s)$ given by (3.41) may not be even locally optimum within the interval $\mathcal{Q}_l$. Therefore, the water-filling power allocation (3.58) must be used with some caution.

Moreover, because $R_L(\mathbf{r}, \mathbf{q})$ can have multiple local maxima, ascent techniques based alternating optimization of the partition $\mathbf{q}$ and rates $\mathbf{r}$ can at best only reach a local maxima. In the following, we show how the necessary conditions for optimality and q_1^*, the first element of $\mathbf{q}^*$ uniquely specify an optimum policy. Thus, the search for optimum policies is reduced to a line search over q_1.

3.4.2 Necessary Conditions for Optimum Policies

The corresponding Lagrangian function for (3.50) is

$$L(\mathbf{r}, \mathbf{q}, \lambda) = R_L(\mathbf{r}, \mathbf{q}) + \lambda[\overline{p} - \rho_L(\mathbf{r}, \mathbf{q})]. \tag{3.62}$$

The Lagrangian function has a positive multiplier λ. The unit of the Lagrangian function is the same as ART.

Given fixed $\mathbf{r}$ and variable $\mathbf{q}$, (3.50) becomes the partition optimization problem studied in Section 3.3. Formally, the partial derivative of the Lagrangian function with respect to q_l is

$$\frac{\partial L(\mathbf{r}, \mathbf{q}, \lambda)}{\partial q_l} = f(q_l) \left\{ -(r_l - r_{l-1}) + \lambda \left(e^{r_l} - e^{r_{l-1}} \right) \frac{N_0}{q_l} \right\}. \tag{3.63}$$

Note that it is here where $F(s)$ is required to be continuous and differentiable. However, this is not required for the MPEQ in Section 3.3. The necessary conditions for optimality

$$\left. \frac{\partial L(\mathbf{r}, \mathbf{q}, \lambda)}{\partial q_l} \right|_{\mathbf{r}=\mathbf{r}^*, \mathbf{q}=\mathbf{q}^*, \lambda=\lambda^*} = 0, \qquad l = 1, \ldots, L, \tag{3.64}$$

imply

$$\lambda^* = \frac{r_l^* - r_{l-1}^*}{\frac{N_0}{q_l^*} \left(e^{r_l^*} - e^{r_{l-1}^*} \right)} = \lambda_{\mathrm{M}}(\mathbf{r}^*), \tag{3.65}$$

where the last equality is from (3.35). Therefore, the optimum Lagrange multiplier λ^* is equal to the incremental efficiency at q_l^* defined in (3.31).

On the other hand, a similar optimization problem can be formulated from (3.50) for given $\mathbf{q}$ and variable $\mathbf{r}$. This problem is the power allocation problem studied in Subsection 3.4.1. The necessary conditions for the optimum policies are

$$\left. \frac{\partial L(\mathbf{r}, \mathbf{q}, \lambda)}{\partial r_l} \right|_{\mathbf{r}=\mathbf{r}^*, \mathbf{q}=\mathbf{q}^*, \lambda=\lambda^*} = \left. F_l \left\{ 1 - \lambda e^{r_l} \frac{1}{\omega_l} \right\} \right|_{\mathbf{r}=\mathbf{r}^*, \mathbf{q}=\mathbf{q}^*, \lambda=\lambda^*} = 0, \qquad l = 1, \ldots, L, \tag{3.66}$$

implying

$$\lambda^* = e^{-r_l^*} \omega_l^* = \lambda_{\mathrm{W}}(\mathbf{q}^*), \tag{3.67}$$

where the last equality is due to (3.61). Therefore, the optimum Lagrange multiplier λ^* is the reciprocal of the water level of the water-filling power allocation obtained in Subsection 3.4.1.

Both (3.65) and (3.67) imply that for an optimum policy $(\mathbf{r}^*, \mathbf{q}^*)$,

$$\lambda_{\mathrm{M}}(\mathbf{r}^*) = \lambda_{\mathrm{W}}(\mathbf{q}^*). \tag{3.68}$$

For convenience, we define $\omega_0^* = \lambda_{\mathrm{W}}(\mathbf{q}^*)$. Although it is inconsistent with (3.48), the definition is conceptually appealing since $P_0 = 0$ can be obtained from (3.58). This is in that regardless how bad the channel quality is in the channel state interval $[0, q_1)$, the average channel quality can be always considered as good as $\lambda_{\mathrm{W}}(\mathbf{q}^*)$ in terms of optimizing power allocation. Thus, we obtain a set of necessary conditions for optimum policies as,

$$q_l^* = N_0 \lambda_{\mathrm{W}}(\mathbf{q}^*) \left[\frac{\left(e^{r_l^*} - e^{r_{l-1}^*} \right)}{r_l^* - r_{l-1}^*} \right] \tag{3.69}$$

$$\stackrel{(a)}{=} N_0 \frac{\omega_l^* - \omega_{l-1}^*}{\log\left(\omega_l^*\right) - \log\left(\omega_{l-1}^*\right)}, \qquad l = 1, 2, \ldots, L, \tag{3.70}$$

by substituting (3.61).

3.4.3 Search for Optimum Policies

Since ω_l^* and $\lambda_{\mathrm{W}}(\mathbf{q}^*)$ depend on $\mathbf{q}^*$, we have only L unknown $\mathbf{q}^*$ and L equations (3.70). The only inconvenience is that the equations of (3.70) are nonlinear and the q_l^* are limits of integrals defining ω_l^* and $\lambda_{\mathrm{W}}(\mathbf{q}^*)$. Therefore, we can not solve for q_l^* using direct substitution. However, in the following, based on the monotonic relationship between q_l^* and ω_l^*, a simple algorithm can be found to search for the optimum solution.

Naturally, the average channel quality ω_l becomes better when $\mathcal{Q}_l$ is enlarged by adding more good channel states. Such an intuition brings two monotonic relations below.

Lemma 5 *q_l^* is strictly increasing in ω_l^*.*

Lemma 6 *ω_{l-1} is strictly increasing in q_l.*

In particular, Lemma 6 follows from the following restatement of (3.48),

$$\frac{1}{\omega_{l-1}} = \int_{q_{l-1}}^{q_l} \frac{N_0}{s} f_l(s)\, ds, \qquad l = 1, \ldots, L. \tag{3.71}$$

These two lemmas are the building blocks of a constructive procedure for finding the optimum policy given q_1^* described in Fig. 3.4. In particular, in steps 1b and 2b, Lemma

1) $l = 1$:
   a) Given q_1^*, use (3.60) to uniquely solve for $\lambda_{\mathrm{W}}(\mathbf{q}^*)$ and, thus, obtain $w_0^* = \lambda_{\mathrm{W}}(\mathbf{q}^*)$;
   b) Given q_1^* and w_0^*, use (3.70) to uniquely solve for w_1^* [Lemma 5].

2) $l > 1$:
   a) Given q_{l-1}^* and ω_{l-1}^*, use (3.71) to uniquely solve for q_l^* [Lemma 6];
   b) Given q_l^* and ω_{l-1}^*, use (3.70) to uniquely solve for ω_l^* [Lemma 5].

3) Repeat step 2) from $l = 2$ to $l = L$, we will have $\mathbf{q}^*$.

Figure 3.4: An algorithm searching for optimum policies started from q_1^*.

5 implies the existence of a unique solution ω_l^*. Further, in step 2a, Lemma 6 implies the existence of a unique solution q_l^*.

Theorem 6 *For each value of q_1^*, there exists a unique policy $(\mathbf{r}^*, \mathbf{q}^*)$ that satisfies the necessary conditions (3.70).*

Overall, (3.48) and (3.70) offer $2L$ equalities. With q_1^*, we repeatedly use (3.48) and (3.70) to uniquely determine other q_l^* and all ω_l^*. Indeed, at the step $l = L$, given q_L^*, ω_L^* can be determined by either (3.70) or by (3.48) with $q_{L+1} = \infty$. These two approaches are distinct and must agree for the optimum policies. In this sense, the $2L$ equalities (3.48) and (3.70) implicitly restrict q_1^*. However, since (3.70) are only necessary conditions obtained from (3.68), it implies that all local optimum satisfy all $2L$ equalities, i.e., (3.48) and (3.70). Therefore, the optimum q_1^* will not be unique when multiple local maxima achieve the same objective value. In this case, there will be multiple optimum policies $(\mathbf{r}^*, \mathbf{q}^*)$.

Since there is only one undetermined parameter q_1^*, a line search over q_1 solves the problem of maximizing ART subject to a power constraint and a finite code rate set.

Finally, let the maximum ART for any L-level policy be

$$\bar{C}_L = R_L(\mathbf{r}^*, \mathbf{q}^*), \tag{3.72}$$

where $(\mathbf{r}^*, \mathbf{q}^*)$ is any L-level optimum policy. The following theorem verifies that increasing the number of levels L yields improved performance.

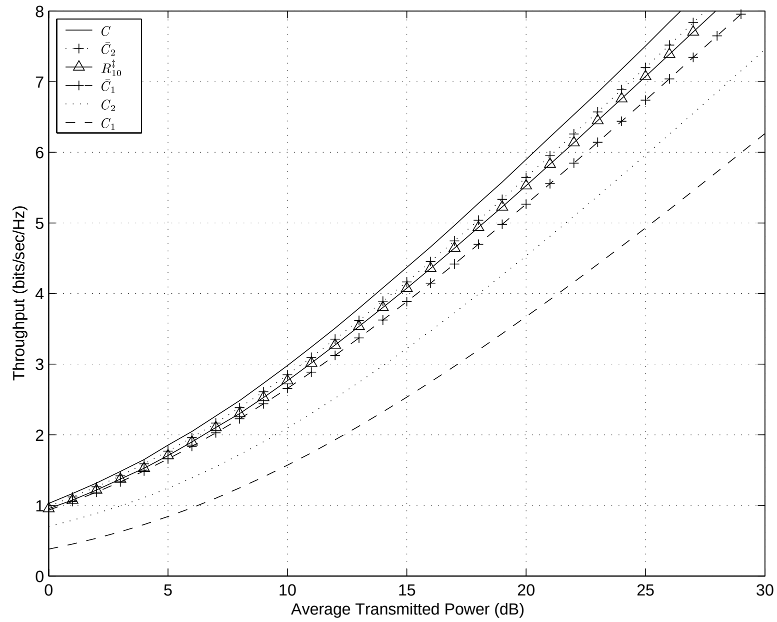

Figure 3.5: A comparison of C, C_L, $R^{\ddagger}_L$, and $\bar{C}_L$ for Rayleigh fading.

Theorem 7 *For all $L > 1$,*

$$\bar{C}_L > \bar{C}_{L-1}. \tag{3.73}$$

3.5 Numerical Results

In this section, following the approach in [10], we let $N_0 = 1$ and evaluate the system performance numerically.

3.5.1 Optimum Code Rates and Partition

Theorem 6 suggests that it is possible to perform a line-search on the single parameter q_1 to find $(\mathbf{r}^*, \mathbf{q}^*)$ and obtain $\bar{C}_L$. In Fig. 3.5, we present a comparison between $\bar{C}_L$ and several known throughputs for a Rayleigh fading channel. C is the ergodic capacity given in [22]. C_L, proposed in Chapter 2,is the maximum ART corresponding to a discrete adaptive transmission policy with L code rates and L power levels. Fig. 3.5, C_2 requires about 3 dB less average transmitted power than C_1 to achieve the same throughput.

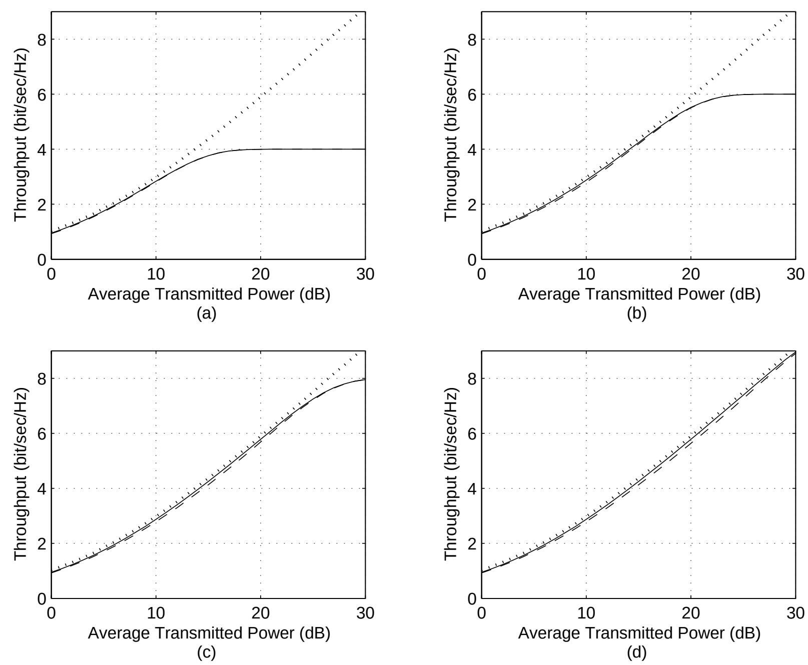

Figure 3.6: A comparison of partitions with $r_0 = 0$ and $r_l = 2l$, $l = 1, 2, \ldots, L-1$ for Rayleigh fading. (a) $L = 3$, (b) $L = 4$, (c) $L = 5$, and (d) $L = 6$. The dotted line indicates the ergodic capacity C, the solid line corresponds to the optimum partition, and the dash line corresponds to the suboptimal partition in [21].

Since C_L is obtained through exhaustive search, the required computation complexity increases exponentially with respect to L and it is not desirable to evaluate C_L for large L. In Chapter 2, by employing a greedy iterative algorithm initialized with an asymptotically optimum solution of C_L, a good lower bound $R_L^{\ddagger}$ of C_L is found. In Fig. 3.5, $R_{10}^{\ddagger}$ is about 1 dB away from C.

We observed that $\bar{C}_2 > R_{10}^{\ddagger}$, which suggests employing adaptive systems with a small number of code rates and a large number of power levels. Finally, it can be shown that $\bar{C}_4$ is only a fraction of a dB away from C for throughput values less than 8 bits/sec/Hz.

3.5.2 Optimum Partition for Preset Code Rates

Here, we have a comparison between the performance of the optimum partition and that of a partition motivated by the design in [21]. Given a code rate set $\mathbf{r}$ with $r_0 = 0$, a suboptimal partition is

$$q_l = \begin{cases} 0, & l = 0 \\ qe^{r_l}, & l = 1, 2, \ldots, L-1 \end{cases}, \tag{3.74}$$

where q is a parameter tuned to satisfy the power constraint (3.22a) since $p(s) = \psi(s, r_l),\ s \in \mathcal{Q}_l$.

Fig. 3.6 shows the results of two partition methods for

$$r_l = \begin{cases} 0, & l = 0 \\ 1, & l = 1 \\ 2(l-1), & l = 2, 3, \ldots, L-1 \end{cases}. \tag{3.75}$$

Using the optimum partition can reduce the transmitted power requirement by as much as 0.5 dB in comparison with using the other suboptimal partition for $L = 5$. However, for a small L, there is only a negligible difference in the performance of optimum and suboptimal partitions.

3.5.3 Spectral Efficiency

In Fig. 3.8, we show the results of spectral efficiency of adaptive MQAM systems with a $P_{\mathrm{b}} = 10^{-6}$. The results show that for 3 regions ($L = 3$), there is a negligible difference between the spectral efficiency corresponding to the optimum partition and the suboptimal partition [21]. The difference becomes distinguishable when there are more regions (larger L).

3.6 Conclusion

In this chapter, the average reliable throughput maximization problem for adaptive transmission systems with a finite number of code rates and continuously varying power level is formulated. By exploring the properties of the optimum policies, we can obtain $\bar{C}_L$ through a simple line-search algorithm over one parameter q_1.

Moreover, while studying the properties of the optimum policies, we discovered the optimum partition given an increasing rate assignment. This is particularly useful for designing adaptive systems with the channel codes selected from a limited set of good

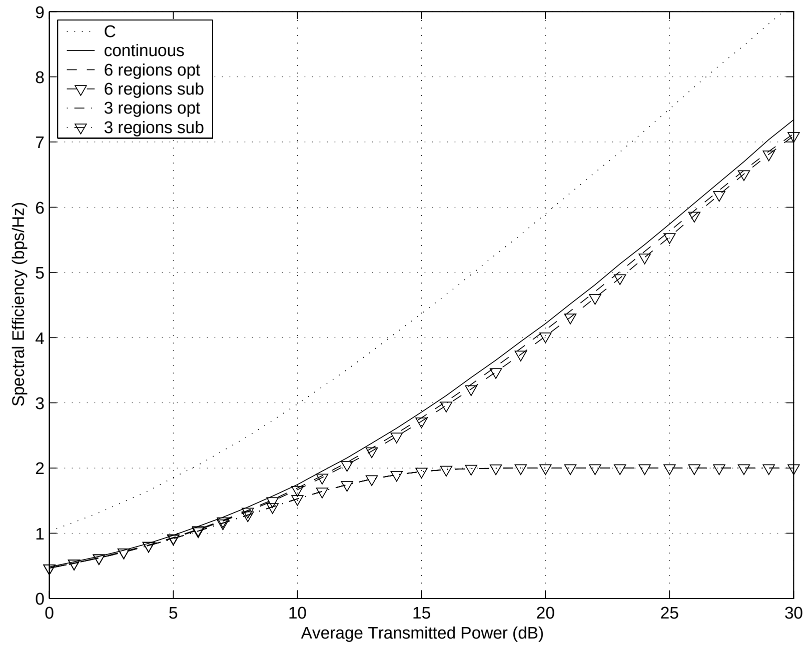

Figure 3.7: Comparison of spectral efficiency obtained by the optimum partition and the suboptimal partition in [21] for $P_{\mathrm{b}} = 10^{-3}$.

codes. The approach used in solving our maximization problem can be applied to solve a spectrum efficiency maximization problem formulated in [21].

3.A Proofs

Proof: Lemma 4

Claim (a) is straightforward since, regardless how much the transmitted power is, we cannot have non-zero rate for reliable communication when $s = 0$.

Combining (3.13), (3.24), (3.23), and (3.31), we have

$$\eta_l(s) = \frac{s}{N_0} \frac{r_l - r_{l-1}}{e^{r_l} - e^{r_{l-1}}}. \tag{3.76}$$

Hence, claim (b) follows.

For claim (c), we notice that e^{r_l} is strictly convex in r_l. For any strictly convex function $g(x)$, we have for $x_2 > x_1$

$$\left.\frac{\partial g(x)}{\partial x}\right|_{x=x_1} < \frac{g(x_2) - g(x_1)}{x_2 - x_1} < \left.\frac{\partial g(x)}{\partial x}\right|_{x=x_2}. \tag{3.77}$$

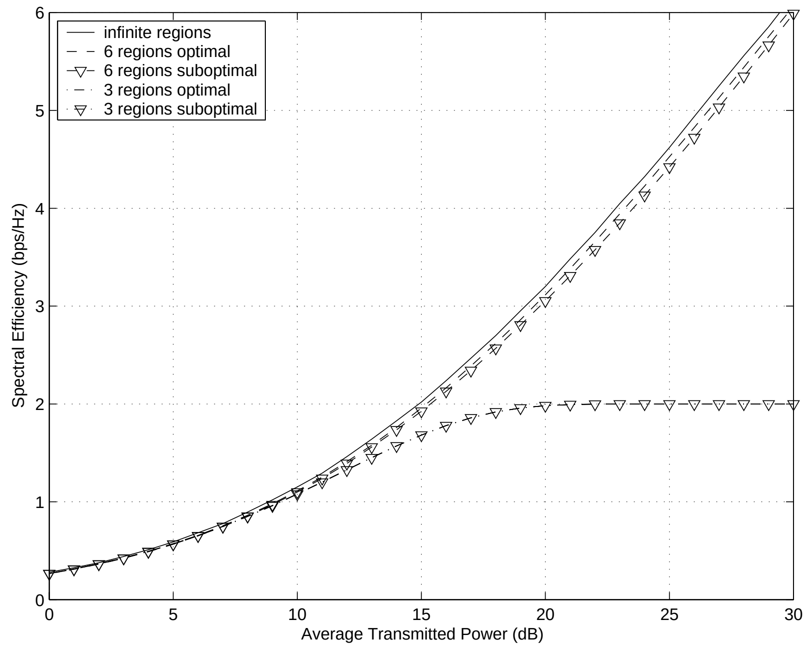

Figure 3.8: Comparison of spectral efficiency obtained by the optimum partition and the suboptimal partition in [21] for $P_{\mathrm{b}} = 10^{-6}$.

Consequently, for any $x_1 < x_2 < x_3$,

$$\frac{g(x_2) - g(x_1)}{x_2 - x_1} < \left.\frac{\partial g(x)}{\partial x}\right|_{x=x_2} < \frac{g(x_3) - g(x_2)}{x_3 - x_2}. \tag{3.78}$$

Therefore, with $g(x) = e^x$, it follows from (3.76)

$$\frac{s}{N_0 \eta_l(s)} = \frac{e^{r_l} - e^{r_{l-1}}}{r_l - r_{l-1}} < \frac{e^{r_{l+1}} - e^{r_l}}{r_{l+1} - r_l} = \frac{s}{N_0 \eta_{l+1}(s)}. \tag{3.79}$$

Since $\eta_l(s)$ is positive, we have (c).

$\square$

Proof: Theorem 4

Given an arbitrary power allocation $I_l(s)$ and the MPEQ allocation $I_l^*(s)$, we observe that

$$I_l^*(s) = I_l^*(s) I_l(s) + I_l^*(s)\left[1 - I_l(s)\right], \tag{3.80}$$

For $\rho_L\left(\mathcal{I}\right)$ given by (3.29),

$$\rho_L\left(\mathcal{I}^*\right) = \rho_L\left(\{I_l^*(s)\}\right) = \rho_L\left(\{I_l^*(s) I_l(s)\}\right) + \rho_L\left(\{I_l^*(s)\left[1 - I_l(s)\right]\}\right). \tag{3.81}$$

It follows from (3.32) and (3.80) that the MPEQ policy achieves ART

$$R_L(\mathcal{I}^*) = \sum_{l=1}^{L} \int_0^\infty I_l^*(s)\eta_l(s)p_l^\Delta(s)f(s)\,ds \tag{3.82}$$

$$= \kappa + \sum_{l=1}^{L} \int_0^\infty I_l^*(s)\left[1 - I_l(s)\right]\eta_l(s)p_l^\Delta(s)f(s)\,ds \tag{3.83}$$

where

$$\kappa = \sum_{l=1}^{L} \int_0^\infty I_l^*(s)I_l(s)\eta_l(s)p_l^\Delta(s)f(s)ds. \tag{3.84}$$

Consequently,

$$
\begin{aligned}
R_L(\mathcal{I}^*) &\overset{(a)}{\geq} \kappa + \sum_{l=1}^{L} \int_0^\infty I_l^*(s)\left[1 - I_l(s)\right]\lambda_{\mathrm{M}}(\mathbf{r})p_l^\Delta(s)f(s)\,ds && (3.85)\\
&\overset{(b)}{=} \kappa + \lambda_{\mathrm{M}}(\mathbf{r})\left(\rho_L\left(\mathcal{I}^*\right) - \rho_L\left(\{I_l^*(s)I_l(s)\}\right)\right) && (3.86)\\
&\overset{(c)}{\geq} \kappa + \lambda_{\mathrm{M}}(\mathbf{r})\left(\rho_L\left(\mathcal{I}\right) - \rho_L\left(\{I_l(s)I_l^*(s)\}\right)\right) && (3.87)\\
&\overset{(d)}{=} \kappa + \sum_{l=1}^{L} \int_0^\infty I_l(s)\left[1 - I_l^*(s)\right]\lambda_{\mathrm{M}}(\mathbf{r})p_l^\Delta(s)f(s)\,ds && (3.88)\\
&\overset{(e)}{\geq} \kappa + \sum_{l=1}^{L} \int_0^\infty I_l(s)\left[1 - I_l^*(s)\right]\eta_l(s)p_l^\Delta(s)f(s)\,ds && (3.89)\\
&\overset{(f)}{=} \sum_{l=1}^{L} \int_0^\infty I_l(s)\eta_l(s)p_l^\Delta(s)f(s)\,ds && (3.90)\\
&= R_L(\mathcal{I}), && (3.91)
\end{aligned}
$$

where

(a) for channel states s with $I_l^*(s) = 1$, $\eta_l(s) \geq \lambda_{\mathrm{M}}(\mathbf{r})$;

(b) is due to (3.29) and (3.81);

(c) is from $\overline{p} = \rho_L(\mathcal{I}^*) \geq \rho_L(\mathcal{I})$;

(d) we apply (3.81) on $\mathcal{I}$ instead of $\mathcal{I}^*$;

(e) for the channel states with $[1 - I_l^*(s)] = 1$, $\eta_l(s) < \lambda_{\mathrm{M}}(\mathbf{r})$;

(f) we substitute κ from (3.84) and apply the relation of (3.80) on $\mathcal{I}$ again.

$\square$

Proof: Proposition 1

The claim 1) is a direct consequence of Definition 1. Specifically, Definition 1 implies that $I_l^{*'}(s) = 1$ wherever $I_l^*(s) = 1$ if $\lambda_{\mathrm{M}}(\mathbf{r}) > \lambda'_{\mathrm{M}}(\mathbf{r})$. Therefore,

$$I_l^*(s) = I_l^{*'}(s) I_l^*(s). \tag{3.92}$$

Consequently,

$$\begin{aligned}
\rho_L(\mathcal{I}^*) &= \sum_{l=1}^{L} \int_0^\infty I_l^*(s) p_l^\Delta(s) f(s)\, ds && (3.93)\\
&= \sum_{l=1}^{L} \int_0^\infty I_l^{*'}(s) I_l^*(s) p_l^\Delta(s) f(s)\, ds && (3.94)\\
&\leq \sum_{l=1}^{L} \int_0^\infty I_l^{*'}(s) I_l^*(s) p_l^\Delta(s) f(s)\, ds && (3.95)\\
&\quad + \sum_{l=1}^{L} \int_0^\infty I_l^{*'}(s)(1 - I_l^*(s)) p_l^\Delta(s) f(s)\, ds && (3.96)\\
&= \sum_{l=1}^{L} \int_0^\infty I_l^{*'}(s) p_l^\Delta(s) f(s)\, ds && (3.97)\\
&= \rho_L(\mathcal{I}^{*\prime}), && (3.98)
\end{aligned}$$

where (3.95) is due to that the second term on the right hand side is the non-negative. Hence, the claim 2) holds.

Note that (3.35) implies

$$q_l = \frac{\lambda_{\mathrm{M}}(\mathbf{r}) N_0 \left(e^{r_l} - e^{r_{l-1}}\right)}{r_l - r_{l-1}} > \frac{\lambda'_{\mathrm{M}}(\mathbf{r}) N_0 \left(e^{r_l} - e^{r_{l-1}}\right)}{r_l - r_{l-1}} = q'_l. \tag{3.99}$$

This leads to the claim 3).

For $\lambda_{\mathrm{M}}(\mathbf{r}) > \lambda'_{\mathrm{M}}(\mathbf{r})$, if $\rho_L(\mathcal{I}^*) = \rho_L(\mathcal{I}^{*\prime}) = \overline{p}$, according to (3.95), we have

$$\sum_{l=1}^{L} \int_0^\infty I_l^{*'}(s)(1 - I_l^*(s)) p_l^\Delta(s) f(s)\, ds = 0. \tag{3.100}$$

Since $p_l^\Delta(s) > 0$, we have

$$\sum_{l=1}^{L} \int_0^\infty I_l^{*'}(s)(1 - I_l^*(s)) f(s)\, ds = 0. \tag{3.101}$$

For $s \in [q'_l, q_l)$, $I_l^{*'}(s) = 1$ and $I_l^*(s) = 0$. Therefore, (3.101) implies

$$f(s) = 0, \qquad s \in [q'_l, q_l), l = 1, \ldots, L. \tag{3.102}$$

Consequently, the claim 4) holds.

Proof: Lemma 5

Note that $\frac{x}{e^x-1}$ decreases in x for $x \geq 0$.

Let $x = [\log(\omega_l) - \log(\omega_{l-1})]$, (3.70) can be written as

$$q_l = N_0 \frac{e^x - 1}{x} \omega_{l-1}. \tag{3.103}$$

Since x strictly increases in ω_l, the proof arrives.

□

Proof: Lemma 6

From (3.48), let $g(q_l) = \omega_{l-1} = \frac{1}{\int_{\mathcal{Q}_{l-1}} f_{l-1}(s)/s \, ds}$.

$$\frac{dg(q_l)}{dq_l} = \frac{f(q_l)}{\left[\int_{\mathcal{Q}_{l-1}} f(s)/s \; ds\right]^2} \left[\int_{\mathcal{Q}_{l-1}} \left(\frac{1}{s} - \frac{1}{q_l}\right) f(s) \; ds\right] > 0, \tag{3.104}$$

since $F(s)$ strictly increases in s.

Therefore, ω_{l-1} strictly increases in q_l.

□

Proof: Theorem 7

An optimum $(\mathbf{r}^*, \mathbf{q}^*)$ policy with $L-1$ distinct non-zero q_l^* can be expanded to a $(\mathbf{r}, \mathbf{q})$ policy with L distinct non-zero q_l by splitting any of its L intervals (including zero-power interval $[0, q_1^*)$) into two intervals.

If the first interval $[0, q_1^*)$ is split, without changing $p(s)$, the new $(\mathbf{r}, \mathbf{q})$ policy will have the same ART as the original one.

On the other hand, we can split any intervals of the original $(\mathbf{r}^*, \mathbf{q}^*)$ policy with non-zero power and re-allocate power according to (3.58). The newly split intervals will increase their contribution to ART while the contribution from the other intervals stays the same. Therefore, the new $(\mathbf{r}, \mathbf{q})$ policy has a higher ART than that one corresponding to the original $(\mathbf{r}^*, \mathbf{q}^*)$ policy. We have

$$\bar{C}_L \geq R_L(\mathbf{r}, \mathbf{q}) > R_{L-1}(\mathbf{r}^*, \mathbf{q}^*) = \bar{C}_{L-1}. \tag{3.105}$$

□

Chapter 4

Adaptive Transmission with a Finite Set of Code Rates and Channel Uncertainty

4.1 Introduction

In Chapter 2 and Chapter 3, we have studied the performance of adaptive transmission systems with a finite set of code rates and/or transmitted power levels. However, the results are only applicable for the case where both CSIT and CSIR are perfect. In this chapter, we investigate the performance of adaptive transmission systems with a finite set of code rates and channel uncertainty at the transmitter side.

We assume a slow multiplicative fading environment with additive white Gaussian noise (AWGN). The channel response is constant during the transmission of a codeword. Perfect CSI and a channel measurement are available at the receiver and the transmitter, respectively. The joint distribution of the channel measurement and the current channel state is assumed to be known. For each transmission, a message is encoded at a rate selected from a finite rate set based on the channel measurement and the resulting codeword is transmitted at a power level based on the same channel estimate. Since each codeword experiences an additive white Gaussian noise (AWGN) channel, random Gaussian codes with multiple codebooks are employed.

For the proposed discrete adaptive system, it is possible that the instantaneous mutual information corresponding to a channel state is less than the assigned code rate. Since typical communication services must sustain a certain QoS requirement in terms of the error performance, we introduce a constant information outage constraint to the traditional power constrained throughput maximization. In this chapter, we demonstrate that the posed constrained throughput maximization problem can be solved in a general case.

In this chapter, the throughput maximization is solved by the similar techniques used in Chapter 3. Moreover, with the introduction of the model of imperfect channel state information at the transmitter, this chapter serves as a transition to the more general approach of Chapter 5.

4.2 System Model and Problem Formulation

4.2.1 System Model

We consider a multiplicative flat fading channel model similar to that in [22]. The complex received signal

$$Y = \sqrt{S}X + W, \tag{4.1}$$

where S is the channel (fading) state, X is the complex transmitted signal, and W is a circularly symmetric additive white Gaussian noise (AWGN) with variance N_0. The channel state S is a real random variable of unit mean with a probability density function (PDF) $f_S(s)$, a cumulative distribution function (CDF) $F_S(s)$, and a domain $\mathcal{S} = \{s|s \geq 0\}$. It is also assumed that the fading is sufficiently *slow* that the channel state is constant during transmission of a codeword.

The transmission is designed in such a way that before transmitting a data message, the transmitter obtains a channel measurement vector $\mathbf{u} = [u_0, \ldots, u_{M-1}]^\top$, where M is the number of measurements. $\mathbf{U}$ is the corresponding random vector with a domain $\mathcal{U}$. We assume that S and $\mathbf{U}$ have a joint PDF $f_{S,\mathbf{U}}(s,\mathbf{u}) = f_{S|\mathbf{U}}(s|\mathbf{u})f_{\mathbf{U}}(\mathbf{u})$. In addition, all corresponding CDFs are continuous and differentiable. Our assumption is rather general. For example, in a block fading channel, $\mathbf{U}$ may be based on the observation corresponding to training symbols transmitted during a finite set of past channel states. Furthermore, measurements can be a function of the stochastic process that describes the channel state evolution.

Given a measurement $\mathbf{u}$, the transmitter selects a code rate $r(\mathbf{u})$ from $\mathcal{R} = \{r_0 = 0, r_1, \ldots, r_L\}$ and a power level $p(\mathbf{u}) = E\left\{|X|^2 \,|\mathbf{u}\right\}$, where $E\{\cdot\}$ denotes expectation, to transmit the data message. It is assumed that perfect CSI is available at the receiver side. Here, without loss of generality, we assume that $r_l \leq r_{l+1}$ for $l = 0, \ldots, L-1$.

4.2.2 Outage, Policy, and Average Reliable Throughput

Since $\mathbf{u}$ is only a measurement, it is possible that the corresponding channel state does not support the assigned transmitted rate $r(\mathbf{u})$ given $p(\mathbf{u})$ and, thus, an outage occurs. The corresponding conditional outage probability is

$$P_{\text{out}}(\mathbf{u}) \quad = \quad \Pr\left[\log\left(1+\frac{p(\mathbf{u})S}{N_0}\right) < r(\mathbf{u})\right] = \Pr\left[S < \frac{N_0}{p(\mathbf{u})}\left(e^{r(\mathbf{u})}-1\right)\right], \quad (4.2)$$

and the worst channel state that supports the rate $r(\mathbf{u})$ is

$$q(\mathbf{u}) = \frac{N_0}{p(\mathbf{u})}\left(e^{r(\mathbf{u})}-1\right). \quad (4.3)$$

Note that (4.3) is only meaningful when $p(\mathbf{u}) > 0$ and, thus, a meaningful $q(\mathbf{u})$ is strictly positive. Consequently,

$$P_{\text{out}}(\mathbf{u}) = F_{S|\mathbf{U}}(q(\mathbf{u})|\mathbf{u}), \quad (4.4)$$

where $F_{S|\mathbf{U}}(s|\mathbf{u})$ denotes the conditional CDF of $S = s$ given $\mathbf{U} = \mathbf{u}$ and is assumed to be strictly increasing in s for all $\mathbf{u}$.

An adaptive transmission policy is uniquely identified by $(p(\mathbf{u}), r(\mathbf{u}))$ or equivalently, $(p(\mathbf{u}), q(\mathbf{u}))$, where $q(\mathbf{u})$ is the worst channel state s that still allows for $r(\mathbf{u})$ to be achieved given $\mathbf{u}$. The system throughput is

$$\bar{R}\left(p(\cdot), q(\cdot)\right) = \int_{\mathcal{U}} \log\left(1+\frac{p(\mathbf{u})q(\mathbf{u})}{N_0}\right) f_{\mathbf{U}}(\mathbf{u})\, d\mathbf{u}. \quad (4.5)$$

The corresponding average power is

$$\rho\left(p(\cdot), q(\cdot)\right) \quad = \quad \int_{\mathcal{U}} p(\mathbf{u}) f_{\mathbf{U}}(\mathbf{u})\, d\mathbf{u}. \quad (4.6)$$

Since communication in a wireless system is typically power limited, we consider the average transmitted power constraint

$$\rho\left(p(\cdot), q(\cdot)\right) \leq \overline{p}. \quad (4.7)$$

In addition to (4.7), we also assume an equality outage constraint,

$$P_{\text{out}}(\mathbf{u}) = P_{\text{out}}, \quad \mathbf{u} \in \{\mathbf{u}'|p(\mathbf{u}') > 0\}, \quad (4.8)$$

where P_{out} is the positive outage probability requirement. Note that the equality (4.8) is a simplification of the practical QoS requirement in terms of the error performance. For instance, it is normally assumed that voice communication with an error probability up to 1% is acceptable. Note that (4.8) is an instantaneous constraint in the sense that it is for each channel measurement $\mathbf{u}$. However, (4.8) can b e interpreted as an average constraint since the outage probability corresponding to each measurement $\mathbf{u}$ is an average effect over the true channel states.

In this chapter, we try to maximize the throughput in (4.5) over a discrete rate set $\mathcal{R}$ and a continuously varying power allocation subject to both the average transmitted power constraint (4.7) and the outage constraint (4.8). The throughput-maximizing policies are referred as *optimum policies*.

4.3 Properties of Optimum Policies

4.3.1 Local Properties

Due to the outage constraint (4.8), the worst supportable channel state satisfies

$$P_{\text{out}} = F_{S|\mathbf{U}}(q(\mathbf{u})|\mathbf{u}), \qquad \mathbf{u} \in \mathcal{U}. \tag{4.9}$$

Within the scope of this chapter, we assume that $\Pr[S = 0|\mathbf{U} = \mathbf{u}] \leq P_{\text{out}}$, for any $\mathbf{u} \in \mathcal{U}$. Thus, since $F_{S|\mathbf{U}}(s|\mathbf{u})$ is strictly increasing in s, $q(\mathbf{u})$ is uniquely specified by P_{out} and $F_{S|\mathbf{U}}(s|\mathbf{u})$. Let $G_{\mathbf{u}}(s) = F_{S|\mathbf{U}}(s|\mathbf{u})$ and, then, we have

$$q(\mathbf{u}) = G_{\mathbf{u}}^{-1}(P_{\text{out}}). \tag{4.10}$$

Thus, with (4.3), $p(\mathbf{u})$ and $r(\mathbf{u})$ have a one-to-one relationship.

In order to understand the structure of our problem, we define the *incremental efficiency* or *efficiency*

$$\begin{aligned}
\eta_l(\mathbf{u}) &= \frac{r_l - r_{l-1}}{\frac{N_0}{q(\mathbf{u})}(e^{r_l} - 1) - \frac{N_0}{q(\mathbf{u})}(e^{r_{l-1}} - 1)} && (4.11)\\
&= \frac{r_l - r_{l-1}}{\frac{N_0}{q(\mathbf{u})}(e^{r_l} - e^{r_{l-1}})} && (4.12)\\
&= \left(\frac{r_l - r_{l-1}}{e^{r_l - r_{l-1}} - 1}\right)\frac{q(\mathbf{u})}{e^{r_{l-1}}N_0}, && (4.13)
\end{aligned}$$

where $l = 1, \ldots, L$. The efficiency is the ratio of the rate increment of the adjacent rates over the corresponding increment in the transmitted power required to sustain the rate increment.

Proposition 2 *The efficiency has the following properties.*

(a) $\eta_l(\mathbf{u})$ *is non-negative;*

(b) $\eta_l(\mathbf{u})$ *decreases in* l.

From Proposition 2, it is clear that we pay a heavier penalty in terms of power to transmit at a higher rate regardless of the measurement $\mathbf{u}$. Consequently, it is intuitive that reducing the average power dictates that lower rates should be transmitted.

4.3.2 Global Properties

We define the most power efficient allocation (MPEA) as a policy with $r(\mathbf{u}) = r_l$ if and only if $\eta_l(\mathbf{u}) \geq \lambda$ and $\eta_{l+1}(\mathbf{u}) < \lambda$ for some $\lambda > 0$ and $l = 1, \ldots, L$.

Theorem 8 *For some* $\lambda > 0$*, MPEA is an optimum policy.*

Due to the constant outage constraint (4.8), $q(\mathbf{u})$ is directly determined by $F_{S|\mathbf{U}}(s|\mathbf{u})$. Consequently, considering the definition of both $\eta_l(\mathbf{u})$ and MPEA, we have the following following corollary.

Corollary 3 *An optimum policy has the same rate/power allocation at* $\mathbf{u}_1$ *and* $\mathbf{u}_2$ *if* $f_{S|\mathbf{U}}(s|\mathbf{u}_1) = f_{S|\mathbf{U}}(s|\mathbf{u}_2)$.

Let $T(\mathbf{u})$ be a sufficient statistic for s. Then [56],

$$f_{\mathbf{U}|S}(\mathbf{u}|s) = f_{T(\mathbf{U})|S}(T(\mathbf{u})|s)h(\mathbf{u}), \quad h(\mathbf{u}) \geq 0, \tag{4.14}$$

where $h(\mathbf{u})$ is a deterministic function which is independent of s and

$$f_{S|\mathbf{U}}(s|\mathbf{u}) \quad = \quad \frac{f_{T(\mathbf{U})|S}(T(\mathbf{u})|s)f_S(s)}{\int_0^\infty f_{T(\mathbf{U})|S}(T(\mathbf{u})|s)f_S(s)\,ds}. \tag{4.15}$$

Corollary 3 and (4.15) imply that for any measurements $\mathbf{u}_1$ and $\mathbf{u}_2$ with $T(\mathbf{u}_1) = T(\mathbf{u}_2)$, an optimum policy has

$$r(\mathbf{u}_1) = r(\mathbf{u}_2), \qquad p(\mathbf{u}_1) = p(\mathbf{u}_2). \tag{4.16}$$

Therefore, in the rest of this chapter, we will concentrate on rate/power allocations over values of $T(\mathbf{U})$ instead of $\mathbf{U}$. In particular, let

$$V = T(\mathbf{U}), \tag{4.17}$$

where V is a random variable with a domain $\mathcal{V}$. It is sufficient to specify policies of interest by $(r(v), p(v))$, where $r(v)$ and $p(v)$ are the rate and power allocations corresponding to any measurement $\mathbf{u}$ satisfying $v = T(\mathbf{u})$. Equivalently, v indicates a set of measurements, $\{\mathbf{u}|T(\mathbf{u}) = v\}$ which share the same rate $r(v)$ and power $p(v)$.

Accordingly, the following functions are re-defined accordingly,

$$\begin{aligned} q(v) &= \frac{N_0}{p(v)}\left(e^{r(v)} - 1\right), & (4.18)\\ \bar{R}\left(p(\cdot), q(\cdot)\right) &= \int_{\mathcal{V}} \log\left(1 + \frac{p(v)q(v)}{N_0}\right) f_V(v)\, dv, & (4.19)\\ \rho\left(p(\cdot), q(\cdot)\right) &= \int_{\mathcal{V}} p(v) f_V(v)\, dv, & (4.20)\\ \eta_l(v) &= \left(\frac{r_l - r_{l-1}}{e^{r_l - r_{l-1}} - 1}\right)\frac{q(v)}{e^{r_{l-1}} N_0}. & (4.21) \end{aligned}$$

We assume that the upper limit of $\mathcal{V}$ is the same as that of $\mathcal{S}$ which is infinity.

Let $\mathcal{V}_l = \{v|r(v) = r_l\}$, $l = 0, \dots, L$, be a non-overlapping partition of $\mathcal{V}$. Then, the average power allocated over $\mathcal{V}_l$ is

$$\begin{aligned} P_l &= \int_{\mathcal{V}_l} p(v) f_l(v) dv & (4.22)\\ &= \int_{\mathcal{V}_l} \frac{N_0}{q(v)}(e^{r_l} - 1) f_l(v) dv & (4.23)\\ &= \frac{N_0}{\omega_l}(e^{r_l} - 1). & (4.24) \end{aligned}$$

where

$$\begin{aligned} \frac{1}{\omega_l} &= \int_{\mathcal{V}_l} \frac{1}{q(v)} f_l(v) dv, & (4.25)\\ f_l(v) &= \frac{f_V(v)}{F_l}, & (4.26)\\ F_l &= \int_{\mathcal{V}_l} f_V(v) dv. & (4.27) \end{aligned}$$

Note that ω_l can be regarded as the average channel quality.

Clearly, if $\mathcal{V}_l$ is known, an optimum policy solves

$$\max_{P_l} \sum_{l=1}^{L} \log\left(1 + \frac{P_l \omega_l}{N_0}\right) F_l \tag{4.28}$$

$$\text{subject to} \qquad \sum_{l=1}^{L} P_l F_l \leq \overline{p}. \tag{4.29}$$

The optimization (4.28) is identical to the well-known rate maximization in the parallel Gaussian channel [16] and the following theorem can be obtained by directly applying the Karush-Kuhn-Tucker conditions.

Theorem 9 *The optimum power allocation is in the water-filling form*

$$P_l = N_0 \left[K_0 - \frac{1}{\omega_l}\right]^+, \tag{4.30}$$

where K_0 is a constant ensuring the power constraint $\sum_{l=1}^{L} P_l F_l = \overline{p}$. The corresponding optimum rates are

$$r_l = \log\left(1 + [K_0 \omega_l - 1]^+\right). \tag{4.31}$$

In addition, it can be shown that the allocated power is zero only for $v \in \mathcal{V}_0$. This means that if the water-filling allocation results in $P_l = 0$ for $l > 0$, an inefficient set of rates $\mathcal{R}$ has been chosen.

Proposition 3 *For optimum policies,*

$$P_l > 0, \qquad l = 1, \ldots, L, \tag{4.32}$$

and

$$r_l = \log\left(K_0 \omega_l\right), \tag{4.33}$$

where

$$K_0 = \frac{\overline{p}/N_0 + \sum_{l=1}^{L} F_l/\omega_l}{\sum_{l=1}^{L} F_l}. \tag{4.34}$$

$$= \frac{\overline{p}/N_0 + \int_{v_1}^{\infty} \frac{1}{q(v)} f_V(v)\, dv}{\int_{v_1}^{\infty} f_V(v)\, dv} \tag{4.35}$$

4.3.3 Sufficient Statistic with Stochastic Ordering

In this subsection, we concentrate on a special class of problems with a sufficient statistic V such that $F_{S|V}(s|v)$ is a collection of distributions with stochastic ordering [36]. Moreover, we concentrate on the case of stochastically increasing scenarios where $v < v'$ implies $F_{S|V}(s|v) > F_{S|V}(s|v')$ for all $s \in \mathcal{S}$. This condition states that the channel state s corresponding to $v < v'$ is statistically worse than that corresponding to v'. The stochastic ordering assumption implies the following properties.

Lemma 7 *$q(v)$ is strictly increasing in v.*

Proposition 4 *The efficiency corresponding to a sufficient statistic v satisfies*

(a) $\eta_l(v)$ is non-negative;

(b) $\eta_l(v)$ decreases in l;

(c) $\eta_l(v)$ increases in v.

The claim (c) of Proposition 4 is a direct consequence of Lemma 7 and can be observed directly from (4.21). Moreover, the claim (c) of Proposition 4 and Theorem 8 directly implies the following proposition.

Proposition 5 *The optimum $\mathcal{V}_l$'s are*

$$\mathcal{V}_l = [v_l, v_{l+1}), \qquad l = 0, \ldots, L-1, \tag{4.36}$$

where $v_{L+1} = \infty$ and $v_l < v_{l+1}$, $l = 0, \ldots, L$.

Following Proposition 3 and Proposition 5, the policy of interest can uniquely be specified by $(\mathbf{r}, \mathbf{v})$, where

$$\mathbf{r} = \{r_1, \ldots, r_L\}, \tag{4.37}$$

$$\mathbf{v} = \{v_1, \ldots, v_L\}. \tag{4.38}$$

The corresponding throughput and the average power are

$$R(\mathbf{r},\mathbf{v}) = \sum_{l=1}^{L} r_l \int_{\mathcal{V}_l} f_V(v)\,dv, \tag{4.39}$$

$$\rho(\mathbf{r},\mathbf{v}) = \sum_{l=1}^{L} (e^{r_l}-1) \int_{\mathcal{V}_l} \frac{N_0}{q(v)} f_V(v)\,dv. \tag{4.40}$$

Then, the throughput maximization problem is

$$\max_{\mathbf{r},\mathbf{v}} \quad R(\mathbf{r},\mathbf{v}) \tag{4.41}$$

$$\text{subject to} \quad \rho_L(\mathbf{r},\mathbf{v}) \le \overline{p}, \tag{4.41a}$$

and an optimum policy $(\mathbf{r}^*,\mathbf{v}^*)$ achieves the maximum ART. The Lagrangian function for (4.41) is

$$L(\mathbf{r},\mathbf{v},\mu) = R(\mathbf{r},\mathbf{v}) + \mu(\overline{p} - \rho_L(\mathbf{r},\mathbf{v})). \tag{4.42}$$

The necessary conditions for an optimum policy $(\mathbf{r}^*,\mathbf{v}^*)$ are

$$\left.\frac{\partial L(\mathbf{r},\mathbf{v},\mu)}{\partial r_l}\right|_{\mathbf{r}=\mathbf{r}^*,\mathbf{v}=\mathbf{v}^*,\mu=\mu^*} = 0, \tag{4.43}$$

and

$$\left.\frac{\partial L(\mathbf{r},\mathbf{v},\mu)}{\partial v_l}\right|_{\mathbf{r}=\mathbf{r}^*,\mathbf{v}=\mathbf{v}^*,\mu=\mu^*} = 0. \tag{4.44}$$

Here, (4.43) implies

$$\mu^* = \frac{\omega_l^*}{e^{r_l^*}} = \frac{1}{K_0^*}, \tag{4.45}$$

where the last equality is from (4.33) since clearly, any optimum policy $(\mathbf{r}^*,\mathbf{v}^*)$ must satisfy the optimum solution of (4.28). In addition, (4.44) implies

$$\mu^* = \frac{r_l^* - r_{l-1}^*}{e^{r_l^*} - e^{r_{l-1}^*}} \frac{q(v_l^*)}{N_0}. \tag{4.46}$$

Therefore,

$$\frac{r_l^* - r_{l-1}^*}{e^{r_l^*} - e^{r_{l-1}^*}} \frac{q(v_l^*)}{N_0} = \frac{1}{K_0^*}. \tag{4.47}$$

1) $l = 1$:
   a) Given v_1^*, use (4.34) to uniquely solve for K_0^* and, thus, obtain $w_0^* = 1/K_0^*$;
   b) Given v_1^* and w_0^*, use (4.49) to uniquely solve for w_1^* [Lemma 8].

2) $l > 1$:
   a) Given v_{l-1}^* and ω_{l-1}^*, use (4.50) to uniquely solve for v_l^* [Lemma 9];
   b) Given v_l^* and ω_{l-1}^*, use (4.49) to uniquely solve for ω_l^* [Lemma 8].

3) Repeat step 2) from $l = 2$ to $l = L$, we will have $\mathbf{v}^*$.

Figure 4.1: An algorithm searching for optimum policies started from v_1^*.

Hence,

$$q(v_l^*) = \frac{N_0}{K_0^*}\frac{e^{r_l^*} - e^{r_{l-1}^*}}{r_l^* - r_{l-1}^*} \tag{4.48}$$

$$= N_0 \frac{\omega_l^* - \omega_{l-1}^*}{\log\left(\omega_l^*\right) - \log\left(\omega_{l-1}^*\right)}, \tag{4.49}$$

where $\omega_0^* = 1/K_0^*$. Note that (4.49) is meaningful since $q(v_l^*)$ is positive.

Due to the stochastic ordering requirement, $q(v)$ is strictly increasing in v. Therefore, the average channel quality ω_l becomes better when $\mathcal{V}_l$ is enlarged by adding more good channel states corresponding to large v. Then, we will have the following two lemmas and construct an algorithm to find the optimum policies.

Lemma 8 *v_l^* is strictly increasing in ω_l^*.*

Lemma 9 *ω_{l-1} is strictly increasing in v_l.*

In particular, Lemma 9 follows from the following restatement of (4.25),

$$\frac{1}{\omega_{l-1}} = \int_{v_{l-1}}^{v_l} \frac{1}{q(v)} f_l(s)\, ds, \qquad l = 2, \ldots, L. \tag{4.50}$$

With Lemma 8 and Lemma 9, we derive the algorithm for finding the optimum policy given v_1^* described in Fig. 4.1. Particularly, in steps 1b and 2b, Lemma 8 implies the existence of a unique solution ω_l^* and in the step 2a, Lemma 9 implies the existence of the unique solutions v_l^*. Hence, the following theorem arrives.

Theorem 10 *An optimum policy can be uniquely specified by v_1^*.*

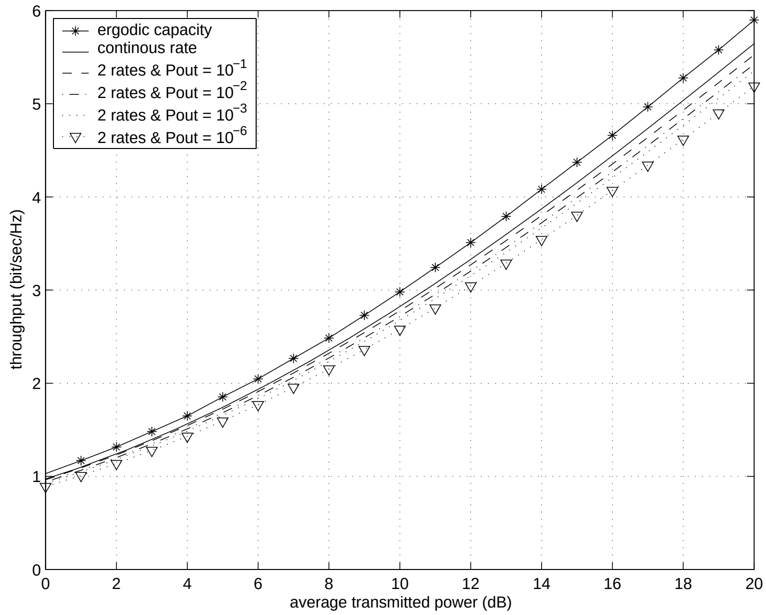

Figure 4.2: Sample performance in a Rayleigh fading channel with $\gamma_{\mathrm{T}} = 1000$.

4.4 Training Based Adaptive Transmission

The transmission is designed in such a manner that, before transmitting a data message, the channel is probed by a training sequence consisting of M identical symbols with an amplitude $x_{\mathrm{T}} = \sqrt{\mathcal{E}_{\mathrm{T}}}$. Note that $\mathcal{E}_{\mathrm{T}}$ is the transmitted energy corresponding to a training symbol. Based on the corresponding M received symbols $\mathbf{u} = [u_0, \ldots, u_{M-1}]^{\top}$, the transmitter selects a code rate $r(\mathbf{u})$ to encode and a power level $p(\mathbf{u})$ to transmit the data message. The signal-to-noise ratio (SNR) for the training sequence is

$$\gamma_{\mathrm{T}} = M \frac{E\{\mathcal{E}_{\mathrm{T}} S\}}{N_0}. \tag{4.51}$$

For simplicity, we assume that the training symbols experience the same channel s as the transmitted information (payload) (4.1). In addition, we assume that the phase of the received signals is accurately known at the receiver. In this case,

$$f_{\mathbf{U}|S}(\mathbf{u}|s) = (\pi N_0)^{-M} \exp\left\{ -\frac{||\mathbf{u} - \sqrt{s}\mathbf{x}_{\mathrm{T}}||^2}{N_0} \right\}, \tag{4.52}$$

where

$$\mathbf{x}_{\mathrm{T}} \;\; = \;\; [x_{\mathrm{T}}, \ldots, x_{\mathrm{T}}]\,. \tag{4.53}$$

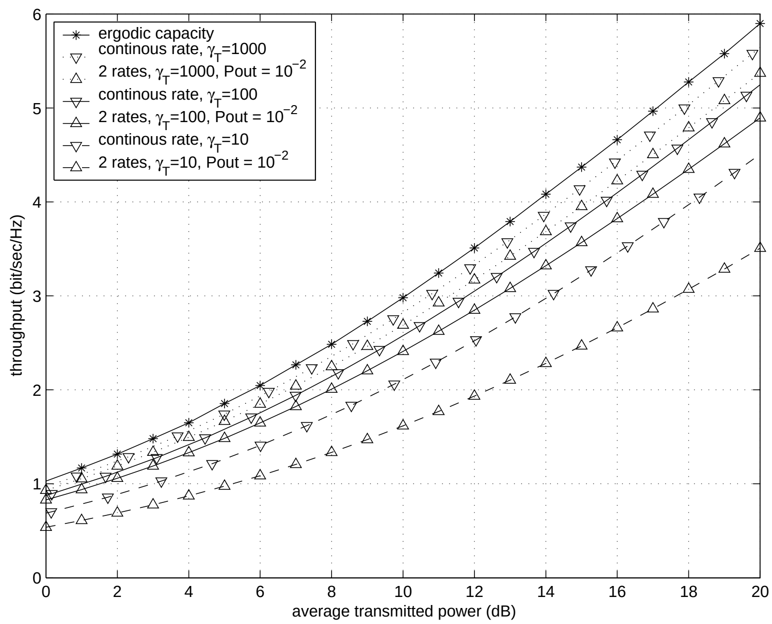

Figure 4.3: Sample performance in a Rayleigh fading channel with various γ_{T}.

A sufficient statistic for s given $\mathbf{u}$ is

$$T(\mathbf{u}) \quad = \quad \Re(\mathbf{u}^{\dagger}\mathbf{x}_{\mathrm{T}}). \tag{4.54}$$

Hence, with (4.17), we have

$$f_{V|S}(v|s) = (\pi M\mathcal{E}_{\mathrm{T}}N_0)^{-\frac{1}{2}} \exp\left[-\frac{(v - M\mathcal{E}_{\mathrm{T}}\sqrt{s})^2}{M\mathcal{E}_{\mathrm{T}}N_0}\right]. \tag{4.55}$$

For a Rayleigh fading channel with

$$f_S(s) = e^{-s}, \tag{4.56}$$

we have

$$f_{S,V}(s,v) = c_a^{-\frac{1}{2}} \left[\pi N_0 c_b\right]^{-\frac{1}{2}} \exp\left[\frac{(c_a c_b^2 - 1)v^2}{N_0 M\mathcal{E}_{\mathrm{T}}}\right] \exp\left[-\frac{(\sqrt{s} - c_b v)^2}{N_0 c_b}\right]$$

where $c_a = M\mathcal{E}_{\mathrm{T}}(M\mathcal{E}_{\mathrm{T}} + N_0)$ and $c_b = M\mathcal{E}_{\mathrm{T}}/c_a$.

Then, we can find

$$f_V(v) \quad = \quad c_a^{-\frac{1}{2}} \exp\left[\frac{(c_a c_b^2 - 1)v^2}{N_0 M\mathcal{E}_{\mathrm{T}}}\right] \left[c_b v \,\mathrm{erfc}\left(-\sqrt{\frac{c_b}{N_0}}v\right) + \sqrt{\frac{N_0 c_b}{\pi}} \exp\left(-\frac{c_b}{N_0}v^2\right)\right],$$

where

$$\mathrm{erfc}\,(\phi) = \frac{2}{\sqrt{\pi}} \int_{\phi}^{\infty} e^{-t^2} dt,$$

and

$$1 - F_{S|V}(q|v) \quad = \quad \frac{\sqrt{c_b} v \,\mathrm{erfc}\left(\sqrt{\frac{q}{N_0 c_b}} - \sqrt{\frac{c_b}{N_0}} v\right) + \sqrt{\frac{N_0}{\pi}} \exp\left(-\left(\sqrt{\frac{q}{N_0 c_b}} - \sqrt{\frac{c_b}{N_0}} v\right)^2\right)}{\sqrt{c_b} v \,\mathrm{erfc}\left(-\sqrt{\frac{c_b}{N_0}} v\right) + \sqrt{\frac{N_0}{\pi}} \exp\left(-\frac{c_b}{N_0} v^2\right)}.$$

It is straightforward to verify that V is a sufficient statistic satisfying the stochastic ordering requirement.

Note that in the throughput maximization problem (4.5), we have not considered the expense of the power used in training. Therefore, our results are a good reference for systems with small training overhead.

4.5 Sample Result

In Fig. 4.2, we present a sample numerical result for a Rayleigh fading channel. Here, following [10], we let $N_0 = 1$ in order to evaluate the system performance numerically.

The ergodic capacity curve is reproduced according to [22]. For the rest of the curves, we fix $\gamma_{\mathrm{T}} = 1000$. When there is an infinite number of rate levels ($L = \infty$), the throughput is within 1 dB of the ergodic capacity.

The last four curves correspond to $L = 2$ and various values of P_{out}. From these curves, we can see that reducing P_{out} makes the constraint more stringent and, thus, lowers the overall throughput. Particularly, the results with $P_{\mathrm{out}} = 10^{-6}$ and $L = 2$ is about 1 to 1.5 dB worse than the results with $L = \infty$.

In Fig. 4.3, it is shown that for high γ_{T}, the results with $L = 2$ and $P_{\mathrm{out}} = 10^{-2}$ are very close to those with $L = \infty$. However, when $\gamma_{\mathrm{T}} = 10$, there is a large gap between the curves corresponding to $L = 2$ with $P_{\mathrm{out}} = 10^{-2}$ and $L = \infty$. Therefore, it may require systems to have a large L to offset the loss due to low quality channel estimates.

In addition, even when $L = \infty$, there is a 3 to 4 dB gap between the curves with $\gamma_{\mathrm{T}} = 1000$ and those with $\gamma_{\mathrm{T}} = 10$. Consequently, for systems with a constraint on the overall power budget including both the training and the payload power expense, increasing γ_{T} implies decreasing the average transmitted power for the payload and, thus, there is a tradeoff in choosing γ_{T}.

4.A Proofs

Proof: Proposition 2

The claim (a) is straightforward since $r_l - r_{l-1}$, r_{l-1}, $q(\mathbf{u})$ are all non-negative in (4.13).

For the claim (b), we notice that e^{r_l} is strictly convex in r_l. For any strictly convex function $g(x)$, we have for $x_2 > x_1$

$$\left.\frac{\partial g(x)}{\partial x}\right|_{x=x_1} < \frac{g(x_2) - g(x_1)}{x_2 - x_1} < \left.\frac{\partial g(x)}{\partial x}\right|_{x=x_2}. \tag{4.57}$$

Consequently, for any $x_1 < x_2 < x_3$,

$$\frac{g(x_2) - g(x_1)}{x_2 - x_1} < \left.\frac{\partial g(x)}{\partial x}\right|_{x=x_2} < \frac{g(x_3) - g(x_2)}{x_3 - x_2}. \tag{4.58}$$

Therefore, with $g(x) = e^x$, it follows from (3.76)

$$\frac{q(\mathbf{u})}{N_0 \eta_l(\mathbf{u})} = \frac{e^{r_l} - e^{r_{l-1}}}{r_l - r_{l-1}} < \frac{e^{r_{l+1}} - e^{r_l}}{r_{l+1} - r_l} = \frac{q(\mathbf{u})}{N_0 \eta_{l+1}(\mathbf{u})}. \tag{4.59}$$

Hence, we have the claim (b).

$\square$

Proof: Theorem 8

Let

$$\begin{aligned} p_0(\mathbf{u}) &= 0, & \quad (4.60)\\ p_l(\mathbf{u}) &= \frac{N_0}{q(\mathbf{u})}(e^{r_l} - 1), \qquad l = 1, \ldots, L. & \quad (4.61) \end{aligned}$$

Then, we have

$$p(\mathbf{u}) = \sum_{l=1}^{L} \left[p_l(\mathbf{u}) - p_{l-1}(\mathbf{u})\right] I_l(\mathbf{u}), \tag{4.62}$$

where $\{I_l(\mathbf{u}) | l = 1, \ldots, L\}$ is a set of binary 0/1-value functions. In addition,

$$r(\mathbf{u}) = \sum_{l=1}^{L} \left[r_l - r_{l-1}\right] I_l(\mathbf{u}). \tag{4.63}$$

Since given $p(\mathbf{u})$, $r(\mathbf{u})$ and $q(\mathbf{u})$ have a one-to-one mapping, any $p(\mathbf{u})$ and $q(\mathbf{u})$ pair can be described by the corresponding $\{I_l(\mathbf{u})|l = 1,\ldots,L\}$ and vice versa. Note that such mapping implies a constraint

$$I_l(\mathbf{u}) \geq I_{l+1}(\mathbf{u}), \qquad l = 1,\ldots,L-1. \tag{4.64}$$

Consequently, any $(p(\mathbf{u}), q(\mathbf{u}))$ policy can be described by the corresponding $\mathcal{I} = \{I_l(\mathbf{u})|\mathbf{u} \in \mathcal{U},\ l = 1,\ldots,L\}$ and vice versa. Then, for any policy $\mathcal{I}$, the average power and the average throughput are given by

$$\begin{aligned}
\rho(\mathcal{I}) &= \int_{\mathcal{U}} \sum_{l=1}^{L} \left[p_l(\mathbf{u}) - p_{l-1}(\mathbf{u})\right] I_l(\mathbf{u}) f_{\mathbf{U}}(\mathbf{u})\, d\mathbf{u}, && (4.65)\\
\bar{R}(\mathcal{I}) &= \int_{\mathcal{U}} \sum_{l=1}^{L} \left[r_l - r_{l-1}\right] I_l(\mathbf{u}) f_{\mathbf{U}}(\mathbf{u})\, d\mathbf{u} && (4.66)\\
&= \int_{\mathcal{U}} \sum_{l=1}^{L} \eta_l(\mathbf{u}) \left[p_l(\mathbf{u}) - p_{l-1}(\mathbf{u})\right] I_l(\mathbf{u}) f_{\mathbf{U}}(\mathbf{u})\, d\mathbf{u} && (4.67)
\end{aligned}$$

We use $\mathcal{I}^*$ to denote an MPEA with $\rho(\mathcal{I}^*) = \overline{p}$ from some $\lambda > 0$. Here, the claim (b) of Proposition 2 guarantees that (4.64) is satisfied by $\mathcal{I}^*$. On the other hand, for any arbitrary $\mathcal{I}$, which may not necessarily satisfy (4.64), Theorem 8 claims that $\bar{R}(\mathcal{I}^*) \geq \bar{R}(\mathcal{I})$ for any $\mathcal{I}$ with $\rho(\mathcal{I}) \leq \overline{p}$.

Given an arbitrary power allocation $I_l(\mathbf{u})$ and the MPEA allocation $I_l^*(\mathbf{u})$, we observe that

$$I_l^*(\mathbf{u}) = I_l^*(\mathbf{u}) I_l(\mathbf{u}) + I_l^*(\mathbf{u}) \left[1 - I_l(\mathbf{u})\right], \tag{4.68}$$

Therefore,

$$\rho\left(\mathcal{I}^*\right) = \rho\left(\{I_l^*(\mathbf{u})\}\right) = \rho\left(\{I_l^*(\mathbf{u}) I_l(\mathbf{u})\}\right) + \rho_L\left(\{I_l^*(\mathbf{u}) \left[1 - I_l(\mathbf{u})\right]\}\right). \tag{4.69}$$

It follows (4.68) that the MPEA policy achieves ART

$$\begin{aligned}
\bar{R}(\mathcal{I}^*) &= \sum_{l=1}^{L} \int_0^\infty I_l^*(\mathbf{u}) \eta_l(\mathbf{u}) p_l(\mathbf{u}) f(\mathbf{u})\ d\mathbf{u} && (4.70)\\
&= \kappa + \sum_{l=1}^{L} \int_0^\infty I_l^*(\mathbf{u}) \left[1 - I_l(\mathbf{u})\right] \eta_l(\mathbf{u}) p_l(\mathbf{u}) f(\mathbf{u})\ d\mathbf{u}, && (4.71)
\end{aligned}$$

where

$$\kappa = \sum_{l=1}^{L} \int_0^\infty I_l^*(\mathbf{u}) I_l(\mathbf{u}) \eta_l(\mathbf{u}) p_l(\mathbf{u}) f(\mathbf{u})\ d\mathbf{u}. \tag{4.72}$$

Consequently,

$$\begin{aligned}
\bar{R}(\mathcal{I}^*) &\overset{(a)}{\geq} \kappa + \sum_{l=1}^{L} \int_0^\infty I_l^*(\mathbf{u}) \left[1 - I_l(\mathbf{u})\right] \lambda p_l(\mathbf{u}) f(\mathbf{u}) \; d\mathbf{u} && (4.73)\\
&\overset{(b)}{=} \kappa + \lambda \left(\rho\left(\mathcal{I}^*\right) - \rho\left(\{I_l^*(\mathbf{u}) I_l(\mathbf{u})\}\right)\right) && (4.74)\\
&\overset{(c)}{\geq} \kappa + \lambda \left(\rho\left(\mathcal{I}\right) - \rho\left(\{I_l^*(\mathbf{u}) I_l(\mathbf{u})\}\right)\right) && (4.75)\\
&\overset{(d)}{=} \kappa + \sum_{l=1}^{L} \int_0^\infty I_l(\mathbf{u}) \left[1 - I_l^*(\mathbf{u})\right] \lambda p_l(\mathbf{u}) f(\mathbf{u}) \; d\mathbf{u} && (4.76)\\
&\overset{(e)}{\geq} \kappa + \sum_{l=1}^{L} \int_0^\infty I_l(\mathbf{u}) \left[1 - I_l^*(\mathbf{u})\right] \eta_l(\mathbf{u}) p_l(\mathbf{u}) f(\mathbf{u}) \; d\mathbf{u} && (4.77)\\
&\overset{(f)}{=} \sum_{l=1}^{L} \int_0^\infty I_l(\mathbf{u}) \eta_l(\mathbf{u}) p_l(\mathbf{u}) f(\mathbf{u}) \; d\mathbf{u} && (4.78)\\
&= \bar{R}(\mathcal{I}), && (4.79)
\end{aligned}$$

where

(a) for any measurement $\mathbf{u}$ with $I_l^*(\mathbf{u}) = 1$, $\eta_l(\mathbf{u}) \geq \lambda$;

(b) is due to (4.69);

(c) is from $\overline{p} = \rho(\mathcal{I}^*) \geq \rho(\mathcal{I})$;

(d) we apply (4.69) on $\mathcal{I}$ instead of $\mathcal{I}^*$;

(e) for the channel states with $[1 - I_l^*(\mathbf{u})] = 1$, $\eta_l(\mathbf{u}) < \lambda$;

(f) we substitute κ from (4.72) and apply the relation of (4.68) on $\mathcal{I}$ again.

$\square$

Proof: Lemma 7

Due to the stochastic ordering, for any $v < v'$, we have $F_{S|V}(q(v)|v) > F_{S|V}(q(v)|v')$. This together with the constant outage constraint (4.8) implies

$$F_{S|V}(q(v')|v') = P_{\text{out}} = F_{S|V}(q(v)|v) > F_{S|V}(q(v)|v'). \tag{4.80}$$

Thus, $q(v) < q(v')$ since $F_{S|V}(s|v)$ is strictly increasing in s. Hence, $q(v)$ is strictly increasing in v.

$\square$

Proof: Lemma 8

In Lemma 7, we show that $q(v)$ is strictly increasing in v. On the other hand, we observe that $\frac{x}{e^x-1}$ decreases in x for $x \geq 0$. Let $x = \left[\log\left(\omega_l^*\right) - \log\left(\omega_{l-1}^*\right)\right]$, (4.49) can be written as

$$q(v_l^*) = \frac{e^x - 1}{x}\omega_{l-1}^*. \tag{4.81}$$

Since x is strictly increasing in ω_l^*, we have $q(v_l^*)$ is strictly increasing in ω_l^*.

Finally, $q(v_l^*)$ is strictly increasing in both v_l^* and ω_l^*. The lemma holds.

□

Proof: Lemma 9

Shown in Lemma 7, $q(v)$ is strictly increasing in v. Therefore, we only need to consider how ω_{l-1} changes according to $q(v_l)$.

From (4.25), let $g(q(v_l)) = \omega_{l-1} = \frac{1}{\int_{v_{l-1}}^{v_l} f_{l-1}(v)/q(v)\,dv}$.

$$\frac{dg(q(v_l))}{dq(v_l)} = \frac{f(q(v_l))}{\left[\int_{v_{l-1}}^{v_l} f(v)/q(v)\,dv\right]^2}\left[\int_{\mathcal{V}_{l-1}}\left(\frac{1}{q(v)} - \frac{1}{q(v_l)}\right) f(v)\,dv\right] > 0, \tag{4.82}$$

since $q(v) < q(v_l)$ for any $v \in [v_{l-1}, v_l)$ due to Lemma 7.

Therefore, ω_{l-1} is strictly increasing in $q(v_l)$. This together with that $q(v)$ is strictly increasing in v clarifies the lemma.

□

Chapter 5

Adaptive Transmission with Channel State Uncertainty

5.1 Introduction

A key element in the realization of adaptive transmission is the advance of channel estimation techniques. However, since channel estimation techniques cannot provide the exact channel state information (CSI), it is very important to examine the achievable performance with uncertain CSI.

In this chapter, we assume a slow multiplicative fading environment with additive white Gaussian noise (AWGN). The channel response is constant during the transmission of a codeword. We assume perfect CSIR and general CSIT that is a channel measurement/estimate. The joint distribution of the channel measurement and current channel state is assumed to be known. For each transmission, a message is both encoded at a rate and transmitted at a power level corresponding to the channel measurement. Since each codeword experiences an additive white Gaussian noise (AWGN) channel, random Gaussian codes organized in multiple codebooks are employed.

Since the code rate and the transmitted power level are based on channel estimates, it is possible that the instantaneous mutual information corresponding to the channel state is less than the assigned code rate. Consequently, we characterize the performance of a system design based on the concept of *average reliable throughput* (ART), defined as the average data rate **assuming zero rate when the channel is in outage** Chapter 2.

In general, the results in [9,48] are not suitable for the delay-limited scenario, where there is a delay-limited constraint or the codewords are not long enough to experience the ergodicity of the slow fading channel. On the other hand, the solution proposed in

[22], which is to multiplex codewords with various rates according to CSI, is meaningful for the delay-limited scenario. However, the multiplexing is only limited to the case with the perfect channel information at both the transmitter and the receiver. Therefore, an important question is how to relax the limitation on the perfect CSI assumption.

When we try to implement an adaptive transmission system, the problems related to channel estimation become even more challenging since there are two distinct types of channel estimation,

- *pre-estimation*: channel estimation used as a basis for adjusting the transmission parameters, e.g. the transmitted power and the encoded rates;
- *post-estimation*: channel estimation used for the purpose of coherence detection.

Here, we do not specify where the pre- or the post-estimation is performed and this enables the following arguments suitable for both frequency division duplex (FDD) and time division duplex (TDD) structures. Note that both pre- and post-estimation can be the same and can be performed at either the transmitter or the receiver if desired.

However, it is very important to realize that pre-estimation and post-estimation differ in their timing requirements. Since pre-estimation is the basis for adjusting the transmission parameters, it is preferable to have the estimate of the pre-estimation available as soon as possible. On the other hand, the post-estimation does not have a very stringent timing requirement. These distinct timing requirements can lead to different algorithmic realizations. For pre-estimation, the algorithm must be fast and, thus, typically simple. It results in coarse estimates. In particular, since the transmitter can exploit only pre-estimation, the transmitter adaptatio must be done in the presence of channel uncertainty. For post-estimation, complex iterative algorithms that combine channel estimation and coded bit detection can be used to produce accurate estimates. Thus, we make the idealized assumption that the receiver has perfect CSI.

In this chapter, a valid policy consists of a space of channel measurements/estimates together with the transmitted power and the encoded code rates or the worst channel states support the code rates corresponding to each channel measurement/estimate. For an ART-maximizing (optimum) policy, it is necessary to be optimum either given the transmitted power or given the worst channel states. This requirement leads to a set of characteristics of the candidates of the optimum policies. One of the important observation is that when increasing the transmitted power, even though the throughput increases accordingly, the assigned code rate decreases. This phenomenon is quite different from the scenario of the adaptive transmission with perfect CSIT that the assigned code rate is strictly increasing in the transmitted power.

For an optimum policy, given a measurement, there may be only a finite number of choices of transmitted power/rate. If this number is less than three, we derived in this chapter an algorithm to search for the optimum policy denoted by the most power efficient allocation (MPEA). In this case, even though we assume that the transmitted power and encoded rates are functions of channel measurements, it turns out that they are only functions of the sufficient statistic for the channel states corresponding to the measurements.

Given a measurement, if there are more than three choices of transmitted power/rate, MPEA may still be the optimum as long as the incremental efficiency, defined as the ratio of the throughput increment over the corresponding power increment, is strictly decreasing. This condition can be simply interpreted as that the throughput gain is diminishing when the transmitted power increases.

The condition for MPEA to be optimum may not be able to hold if there is no constraint over the joint distribution between the channel states and the measurements. An example is constructed in the appendix. For such cases, some approximation techniques are proposed.

5.2 System Model and Problem Formulation

5.2.1 System Model

We consider a multiplicative flat fading channel model similar to that in [22]. The complex received signal

$$Y = \sqrt{S}X + W, \tag{5.1}$$

where S is the channel (fading) state, X is the complex transmitted signal, and W is a circularly symmetric additive white Gaussian noise (AWGN) with variance N_0. The channel state S is a real random variable of unit mean with a probability density function (PDF) $f_S(s)$, a cumulative distribution function (CDF) $F_S(s)$, and a domain $\mathcal{S} = \{s|s \geq 0\}$. It is also assumed that the fading is sufficiently *slow* that the channel state is constant during transmission of a codeword.

The transmission is designed in such a way that before transmitting a data message, the transmitter obtains a channel measurement vector $\mathbf{u} = [u_0, \ldots, u_{M-1}]^\top$. $\mathbf{U}$ is the corresponding random vector and we assume that S and $\mathbf{U}$ have a joint PDF $f_{S,\mathbf{U}}(s, \mathbf{u})$ and the marginal PDF $f_{\mathbf{U}}(\mathbf{u})$ exists. Our assumption is rather general. For example, in a block fading channel, $\mathbf{U}$ may be based on the observation of training symbols transmitted over some finite set of past channel states. Of course, when $\mathbf{U}$ employs measurements outside the current fading block, $f_{S,\mathbf{U}}(s, \mathbf{u})$ will depend on the stochastic process that describes the channel state evolution.

Given a measurement $\mathbf{U} = \mathbf{u}$, the transmitter selects a code rate $r(\mathbf{u})$ to encode and a power level $p(\mathbf{u}) = E\left\{|X|^2|\mathbf{u}\right\}$, where $E\{\cdot\}$ denotes expectation, to transmit the data message. It is assumed that perfect CSI is available at the receiver side.

5.2.2 Outage, Policy, and Average Reliable Throughput

Given a measurement vector $\mathbf{u}$ and the corresponding allocated $p(\mathbf{u}) > 0$, it is possible the corresponding channel state s does not support the assigned code rate $r(\mathbf{u}) > 0$ and, thus, an outage occurs. The corresponding outage probability is

$$P_{\text{out}}(\mathbf{u}) = \Pr\left[\log\left(1 + \frac{p(\mathbf{u})S}{N_0}\right) < r(\mathbf{u})\middle|\mathbf{u}\right] = \Pr\left[S < \frac{N_0}{p(\mathbf{u})}\left(e^{r(\mathbf{u})} - 1\right)\middle|\mathbf{u}\right]. \quad (5.2)$$

In addition, the worst channel state that supports the rate $r(\mathbf{u})$ is

$$q(\mathbf{u}) = \frac{N_0}{p(\mathbf{u})}\left(e^{r(\mathbf{u})} - 1\right). \quad (5.3)$$

Consequently,

$$P_{\text{out}}(\mathbf{u}) = F_{S|\mathbf{U}}(q(\mathbf{u})|\mathbf{u}), \quad (5.4)$$

where $F_{S|\mathbf{U}}(\cdot|\mathbf{u})$ is the conditional CDF of S given $\mathbf{U} = \mathbf{u}$.

In this case, a simple rate assignment [22] which is completely determined by the power allocation does not exist. We have to jointly optimize the power allocation $p(\mathbf{u})$ and the rate assignment $r(\mathbf{u})$ when the perfect channel state information is not available.

An adaptive transmission policy is uniquely identified by $(p(\mathbf{u}), r(\mathbf{u}))$ or equivalently, $(p(\mathbf{u}), q(\mathbf{u}))$, where $q(\mathbf{u})$ is the preset worst channel state given $\mathbf{u}$. Assuming that information is successful received only if the channel is not in outage, we define the reliable throughput given $\mathbf{u}$

$$R(\mathbf{u}) = \left[1 - F_{S|\mathbf{U}}(q(\mathbf{u})|\mathbf{u})\right] \log\left(1 + \frac{p(\mathbf{u})q(\mathbf{u})}{N_0}\right), \tag{5.5}$$

and the average reliable throughput (ART) Chapter 2

$$\bar{R}\left(p(\cdot), q(\cdot)\right) \quad = \quad \int_{\mathcal{U}} R(\mathbf{u}) f(\mathbf{u})\, d\mathbf{u} \tag{5.6}$$

where $\mathcal{U}$ is the domain of $\mathbf{u}$. The corresponding average power

$$\rho\left(p(\cdot), q(\cdot)\right) = \int_{\mathcal{U}} p(\mathbf{u}) f(\mathbf{u}) d\mathbf{u}. \tag{5.7}$$

Within the scope of this chapter, we study the problem of ART-maximization subject to a power constraint:

$$\max_{p(\mathbf{u}), q(\mathbf{u})} \qquad \bar{R}\left(p(\cdot), q(\cdot)\right) \tag{5.8}$$

$$\text{subject to} \qquad \rho\left(p(\cdot), q(\cdot)\right) \leq \overline{p}, \tag{5.8a}$$

A corresponding ART-maximizing policy is referred to as an optimum policy and is denoted by $(p^*(\mathbf{u}), q^*(\mathbf{u}))$.

5.3 Properties of Optimum Policies

5.3.1 Subproblems

Since $F_{S|\mathbf{U}}(q(\mathbf{u})|\mathbf{u})$ is not specified, it is relatively difficult to pursue the optimum result directly. Therefore, two relatively easy subproblems are solved in this subsection and, then, further characterization of optimum policies is made.

With (5.6), given an arbitrary power allocation $p(\mathbf{u})$, the local optimality guarantees that there are $\bar{R}$-maximizing $q(\mathbf{u})$'s. Therefore, the first subproblem is

$$\max_{q(\mathbf{u})} \quad \bar{R}\left(p(\cdot), q(\cdot)\right) \tag{5.9}$$

$$\text{subject to} \quad \rho\left(p(\cdot), q(\cdot)\right) \leq \overline{p}. \tag{5.9a}$$

Note that (5.9a) is a redundant constraint since $p(\mathbf{u})$ is set and, thus, varying $q(\mathbf{u})$ will not result any policy violating (5.9a) if the preset $p(\mathbf{u})$ is feasible. With the non-negativity of $f(\mathbf{u})$ in (5.6), the local optimality is stated in the following theorem.

Theorem 11 *Given an arbitrary* $p(\mathbf{u})$*, the corresponding* $\bar{R}$*-maximizing worst channel state satisfies*

$$q^{\ddagger}(\mathbf{u}, p(\mathbf{u})) = \arg\max_{q} \left[1 - F_{S|\mathbf{U}}(q|\mathbf{u})\right] \log\left(1 + \frac{qp(\mathbf{u})}{N_0}\right), \tag{5.10}$$

and the corresponding reliable throughput is $R^{\ddagger}(\mathbf{u}, p(\mathbf{u}))$.

According to Theorem 11, $p(\mathbf{u})$ is the least power sufficient to achieve $R^{\ddagger}(\mathbf{u}, p(\mathbf{u}))$.

When CSIT is perfect, $\mathbf{u} \to s$ and $F_{S|\mathbf{U}}(q|\mathbf{u}) \to U(q-s)$ where $U(x)$ is the step function defined as

$$U(x) = \begin{cases} 0, & x \leq 0, \\ 1, & x > 0. \end{cases} \tag{5.11}$$

Hence,

$$R^{\ddagger}(s, p(s)) = \max_{q \leq s} \log\left(1 + \frac{qp(s)}{N_0}\right). \tag{5.12}$$

Since both q and $p(s)$ are non-negative,

$$q^{\ddagger}(s, p(s)) = s, \tag{5.13}$$

$$R^{\ddagger}(s, p(s)) = \log\left(1 + \frac{sp(s)}{N_0}\right). \tag{5.14}$$

On the other hand, the second subproblem is to optimize the power allocation $p(\mathbf{u})$ given $q(\mathbf{u})$ to maximize $\bar{R}\left(p(\cdot), q(\cdot)\right)$. Mathematically, this problem is

$$\max_{p(\mathbf{u})} \quad \bar{R}\left(p(\cdot), q(\cdot)\right) \tag{5.15}$$

$$\text{subject to} \quad \rho\left(p(\cdot), q(\cdot)\right) \leq \overline{p}. \tag{5.15a}$$

This problem can be solved by directly applying the standard Kuhn-Karush-Tacker conditions.

Theorem 12 *Given an arbitrary worst channel state assignment $q(\mathbf{u})$, the optimum power is a water-filling solution given by*

$$p^{\ddagger}(\lambda, \mathbf{u}, q(\mathbf{u})) = \left[\frac{1 - F_{S|\mathbf{U}}(q(\mathbf{u})|\mathbf{u})}{\lambda} - \frac{N_0}{q(\mathbf{u})}\right]^{+}, \tag{5.16}$$

where the Lagrange multiplier λ is a positive parameter chosen to satisfy

$$\int_{\mathcal{U}} p^{\ddagger}(\lambda, \mathbf{u}, q(\mathbf{u})) f_{\mathbf{U}}(\mathbf{u}) d\mathbf{u} = \overline{p}. \tag{5.17}$$

In (5.16), to increase λ without changing $q(\mathbf{u})$ leads to a pointwise reduction in $p^{\ddagger}(\lambda, \mathbf{u}, q(\mathbf{u}))$ for each $\mathbf{u}$ where $p^{\ddagger}(\lambda, \mathbf{u}, q(\mathbf{u})) > 0$. Consequently, the average power is also decreased.

Proposition 6 *For policies with an arbitrary $q(\mathbf{u})$ and the corresponding $p^{\ddagger}(\lambda, \mathbf{u}, q(\mathbf{u}))$, the total power $\rho(p^{\ddagger}(\lambda, \cdot, q(\cdot)), q(\cdot))$ strictly decreases in λ.*

Corollary 4 *For policies with an arbitrary $q(\mathbf{u})$ and the corresponding $p^{\ddagger}(\lambda, \mathbf{u}, q(\mathbf{u}))$, the average reliable throughput $\bar{R}(p^{\ddagger}(\lambda, \cdot, q(\cdot)), q(\cdot))$ strictly decreases in λ.*

When CSIT is perfect, (5.13), (5.14), and (5.16) imply

$$p^{\ddagger}(\lambda, s, q^{\ddagger}(s, p(s))) = \left[\frac{1}{\lambda} - \frac{N_0}{s}\right]^{+}. \tag{5.18}$$

Clearly, (5.13), (5.14), and (5.18) agree with the results for perfect CSIT shown in [22].

Neither Theorem 11 nor Theorem 12 explicitly reveals much information except they both agree that for a given $q(\mathbf{u})$, the throughput will increase in $p(\mathbf{u})$. However, we have observed that some confusing scenarios are possible. Particularly, for Theorem 12, due to the fact that $F_{S|\mathbf{U}}(s|\mathbf{u})$ is arbitrary and increases in s, it seems possible that for $q(\mathbf{u}) \neq q'(\mathbf{u})$, (5.16) allows $p^{\ddagger}(\lambda, \mathbf{u}, q(\mathbf{u})) = p^{\ddagger}(\lambda, \mathbf{u}, q'(\mathbf{u}))$. Similarly, according to Theorem 11, for $q(\mathbf{u}) \neq q'(\mathbf{u})$, there might be two choices of the transmitted code rates, $\log(1 + p^{\ddagger}(\lambda, \mathbf{u}, q(\mathbf{u}))q(\mathbf{u})/N_0) \neq \log(1 + p^{\ddagger}(\lambda, \mathbf{u}, q'(\mathbf{u}))q'(\mathbf{u})/N_0)$, which enable us to achieve the same $R^{\ddagger}(\mathbf{u}, p^{\ddagger}(\lambda, \mathbf{u}, q(\mathbf{u})))$.

5.3.2 Pointwise Property

Since each policy must decide its own Lagrange multiplier λ and the rate/power allocation given any $\mathbf{u} \in \mathcal{U}$, we assume that λ is known and derive the relations of the possible optimum solutions at any specific $\mathbf{u}$.

Since we are only interested in the optimum policies, we only need to concentrate on the polices satisfy both Theorem 11 and Theorem 12, Thus, a police of interest $(p(\cdot), q(\cdot))$ satisfies

$$p(\mathbf{u}) = p^{\ddagger}(\lambda, \mathbf{u}, q(\mathbf{u})), \tag{5.19}$$

$$q(\mathbf{u}) \in \left\{ q^{\ddagger}(\mathbf{u}, p(\mathbf{u})) \right\}. \tag{5.20}$$

Let the transmitted code rate corresponding to $p(\mathbf{u})$ and $q(\mathbf{u})$ be

$$r(\mathbf{u}) = \log\left(1 + \frac{p(\mathbf{u})q(\mathbf{u})}{N_0}\right). \tag{5.21}$$

According to (5.10), for $p(\mathbf{u}) > 0$,

$$1 - F_{S|\mathbf{U}}(q(\mathbf{u})|\mathbf{u}) = \frac{R^{\ddagger}(\mathbf{u}, p(\mathbf{u}))}{r(\mathbf{u})} = \frac{R^{\ddagger}(\mathbf{u}, p(\mathbf{u}))}{\log\left(1 + p(\mathbf{u})q(\mathbf{u})/N_0\right)}. \tag{5.22}$$

On the other hand, according to (5.16), for $p(\mathbf{u}) > 0$,

$$1 - F_{S|\mathbf{U}}(q(\mathbf{u})|\mathbf{u}) = \lambda\left(p(\mathbf{u}) + \frac{N_0}{q(\mathbf{u})}\right) = \lambda\frac{e^{r(\mathbf{u})}}{q(\mathbf{u})} = \lambda\frac{p(\mathbf{u})}{1 - e^{-r(\mathbf{u})}}. \tag{5.23}$$

From (5.22) and (5.23), we will show that there are no two $q(\mathbf{u})$'s sharing the same $p(\mathbf{u})$ for any $\mathbf{u}$ in the following lemma.

Lemma 10 *Give a policy satisfying (5.19) and (5.20), for any* $\mathbf{u}$ *and* λ*, there is a unique* $q(\mathbf{u})$ *corresponding to* $p(\mathbf{u}) > 0$ *for any* $\mathbf{u}$*. When* $p(\mathbf{u}) = 0$*, it is not of interest to decide the corresponding* $q(\mathbf{u})$*'s.*

The pointwise uniqueness between $p(\mathbf{u})$ and $q(\mathbf{u})$ (pointwise in the sense of $\mathbf{u}$) reduces the number of elements in the feasible set for searching for the optimum solutions since the policies of interest can be uniquely specified by either $p(\mathbf{u})$ or $q(\mathbf{u})$.

A serious issue within our characterization is that the number of feasible $p(\mathbf{u})$ or $q(\mathbf{u})$ solutions, which satisfy both (5.10) and (5.16), at any $\mathbf{u}$ is unknown. For perfect CSIT,

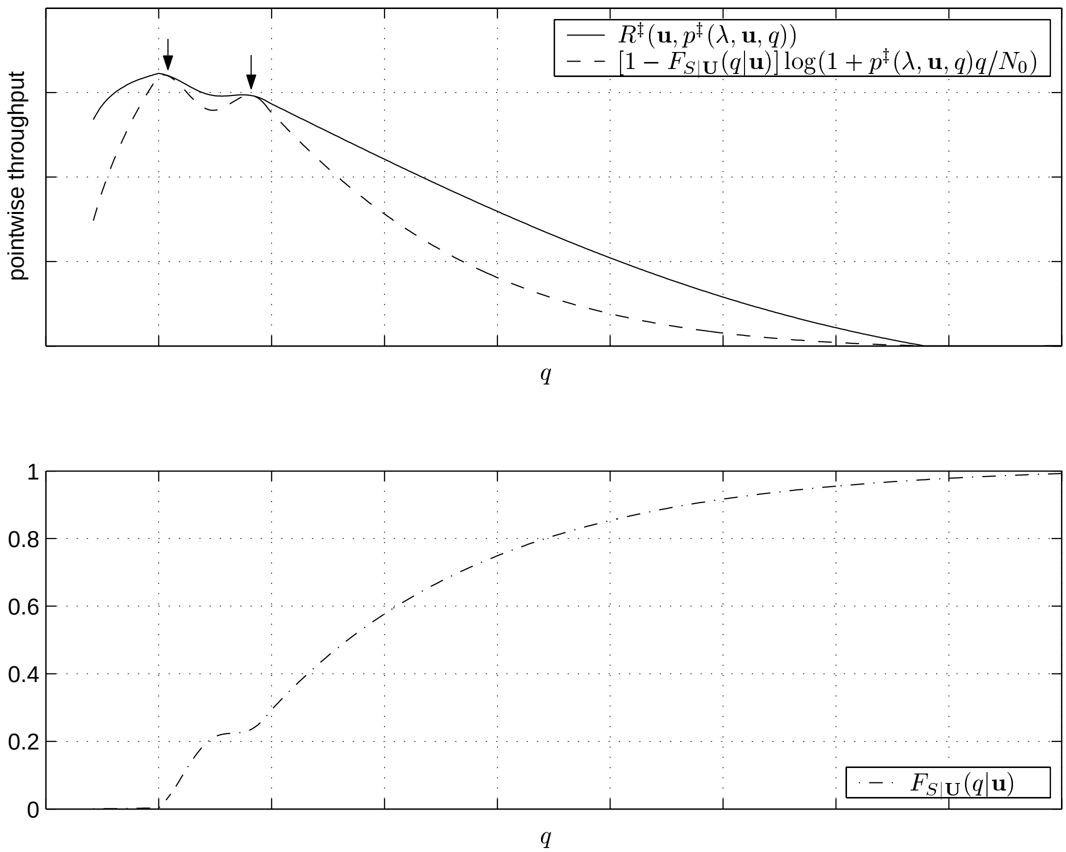

Figure 5.1: An example of two solutions satisfying both (5.10) and (5.16). The corresponding q's are indicated by arrows. Both curves in the top figure are throughputs corresponding to $p^\ddagger(\lambda, \mathbf{u}, q)$ which is the optimum given q. However, the solid line is the throughput maximized over all possible q and the dash line is the one obtained by simply choosing q as the worst channel state. Clearly, in this example, $q^\ddagger(\mathbf{u}, p^\ddagger(\lambda, \mathbf{u}, q))$ is not q typically.

(5.13), (5.14), and (5.18) imply that there is only one candidate optimum solution. However, the number of feasible solutions can be multiple. In Fig. 5.1, we have an example of two solutions. Since we do not constrain the related PDFs, the number of feasible $p(\mathbf{u})$ or $q(\mathbf{u})$ could be as many as uncountable.

However, without loss of generality, we limit our scope only to a countable number of feasible solutions at each $\mathbf{u}$. Consequently, we can add index to identify the possible solutions by $p_i(\mathbf{u})$ or $q_i(\mathbf{u})$. Moreover, with Lemma 10, it is possible to rank the solutions, which satisfy both (5.16) and (5.10), according to the allocated power $p(\mathbf{u})$ for each $\mathbf{u}$ given λ and preserves the order of allocated power. Hence, $p_{i_1}(\mathbf{u}) < p_{i_2}(\mathbf{u})$ for $i_1 < i_2$. On the other hand, it is also possible that the numbers of potential $p(\mathbf{u})$

	definition	claim 1)	claim 2)	claim 3)	claim 4)	claim 5)
$i\uparrow$	$p_i(\mathbf{u})\uparrow$	$R_i(\mathbf{u})\uparrow$	$R_i(\mathbf{u})/p_i(\mathbf{u})\downarrow$	$r_i(\mathbf{u})\downarrow$	$q_i(\mathbf{u})\downarrow$	$F_{S\vert\mathbf{U}}(q_i(\mathbf{u})\vert\mathbf{u})\downarrow$

Table 5.1: Monotonic relations for fixed $\mathbf{u}$

or $q(\mathbf{u})$ solutions differ for different $\mathbf{u}$. Therefore, we denote the number of solutions by $N(\mathbf{u})$. Consequently, at each $\mathbf{u}$, the possible values of an optimum policy can be enumerated by $\{(p_i(\mathbf{u}), q_i(\mathbf{u})),\ i \le N(\mathbf{u})\}$.

Let

$$\begin{aligned} r_i(\mathbf{u}) &= \log\left(1+\frac{p_i(\mathbf{u})q_i(\mathbf{u})}{N_0}\right), &(5.24)\\ R_i(\mathbf{u}) &= [1-F_{S|\mathbf{U}}(q_i(\mathbf{u})|\mathbf{u})]r_i(\mathbf{u}). &(5.25)\end{aligned}$$

Here, $R_i(\mathbf{u}) = R^{\ddagger}(\mathbf{u}, p_i(\mathbf{u}))$. In order to simplify our further discussions, we introduce

$$\begin{aligned} r_i'(\mathbf{u},p) &= \log\left(1+\frac{pq_i(\mathbf{u})}{N_0}\right), &(5.26)\\ R_i'(\mathbf{u},p) &= [1-F_{S|\mathbf{U}}(q_i(\mathbf{u})|\mathbf{u})]r_i'(\mathbf{u},p). &(5.27)\end{aligned}$$

So, $r_i(\mathbf{u}) = r_i'(\mathbf{u}, p_i(\mathbf{u}))$ and $R_i(\mathbf{u}) = R_i'(\mathbf{u}, p_i(\mathbf{u}))$. We observe that both $r_i'(\mathbf{u}, p)$ and $R_i'(\mathbf{u}, p)$ are strictly concave with respect to p. This concavity together with the (5.10) and (5.16) leads to the following theorem.

Theorem 13 *Given $\mathbf{u}$ and λ, the optimum policies have the following characteristics.*

1) $R_i(\mathbf{u})$ strictly increases in i,

2) $R_i(\mathbf{u})/p_i(\mathbf{u})$ strictly decreases in i,

3) $r_i(\mathbf{u})$ strictly decreases in i,

4) $q_i(\mathbf{u})$ strictly decreases in i,

5) $F_{S|\mathbf{U}}(q_i(\mathbf{u})|\mathbf{u})$ strictly decreases in i.

The claims in Theorem 13 are also listed in Table 5.1.

Since $F_{S|\mathbf{U}}(s|\mathbf{u}) \le 1$, the claim 1) and 3) in Theorem 13 suggest the following chain rule.

$$r_1(\mathbf{u}) > r_2(\mathbf{u}) > \cdots > r_{N(\mathbf{u})}(\mathbf{u}) > R_{N(\mathbf{u})}(\mathbf{u}) > \cdots > R_2(\mathbf{u}) > R_1(\mathbf{u}). \tag{5.28}$$

The meaning of the chain rule is that in order to increase the throughput return $R_i(\mathbf{u})$, it is required to reduce the assigned transmitted code rate $r_i(\mathbf{u})$. This result is very counter intuitive since it is quite different from those obtained with perfect CSIT where the assigned transmitted code rate is strictly increasing in the allocated transmitted power.

In practice, $N(\mathbf{u})$ may be a large number. This leads to high complexity in searching for $\{(p_i(\mathbf{u}), q_i(\mathbf{u}))\}$ as well as the policy design. However, (5.28) suggests that if there is an L such that $r_L(\mathbf{u})$ is very close to $R_L(\mathbf{u})$, we can upper bound the searching range of i by L with very little overall loss.

5.3.3 Pointwise Efficient Policy

It is shown that for any policy of interest, given $\mathbf{u}$ and λ, $R(\mathbf{u})$ and $p(\mathbf{u})$ can be chosen from $\{R_i(\mathbf{u})|i = 0, \ldots, N(\mathbf{u})\}$ and $\{p_i(\mathbf{u})|i = 0, \ldots, N(\mathbf{u})\}$, respectively. Assuming that both $\{R_i(\mathbf{u})|i = 0, \ldots, N(\mathbf{u})\}$ and $\{p_i(\mathbf{u})|i = 0, \ldots, N(\mathbf{u})\}$ are known for any $\mathbf{u}$ and λ, we want to find what the ART-maximizing policy is.

With the previous development in Subsection 5.3.2, we have

$$R(\mathbf{u}) = \sum_{i=1}^{N(\mathbf{u})} \left[R_i(\mathbf{u}) - R_{i-1}(\mathbf{u})\right] I_i(\mathbf{u}), \tag{5.29}$$

$$p(\mathbf{u}) = \sum_{i=1}^{N(\mathbf{u})} \left[p_i(\mathbf{u}) - p_{i-1}(\mathbf{u})\right] I_i(\mathbf{u}), \tag{5.30}$$

where the coefficients, $\{I_i(\mathbf{u})|i = 1, \ldots, N(\mathbf{u})\}$ is a set of binary 0/1-valued functions. Thus, given λ, $\mathcal{I} = \{I_i(\mathbf{u})|\mathbf{u} \in \mathcal{U}, i = 1, \ldots, N(\mathbf{u})\}$ describes a policy of interest. The corresponding ART and average power are

$$\bar{R}(\mathcal{I}) = \int_{\mathcal{U}} \left\{ \sum_{i=1}^{N(\mathbf{u})} \left[R_i(\mathbf{u}) - R_{i-1}(\mathbf{u})\right] I_i(\mathbf{u}) \right\} f_{\mathbf{U}}(\mathbf{u}) d\mathbf{u}, \tag{5.31}$$

$$\rho(\mathcal{I}) = \int_{\mathcal{U}} \left\{ \sum_{i=1}^{N(\mathbf{u})} \left[p_i(\mathbf{u}) - p_{i-1}(\mathbf{u})\right] I_i(\mathbf{u}) \right\} f_{\mathbf{U}}(\mathbf{u}) d\mathbf{u}. \tag{5.32}$$

Consequently, (5.8) can be rewritten as

$$\max_{\mathcal{I}} \quad \bar{R}(\mathcal{I}) \tag{5.33}$$

$$\text{subject to} \quad \rho(\mathcal{I}) \leq \overline{p}, \tag{5.33a}$$

Since both $R_i(\mathbf{u})$ and $p_i(\mathbf{u})$ strictly increase in i, a valid policy is required to satisfy a precedence constraint

$$I_{i'}(\mathbf{u}) \geq I_i(\mathbf{u}), \qquad i' \leq i. \tag{5.34}$$

The precedence constraint simply says that if $I_i(\mathbf{u}) = 1$, then $I_{i'}(\mathbf{u}) = 1$ for all $i' \leq i$. In addition, if $I_i(\mathbf{u}) = 0$, then $I_{i'}(\mathbf{u}) = 0$ for all $i' > i$.

Here, we can introduce a pointwise incremental efficiency or simply efficiency

$$\eta_i(\mathbf{u}) = \frac{R_i(\mathbf{u}) - R_{i-1}(\mathbf{u})}{p_i(\mathbf{u}) - p_{i-1}(\mathbf{u})}, \tag{5.35}$$

which is a ratio between an increment in the throughput from $R_{i-1}(\mathbf{u})$ to $R_i(\mathbf{u})$ given $\mathbf{u}$ and the corresponding power expenditure $p_i(\mathbf{u}) - p_{i-1}(\mathbf{u})$. Note that $\eta_i(\mathbf{u})$ is a concept depending on the fading model and channel estimate techniques. Later, we will give an algorithm which searching for the optimum policies based on $\eta_i(\mathbf{u})$.

The definition of $\eta_i(\mathbf{u})$ requires at least two solutions for satisfying both (5.16) and (5.10). It is clear that there is at least one such solution with $p_0(\mathbf{u}) = 0$ and $r_0(\mathbf{u}) = 0$. For some $\mathbf{u}$, given λ, it is possible that there is no feasible solutions with $p_i(\mathbf{u}) > 0$. However, we will see that for these $\mathbf{u}$'s, the optimum power allocated is zero.

Lemma 11 *Given any* $\mathbf{u}$ *and* λ*, we have*

1) $\eta_i(\mathbf{u})$ *is non-negative,*

2) $\eta_2(\mathbf{u}) < \eta_1(\mathbf{u})$.

Note that the claim 2) in Lemma 11 can be generalized to

$$\frac{R_{i_2}(\mathbf{u}) - R_{i_1}(\mathbf{u})}{p_{i_2}(\mathbf{u}) - p_{i_1}(\mathbf{u})} < \frac{R_{i_1}(\mathbf{u})}{p_{i_1}(\mathbf{u})}, \qquad i_2 > i_1. \tag{5.36}$$

And the proof stays the same.

With the definition of $\eta_i(\mathbf{u})$,

$$\bar{R}(\mathcal{I}) \quad = \quad \int_{\mathcal{U}} \left\{ \sum_{i=1}^{N(\mathbf{u})} \left[p_i(\mathbf{u}) - p_{i-1}(\mathbf{u})\right] \eta_i(\mathbf{u}) I_i(\mathbf{u}) \right\} f_{\mathbf{U}}(\mathbf{u})\, d\mathbf{u}, \tag{5.37}$$

Definition 2 The most power efficient allocation (MPEA) is a policy $\mathcal{I}^* = \left\{I_i^*(\mathbf{u})\right\}$ where

$$I_i^*(\mathbf{u}) = \begin{cases} 1 & \eta_i(\mathbf{u}) \geq \eta_{\text{lo}} \\ 0 & otherwise \end{cases}. \tag{5.38}$$

where the positive constant η_{lo} is determined by the average power constraint $\rho_L(\mathcal{I}^*) \leq \overline{p}$.

A very important remark about the definition of MPEA is that even though MPEA is unique with respect to η_{lo}, it may not be necessary that MPEA is unique with respect to $\overline{p}$.

Theorem 14 *If $\eta_i(\mathbf{u})$ is strictly decreasing in i, for any policy $\mathcal{I} = \{I_i(\mathbf{u})\}$ meeting the average constraint $\rho(I) \leq \overline{p}$,*

$$\bar{R}(\mathcal{I}) \leq \bar{R}(\mathcal{I}^*) \tag{5.39}$$

Therefore, MPEA is the solution of our ART maximization (5.33) if $\eta_i(\mathbf{u})$ is strictly decreasing in i. Moreover, we emphasize that Theorem 14 holds whether the policy $I_i(\mathbf{u})$ satisfies the precedence constraint. Due to the claim 2) in Lemma 11, MPEA is optimum for any distributions and estimation techniques such that $N(\mathbf{u}) < 3$ for all $\mathbf{u} \in \mathcal{U}$. Indeed, this condition can be satisfied. For instance, if the estimation is close to the perfect, $N(\mathbf{u})$ will be close to 1 given the fact that with perfect CSIT, $N(\mathbf{u}) = 1$.

On the other hand, it is possible that $\eta_i(\mathbf{u})$ is not strictly decreasing in i as described in Section 5.B. In this case, some dynamic programming techniques may be required to search for the optimum solutions. However, such cases are not covered in this chapter.

5.3.4 Approximation of the Maximum ART

If $\eta_i(\mathbf{u})$ is not strictly decreasing in i, we can still obtain some approximations of the maximum ART.

Regardless of $\mathbf{u}$ and λ, due to (5.28), for an optimum policy $\mathcal{I}^*$, the non-zero $R(\mathbf{u})$ satisfies

$$r_1(\mathbf{u}) > R(\mathbf{u}) \geq R_1(\mathbf{u}). \tag{5.40}$$

Then, we can construct a suboptimum MPEA policy $\mathcal{I}'$ with $R'(\mathbf{u}) \in \{0, R_1(\mathbf{u})\}$ and the policy has an efficiency lower bound η'_{lo}. Let

$$\mathcal{U}^* = \{\mathbf{u}|R(\mathbf{u}) > 0\}, \tag{5.41}$$

$$\mathcal{U}' = \left\{\mathbf{u}|R'(\mathbf{u}) > 0\right\}. \tag{5.42}$$

We can find that $\mathcal{U}^*$ is contained within $\mathcal{U}'$.

Proposition 7

$$\mathcal{U}^* \subset \mathcal{U}'. \tag{5.43}$$

Therefore, since $\bar{R}(\mathcal{I}^*) = \int_{\mathbf{u}\in\mathcal{U}} R(\mathbf{u})f(\mathbf{u})\,d\mathbf{u} = \int_{\mathbf{u}\in\mathcal{U}^*} R(\mathbf{u})f(\mathbf{u})\,d\mathbf{u}$,

$$\int_{\mathbf{u}\in\mathcal{U}'} r_1(\mathbf{u})f(\mathbf{u})\,d\mathbf{u} > \int_{\mathbf{u}\in\mathcal{U}^*} r_1(\mathbf{u})f(\mathbf{u})\,d\mathbf{u} > \bar{R}(\mathcal{I}^*) \geq \bar{R}(\mathcal{I}') = \int_{\mathbf{u}\in\mathcal{U}'} R_1(\mathbf{u})f(\mathbf{u})\,d\mathbf{u}. \tag{5.44}$$

Clearly, the bounds given in (5.44) is tight if $F(q_1(\mathbf{u})|\mathbf{u})$ or $F(q_1(\mathbf{u})|\mathbf{u})f(\mathbf{u})$ is small.

On the other hand, the bounds in (5.44) can be improved by considering more $R_i(\mathbf{u})$'s. For instance, if $R_2(\mathbf{u})$ and $r_2(\mathbf{u})$ are known, we can find $R''(\mathbf{u})$ given by MPEA determined by limiting $R''(\mathbf{u}) \in \{0, R_1(\mathbf{u}), R_2(\mathbf{u})\}$. Then,

$$\bar{R}(\mathcal{I}^*) \geq \int_{\mathbf{u}\in\mathcal{U}} R''(\mathbf{u})f(\mathbf{u})\,d\mathbf{u}. \tag{5.45}$$

In addition, since

$$r_2(\mathbf{u}) > R(\mathbf{u}), \tag{5.46}$$

from the chain rule (5.28) and (5.43), we have

$$\int_{\mathbf{u}\in\mathcal{U}'} r_2(\mathbf{u})f(\mathbf{u})\,d\mathbf{u} > \bar{R}(\mathcal{I}^*) \geq \int_{\mathbf{u}\in\mathcal{U}} R''(\mathbf{u})f(\mathbf{u})\,d\mathbf{u}. \tag{5.47}$$

Surely, the tightest upper bound in this analogy is given by

$$\int_{\mathbf{u}\in\mathcal{U}'} R_{N(\mathbf{u})}(\mathbf{u})f(\mathbf{u})\,d\mathbf{u} > \bar{R}(\mathcal{I}^*). \tag{5.48}$$

5.3.5 Searching for Optimum Policies

The concept of MPEA indeed shows a mean to search for the optimum solution of (5.33) if $\eta_i(\mathbf{u})$ is strictly decreasing in i. However, the complexity of such a search is extremely high due to the fact that the measurement $\mathbf{u}$ is a vector instead of a scalar.

An observation from (5.10) and (5.16) is that both $p(\mathbf{u})$ and $q(\mathbf{u})$ of interest are determined by $f_{S|\mathbf{U}}(s|\mathbf{u})$ or $F_{S|\mathbf{U}}(s|\mathbf{u})$ given λ. So, in the following, we will apply the concept of sufficient statistic to simplify out derivation solution. When $T(\mathbf{u})$ is a sufficient statistic for s, according to [56]

$$f_{\mathbf{U}|S}(\mathbf{u}|s) = f_{T(\mathbf{U})|S}(T(\mathbf{u})|s)h(\mathbf{u}), \quad h(\mathbf{u}) \geq 0, \tag{5.49}$$

where $h(\mathbf{u})$ is a deterministic function which is independent of s, and

$$f_{S|\mathbf{U}}(s|\mathbf{u}) \quad = \quad \frac{f_{T(\mathbf{U})|S}(T(\mathbf{u})|s)f_S(s)}{\int_0^\infty f_{T(\mathbf{U})|S}(T(\mathbf{u})|s)f_S(s)ds}. \tag{5.50}$$

Proposition 8 *If $\eta_i(\mathbf{u})$ is strictly decreasing in i, for an optimum policy $(p^*(\mathbf{u}), q^*(\mathbf{u}))$, the power allocation $p^*(\mathbf{u})$ and the worst state assignment $q^*(\mathbf{u})$ are determined by $T(\mathbf{u})$, which is the sufficient statistic for s.*

This theorem is the consequence of that for any $\mathbf{u}$ with the same $f_{\mathbf{U}|S}(\mathbf{u}|s)$, MPEA has the same set of $\{(p(\mathbf{u}), q(\mathbf{u}))\}$ satisfying both (5.19) and (5.20). Therefore, given λ, the search for MPEA can be simplified from a search of $\mathbf{u}$ over $\mathcal{U}$ to a search of v where v is a value of a random variable $V = T(\mathbf{U})$ with a domain $\mathcal{V} = (-\infty, +\infty)$. Equivalently, v indicates a set of measurements, $\{\mathbf{u}|T(\mathbf{u}) = v\}$, which share the same $p(v)$ and $q(v)$.

And a series of functions are re-defined accordingly,

$$
\begin{aligned}
p(v) &= p(\mathbf{u}), \qquad \{\mathbf{u}|T(\mathbf{u}) = v\}, && (5.51)\\
r(v) &= r(\mathbf{u}), \qquad \{\mathbf{u}|T(\mathbf{u}) = v\}, && (5.52)\\
q(v) &= \frac{N_0}{p(v)}\left(e^{r(v)} - 1\right), && (5.53)\\
\bar{R}\left(p(\cdot), q(\cdot)\right) &= \int_{\mathcal{V}} \left[1 - F_{S|V}(q(v)|v)\right] \log\left(1 + \frac{p(v)q(v)}{N_0}\right) f_V(v)\, dv, && (5.54)\\
\rho\left(p(\cdot), q(\cdot)\right) &= \int_{\mathcal{V}} p(v) f_V(v)\, dv, && (5.55)\\
\eta_i(v) &= \frac{R_i(v) - R_{i-1}(v)}{p_i(v) - p_{i-1}(v)}, && (5.56)
\end{aligned}
$$

where $R_i(v)$, $r_i(v)$, $q_i(v)$, and $p_i(v)$ are the i-th candidate solution given v.

An algorithm for finding optimum policies, if $\eta_i(\mathbf{u}) < \eta_{i-1}(\mathbf{u})$ for all $i \leq N(v)$, is shown in Fig. 5.2. The domain $\mathcal{S}_\lambda$ is not easy to decide. However, due to (5.16), we have

$$p(v) < \frac{1}{\lambda}. \tag{5.57}$$

Therefore, $\lambda \in \mathcal{S}_\lambda$ must satisfy

$$\int_{v\in\mathcal{V}} \frac{1}{\lambda} f_V(\,dv) \geq \overline{p}. \tag{5.58}$$

So, the upper bound of $\mathcal{S}_\lambda$ can be set to the value when (5.58) achieves equality. On the other hand, due to the continuity of all functions, the lower bound of $\mathcal{S}_\lambda$ can be determined by reducing the lower bound progressively until the increment in the corresponding $\bar{R}(\mathcal{I}^*)$ is fractional.

Normally, we may not able to fully search over $\mathcal{V}$ and all possible λ for optimum $\bar{R}(\mathcal{I}^*)$ since given each v and λ, we need to find $p(v)$ and $q(v)$. In practice, a search with a good resolution is assumed.

5.4 Training Based Adaptation for Rayleigh Fading

The transmission is designed in such a manner that, before transmitting a data message, the channel is probed by a training sequence consisting of M identical symbols x_{T}. Based on the corresponding M received symbols $\mathbf{u} = [u_0, \ldots, u_{M-1}]^\top$, the transmitter

0. Let $\bar{R} = 0$, $l = 0$, and $\mathcal{I} = \Phi$;
1. Specify a domain $\mathcal{V}_e$ such that $1 - \Pr[v \in \mathcal{V}_e]$ is negligible;
2. Specify a domain $\mathcal{S}_\lambda$ for possible λ's;
3. Choose a $\lambda \in \mathcal{S}_\lambda$ and let $\mathcal{S}_\lambda = \mathcal{S}_\lambda \backslash \{\lambda\}$;
4. Find the number of candidate solutions $N(v)$ and $\{(p_i(v), q_i(v))\}$ for all $v \in \mathcal{V}_e$;
5. Find MPEA $\mathcal{I}^*$ and corresponding maximum ART $\bar{R}(\mathcal{I}^*)$;
6. If $\bar{R}(\mathcal{I}^*) > \bar{R}$, let $\bar{R} = \bar{R}(\mathcal{I}^*)$, $l = \lambda$, and $\mathcal{I} = \mathcal{I}^*$;
7. Repeat 3, 4, 5, and 6 until $\mathcal{S}_\lambda = \Phi$.

Figure 5.2: An algorithm searching for optimum policies if $\eta_i(v) < \eta_{i-1}(v)$ for all $i \le N(v)$.

selects a code rate $r(\mathbf{u})$ to encode and a power level $p(\mathbf{u})$ to transmit the data message. In this context, we have

$$\mathbf{x}_{\mathrm{T}} = [x_{\mathrm{T}}, \ldots, x_{\mathrm{T}}], \tag{5.59}$$

and without a loss of generality, we assume $x_{\mathrm{T}} = \sqrt{\mathcal{E}_{\mathrm{T}}}$. Note that $\mathcal{E}_{\mathrm{T}}$ is an amount of energy and the signal-to-noise ratio (SNR) for the training sequence is

$$\gamma_{\mathrm{T}} = M \frac{E\{\mathcal{E}_{\mathrm{T}} S\}}{N_0}. \tag{5.60}$$

5.4.1 Coherent Model

When the phase of the received signals is accurately known at the receiver, we can assume that the training symbols experience the same channel s as the transmitted information (payload), we have

$$f_{\mathbf{U}|S}(\mathbf{u}|s) = (\pi N_0)^{-M} \exp\left\{ -\frac{||\mathbf{u} - \sqrt{s}\mathbf{x}_{\mathrm{T}}||^2}{N_0} \right\}. \tag{5.61}$$

A sufficient statistic for s is

$$v = \Re(\mathbf{u}^\dagger \mathbf{x}_{\mathrm{T}}). \tag{5.62}$$

Hence, we have

$$f_{V|S}(v|s) = (\pi M\mathcal{E}_{\mathrm{T}} N_0)^{-\frac{1}{2}} \exp\left[-\frac{(v - M\mathcal{E}_{\mathrm{T}}\sqrt{s})^2}{M\mathcal{E}_{\mathrm{T}} N_0}\right]. \tag{5.63}$$

For a Rayleigh fading channel with

$$f_S(s) = e^{-s}, \tag{5.64}$$

we have

$$f_{S,V}(s,v) = c_a^{-\frac{1}{2}} \left[\pi N_0 c_b\right]^{-\frac{1}{2}} \exp\left[\frac{(c_a c_b^2 - 1)v^2}{N_0 M\mathcal{E}_{\mathrm{T}}}\right] \exp\left[-\frac{(\sqrt{s} - c_b v)^2}{N_0 c_b}\right] \tag{5.65}$$

where $c_a = M\mathcal{E}_{\mathrm{T}}(M\mathcal{E}_{\mathrm{T}} + N_0)$ and $c_b = M\mathcal{E}_{\mathrm{T}}/c_a$.

Then,

$$f_V(v) = c_a^{-\frac{1}{2}} \exp\left[\frac{(c_a c_b^2 - 1)v^2}{N_0 M\mathcal{E}_{\mathrm{T}}}\right] \left[c_b v \operatorname{erfc}\left(-\sqrt{\frac{c_b}{N_0}} v\right) + \sqrt{\frac{N_0 c_b}{\pi}} \exp\left(-\frac{c_b}{N_0} v^2\right)\right] \tag{5.66}$$

$$1 - F_{S|V}(q|v) = \frac{\sqrt{c_b} v \operatorname{erfc}\left(\sqrt{\frac{q}{N_0 c_b}} - \sqrt{\frac{c_b}{N_0}} v\right) + \sqrt{\frac{N_0}{\pi}} \exp\left(-\left(\sqrt{\frac{q}{N_0 c_b}} - \sqrt{\frac{c_b}{N_0}} v\right)^2\right)}{\sqrt{c_b} v \operatorname{erfc}\left(-\sqrt{\frac{c_b}{N_0}} v\right) + \sqrt{\frac{N_0}{\pi}} \exp\left(-\frac{c_b}{N_0} v^2\right)}, \tag{5.67}$$

where

$$\operatorname{erfc}(\phi) = \frac{2}{\sqrt{\pi}} \int_{\phi}^{\infty} e^{-t^2} dt. \tag{5.68}$$

Here, we first numerically search for $p_i(v)$ and $q_i(v)$. It turns out that given any v and λ within the range of interest, $N(v) = 1$. Then, we numerically search for the optimum policies using the algorithm in the previous section. In Fig. 5.3, we compare the maximum $\bar{R}$ corresponding to different γ_{T} with the ergodic capacity C [22] and C_1, which is the capacity for non-adaptive transmission Chapter 2.For $\gamma_{\mathrm{T}} = 0$ dB, the maximum $\bar{R}$ is about 1 dB better than C_1. Increasing γ_{T} by 10 or 20 dB, the performance gap between C and the maximum $\bar{R}$ can be reduced by 2 or 4 dB, respectively. In order to achieve any the maximum $\bar{R}$ within 1 dB from C, we need to have a substantially large γ_{T}.

5.4.2 Non-coherent Model

In general, the knowledge of the phase of the received signals is with uncertainty. In this subsection, we assume the worst scenario that the phase is uniformly distributed.

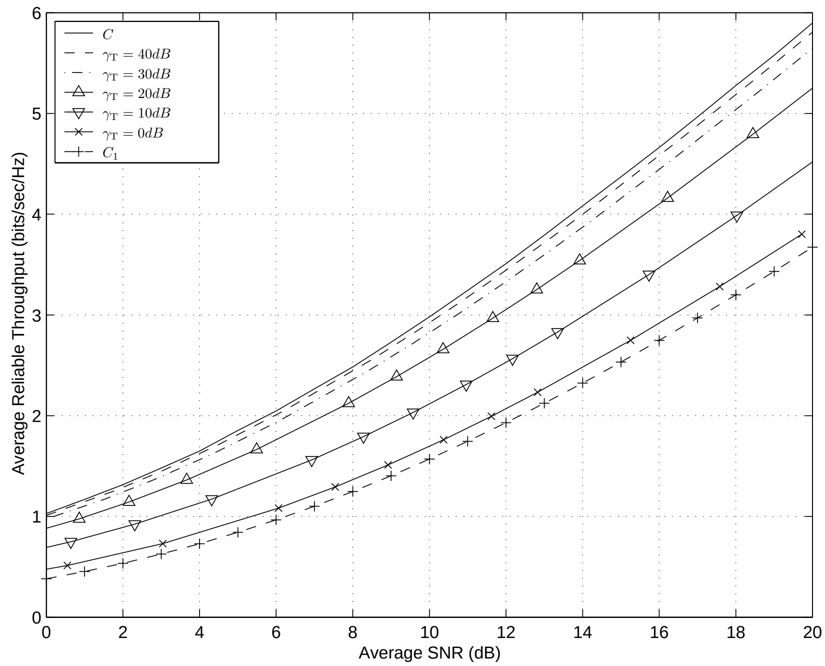

Figure 5.3: Sample performance for coherent model in Rayleigh fading.

However, to alleviate the difficulty, we also assume that all received training symbols are with the same phase Θ, which is a random variable uniformly distributed over $[0, 2\pi)$, and experience the same channel state as the transmitted information (payload). Therefore, we have a channel model

$$Y = \sqrt{S}e^{j\Theta}X + W, \tag{5.69}$$

and, so,

$$f_{\mathbf{U}|S,\Theta}(\mathbf{u}|s,\theta) = (\pi N_0)^{-M} \exp\left\{-\frac{||\mathbf{u} - \sqrt{s}e^{j\theta}\mathbf{x}_{\mathrm{T}}||^2}{N_0}\right\}. \tag{5.70}$$

Since $f_\Theta(\theta) = 1/2\pi$,

$$f_{\mathbf{U}|S}(\mathbf{u}|s) = (\pi N_0)^{-M} \exp\left\{-\frac{||\mathbf{u}||^2 + M\mathcal{E}_{\mathrm{T}}s}{N_0}\right\} I_0\left(\frac{\sqrt{s}}{\sigma^2}|\mathbf{u}^\dagger\mathbf{x}_{\mathrm{T}}|\right). \tag{5.71}$$

A sufficient statistic for s is

$$v \quad = \quad |\mathbf{u}^\dagger\mathbf{x}_{\mathrm{T}}|. \tag{5.72}$$

Hence, we have

$$f_{V|S}(v|s) = \frac{v}{\Sigma^2}e^{-(M^2\mathcal{E}_{\mathrm{T}}^2 s + v^2)/2\Sigma^2} I_0\left(\frac{\sqrt{s}v}{\sigma^2}\right), \tag{5.73}$$

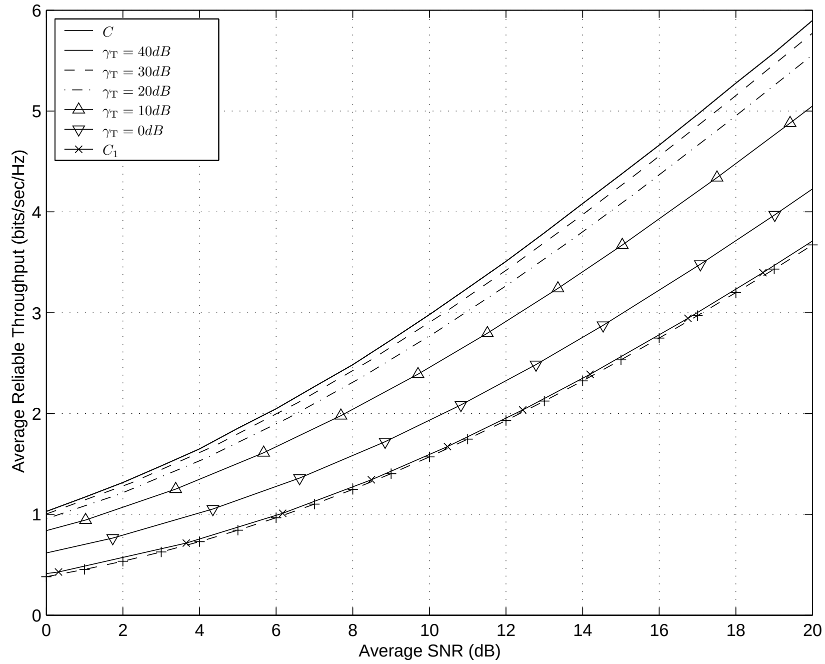

Figure 5.4: Sample performance for non-coherent model in Rayleigh fading.

where

$$\Sigma^2 = M\mathcal{E}_{\mathrm{T}}\sigma^2. \tag{5.74}$$

For a Rayleigh fading channel with

$$f_S(s) = e^{-s}, \tag{5.75}$$

we have

$$f_{S,V}(s,v) = \frac{v}{\Sigma^2} e^{-((M^2\mathcal{E}_{\mathrm{T}}^2+2\Sigma^2)s+v^2)/2\Sigma^2} I_0\left(\frac{\sqrt{s}v}{\sigma^2}\right). \tag{5.76}$$

Unfortunately, we are not able to obtain the close form of either $f_V(v)$ or $f_{S|V}(s|v)$. However, it is still possible for us to evaluate the maximum $\bar{R}$ numerically for this case as shown in Fig. 5.4. Similar to the coherent model, given v and λ within the range of interest, $N(v) = 1$. In comparison with Fig. 5.3, we notice that there is a loss on the capacity when the phase of the received training signals is unknown. The loss is extremely large for small γ_{T}. For instance, if $\gamma_{\mathrm{T}} = 1$ (0dB), there is only a negligible difference between the capacity of adaptive transmission and non-adaptive one. The difference between Fig. 5.3 and Fig. 5.4 suggests whether phase information is important or not. It is shown that with phase information, the maximum $\bar{R}$ can be improved by 0.5 dB to 1 dB for $\gamma_{\mathrm{T}} \leq 20$ dB. For high γ_{T}, the phase information is not that important.

5.5 Discussion

Note that the maximum ART is a meaningful measurement of capacity only when a training sequence consumes a negligible fraction of both the communication bandwidth and power. Therefore, the meaning of the results in Fig. 5.3 is twofold. Firstly, it implies that some types of decision feedback channel estimation techniques must be used in order to achieve ART close to C since it may not be possible to have a training sequence with very high power/energy or as many as thousands of symbols of small symbol energy. On the other hand, these results also imply that there is a good reason for not using a training sequence with fixed energy/power. A possible approach is to adapt the training energy/power based on a predicted CSI.

There are still a lot of unknowns within the framework of this chapter. For instant, we do not know the exact $N(\mathbf{u})$, and how to find $\{p_i(\mathbf{u})\}$ and $\{q_i(\mathbf{u})\}$ efficiently. In addition, we do not have an algorithm to find optimum policies for arbitrary $f_{S|\mathbf{U}}(s, \mathbf{u})$.

5.A Proofs

Proof: Lemma 10

Given $\mathbf{u}$ and λ, let (p, q) and (p, q') be two pairs satisfy both (5.16) and (5.10) though $q \neq q'$. Let

$$r = \log\left(1 + \frac{pq}{N_0}\right), \tag{5.77}$$

$$r' = \log\left(1 + \frac{pq'}{N_0}\right). \tag{5.78}$$

Since (5.10) is satisfied by both (p, q) and (p, q'),

$$R^{\ddagger}(\mathbf{u}, p) = \left[1 - F_{S|\mathbf{U}}(q|\mathbf{u})\right] r = \left[1 - F_{S|\mathbf{U}}(q'|\mathbf{u})\right] r'. \tag{5.79}$$

On the other hand, according to (5.23),

$$1 - F_{S|\mathbf{U}}(q|\mathbf{u}) = \lambda \left[p + \frac{N_0}{q} \right] = \lambda \frac{p}{1 - e^{-r}}, \tag{5.80}$$

$$1 - F_{S|\mathbf{U}}(q'|\mathbf{u}) = \lambda \left[p + \frac{N_0}{q'} \right] = \lambda \frac{p}{1 - e^{-r'}}. \tag{5.81}$$

Hence, (5.79) implies

$$\frac{r}{1 - e^{-r}} = \frac{r'}{1 - e^{-r'}}. \tag{5.82}$$

Since $x/(1 - \exp(-x))$ strictly increases in x when $x > 0$, we have

$$r = r'. \tag{5.83}$$

Therefore, (5.79) implies

$$F_{S|\mathbf{U}}(q|\mathbf{u}) = F_{S|\mathbf{U}}(q'|\mathbf{u}). \tag{5.84}$$

Since $F_{S|\mathbf{U}}(s|\mathbf{u})$ is assumed to strictly increase in s, $q' = q$. This controdicts the original assumption and the lemma holds.

□

Proof: Theorem 13

The claim 1) is straightforward from the definition of $p_i(\mathbf{u})$.

Given $\mathbf{u}$ and $i_1 < i_2$,

$$p_{i_1}(\mathbf{u}) < p_{i_2}(\mathbf{u}), \tag{5.85}$$

and

$$R_{i_1}(\mathbf{u}) \overset{(a)}{<} r'_{i_1}(\mathbf{u}, p_{i_2}(\mathbf{u})) \overset{(b)}{<} R_{i_2}(\mathbf{u}), \tag{5.86}$$

where (a) results from the monotonicity of the logrithm function and (b) is due to (5.10). Therefore, the claim 2) holds.

Due to (5.10), $R'_{i_1}(\mathbf{u}, p)$ achieves $R_{i_1}(\mathbf{u})$ by spending the minimum amount of power $p_{i_1}(\mathbf{u})$. Therefore,

$$\frac{R_{i_1}(\mathbf{u})}{p_{i_1}(\mathbf{u})} = \frac{R'_{i_1}(\mathbf{u}, p_{i_1}(\mathbf{u}))}{p_{i_1}(\mathbf{u})} > \frac{R'_{i_2}(\mathbf{u}, p_{i_1}(\mathbf{u}))}{p_{i_1}(\mathbf{u})}. \tag{5.87}$$

Since $\log(1+\alpha x)/x$ is strictly decreasing in x for $\alpha > 0$ and $x > 0$, we have

$$\frac{R'_{i_2}(\mathbf{u}, p_{i_1}(\mathbf{u}))}{p_{i_1}(\mathbf{u})} > \frac{R'_{i_2}(\mathbf{u}, p_{i_2}(\mathbf{u}))}{p_{i_2}(\mathbf{u})}. \tag{5.88}$$

Since $R_{i_2}(\mathbf{u}) = R'_{i_2}(\mathbf{u}, p_{i_2}(\mathbf{u}))$, (5.87) and (5.88) implies

$$\frac{R_{i_1}(\mathbf{u})}{p_{i_1}(\mathbf{u})} > \frac{R_{i_2}(\mathbf{u})}{p_{i_2}(\mathbf{u})}, \qquad i_1 < i_2 \tag{5.89}$$

and, thus, the claim 3) holds.

According to (5.23),

$$p_i(\mathbf{u}) = \frac{[1 - F_{S|\mathbf{U}}(q_i(\mathbf{u})|\mathbf{u})]}{\lambda}\left\{1 - e^{-r_i(\mathbf{u})}\right\}. \tag{5.90}$$

Therefore, (5.25) implies

$$\frac{R_i(\mathbf{u})}{p_i(\mathbf{u})} = \frac{[1 - F_{S|\mathbf{U}}(q_i(\mathbf{u})|\mathbf{u})]r_i(\mathbf{u})}{p_i(\mathbf{u})} = \lambda\frac{r_i(\mathbf{u})}{1 - \exp(-r_i(\mathbf{u}))}. \tag{5.91}$$

Since $x/(1-\exp(-x))$ strictly increases in x for any $x > 0$, (5.91) implies that $R_i(\mathbf{u})/p_i(\mathbf{u})$ increases in $r_i(\mathbf{u})$. Consequently, the claim 3) leads to the claim 4).

With (5.3),

$$q_i(\mathbf{u}) = N_0\frac{e^{r_i(\mathbf{u})} - 1}{p_i(\mathbf{u})}. \tag{5.92}$$

Hence, the claim 1) and 4) imply the claim 5).

With (5.23), the claim 6) results from the claim 1) and 5). On the other hand, the claim 6) can also be directly obtained from claim 5) and the monotonicity of $F_{S|\mathbf{U}}(s|\mathbf{u})$.

□

Proof: Lemma 11

The claim 1) is trivial since we know both $p_i(\mathbf{u})$ and $R_i(\mathbf{u})$ strictly increases in i stated by the claim 1) and 2) in Theorem 13.

Let

$$\alpha = \frac{R_1(\mathbf{u})}{p_1(\mathbf{u})}, \tag{5.93}$$

$$\beta = \frac{R_2(\mathbf{u})}{p_2(\mathbf{u})}. \tag{5.94}$$

According to the claim 3) in Theorem 13, $\alpha > \beta$.

$$\begin{aligned} \eta_2(\mathbf{u}) &= \frac{R_2(\mathbf{u}) - R_1(\mathbf{u})}{p_2(\mathbf{u}) - p_1(\mathbf{u})} && (5.95)\\ &= \frac{\beta p_2(\mathbf{u}) - \alpha p_1(\mathbf{u})}{p_2(\mathbf{u}) - p_1(\mathbf{u})} && (5.96)\\ &= \frac{(\beta - \alpha) p_2(\mathbf{u})}{p_2(\mathbf{u}) - p_1(\mathbf{u})} + \alpha && (5.97)\\ &\overset{(a)}{<} \alpha, && (5.98) \end{aligned}$$

where (a) is due to $\alpha > \beta$ and $p_2(\mathbf{u}) > p_1(\mathbf{u})$ implied by the claim 1) in Theorem 13. Therefore, the claim 2) holds.

□

Proof: Theorem 14

Given an arbitrary power allocation $\mathcal{I}$ and the MPEA allocation $\mathcal{I}^*$, we observe that

$$I_i^*(\mathbf{u}) = I_i^*(\mathbf{u}) I_i(\mathbf{u}) + I_i^*(\mathbf{u}) \left[1 - I_i(\mathbf{u})\right], \tag{5.99}$$

For $\rho_L(\mathcal{I})$ given by (5.32),

$$\rho_L(\mathcal{I}^*) = \rho_L\left(\left\{I_i^*(\mathbf{u})\right\}\right) = \rho_L\left(\left\{I_i^*(\mathbf{u}) I_i(\mathbf{u})\right\}\right) + \rho_L\left(\left\{I_i^*(\mathbf{u})\left[1 - I_i(\mathbf{u})\right]\right\}\right). \tag{5.100}$$

It follows from (5.31) and (5.99) that the MPEA policy achieves ART

$$\begin{aligned} R_L(\mathcal{I}^*) &= \int_{\mathcal{U}} \sum_{i=1}^{N(\mathbf{u})} I_i^*(\mathbf{u}) \eta_i(\mathbf{u}) \left(p_i(\mathbf{u}) - p_{i-1}(\mathbf{u})\right) f(\mathbf{u})\, d\mathbf{u} && (5.101)\\ &= \kappa + \int_{\mathcal{U}} \sum_{i=1}^{N(\mathbf{u})} I_i^*(\mathbf{u}) \left[1 - I_i(\mathbf{u})\right] \eta_i(\mathbf{u}) \left(p_i(\mathbf{u}) - p_{i-1}(\mathbf{u})\right) f(\mathbf{u})\, d\mathbf{u} && (5.102) \end{aligned}$$

where

$$\kappa = \int_{\mathcal{U}} \sum_{i=1}^{N(\mathbf{u})} I_i^*(\mathbf{u}) I_i(\mathbf{u}) \eta_i(\mathbf{u}) \left(p_i(\mathbf{u}) - p_{i-1}(\mathbf{u})\right) f(\mathbf{u})\, d\mathbf{u}. \tag{5.103}$$

Consequently,

$$
\begin{aligned}
R_L(\mathcal{I}^*) &\overset{(a)}{\geq} \kappa + \int_{\mathcal{U}} \sum_{i=1}^{N(\mathbf{u})} I_i^*(\mathbf{u}) \left[1 - I_i(\mathbf{u})\right] \eta_{\mathrm{lo}} \left(p_i(\mathbf{u}) - p_{i-1}(\mathbf{u})\right) f(\mathbf{u})\, d\mathbf{u} && (5.104)\\
&\overset{(b)}{=} \kappa + \eta_{\mathrm{lo}} \left(\rho_L\left(\mathcal{I}^*\right) - \rho_L\left(\left\{I_i^*(\mathbf{u}) I_i(\mathbf{u})\right\}\right)\right) && (5.105)\\
&\overset{(c)}{\geq} \kappa + \eta_{\mathrm{lo}} \left(\rho_L\left(\mathcal{I}\right) - \rho_L\left(\left\{I_i(\mathbf{u}) I_i^*(\mathbf{u})\right\}\right)\right) && (5.106)\\
&\overset{(d)}{=} \kappa + \int_{\mathcal{U}} \sum_{i=1}^{N(\mathbf{u})} I_i(\mathbf{u}) \left[1 - I_i^*(\mathbf{u})\right] \eta_{\mathrm{lo}} \left(p_i(\mathbf{u}) - p_{i-1}(\mathbf{u})\right) f(\mathbf{u})\, d\mathbf{u} && (5.107)\\
&\overset{(e)}{\geq} \kappa + \int_{\mathcal{U}} \sum_{i=1}^{N(\mathbf{u})} I_i(\mathbf{u}) \left[1 - I_i(\mathbf{u})\right] \eta_i(\mathbf{u}) \left(p_i(\mathbf{u}) - p_{i-1}(\mathbf{u})\right) f(\mathbf{u})\, d\mathbf{u} && (5.108)\\
&\overset{(f)}{=} \int_{\mathcal{U}} \sum_{i=1}^{N(\mathbf{u})} I_i(\mathbf{u}) \eta_i(\mathbf{u}) \left(p_i(\mathbf{u}) - p_{i-1}(\mathbf{u})\right) f(\mathbf{u})\, d\mathbf{u} && (5.109)\\
&= R_L(\mathcal{I}), && (5.110)
\end{aligned}
$$

where

(a) for $\mathbf{u}$ with $I_i^*(\mathbf{u}) = 1$, $\eta_i(\mathbf{u}) \geq \eta_{\mathrm{lo}}$;

(b) is due to (5.32) and (5.100);

(c) is from $\overline{p} = \rho_L(\mathcal{I}^*) \geq \rho_L(\mathcal{I})$;

(d) we apply (5.100) on $\mathcal{I}$ instead of $\mathcal{I}^*$;

(e) for the channel states with $[1 - I_i^*(\mathbf{u})] = 1$, $\eta_i(\mathbf{u}) < \eta_{\mathrm{lo}}$;

(f) we substitute κ from (5.103) and apply the relation of (5.99) on $\mathcal{I}$ again.

□

Proof: Proposition 7

Let

$$\mathcal{A} = \mathcal{U}^* \cap \mathcal{U}', \tag{5.111}$$

$$\mathcal{B} = \mathcal{U}^* \backslash \mathcal{A}, \tag{5.112}$$

$$\mathcal{C} = \mathcal{U}' \backslash \mathcal{A}. \tag{5.113}$$

Given a set of channel states Φ, define

$$\rho^*(\Phi) = \int_{\mathbf{u} \in \Phi} p(\mathbf{u})\, d\mathbf{u}, \tag{5.114}$$

$$\rho'(\Phi) = \int_{\mathbf{u} \in \Phi} p'(\mathbf{u})\, d\mathbf{u}, \tag{5.115}$$

where $p'(\mathbf{u})$ is the power allocation for the policy $\mathcal{I}'$.

Since for any $\mathbf{u} \in \mathcal{A}$, $R(\mathbf{u}) \geq R'(\mathbf{u})$, due to the claim 1) in Theorem 13, we have $p(\mathbf{u}) \geq p'(\mathbf{u})$. Thus,

$$\rho^*(\mathcal{A}) \geq \rho'(\mathcal{A}). \tag{5.116}$$

In addition, since both $\mathcal{I}^*$ and $\mathcal{I}'$ have the same average power $\overline{p}$,

$$\rho^*(\mathcal{B}) = \overline{p} - \rho^*(\mathcal{A}) \leq \overline{p} - \rho'(\mathcal{A}) = \rho'(\mathcal{C}). \tag{5.117}$$

Hence, it is always possible to find a subset $\mathcal{C}' \subset \mathcal{C}$ such that

$$\rho'(\mathcal{C}') = \rho^*(\mathcal{B}). \tag{5.118}$$

On the other hand, given a set of channel states Φ, we can define

$$\bar{R}^*(\Phi) = \int_{\mathbf{u}\in\Phi} R(\mathbf{u})\, d\mathbf{u}, \tag{5.119}$$

$$\bar{R}'(\Phi) = \int_{\mathbf{u}\in\Phi} R'(\mathbf{u})\, d\mathbf{u}. \tag{5.120}$$

Hence, we have

$$\begin{aligned}
\bar{R}^*(\mathcal{B}) &= \int_{\mathbf{u}\in\mathcal{B}} R(\mathbf{u})\, d\mathbf{u} && (5.121)\\
&= \int_{\mathbf{u}\in\mathcal{B}} \frac{R(\mathbf{u})}{p(\mathbf{u})} p(\mathbf{u})\, d\mathbf{u} && (5.122)\\
&\overset{(a)}{\leq} \int_{\mathbf{u}\in\mathcal{B}} \frac{R_1(\mathbf{u})}{p_1(\mathbf{u})} p(\mathbf{u})\, d\mathbf{u} && (5.123)\\
&\overset{(b)}{<} \int_{\mathbf{u}\in\mathcal{B}} \eta'_{\mathrm{lo}} p(\mathbf{u})\, d\mathbf{u} && (5.124)\\
&\overset{(c)}{=} \int_{\mathbf{u}\in\mathcal{C}'} \eta'_{\mathrm{lo}} p_1(\mathbf{u})\, d\mathbf{u} && (5.125)\\
&\overset{(d)}{\leq} \int_{\mathbf{u}\in\mathcal{C}'} \frac{R_1(\mathbf{u})}{p_1(\mathbf{u})} p_1(\mathbf{u})\, d\mathbf{u} && (5.126)\\
&= \int_{\mathbf{u}\in\mathcal{C}'} R_1(\mathbf{u})\, d\mathbf{u} && (5.127)\\
&= \bar{R}'(\mathcal{C}'), && (5.128)
\end{aligned}$$

where

(a) is due to the claim 3) of Theorem 13;

(b) is from the fact that due to $\mathcal{I}'$ is an MPEA with the efficiency lower bound η'_{lo}, $\mathcal{U}'$

includes all $\mathbf{u}$ such that $R_1(\mathbf{u})/p_1(\mathbf{u}) \geq \eta'_{\text{lo}}$ and $\mathcal{B} \cap \mathcal{U}'$ is the empty set;

(c) results from $\rho^*(\mathcal{B}) = \rho'(\mathcal{C})$;

(d) is again due to that $\mathcal{I}'$ is a MPEA.

Consequently, we demonstrate that

$$\bar{R}^*(\mathcal{I}^*) < \bar{R}^*(\mathcal{A}) + \bar{R}'(\mathcal{C}'), \tag{5.129}$$

where the throughput on the right hand side can be achieved by using the average power $\overline{p} = \rho^*(\mathcal{A}) + \rho'(\mathcal{C}')$. Thus, $\mathcal{I}^*$ is not an optimum policy and this contradicts the assumption. Hence, the proposition holds.

□

Proof: Proposition 8

For two different measurements, $\mathbf{u}$ and $\mathbf{u}'$, if $f(\mathbf{u}|s) \equiv f(\mathbf{u}'|s)$, the set of policies of interest $\{(p_i(\mathbf{u}), q_i(\mathbf{u}))\}$ is the same as $\{(p_i(\mathbf{u}'), q_i(\mathbf{u}'))\}$ since these policies must satisfy both (5.19) and (5.20). Consequently, the set of incremental efficiencies $\{\eta_i(\mathbf{u})\}$ is the same as $\{\eta_i(\mathbf{u}')\}$. From (3.33), we know that the power/rate allocations for $\mathbf{u}$ and $\mathbf{u}'$ are the same.

□

5.B When MPEA is not Optimum

We only prove the optimality of MPEA when $\eta_i(\mathbf{u})$ is strictly decreasing in i. However, in this subsection, we will demonstrate by some geometric constructions that since we do not limit the distribution of $f(s, \mathbf{u})$, $\eta_i(\mathbf{u})$ may not be always decreasing in i. An example is shown in Fig. 5.5 and we will explain this example in the following.

Here, we still assume that $\mathbf{u}$ and λ are given. Let us take a close look at $R'_i(\mathbf{u}, p)$ defined in (5.27) and consider

$$D_i(x) = \frac{\partial R'_i(\mathbf{u}, p)}{\partial p} = \frac{1 - F_{S|\mathbf{U}}(q_i(\mathbf{u})|\mathbf{u})}{p + \frac{N_0}{q_i(\mathbf{u})}}. \tag{5.130}$$

Clearly, with (5.23),

$$D_i(0) \quad = \quad \frac{q_i(\mathbf{u})}{N_0}\left[1 - F_{S|\mathbf{U}}(q_i(\mathbf{u})|\mathbf{u})\right] = \frac{\lambda}{N_0} e^{r_i(\mathbf{u})}, \tag{5.131}$$

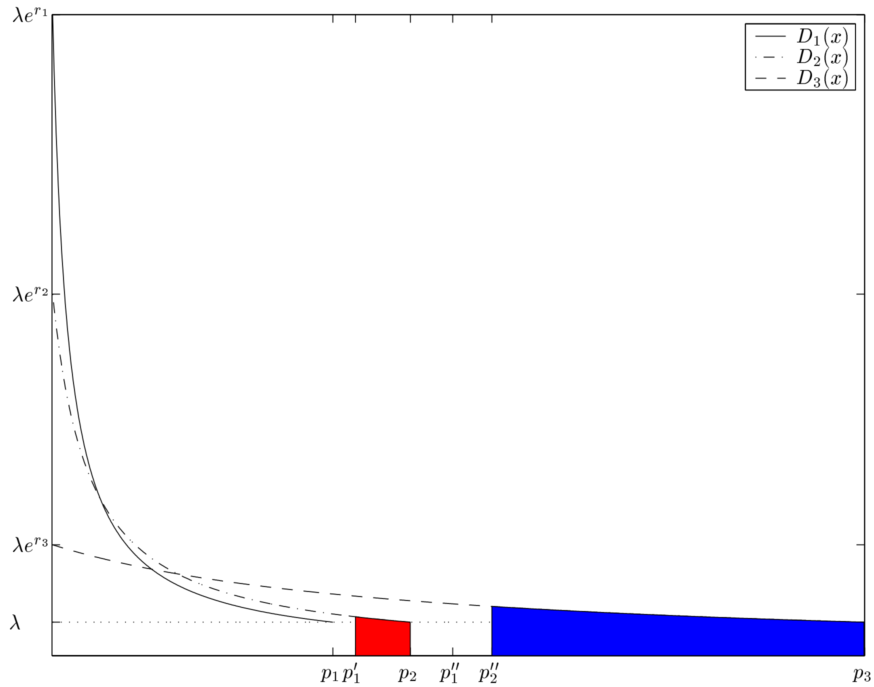

Figure 5.5: An example has $\eta_3(\mathbf{u}) > \eta_2(\mathbf{u})$ given $\mathbf{u}$ and λ. For simplicity, $(\mathbf{u})$ is omitted for functions depending on $\mathbf{u}$. The area of the shaded region between $p_1'(\mathbf{u})$ and $p_2(\mathbf{u})$ is $R_2(\mathbf{u}) - R_1(\mathbf{u})$. Similarly, the area of the other shaded region is $R_3(\mathbf{u}) - R_2(\mathbf{u})$.

Therefore, $D_i(0) > D_{i+1}(0)$ for any $i < N(\mathbf{u})$ due to the claim 3) in Theorem 13. Similarly,

$$D_i(p_i(\mathbf{u})) \quad = \quad \frac{1 - F_{S|\mathbf{U}}(q_i(\mathbf{u})|\mathbf{u})}{p_i(\mathbf{u}) + \frac{N_0}{q_i(\mathbf{u})}} = \lambda. \tag{5.132}$$

Hence, $D_i(p_i(\mathbf{u})) = D_{i+1}(p_{i+1}(\mathbf{u}))$ for any $i < N(\mathbf{u})$. In addition,

$$R_i(\mathbf{u}) \quad = \quad R_i'(\mathbf{u}, p_i(\mathbf{u})) = \int_0^{p_i(\mathbf{u})} D_i(x)dx. \tag{5.133}$$

Consequently, $\{D_i(x)\}$ are a set of $N(\mathbf{u})$ convex curves starting from $\{D_i(0)\}$ and ending, within our interests, at $\{D_i(p_i(\mathbf{u}))\}$. The areas under these curves are the corresponding rates $\{R_i(\mathbf{u})\}$.

Since $D_i(p_i(\mathbf{u})) = D_{i+1}(p_{i+1}(\mathbf{u}))$, due to the monotonicity of $D_i(x)$, we have $D_i(p_i(\mathbf{u})) < D_{i+1}(p_i(\mathbf{u}))$. In addition, since $D_i(0) > D_{i+1}(0)$, there exist at least a crossing point $p^{\mathrm{x}}_{i+1}(\mathbf{u}) \in (0, p_i(\mathbf{u}))$ such that

$$D_i(p^{\mathrm{x}}_{i+1}(\mathbf{u})) = D_{i+1}(p^{\mathrm{x}}_{i+1}(\mathbf{u})). \tag{5.134}$$

Moreover, it is straightforward to obtain that $p^{\mathrm{x}}_{i+1}(\mathbf{u})$ is unique by checking $D_i(x)/D_{i+1}(x)$ and, therefore,

$$D_i(x) > D_{i+1}(x), \qquad x \in [0, p^{\mathrm{x}}_{i+1}(\mathbf{u})), \tag{5.135}$$

$$D_i(x) < D_{i+1}(x), \qquad x \in (p^{\mathrm{x}}_{i+1}(\mathbf{u}), \infty). \tag{5.136}$$

So, any pair of curves within the set $\{D_i(x)\}$ will cross over each other once and once only.

Let $p'_{i-1}(\mathbf{u})$ satisfy

$$R'_i(\mathbf{u}, p'_{i-1}(\mathbf{u})) = R_{i-1}(\mathbf{u}). \tag{5.137}$$

Note that due to (5.10) and the uniqueness of $p_{i-1}(\mathbf{u})$ and $q_{i-1}(\mathbf{u})$ pair, $p'_{i-1}(\mathbf{u}) > p_{i-1}(\mathbf{u})$. In addition, since $R'_i(\mathbf{u}, p)$ is strictly increasing in p, $p'_{i-1}(\mathbf{u}) < p_i(\mathbf{u})$ due to $R'_i(\mathbf{u}, p'_i(\mathbf{u})) = R_{i-1}(\mathbf{u}) < R_i(\mathbf{u}) = R'_i(\mathbf{u}, p_i(\mathbf{u}))$. Overall, we have

$$p_{i-1}(\mathbf{u}) < p'_{i-1}(\mathbf{u}) < p_i(\mathbf{u}). \tag{5.138}$$

Then, according to (5.35),

$$\eta_i(\mathbf{u}) = \frac{R_i(\mathbf{u}) - R_{i-1}(\mathbf{u})}{p_i(\mathbf{u}) - p_{i-1}(\mathbf{u})} = \frac{R'_i(\mathbf{u}, p_i(\mathbf{u})) - R'_i(\mathbf{u}, p'_{i-1}(\mathbf{u}))}{p_i(\mathbf{u}) - p_{i-1}(\mathbf{u})} = \frac{\int_{p'_{i-1}(\mathbf{u})}^{p_i(\mathbf{u})} D_i(x)dx}{p_i(\mathbf{u}) - p_{i-1}(\mathbf{u})}. \tag{5.139}$$

So, $\eta_i(\mathbf{u})$ is the ratio of the area under the curve $D_i(x)$ between $p'_{i-1}(\mathbf{u})$ and $p_i(\mathbf{u})$ over the difference between $p_{i-1}(\mathbf{u})$ and $p_i(\mathbf{u})$.

Similarly, let $p''_{i-1}(\mathbf{u})$ and $p''_i(\mathbf{u})$ satisfy

$$R'_{i+1}(\mathbf{u}, p''_{i-1}(\mathbf{u})) = R_{i-1}(\mathbf{u}). \tag{5.140}$$

$$R'_{i+1}(\mathbf{u}, p''_i(\mathbf{u})) = R_i(\mathbf{u}). \tag{5.141}$$

respectively. Then, we have

$$p_{i-1}(\mathbf{u}) < p''_{i-1}(\mathbf{u}) < p''_i(\mathbf{u}) < p_{i+1}(\mathbf{u}), \tag{5.142}$$

and

$$\eta_i(\mathbf{u}) = \frac{\int_{p''_{i-1}(\mathbf{u})}^{p''_i(\mathbf{u})} D_i(x)dx}{p_i(\mathbf{u}) - p_{i-1}(\mathbf{u})}, \tag{5.143}$$

$$\eta_{i+1}(\mathbf{u}) = \frac{\int_{p''_i(\mathbf{u})}^{p_{i+1}(\mathbf{u})} D_i(x)dx}{p_{i+1}(\mathbf{u}) - p_i(\mathbf{u})}, \tag{5.144}$$

Note that $p_i(\mathbf{u}) < p''_i(\mathbf{u})$ due to (5.10). Also, we can show that $p'_{i-1}(\mathbf{u}) < p''_{i-1}(\mathbf{u})$. Otherwise, we must have $p^{\mathrm{x}}_{i+1}(\mathbf{u}) < p'_{i-1}(\mathbf{u})$ and, thus, $p''_i(\mathbf{u}) < p_i(\mathbf{u})$ which contradicts (5.10).

With the knowledge of the curves $\{D_i(x)\}$, we will try to construct an example with $N(\mathbf{u}) = 3$ and $\eta_3(\mathbf{u}) > \eta_2(\mathbf{u})$. Specifically, we want to construct a scenario satisfying

1) $p_3(\mathbf{u}) \to \infty$ and, thus, $\eta_3(\mathbf{u}) > \lambda$;

2) $p'_1(\mathbf{u}) > p_1(\mathbf{u})$, $p_2(\mathbf{u}) \tilde{=} p'_1(\mathbf{u})$, and, therefore, $\eta_2(\mathbf{u}) \to 0 < \lambda$.

Hence, $\eta_3(\mathbf{u}) > \eta_2(\mathbf{u})$.

Given any $r_2(\mathbf{u})$ and $p_2(\mathbf{u})$, if

$$r_3(\mathbf{u}) < \log\left(\frac{r_2(\mathbf{u})}{\left[1 - e^{-r_2(\mathbf{u})}\right]}\right), \tag{5.145}$$

we have

$$\begin{aligned} R'_2(\mathbf{u}, p_2(\mathbf{u})) &= [1 - F_{S|\mathbf{U}}(q_3(\mathbf{u})|\mathbf{u})] \log\left(1 + \frac{p_2(\mathbf{u})q_3(\mathbf{u})}{N_0}\right) && (5.146)\\ &= \lambda \frac{p_3(\mathbf{u})}{1 - e^{-r_3(\mathbf{u})}} \log\left(1 + \frac{p_2(\mathbf{u})\left[e^{r_3(\mathbf{u})} - 1\right]}{p_3(\mathbf{u})}\right) && (5.147)\\ &= \lambda \frac{p_2(\mathbf{u})\left(e^{r_3(\mathbf{u})} - 1\right) p_3(\mathbf{u})}{(1 - e^{-r_3(\mathbf{u})}) p_2(\mathbf{u})\left[e^{r_3(\mathbf{u})} - 1\right]} \log\left(1 + \frac{p_2(\mathbf{u})\left[e^{r_3(\mathbf{u})} - 1\right]}{p_3(\mathbf{u})}\right) && (5.148)\\ &\overset{(a)}{<} \lambda p_2(\mathbf{u}) e^{r_3(\mathbf{u})} && (5.149)\\ &< \lambda p_2(\mathbf{u}) \frac{r_2(\mathbf{u})}{\left[1 - e^{-r_2(\mathbf{u})}\right]} && (5.150)\\ &= R_2(\mathbf{u}). && (5.151) \end{aligned}$$

where (a) results from that $x \log(1 + 1/x)$ is strictly increasing in x and upperbounded by 1 for any $x > 0$. Consequently, (5.10) is satisfied and $p_3(\mathbf{u})$ is allowed to approaching ∞. Thus, if $p_3(\mathbf{u}) \to \infty$ and $p''_2(\mathbf{u}) = (1 + \delta)p_2(\mathbf{u})$ for any finite $\delta > 0$,

$$\eta_3(\mathbf{u}) = \frac{\int_{p''_2(\mathbf{u})}^{p_3(\mathbf{u})} D_3(x)dx}{p_3(\mathbf{u}) - p_2(\mathbf{u})} \to \frac{\int_{p''_2(\mathbf{u})}^{p_3(\mathbf{u})} D_3(x)dx}{p_3(\mathbf{u}) - p''_2(\mathbf{u})} > \lambda, \tag{5.152}$$

where the last term is due to $D_i(x) \geq \lambda$ for all $x \in [0, p_i(\mathbf{u})]$.

On the other hand, given $p_1(\mathbf{u})$, let $p'_1(\mathbf{u}) = (1 + \Delta)p_1(\mathbf{u})$ for any $\Delta > 0$ and $p_2(\mathbf{u}) = (1 + \Delta')p'_1(\mathbf{u})$ for any $\Delta' > 0$. Here, with (5.23), $r_2(\mathbf{u})$ is determined by

$$R_1(\mathbf{u}) = R'_2(\mathbf{u}, p'_1(\mathbf{u})) = \lambda \frac{p_2(\mathbf{u})}{1 - e^{-r_2(\mathbf{u})}} \log\left(1 + \frac{p'_2(\mathbf{u})(e^{r_2(\mathbf{u})} - 1)}{p_2(\mathbf{u})N_0}\right). \tag{5.153}$$

In order to have a reasonable $r_2(\mathbf{u})$, $R_1(\mathbf{u})$ must satisfy

$$R_1(\mathbf{u}) \geq \min_{r>0} \lambda \frac{p_2(\mathbf{u})}{1-e^{-r}} \log\left(1 + \frac{p_2'(\mathbf{u})(e^r - 1)}{p_2(\mathbf{u})N_0}\right). \tag{5.154}$$

Indeed, any $R_1(\mathbf{u}) > 0$ is feasible given any $p_1(\mathbf{u}) > 0$ since we can always find $r_1(\mathbf{u})$ satisfying

$$R_1(\mathbf{u}) = \lambda \frac{p_1(\mathbf{u}) r_1(\mathbf{u})}{1 - e^{-r_1(\mathbf{u})}}, \tag{5.155}$$

due to the fact that the range of $r_1(\mathbf{u})/(1 - \exp(-r_1(\mathbf{u})))$ is $(0, \infty)$ for $r_1(\mathbf{u}) > 0$.

In addition, we observe that $R_2'(\mathbf{u}, p)$ is strictly and continuously increasing in $p_2(\mathbf{u})$. Therefore, $R_2(\mathbf{u})$ is strictly and continuously increasing in $p_2(\mathbf{u})$. Hence, if we fix $r_2(\mathbf{u})$ and reduce $p_2(\mathbf{u})$, $R_2(\mathbf{u})$ will decrease accordingly. Then, we can keep reducing $p_2(\mathbf{u})$ until

$$R_2(\mathbf{u}) - R_1(\mathbf{u}) = \int_{p_1'(\mathbf{u})}^{p_2(\mathbf{u})} D_2(x)dx < \Delta'' p_1(\mathbf{u}), \tag{5.156}$$

where $\Delta'' << \Delta$. Consequently, we have

$$\eta_2(\mathbf{u}) = \frac{R_2(\mathbf{u}) - R_1(\mathbf{u})}{p_2(\mathbf{u}) - p_1(\mathbf{u})} < \frac{\Delta'' p_1(\mathbf{u})}{\Delta p_1(\mathbf{u})} \to 0. \tag{5.157}$$

In summary, the construction procedure starts from specifying $p_1(\mathbf{u})$, Δ, and Δ'. Then, $R_1(\mathbf{u})$ needs to satisfy (5.154). Hence, $r_1(\mathbf{u})$ and $r_2(\mathbf{u})$ can be determined by (5.155) and (5.153), respectively. Further, after deciding Δ'' such that $\Delta'' p_1(\mathbf{u}) << \Delta$, $p_1'(\mathbf{u})$ and $p_2(\mathbf{u})$ are set based on (5.156). So, (5.157) can be obtained. In order to have (5.152), we only need to choose any $r_3(\mathbf{u})$ satisfying (5.145) and let $p_3(\mathbf{u}) \to \infty$.

In Fig. 5.5, a set of $\{D_i(x),\ i = 1, 2, 3\}$ satisfying all the requirements discussed within this subsection are shown. For this case, $\eta_2(\mathbf{u})$ is below λ and $\eta_3(\mathbf{u})$ is clearly larger than λ. A question is how to generate the distribution $f(s, \mathbf{u})$ leading to the scenario in Fig. 5.5. Indeed, given $\{p_i(\mathbf{u})\}$ and $\{r_i(\mathbf{u})\}$, $\{q_i(\mathbf{u})\}$ and $\left\{F_{S|\mathbf{U}}(q_i(\mathbf{u})|\mathbf{u})\right\}$ are specified according to (5.24) and (5.23). Then, since only few points are determined, $F_{S|\mathbf{U}}(s|\mathbf{u})$ and, hence, $f(s, \mathbf{u})$ can have a wide range of choices.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

When there is a constraint on the decoding delay, information outage can occur. Based on the concept of information outage, we have studied the problems of the throughput maximization of adaptive transmission system with practical considerations, e.g. channel uncertainty and a finite set of code rates and/or power levels.

Given a flat fading channel, it is shown that in order to achieve the throughput performance within one dB from the ergodic capacity, it is sufficient to have transceivers which are capable of adjusting to any of ten code rates and power levels if the number of code rates and power levels are the same. In addition, if continuous power control is available, the number of required code rates can be reduced substantially. For a Rayleigh flat fading channel, transceivers with two non-zero code rates can achieve a throughput within one dB of the ergodic capacity. Therefore, for adaptive transmission systems, it is desirable to have good power control ability to offset the complexity of channel codecs.

A natural generalization of this work is to investigate the throughput maximization with a finite number of code rates and/or power levels and imperfect channel information at the transmitter. We have studied a variation of the problem with an additional equality outage constraint, implementing a quality-of-service requirement. The optimum policies can be found by a simple one-parameter line search.

Moreover, regardless of the channel state distribution and the statistical estimation technique, it can be shown that the candidate solutions to maximize the throughput when the transmitter does not have the perfect channel information have a well ordered structure. For instance, we have derived that increasing the throughput requires more transmitted power and a reduction in the code rate. The structure can be used to obtain good approximations of the optimum policies with less complexity.

Throughout the thesis, a simple flat fading channel is assumed. This choice simplifies our derivations and enables further insight into the problems. In addition, the flat fading channel model can be considered an instance of a block fading channel model [8, 10, 43].

The results in the thesis have been presented in [37–41].

6.2 Future Work

In this thesis, we have not carefully studied the effect of reducing the complexity of the algorithms. Therefore, the results may be limited to off-line scenarios where the transmission policies are not determined in real time. However, due to the popularity of wireless local area networks and ad hoc networks, it is possible that there is a need to determine the transmission polices in real time. In this case, how to simplify the algorithms and how to estimate parameters may become important problems. Surely, we assume that models of channels are known.

An underlying assumption in this thesis is that there is always some message ready to be encoded and transmitted. In reality, this may not be necessarily the case. Based on models of the message arrival process, predictions of future state of the queue at the transmitter can be made. Therefore, extending the work on channel uncertainty to the scenario with both channel uncertainty and the future queue state uncertainty is a possible future research direction.

References

[1] *EIA/TIA-IS95: Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spectrum Cellular System.* 1989.

[2] F. Adachi, M. Sawahashi, and H. Suda. Wideband DS-CDMA for next-generation mobile communications systems. *IEEE Communications Magazine*, 36(9):56–69, Sep. 1998.

[3] S. Alamouti and S. Kallel. Adaptive trellis-coded multiple-phase-shift keying for rayleigh fading channels. *IEEE Transactions on Communications*, 42:2305–2314, June 1994.

[4] K. Balachandran, R. Ejzak, S. Nanda, S. Vitebskiy, and S. Seth. GPRS-136: high-rate packet data service for North American TDMA digital cellular systems. *IEEE Personal Communications*, 6(3):34–47, Jun. 1999.

[5] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushyana, and S. Viterbi. CDMA/HDR: a bandwidth efficient high speed wireless data service for nomadic users. *IEEE Communications Magazine*, 38(7):70–77, Jul. 2000.

[6] C. Berrou, A. Glavieux, and P. Thitimajshima. Near shannon limit error-correcting coding and decoding: Turbo codes. *IEEE Proc. of ICC'93*, pages 1064–1070, 1993.

[7] D. Bertsekas. *Nonlinear Programming: 2nd Edition.* Athena Scientific, 1999.

[8] 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.

[9] G. Caire and S. Shamai. On the capacity of some channels with channel state information. *IEEE Transactions on Information Theory*, 45(6):2007–2019, Sep. 1999.

[10] G. Caire, G. Taricco, and E. Biglieri. Optimum power control over fading channels. *IEEE Transactions on Information Theory*, 45(5):1468–1489, Jul. 1999.

[11] J. Campello. Optimal discrete bit loading for multicarrier modulation systems. *Proceedings of IEEE International Symposium on Information Theory 1998*, page 193, Aug. 1998.

[12] J. Campello and J. Cioffi. Optimal discrete loading. *T1E1.4*, 98(341), 1998.

[13] J. Cavers. An analysis of pilot symbol assisted modulation for Rayleigh fading channels. *IEEE Trans. Veh. Tech.*, 40(4):686–693, Nov. 1991.

[14] J. K. Cavers. Variable rate transmission for rayleigh fading channel. *IEEE Transactions on Communications*, COM-20:15–21, Feb. 1972.

[15] S. Chung and A. Goldsmith. Degrees of freedom in adaptive modulatoin: A unified view. *IEEE Transactions on Communications*, 49(9):1561–1571, Sep. 2001.

[16] T. Cover and J. Thomas. *Elements of Information Theory.* Johe Wiley & Sons, Inc., 1991.

[17] A. Duel-Hallen, S. Hu, and H. Hallen. Long-range prediction of fading signals. *IEEE Transactions on Information Theory*, 17(3):62–75, May 2000.

[18] R. Gallager. Low-desity parity-check codes. *IRE Transactions on Information Theory*, pages 21–28, Jan. 1962.

[19] D. Goeckel. Adaptive coding for time-varying channels using outdated fading estimates. *IEEE Transactions on Communications*, 47(6):844–855, Jun. 1999.

[20] A. Goldsmith and S. Chua. Adaptive coded modulation for fading channels. *IEEE Transactions on Communications*, 46(5):595–602, May 1998.

[21] A. Goldsmith and S. Chua. Variable-rate variable-power MQAM for fading channels. *IEEE Transactions on Communications*, 45(10):1218–1229, Oct. 1997.

[22] A. Goldsmith and P. Varaiya. Capacity of fading channels with channel side information. *IEEE Transactions on Information Theory*, 43(6):1986–1992, Nov. 1997.

[23] A. J. Goldsmith. The capacity of downlink fading channels with variable rate and power. *IEEE Transactions on Vehicular Technology*, 46(3):569–580, Aug. 1997.

[24] S. Hanly and D. Tse. Multi-access fading channels: Part II: Delay-limited capacities. *IEEE Transactions on Information Theory*, 44(7):2816–2831, Nov. 1998.

[25] J. Hayes. Adaptive feedback communications. *IEEE Trans. Commun. Tech.*, COM-16(1):29–34, Feb. 1968.

[26] S. Jafar, S. Vishwanath, and A. Goldsmith. Throughput maximization with multiple codes and partial outages. *IEEE Proc. of GLOBECOM 2001*, pages 1307–1311.

[27] Y. Kamio, S. Sampei, and H. Sasaoka. Performance of modulation level controlled adaptive modulation under limited transmission delay time for land mobile communications. *Proc. of IEEE Vehicular Technology Conf. (VTC95)*, July 1995.

[28] R. S. Kennedy. *Fading Dispersive Communication Channels.* Wiley-Interscience, 1969.

[29] D. Knisely, S. Kumar, S. Laha, and S. Nanda. Evolution of wireless data services: IS-95 to CDMA2000. *IEEE Communications Magazine*, 36(10):140–149, Oct. 1998.

[30] R. Knopp and P. Humblet. On coding for block fading channels. *IEEE Transactions on Information Theory*, 46(1):189–205, Jan. 2000.

[31] B. Korte and J. Vygen. *Combinatorial Optimization: Theory and Algorithms (2nd Ed.).* Springer-Verlag New York Inc., 2002.

[32] C. Kose and D. Goeckel. On power adaptation in adaptive signaling systems. *IEEE Trans. Commun.*, 48(11):1769–1773, Nov. 2000.

[33] K. N. Lau. Variable rate adaptive modulation for ds-cdma. *IEEE Transactions on Communications*, 47:577–589, Apr. 1999.

[34] W. C. Y. Lee. *Mobile Communications Engineering*. McGraw Hill, 1982.

[35] W. C. Y. Lee. *Mobile Communications Design Fundamentals, 2nd Edition*. Wiley-Interscience, 1993.

[36] E. Lehmann. *Testing Statistical Hypotheses (2nd Edition)*. Springer-Verlag, 1997.

[37] L. Lin, R. Yates, and P. Spasojević. Adaptive transmission with channel uncertainty. *IEEE ISIT'2002*.

[38] L. Lin, R. Yates, and P. Spasojević. Adaptive transmission with channel uncertainty and finite code rate. *IEEE GLOBECOM'2003*.

[39] L. Lin, R. Yates, and P. Spasojević. Adaptive transmission with discrete code rate. *IEEE ICC'2002*.

[40] L. Lin, R. Yates, and P. Spasojević. Adaptive transmission with discrete code rate and power level. *IEEE ICC'2001*.

[41] L. Lin, R. Yates, and P. Spasojevic. Discrete adaptive transmission for fading channels. *To appear in IEEE Transactions on Communications*.

[42] E. Malkamäki and H. Leib. Coded diversity on block-fading channels. *IEEE Transactions on Information Theory*, 45(2):771–781, Mar. 1999.

[43] R. McEliece and W. Stark. Channels with block interference. *IEEE Transactions on Information Theory*, IT-30(1):44–53, Jan. 1984.

[44] M. Medard. The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel. *IEEE Transactions on Information Theory*, 46(3):933–946, May 2000.

[45] L. Ozarow, S. Shamai, and A. Wyner. Information theoretic considerations for cellular mobile radio. *IEEE Transactions on Vehicular Technology*, 43(2):359–378, May 1994.

[46] J. G. Proakis. *Digital Communications, 3rd Edition*. McGraw-Hill, 1995.

[47] T. Rappaport. *Wireless Communications: Principles & Practice*. Prentice Hall, 1996.

[48] M. Salehi. Capacity and coding for memories with real-time noisy defect information at encoder and decoder. *Proc. Inst. Elec. Eng. - Pt. I*, 139(2):113–117, Apr. 1992.

[49] S. Sampei, S. Komarki, and N. Morinaga. Adaptive modulation/tdma scheme for large capacity personal multi-media communication systems. *IEICE Transactions on Communications*, E-77B:1096–1103, Sep. 1994.

[50] N. Seshadri. Joint data and channel estimation using blind trellis search techniques. *IEEE Trans. Commun.*, 42(2/3/4):1000–1011, Feb./Mar./Apr. 1994.

[51] S. Shamai. A broadcast strategy for the Gaussian slowly fading channel. *IEEE Proc. of ISIT 1997*, page 150.

[52] S. Shamai and I. Bettesh. Outages, expected rates and delays in multiple-users fading channels. *IEEE Proc. of CISS 2000*, I:WA4–7–WA4–15.

[53] C. E. Shannon. Channels with side information at the transmitter. *IBM J. Res. Devel.*, 2:289–293, 1958.

[54] G. Stuber. *Principles of Mobile Communication*. Kluwer Academic Publishers, 1998.

[55] E. Telatar. Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications*, 10(6):585–595, Nov.-Dec. 1999.

[56] H. L. Van Trees. *Detection Estimation and Modulation Theory, vol. 1-3*. Wiley and Sons, 1971.

[57] D. Tse and S. Hanly. Multi-access fading channels: Part I: Polymatroid structure, optimal resource allocation and throughput capacities. *IEEE Transactions on Information Theory*, 44(7):2796–2815, Nov. 1998.

[58] B. Vucetic. An adaptive coding scheme for time-varying channels. *IEEE Transactions on Communications*, 39:653–663, May 1991.

[59] W. T. Webb and R. Steele. Variable rate qam for mobile radio. *IEEE Transactions on Communications*, 43(7):2223–2230, July 1995.

[60] Wolfowitz. *Coding Theorems of Information Theory*. New York: Springer-Verlag, 2nd ed. edition, 1964.

[61] D. Yellin and B. Porat. Blind identification of FIR systems excited by discrete-alphabet inputs. *IEEE Trans. Signal Processing*, 41(3):1331–1339, Mar. 1993.

Vita

Lang Lin

1989-1994 Attended Peking University, Beijing, P. R. China. Majored in Radio Electronics.

1994 B.S., Peking University.

1995-1997 Graduate work in Electrical and Electronic Engineering, Hong Kong University of Science and Technology, Hong Kong.

1997 L. Lin and R. Cheng, "Improvements on SOVA-Based Decoding for Turbo Codes", IEEE ICC'1997.

1997 L. Lin, C-Y Tsui and R. Cheng, "Low Power Soft Output Viterbi Decoder for Turbo Code Decoding", IEEE ISCAS'1997.

1997 L. Lin and R. Cheng, "On the Tail Effect of SOVA-Based Decoding for Turbo Codes", IEEE GLOBECOM'1997.

1997-2003 Graduate Research Assistant in Electrical Engineering, Rutgers, The State University of New Jersey.

1997 M.Phil., Hong Kong University of Science and Technology.

1998 Summer intern in AT&T Labs-Research.

1999 L. Lin, L. Cimini, and J. Chuang, "Turbo codes for OFDM with antenna diversity", IEEE VTC'1999 spring.

1999 J. Chuang, L. Cimini, Y. Li, B. McNair, N. Sollenberger, H. Zhao, L. Lin, and M. Suzuki, "Highspeed wireless data access based on combining EDGE with wideband OFDM", IEEE Commun. Magazine, vol. 37, num. 11, 1999, pp. 92-98.

2000 L. Lin, Y. Li, and J. Chuang, "Performance of COFDM System with Robust Channel Estimation in Rapid Dispersive Fading Channels", IEEE VTC'2000 spring.

2000 L. Lin, J. Chuang, and Y. Li, "Near optimal joint channel estimation and detection for COFDM system", IEEE GLOBECOM'2000.

2000 J. Chuang, Y. Li, N. Sollenberger, and L. Lin, "OFDM based High-Speed Wireless Access for Internet Applications", IEEE PIMRC'2000.

2000 L. Lin, J. Chuang, and Y. Li, Method for Near Optimal Joint Channel Estimation and Data Detection for COFDM Systems, U.S. Patent Application No. 09/774,875. (pending)

2000 L. Lin, J. Chuang, and Y. Li, System for Near Optimal Joint Channel Estimation and Data Detection for COFDM Systems, U.S. Patent Application No. 09/774,802. (awarded)

2000 L. Lin, L. Cimini, and J. Chuang, "Comparison of Convolutional and Turbo Codes for OFDM with Antenna Diversity", IEEE Commun. Letter, vol. 4, num. 9, 2000, pp. 277-279.

2001 L. Lin, R. Yates, and P. Spasojević, "Discrete Adaptive Transmission for Fading Channels", IEEE ICC'2001.

2001 J. Luo, L. Lin, R. Yates, and P. Spasojević, "Service Outage based Power and Rate Allocation", CISS'2001.

2002 L. Lin, R. Yates, and P. Spasojević, "Adaptive Transmission with Discrete Code Rates", IEEE ICC'2002.

2002 L. Lin, R. Yates, and P. Sapsojević, "Adaptive Transmission with Channel State Uncertainty", IEEE ISIT'2002.

2003 J. Luo, L. Lin, R. Yates, and P. Spasojević, "Service Outage based Power and Rate Allocation", IEEE Trans. Info. Theory, vol. 49, no. 1, pp. 323-330, Jan. 2003.

2003 L. Lin, R. Yates, and P. Spasojević, "Adaptive Transmission with Discrete Code Rates and Power Levels", to appear on IEEE Trans. Commun.

2004 Ph.D. in Electrical Engineering.