PERFORMANCE ANALYSIS OF DIVERSITY WIRELESS SYSTEMS

CAO LE

NATIONAL UNIVERSITY OF SINGAPORE

2011

PERFORMANCE ANALYSIS OF DIVERSITY WIRELESS SYSTEMS

CAO LE

(*M. Sc., National University of Singapore*)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2011

Acknowledgement

During my PhD studies, I have worked with my supervisors and colleagues who have contributed in assorted ways to the research and this thesis. This thesis would not have been possible without their unconditionally kind support. I am more than glad to convey my gratitude to them all in my humble acknowledgment.

In the first place, my sincere gratitude and appreciation undoubtedly go to my supervisor, Professor Kam Pooi Yuen for his supervision, advice, and guidance. Above all and the most important, he provided me unflinching encouragement and support in various ways. It is he who gives me a compass and an interesting book along my research journey. His truly scientific intuition has made him as a constant oasis of ideas and passions in science, which exceptionally inspire and enrich my growth as a student, a researcher and a scientist-to-be. I am indebted to him more than he knows. Secondly, I would like to record my gratitude to Dr. Tao Meixia for her supervision, advice, and guidance in the very early stage of my research journey. Her involvement in the detailed work has triggered and nourished my intellectual maturity that I benefit from.

Collective and individual acknowledgments are also owed to my colleagues at ECE-I2R Wireless Communications Lab whose presence are somehow perpetually refreshing, helpful, and memorable. Many thanks go to Dr. Zhu Yonglan for her valuable suggestions, sharing various thoughts, and patient discussions. I would like to thank Mr. Siow Hong Lin, Eric for the technical support to our lab. Many thanks

go in particular to Dr. Li Yan, Dr. Cao Wei, Dr. Gao Feifei and Dr. Jiang Jianhua for giving me a lot of constant help and advice for my study life and living life since I began my studies in NUS. It is a pleasure to mention: Dr. Lu Yang, Dr. Zhang Xiaolu, Dr. Hou Shengwei, Mr. Chen Qian, Ms. Wu Mingwei, Dr. Shao Xuguang and Mr. Lin Xuzheng for creating such a great friendship at the lab and spending wonderful and memorable time at lunch. Thanks to Ms. Zhou Xiaodan for being such a good colleague and neighbor. I did not feel lonely any more on the one-hour way back home since we became neighbors. It is a pleasure to mention Ms. Tian Zhengmiao who is one of my fellow alumni of Xidian University, China. I am more than happy to become her colleague again at NUS. I also would like to thank Dr. Zhang Qi, Dr. Elisa Mo, Mr. Kang Xin, Mr. Yuan Haifeng, Dr. Mahtab Hossain, Dr. Nitthita Chirdchoo and Dr. Pham The Hanh for giving me such a pleasant time when working at the same lab.

Where would I be without my family? My parents deserve special mention for their inseparable and everlasting support and love. My father, in the first place, is the person who showed me the joy of intellectual pursuit ever since I was a child. My mother is the one who sincerely raised me with her tender care and endless love. Her understanding and support encourage me to work hard and to continue my studies abroad. Her firm and kind-hearted personality has affected me to be steadfast and never bend to difficulties. Last but not least, I am greatly indebted to my devoted husband. He is the backbone and origin of my happiness. His unconditional love, support, company and encouragement make me dedicated to what I want to do. I am so grateful for his presence in my life.

Finally, the support of Singapore MoE AcRF Tier 2 Grant T206B2101 in the form of research scholarship is gratefully acknowledged.

Contents

Summary

Many wireless communication systems make use of the diversity technique: a well-known concept to combat the effects of multipath fading. Diversity reception consists of receiving redundantly the same information-bearing signal over multiple fading channels, (then combining them at the receiver so as to increase the received signal-to-noise ratio (SNR).)

One way by which these multiple replicas can be obtained is using multiple antennas in multiple-input-multiple-output (MIMO) systems for achieving space diversity. The ergodic capacity is a key performance parameter of a MIMO fading channel. We obtain tight bounds on the ergodic capacity over an identical MIMO fading channel, which show explicitly the dependency of the ergodic capacity on the SNR and the number of transmit and receive antennas. The results enable us to determine the optimal number of transmit antennas to be used for a given SNR and a given total number of antennas. Recently, MIMO systems over a non-identical fading channel have attracted great attention because of their applications in cooperative communications and distributed antenna systems. We derive explicit and closed-form expressions of the ergodic mutual information (MI) and the information outage probability. Two simple and near-optimal power-allocation schemes are then proposed for maximizing the ergodic MI and minimizing the information outage, respectively.

Another approach to obtain multiple replicas of the same information-bearing signal is by using multiple time slots separated by at least the coherence time of

the channel in automatic-repeat-request (ARQ) systems, leading to the exploitation of time diversity. With imperfect channel state information at the receiver (CSIR), the performance parameters of ARQ systems are evaluated as a function of the accuracy of the channel estimation. A link between data-link-layer performances and physical-layer parameters is therefore established. An attempt is made to study the inter-relationships among the various relevant system performance parameters and the dependency of these relationships on the CSIR accuracy. For enhancing the throughput, adaptive transmission strategies have been adopted to match the transmission rate to time-varying channel conditions for achieving higher spectral efficiency. Therefore, with regard to maximizing the throughput, in addition to providing a more reliable transmission, ARQ schemes with adaptive transmissions are extensively adopted. Considering a practical case with the imperfect channel state information at the transmitter (CSIT) and the imperfect CSIR, an optimal continuous-rate adaptation scheme is studied so as to achieve a maximum goodput.

List of Figures

Abbreviations

ACK	Positive Acknowledgement
ARQ	Automatic-Repeat-Request
APER	Accepted Packet Error Rate
AWGN	Additive White Gaussian Noise
BEP	Bit Error Probability
BPSK	Binary Phase Shift Keying
BSC	Binary Symmetric Channel
CCDF	Complementary Cumulative Distribution Function
CDF	Cumulative Distribution Function
CRC	Cyclic Redundancy Check
CSI	Channel State Information
CSIT	Channel State Information at the Transmitter
CSIR	Channel State Information at the Receiver
DBPSK	Differential Binary Phase Shift Keying
DF	Decode-and-Forward
DPSK	Differential Phase Shift Keying
FEC	Forward-Error-Control
HARQ	Hybrid ARQ
i.i.d	Independent and Identically Distributed

i.n.d	Independent and Non-identically Distributed
IR	Incremental Redundancy
MAP	Maximum A Posteriori
MAC	Medium Access Control
MI	Mutual Information
MISO	Multiple-Input-Single-Output
MIMO	Multiple-Input-Multiple-Output
ML	Maximum-Likelihood
MMSE	Minimum Mean Square Error
M-QAM	M-Quadrature Amplitude Modulation
MRC	Maximal Ratio Combiner
MSE	Mean Square Error
NAK	Negative Acknowledgment
OSTBC	Orthogonal Space-Time Block Codes
PDF	Probability Distribution Function
PEP	Packet Error Probability
PSAM	Pilot Symbol Assisted Modulation
QPSK	Quadrature Phase Shift Keying
SIMO	Single-Input-Multiple-Output
SISO	Single-Input-Single-Output
SNR	Signal-to-Noise Ratio
STBC	Space-Time Block Codes
TISO	Two-Input-Single-Output

Notations

In this dissertation, scalar variables are written as plain lower-case letters, vectors as bold-face lower-case letters, and matrices as bold-face upper-case letters. Some further used notations and commonly used acronyms are listed in the following:

a	plain lower-case to denote scalars
$\boldsymbol{a}$	boldface lower-case to denote column vectors
$\boldsymbol{A}$	boldface upper-case to denote matrices
$\mathbf{I}_N$	the $N \times N$ identity matrix
$(\cdot)^*$	the conjugate operation
$(\cdot)^T$	the transpose operation
$(\cdot)^H$	the conjugate transpose operation
$\det(\cdot)$	the determinant of a matrix
$\mathrm{tr}(\cdot)$	the trace of a matrix
$\mathbb{E}(\cdot)$	the statistical expectation operation
$\Re(\cdot)$	the real part of the argument
$\otimes$	the Kronecker product
$\|\cdot\|_F^2$	the Frobenius norm square
$\mathrm{erfc}(\cdot)$	the complementary error function
$\Gamma(\cdot)$	the Gamma function
$\Gamma(\cdot,\cdot)$	the upper incomplete Gamma function

$I_m(\cdot)$	the m-th order modified Bessel function of the first kind
$K_m(\cdot)$	the m-th order modified Bessel function of the second kind
$Q_1(\cdot,\cdot)$	the first order Marcum Q-function
$Q_m(\cdot,\cdot)$	the generalized Marcum Q-function

Chapter 1

Introduction

1.1 Introduction to Diversity Wireless Systems

Many of the current and emerging wireless communication systems make use in one form or another of *diversity*: a classic and well-known concept [1–4] that has been used since the early 1950's to combat the effects of multipath fading. Diversity combining consists of receiving redundantly the same information-bearing signal over two or more fading channels, then combining these multiple replicas at the receiver (in order to increase the overall received signal-to-noise-ratio (SNR)). It offers one of the greatest potentials for radio link performance improvement to many of the current and future wireless technologies. The intuition behind this concept is to exploit the low probability of concurrence of deep fades in all the diversity channels to lower the probability of error and of outage. Depending on the domain where replicas of the same information-bearing signal are obtained, diversity techniques can be categorized into three types: time diversity, frequency diversity and space diversity. In this thesis, we will focus on space diversity and time diversity. The space diversity can be achieved by using multiple antennas in MIMO systems while the time diversity can be achieved by using multiple time slots separated by at least the coherence time of the channel in

ARQ systems.

1.1.1 MIMO Systems

A conventional approach to achieving *space diversity* is to employ multiple transmit and/or multiple receive antennas. If the antennas are placed sufficiently far apart, the channel gains between different antenna pairs are independent. For a mobile terminal, a half to one carrier wavelength separation among antennas is sufficient to guarantee that the channel gains are independent. Through transmitting the replicas of the signal through different antennas, and/or combining the different replicas together at the receiver, space diversity can be achieved. Traditionally, space diversity is achieved by employing multiple receive antennas at the receiver in single-input-multiple-output (SIMO) systems, where combining, selection or switching of the received signals is performed. This is so-called receive diversity. By deploying multiple transmit antennas at the transmitter in multiple-input-single-output (MISO) systems, transmit diversity techniques shift the complexity associated with realizing diversity to the transmitter. A multiple-input-multiple-output (MIMO) communication system with multiple transmit and receive antennas provides even greater potential. In addition to the aforementioned diversity benefits, the spectral efficiency is possibly enhanced by spatial multiplexing. The maximum spatial multiplexing order is determined by the minimum of the number of transmit and receive antennas. Therefore, the advantage of an MIMO system can be utilized not only to increase the diversity of the system leading to an improved error performance [5, 6] but also to increase the number of transmitted symbols leading to a high spectral efficiency [7–9].

1.1.2 ARQ/HARQ Systems

As another type of diversity techniques, *time diversity* can be obtained in automatic-repeat-request (ARQ) systems by combining packets transmitted in different time slots. The idea is that the packets that cause retransmission in the current slot can be stored and later combined with additional copies of the same packet transmitted in the successive time slots. The separation between successive time slots equals or exceeds the coherence time of the channel. Therefore, combining the multiple copies of a packet creates a single packet whose constituent symbols are more reliable than those of any of the individual copies. Classified by the mechanisms of transmissions and/or retransmissions, there are three basic types of ARQ schemes: the *selective-repeat* ARQ, the *stop-and-wait* ARQ, and the *go-back N* ARQ [10, 11]. All three basic ARQ schemes achieve the same *reliability*; however, they have different *throughput*s. Taking into account of both the reliability and the throughput, the *goodput* [12, 13], which shows the proportion of the throughput consisting of correct packets, is more meaningful. For further improving the throughput and the system reliability, it is preferred to combine ARQ with a forward-error-control (FEC) system to reduce the frequency of retransmissions. The FEC scheme can be incorporated into any of the three basic ARQ schemes. Such a combination of the ARQ and the FEC is referred to as a hybrid ARQ (HARQ). In the Type-I HARQ scheme, the same coded packet is retransmitted and these multiple packets can be combined in two distinct ways. In the *code combining* scheme, these repeated packets are concatenated to form a single packet at a lower code rate, which is often referred to as *Chase combining* [14–16]. In the *diversity combining* scheme, these repeated packets are combined into a single packet at the same rate with more reliable constituent symbols by using symbol voting schemes [17] or by using symbol averaging schemes [16]. In the Type-II HARQ scheme, instead of re-sending the same packet, the transmitter tries

to construct and sends additional parity bits when a negative acknowledgment (NAK) is received. This is also known as the incremental redundancy (IR) scheme [18].

1.2 Motivations of the Work

1.2.1 MIMO Systems

MIMO systems offer significant increases in data throughput and link reliability without additional bandwidth or transmit power in wireless communications. Substantial efforts have been made on characterizing the ultimate information theoretic limits of the MIMO systems and designing optimal transmission strategies.

Information Theoretic Performance Limits

Before we proceed to discuss the channel capacity in various channel state information (CSI) scenarios, we shall clarify two important concepts of the capacity for fading channels. For ergodic channels, when the code is sufficiently long so that it spans an ergodic fading process, the resulting channel capacity is a nonzero ergodic capacity. The *ergodic capacity* refers to the capacity in Shannon's sense; that is, for any transmission rate smaller than the ergodic capacity, there exists at least one encoder and one decoder that achieves arbitrarily small error probability. However, for non-ergodic channels, there is no significant channel variation across the code. In this circumstance, the channel capacity is viewed as a random variable as it depends on the instantaneous channel state realization. Hence, the ergodic capacity in Shannon's sense of these channels is zero, meaning that no matter how small the transmission rate is, there is no guarantee that the transmission will be error-free. Therefore, instead of looking at the ergodic capacity in Shannon's sense, it is more meaningful to look at the capacity from an outage perspective i.e., the outage capacity at a given outage probability. The *outage*

probability is the cumulative distribution function (CDF) of the mutual information (MI), and measures the tradeoff between the transmission rate and the reliability. There has been substantial work on characterizing the ergodic capacity of MIMO systems under a variety of fading conditions. The ergodic capacity of the MIMO channel has been developed for several different cases which depend on the availability of the channel state information at the transmitter (CSIT) and/or the channel state information at the receiver (CSIR).

Optimal Transmission Strategies

With perfect CSIR, coherent detection can be done, resulting in an enhanced channel capacity compared to the case without any CSI knowledge. When the transmitter has perfect CSIT, power allocation (in both the spatial and temporal dimensions) can be performed at the transmitter which results in an additional enhancement of channel capacity [19]. However, it is too optimistic in practice to assume the availability of the instantaneous CSIT since it impose a heavy signalling burden on the feedback channels. Hence, using partial CSI feedback greatly reduces the signalling burden compared to using instantaneous CSI feedback. It has been shown that even partial CSIT can increase the ergodic capacity of a MIMO system [20]. The fading channel, given the feedback, can be modeled as a complex Gaussian random vector [20, 21]. Two extreme cases are considered: *mean feedback* and *covariance feedback*. For the mean feedback, the partial CSIT resides in the mean of the distribution, with the covariance modeled as white. For the covariance feedback, the fading channel is assumed to be varying too rapidly to track its mean, so that the mean is set to zero, and the partial CSIT regarding the relative geometry of the propagation paths is captured by a covariance matrix. Therefore, depending on the different levels of the feedback information available at the transmitter, it is important to investigate different

transmission strategies that achieve the ergodic capacity in the MIMO systems.

1.2.2 ARQ/HARQ Systems

ARQ/HARQ is an alternative way to mitigate channel fading since the noise burst may have run its course before the retransmitted packet begins to make its way across the channel. Substantial efforts have been made on analyzing the performance of ARQ/HARQ systems and in designing adaptive transmission strategies.

Performance of ARQ/HARQ Schemes

There are two basic parameters by which we can evaluate the performance of an ARQ/HARQ system: reliability and throughput. The reliability is often expressed in terms of the accepted packet error rate (APER) [10]. The APER is the percentage of packets accepted by the receiver that contain one or more bit errors. Throughput is defined as the ratio of the average number of information bits received per unit of time to the total number of bits that could be transmitted per unit of time [10]. The throughput is meaningful only when considered in conjunction with the reliability. Therefore, the goodput, defined as the ratio of the expected number of information bits correctly received per unit of time to the total number of bits that can be transmitted per unit of time, shows the proportion of the throughput consisting of correct packets [12, 13]. The performance parameters in the data-link layer due to ARQ/HARQ, such as the APER, throughput, goodput and drop rate depend not only on the medium access control (MAC) protocol, but also on the physical-layer parameters. Much work has been done on the performance of ARQ/HARQ schemes over fading channels [22–25]. Due to the large number and the complexity of the parameters as well as the protocols across the two layers, in previous works, by and large, perfect CSI in the physical-layer is assumed and the characterization of channel errors is mostly

modeled by using a Markov model with a finite number of states [13, 26–28] []. Nevertheless, the CSI may be outdated or imperfect due to the feedback delays and the channel estimation errors both at the transmitter and the receiver. Since the CSI can be used to perform link adaption, transmit diversity selection [29] and relay selection [30], evaluating the effects of imperfect CSI on the system performance is important to provide insights on system operation and guidelines for designing effective system management schemes. Therefore, we focus on providing a systematic approach whereby the link-layer performance parameters can be evaluated in terms of the parameters at the lowest physical-layer. More importantly, we study the impact of imperfect CSIR on ARQ/HARQ schemes and demonstrate that the accuracy of the CSIR plays a crucial role in determining the performance in the data-link layer.

Adaptive Transmission Strategies

Desire to avoid both low spectral efficiency and unreliable transmissions associated with the use of a fixed transmission strategy over fading channel has motivated the use of adaptive transmission strategies. Adaptive transmission strategies have been studied extensively to match the modulation and coding to time-varying channel conditions for enhancing the throughput [19, 31–35]. However, in order to achieve high reliability, one has to reduce the transmission rate using either small constellations, or powerful but low rate codes. Since ARQ is an alternative way to mitigate channel fading, high reliable adaptive transmission strategies combined with ARQ techniques has been known to offer a higher spectral efficiency, in addition to providing a more reliable transmission [36]. The transmission rates are adapted with respect to the channel conditions. Therefore, the CSI plays a crucial role in determining the performance of the systems and it is more important to study the performance of ARQ schemes with adaptive transmission strategies.

1.3 Research Objectives and Contributions

For the information theoretic performance limits mentioned in Section 1.2.1, with no CSIT and perfect CSIR, the ergodic capacity of MIMO fading channels depends on the joint distribution of the eigenvalues of a Wishart matrix, and is quite complex in general. Such complicated results do not allow one in general to study explicitly the dependency of the ergodic capacity on system parameters. In particular, we are interested in the optimal number of transmit antennas to be used, as a function of the SNR. In Chapter 3, a new approach based on the trace and determinant of a Wishart matrix is proposed to derive upper and lower bounds on the ergodic capacity instead of using the joint distribution of the eigenvalues of a Wishart matrix. Our approach to the ergodic capacity analysis greatly simplifies the computational procedure, and provides easy and accurate ways to deal with ergodic capacity related calculations for MIMO Rayleigh fading channels. The bounds obtained here on the ergodic capacity are expressed in simple closed forms, and show explicitly the effects of the system parameters on the ergodic capacity. The bounds are valid for an arbitrary number of antennas, and they enable us to design an optimal antenna deployment strategy, i.e., to determine the optimal number of transmit antennas for a given SNR and a given total number of antennas in the system.

For the optimal transmission strategy with partial CSIT and perfect CSIR addressed in Section 1.2.1, with covariance feedback of a MISO channel, the optimum solution consists of independent, complex, circular, Gaussian transmit signals along the N eigenvectors of the transmit covariance matrix. However, the powers along the eigenvectors have to be determined through numerical maximization techniques. It is well known that equal power allocation is optimal for a MIMO channel with an identity matrix as the covariance feedback [9], which is actually the case of no

CSIT addressed in [20]. However, it is no longer optimal for the case of covariance feedback over independent, and non-identically distributed (i.n.d) fading channels. To the best of our knowledge, no closed-form power control is available in the literature for the ergodic capacity maximization. The power has to be determined through numerical maximization techniques [21, 37, 38]. In Chapter 4, we are interested in the performance limits and associated power allocation problems in a MIMO system with the covariance feedback of the CSIT and the perfect CSIR [39, 40]. Our first contribution is therefore to obtain the closed-form optimal power allocation for maximizing the ergodic capacity over i.n.d fading channels. For outage probability minimization, the information outage of a MISO system over i.n.d Rayleigh fading is studied in [41]. Therein, a heuristic power control scheme named *equal power allocation with channel selection* is proposed. Generalizing to a MIMO system, we obtain the closed-form expressions of the outage probability over i.n.d Rayleigh fading channels and derive the closed-form power allocation scheme for exploiting the non-identical channel statistics to minimize the outage probability [40].

As addressed in Section 1.2.2, much work has been done on the performance of ARQ/HARQ schemes over fading channels. Due to the large number and the complexity of the parameters as well as the protocols across the physical layer and the data-link layer, in previous works, by and large, perfect CSIR in the physical layer is assumed. Nevertheless, the CSIR may be imperfect due to the channel estimation errors. Therefore, in Chapter 5, we study the impact of imperfect CSIR on ARQ/HARQ schemes and demonstrate that the accuracy of the CSIR plays a crucial role in determining the performance in the data-link layer. Our aim is on establishing a link between network-layer and physical layer performance parameters. We analyze the performance of three basic ARQ schemes as well as three Type-I HARQ schemes with diversity combining over a block fading channel with imperfect

CSIR. The imperfect CSIR is acquired via minimum mean square error (MMSE) channel estimation with the aid of pilot symbols. Three performance parameters: APER, goodput and drop rate are investigated, respectively. We obtain closed-form upper and lower bounds on the APER, the goodput as well as the drop rate. Using numerical results, we compare the impact of the accuracy of the imperfect CSIR on basic ARQs and on Type-I HARQs. In practice, the number of transmissions is limited in Type-I HARQ, which can result in a drop rate of data packets without guaranteeing their error-free delivery. Our work provides a systematic approach whereby the link-layer performance parameters can be evaluated in terms of the parameters at the lowest physical-layer. While closed-form expressions of the bounds on the APER, goodput and drop rate are nonlinear functions of the MMSE, they enable the system designer to study numerically the dependence of the link-layer performance parameters on the MMSE and the effective SNR, for any given (n, m) linear block code and any modulation format for transmitting the code bits. A key physical-layer parameter that plays an implicit but crucial role in the analysis is the channel bandwidth. The bandwidth, together with the code rate, determines the allowable number of pilot symbols per packet, which in turn determines the required SNR for achieving the desired channel estimation MMSE that leads to the target link-layer performance level.

For the adaptive transmission mentioned in Section 1.2.2, much previous work on this topic assume perfect CSIT is available. However, it is too optimistic in practice to assume the availability of perfect CSIT and perfect CSIR. In adaptive transmission, the CSIT used to perform rate adaptation maybe outdated and/or imperfect due to the transmission delay and the processing imperfections both at the transmitter and at the receiver. In Chapter 6, we focus on the imperfect CSIT and imperfect CSIR due to both the channel estimation errors at the transmitter and the prediction errors at the receiver. While a strictly causal channel predictor is employed to predict the

channel state for the transmitter to adapt its transmission rates, a noncausal channel estimator estimates the channel for the receiver to perform coherent demodulation. The goodput is used as the performance measure. It is defined as an amount of data delivered to the receiver correctly per time unit, and it takes into consideration both the throughput and the reliability. Our objective is to maximize the goodput by using adaptive transmission strategies. An optimal continuous-rate adaptation scheme is proposed which takes account of the effect of the imperfect CSIT and imperfect CSIR. The pilot symbol assisted modulation (PASM) scheme is applied at the transmitter to facilitate the channel prediction and channel estimation at the receiver. Based on the predicted channel gain and a utilization factor, the transmitter allocates the optimal transmission rates which maximize the goodput. The utilization factor, which takes into account both the estimation and prediction errors, is to be optimized in order to achieve the maximum goodput.

1.4 Organization of the Thesis

The rest of this dissertation is organized as follows.

In Chapter 2, for both MIMO systems and ARQ/HARQ systems, a comprehensive literature review is provided on performance analysis and transmission strategies with the different levels of CSI availability.

In Chapter 3, bounds on the ergodic capacity of the MIMO Rayleigh fading channel are derived by exploiting the properties and distributions of the trace and the determinant of a Wishart matrix. Thus, three simple and tight bounds on the ergodic capacity are obtained, which show explicitly the dependence of the ergodic capacity on the SNR and the number of transmit and receive antennas. Based on the obtained tight bounds, an optimal number of transmit antennas used for a given SNR and a given

total number of antennas is studied for maximizing the ergodic capacity.

In Chapter 4, the performance limits and associated power-allocation problems in a multiple-antenna diversity system with partial CSIT is investigated. Bounds on ergodic capacity and information outage are obtained in closed-forms. By studying both ergodic capacity and information outage, two simple and near-optimal power-allocation schemes are obtained in closed-form as a function of the partial CSIT for maximizing the ergodic capacity and minimizing the outage probability, respectively.

In Chapter 5, with imperfect CSIR, the performance of basic ARQ and HARQ systems are evaluated as a function of the accuracy of channel estimation. The performance parameters we study in particular are the goodput, APER and the drop rate, as a function of the channel estimation mean square error (MSE) and the factors which affect the MSE. Upper and lower bounds on the APER, the goodput as well as the drop rate are derived. The precise dependence of the APER and the goodput on the channel estimation accuracy is quantified.

In Chapter 6, with imperfect CSIT and imperfect CSIR, a rate adaptation scheme is developed, which takes account of both channel estimation and channel prediction errors. The adaptive transmission strategy adapts the continuous-rate of the transmission relative to the predicted channel gain and a utilization factor. In turn, this utilization factor is optimized as a function of the MSEs of both channel estimation and channel prediction so as to maximize the goodput of the system.

Finally, Chapter 7 summarizes our work, and points out a number of future research directions.

Chapter 2

Literature Review

2.1 MIMO Systems

MIMO systems offer significant increases in data throughput and link reliability without additional bandwidth or transmit power in wireless communications. Substantial efforts have been made on characterizing the ultimate information theoretic limits of the MIMO systems and designing optimal transmission strategies.

2.1.1 Information Theoretic Performance Limits

There has been substantial work on characterizing the ergodic capacity of MIMO systems under a variety of fading conditions. The ergodic capacity of the MIMO channel has been developed for several different cases which depend on the availability of the CSIT and/or the CSIR.

Perfect CSIT and Perfect CSIR

The capacity of a fading channel with perfect CSIT and perfect CSIR is analyzed in [9, 19, 42]. For achieving the capacity on frequency-selective fading channels, the transmit signal is circularly symmetric, zero-mean, complex Gaussian distributed and

the optimal power allocation is a "water-filling" on the eigenvalues of the channel matrix [9,42]. The capacity of a time-varying channel is achieved when the transmitter adapts its power, data rate, and coding scheme to the channel variation, and the optimal power allocation is a "water-filling" in time [19].

No CSIT and Perfect CSIR

For the important case when CSIT is not available but perfect CSIR is known, a lot of work on the ergodic capacity has been done in [7, 9, 19, 43–46]. For independent and identically distributed (i.i.d) Rayleigh fading, the ergodic capacity of a MIMO system is obtained exactly in [9], where the ergodic capacity was expressed in terms of Laguerre polynomials. A lower bound on the capacity of a MIMO system with $N \leq M$ is obtained in an expression of a random variable whose distribution is indicated [7], where N is the number of the transmit antennas and M is the number of the receive antennas. Since the exact expression of the ergodic capacity is either complex or is not given in a closed form, a lower bound is obtained in [47] by applying Minkowski's inequality and Jensen's inequality. By using the expectation of the determinant of a complex central Wishart matrix, a lower bound is obtained that is only tight at low SNR [46]. The outage probability is given in a closed integral form [48], which can only be evaluated numerically.

For i.i.d Rician fading channels, the ergodic capacity has been presented in [44, 46, 49]. By making use of the joint distribution of the eigenvalues of a noncentral Wishart matrix, the exact ergodic capacity is obtained in multiple integral forms that can only be evaluated by numerical integration [49]. By determining the expected value of the determinant and the log-determinant of a complex noncentral Wishart matrix, bounds on ergodic capacity are obtained but in a complicated form consisting of Digamma functions [44]. Following the approach of [47], upper and lower bounds

are derived in [46], that are tight only at low and high SNR, respectively. Exploiting the properties and statistical distributions of the determinant and trace of a noncentral, complex, Wishart matrix, lower and upper bounds on the outage probability are obtained in closed-forms and can be reduced to the case of Rayleigh fading [50].

Partial CSIT and Perfect CSIR

For mean feedback, the partial CSIT resides in the mean of the distribution, with the covariance modeled as white. For covariance feedback, the fading channel is assumed to be varying too rapidly to track its mean, so that the mean is set to zero, and the partial CSIT regarding the relative geometry of the propagation paths is captured by a covariance matrix. The covariance matrix is usually assumed to be nontrivial, i.e., an nonidentity matrix. The ergodic capacity of a MIMO system with partial CSIT has been analyzed in [21, 38, 51, 52].

For the case of mean feedback, the ergodic capacity is obtained in a general form and only solved by a number of numerical algorithms [51]. As an alternative work to [51], analytical expressions of the ergodic capacity are obtained for two cases of partial CSIT feedback [21].

For the case of covariance feedback, the ergodic capacity is obtained in a general form and only solved by a number of numerical algorithms [51]. As an alternative work to [51], analytical expressions of the ergodic capacity are obtained for the two cases of partial CSIT feedback [21]. The results in [21, 51] are only valid for a MISO system. As an extension to [21,51], the ergodic capacity of a two-input-multiple-output (TIMO) system in terms of a single integral is shown analytically in [38]. By applying a method from physics, known as the replica approach, the ergodic capacity of a MIMO system with a large number of antennas, is obtained in a complicated closed-form expression consisting of the trace of matrices [52]. For the outage probability, an

upper bound is obtained in [41] for a MISO system.

2.1.2 Optimal Transmission Strategies

With perfect CSIR, channel-matched decoding can be done, resulting in an enhanced channel capacity compared to the case without any CSI knowledge. When the transmitter has perfect CSIT, an adaptation (in both the spatial and temporal dimensions) can be performed at the transmitter which results in an additional enhancement of channel capacity. However, it is too optimistic in practice to assume the availability of the instantaneous CSIT since it impose a heavy signalling burden on feedback channels. Hence, using partial CSI feedback greatly reduces the signalling burden compared to using the instantaneous CSI feedback. It has been shown that even partial CSIT can increase the ergodic capacity of a MIMO system [20]. Therefore, depending on the different levels of the feedback information available at the transmitter, we have different transmission strategies that achieve the ergodic capacity in the MIMO systems.

Perfect CSIT and Perfect CSTR

For achieving the capacity on frequency-selective fading channels, the transmit signal is circularly symmetric, zero-mean, complex Gaussian distributed and the optimal power allocation is a “water-filling” on the eigenvalues of the channel matrix [9, 42]. The capacity of a time-varying channel is achieved when the transmitter adapts its power, data rate, and coding scheme to the channel variation, and the optimal power allocation is a “water-filling” in time [19]. For the case of single receive antenna, the results in [9, 19] can be reduced in the rank of a matrix. Hence, the capacity-achieving transmit covariance matrix has rank one and, therefore, beamforming achieves the capacity [53].

No CSIT and Perfect CSIR

For i.i.d Rayleigh fading channels, it has been shown that the capacity of the channel is achieved when the transmit signal is independently, circularly symmetric complex Gaussian distributed with mean zero and variance P/N [9, 54], where P is the total transmit power. Hence, the optimal power allocation scheme is equal power allocation. For a special case of $N = M$, there are two conclusions drawn on the ergodic capacity as follows: the capacity scales with increasing SNR for a large but practical number of N and the capacity increases linearly by the number of antenna N [7, 8, 46].

Partial CSIT and Perfect CSIR

It has been shown that even partial CSIT can increase the ergodic capacity of a MIMO system. For any given input covariance matrix, the input distribution that achieve the capacity is shown to be complex vector Gaussian. This leads to the transmitter optimization problem, i.e., finding the transmit covariance matrix that achieves the ergodic capacity subject to a transmit power constraint.

For mean feedback of a MISO channel, the beamforming strategy performs close to the optimal strategy at high feedback SNR [20, 21, 51, 55] since in that case the transmitter can take good advantage of the mean feedback. However, at low feedbck SNR, the optimal strategy is to use N-fold diversity (transmit covariance is full rank), and the power is distributed according to a "water-filling" strategy among the N directions [21,51]. In [21], the optimization procedure involves maximizing an integral over one parameter. For mean feedback of the MIMO channel, results in [56] justify the observations and numerical results for the MISO channel in [51, 55] valid for the MIMO channel.

For covariance feedback of a MISO channel, the optimum solution consists of independent complex circular Gaussian transmit signals along the N eigenvectors

of the transmit covariance matrix. The power along the eigenvectors need to be determined through numerical maximization technologies. The solution resembles "water-filling" principle, i.e., the eigenvectors corresponding to larger eigenvalues receive more power. The power along some of the eigenvectors may be zero so that the optimal diversity order may be less than N [20, 51]. When there is a large enough difference between the two strongest eigenvalues or the amount of water (the SNR) is small enough, then the "waterfilling" just covers the strongest eigenvalue [20], i.e., the beamforming strategy along the corresponding eigenvector performs close to the optimal strategy. This conclusion is also shown in [55] that beamforming in the direction corresponding to the largest eigenvalue is asymptotically optimum as the SNR tends to zero. For determining the power along the eigenvectors, [37] provides an algorithm which computes the optimum power allocation. The maximizing the capacity reduces to an N-parameter maximization over the eigenvalues of the transmit covariance matrix, which can be done with numerical effort [21]. For covariance feedback of a MIMO channel, results in [52,56] justify the observations and numerical results for the MISO channel in [51, 55] valid for the MIMO channel. For the TIMO sytem, optimization over the transmit covariance matrix reduces to a trivial numerical optimization over a single parameter [38].

Substantially different from the results of maximizing the ergodic capacity, minimizing the outage probability for a two-input-single-output (TISO) system with the covariance feedback does not favor the beamforming approach especially for a low number of receive antennas, since the beamforming is highly susceptible to fadings [38]. A near optimum power allocation named as *equal power allocation with channel selection* is derived in [41] for a MISO sytem with the covariance feedback. This near optimum scheme is to select a certain set of transmit antennas and allocate power equally among the selected antennas. As the number of receive antennas increases, the

optimum transmission strategy has a similar trends as that for maximizing the ergodic capacity [38].

2.2 ARQ/HARQ Systems

ARQ/HARQ is an alternative way to mitigate channel fading since the noise burst may have run its course before the retransmitted packet begins to make its way across the channel. In many applications the communication channel is almost error-free except for occasional bursts of noise of short duration. Such noise bursts may be caused, for example, by nearby power machinery, electrical storms, and single-event upsets in digital hardware. In such situations, a simple error detecting ARQ scheme which detects the error burst, discards or stores the affected packet, and requests a retransmission can provide a great deal of protection.

2.2.1 Background of ARQ/HARQ Systems

There are three basic types of ARQ schemes: the stop-and-wait ARQ (SW-ARQ), the go-back-N ARQ (GBN-ARQ), and the selective-repeat ARQ (SR-ARQ) [10, 11]. In a SW-ARQ system, the transmitter sends a codeword to the receiver and waits for an acknowledgement from the receiver. A positive acknowledgement (ACK) from the receiver signals that the codeword has been successfully received, and the transmitter sends the next codeword. A NAK from the receiver indicates that the received has been detected in error, and the transmitter re-sends the codeword. Retransmissions continue until an ACK is received by the transmitter. In a GBN-ARQ system, codewords are transmitted continuously. The transmitter does not wait for an acknowledgement after sending a codeword; as soon as it has completed sending one, it begins sending the next codeword. The acknowledgment for a codeword arrives after a *round-trip delay*

which is defined as the time interval between the transmission of a codeword and the receipt of an acknowledgment for that codeword. During this interval, $N - 1$ other codewords have also been transmitted. When a NAK is received, the transmitter backs up to the codeword that was negatively acknowledged and re-sends that codeword and $N - 1$ succeeding codewords that were transmitted during the round-trip delay. In an SR-ARQ system, codewords are also transmitted continuously; however, the transmitter re-sends only those codewords that are negatively acknowledged.

There is another technique for controlling transmission errors in packet transmission systems: the FEC scheme. In an FEC system, an error-correcting code is used. When the receiver detects the presence of errors in a received vector it attempts to determine the error locations and then corrects the errors. If the exact locations of errors are determined, the received vector will be correctly decoded; if the receiver fails to determine the exact locations of errors, the received vector will be decoded incorrectly, and erroneous data will be delivered to the user. Systems using FEC maintain constant throughput regardless of the channel error rate; however, FEC systems have two drawbacks. First, when a received codeword is detected in error it must be decoded and the decoded message must be delivered to the user regardless of whether it is correct or incorrect. Because the probability of a decoding error is much greater than the probability of an undetected error [10], it is hard to achieve high system reliability with FEC. Second, to obtain high system reliability, a long powerful code must be used and a large collection of error patterns must be corrected. This makes decoding hard to implement and expensive. Comparing the FEC and ARQ systems, ARQ is simple and provides high system reliability. For these reasons, ARQ is often preferred than FEC for error control in communication systems. However, ARQ systems have a severe drawback: their throughput falls rapidly with increasing channel error rate. A combination of the ARQ and the FEC is referred to as a HARQ

system, which overcomes the drawbacks in both the ARQ and the FEC. The function of the FEC subsystem is to reduce the frequency of retransmission by correcting the error patterns that occur most frequently. This increases the system throughput. When a less frequent error pattern occurs and is detected , the receiver requests a retransmission rather than passing the unreliably decoded message to the user. This increases the system reliability. As a result, a HARQ system provides higher reliability than an FEC system alone and a higher throughput than the system with ARQ only.

Considering the transmission mechanisms of the parity-check bits for error correction, the HARQ schemes are classified into Type-I HARQ and Type-II HARQ. In a Type-I HARQ system [57], each packet is encoded for both error detection and error correction. For two-code Type-I HARQ systems, the transmitter is assumed to generate data packets of some fixed length m. The data is first encoded using a high-rate (n', m) error detection code $\mathbf{C}_1$; cyclic redundancy check (CRC) codes are frequently used for $\mathbf{C}_1$. The encoded data is then encoded once again using an (n, n') FEC code $\mathbf{C}_2$. When the packet arrives at the receiver, it is first decoded using the FEC decoder. The resulting n'-bit "message" is then sent to the error detecting decoder. If errors are detected, an retransmission request is sent back to the transmitter. Otherwise, the packet is accepted and the m-bit data packet passes along to the data sink. For single-code Type-I HARQ systems, the FEC decoder is modified to generate retransmission requests using one or both of the following two approaches. The first approach is that if the FEC code is not perfect and the decoder is a bounded-distance decoder, a retransmission request is sent back in the event of a decoder failure. The second approach is that if the FEC decoder is t-error-correcting, a retransmission threshold $t' < t$ is designated such that a retransmission request is generated whenever the number of errors corrected exceeds t'. The design of single-code Type-I HARQ typified by the Golay protocol in [11, Example 15-4] has been applied to a number of

different block and convolutionally encoded FEC systems. Much work has been done on the development of single-code Type-I HARQ baed on the sequential decoding of convolutional codes [58], on the Viterbi decoder [16, 59], and on the majority-logic decoding of both convolutional [17] and cyclic block codes [60]. The most powerful of the single-code Type-I HARQ systems are those based on Reed-Solomon codes [61, 62]. In the Type-I HARQ scheme, the same coded packet is retransmitted and these multiple packets can be combined in two distinct ways. In the code combining scheme, these repeated packets are concatenated to form a single packet at a lower code rate, which is often referred to as Chase combining [14–16]. In the diversity combining scheme, these repeated packets are combined into a single packet at the same rate with more reliable constituent symbols by using symbol voting schemes [17] or by using symbol averaging schemes [16].

In a Type-II HARQ scheme [18,63,64], the data is first encoded using a high-rate error detecting code to form a packet. The packet is then encoded using a systematic invertible code by adding some parity bits. The additional parity check bits are sent only when errors are detected in the packet. The receiver appends these bits to the received packet for increasing the error correction capability. This is also known as the incremental redundancy scheme [18]. Two separate codes can be used in this scheme: a high-rate (n, m) error detecting code $\mathbf{C}_1$ and a $(2n, n)$ systematic invertible code $\mathbf{C}_2$. An m-bit message is first encoded using $\mathbf{C}_1$ to form an n-bit packet $\mathbf{P}_1$. Then $\mathbf{P}_1$ is encoded using $\mathbf{C}_2$. The n parity bits called $\mathbf{P}_2$ from the $\mathbf{C}_2$ code word are saved in a buffer, while the $\mathbf{C}_1$ codeword $\mathbf{P}_1$ is transmitted. The initial packet is checked for errors at the receiver. If it is found to contain errors, a retransmission request is sent back to the transmitter. The transmitter responds by sending $\mathbf{P}_2$. Since $\mathbf{C}_2$ is invertible, the n bits used to create the $\mathbf{C}_2$ codeword can be obtained by inverting the packet $\mathbf{P}_2$. An inverted version of $\mathbf{P}_2$ is created and check for errors. If the inverted

version contains errors, $\mathbf{P}_2$ is appended to $\mathbf{P}_1$ to create a noise corrupted $\mathbf{C}_2$ codeword. After FEC decoding, the resulting message is checked once again for errors. If there are still errors, the process continues, with the transmitter alternating transmission of $\mathbf{P}_1$ and $\mathbf{P}_2$ until one of the three error detection decoding operations is successfully passed. The error detection role can be served by CRC codes while a class of codes based on shortened cyclic codes is selected for the half-rate systematic invertible code.

2.2.2 Performance of Packet ARQ/HARQ Schemes

Much work has been done on the performance of ARQ/HARQ systems over fading channels. The performance analysis has been developed for different cases which depend on the availability of the CSIT and/or the CSIR.

Perfect CSIT and Perfect CSIR

For block fading channels, the goodput performance of the Type-I schemes with code combining and diversity combining are theoretically analyzed [12, 16, 26]. The performance derivation is based on the use of the sphere-packing bound. The bounds on the throughput of Type-II HARQ schemes are obtained by using punctured convolutional coding in [65, 66] and by using block codes in [67]. For correlated channels, the throughput performance of the basic ARQ is presented by using a one-step Markov process [27] and by using finite state channel models in [22]. The throughput performance of the Type-I HARQ scheme with code combining is theoretically analyzed over by using a two state Markov channel model [28]. The goodput performance of the Type-II HARQ scheme is theoretically analyzed by adopting a finite-state Markov chain [12, 26] based on the use of the sphere-packing bound. Sphere-packing bound can be used to evaluate a reasonably accurate approximation for the achievable performance. By using punctured convolutional

coding, the throughput performance is analyzed by adopting a two state Markov channel model [28] and by adopting a M-state Markov channel model [68].

Imperfect CSIT and Perfect CSIR

For block fading channels, bounds on the throughput as well as the reliability are derived using random coding techniques in [69]. The random coding bounds reveal the achievable performance with block codes and maximum-likelihood soft-decision decoding. For correlated fading channels, the throughput of basic ARQ schemes with unreliable feedback is analyzed in [70–72]. The fading channel is modeled by a Gilbert channel model and the patterns of packet and feedback errors follow two independent first-order Markov models.

Imperfect CSIT and Imperfect CSIR

By considering the burst nature of both forward channel errors and feedback channel errors modeled by using a joint hidden Markov model with a finite number of states, the throughput of the GBN-ARQ schemes is analyzed over fading channels in [73].

2.2.3 Adaptive Transmission Strategies

In adaptive transmission, the transmission rates are adapted with respect to the channel conditions. Therefore, the CSI plays a crucial role in determining the performance of the systems. Some works on the performance analysis of ARQ schemes with adaptive transmission strategies have been addressed for different cases which depend on the availability of the CSIT and/or the CSIR.

Perfect CSIT and Perfect CSIR

For block fading channels, the general expressions for the throughput of basic ARQ schemes with adaptive transmissions is provided in [74]. A cross-layer design which combines adaptive modulation and coding with ARQ is developed in order to maximize spectral efficiency under prescribed delay and error performance constraints [75]. The performance of the cross-layer design is analyzed and the throughput is obtained in closed-form. For correlated fading channels, a generic method to analyze the goodput performance of an IEEE 802.11a system is presented in [13]. The general expressions of the goodput as a function of the data payload length, the packet retry count, the wireless channel condition, and the selected data transmission rate are obtained by assuming a two-state discrete time Markov chain channel variation mode. An analytical method that uses a finite-state Markov chain as an error model is used to analyze the performance of ARQ schemes with adaptive transmissions [76].

Imperfect CSIT and Perfect CSIR

For block fading channels, the throughput of the basic ARQ schemes with adaptive transmissions with imperfect CSIT is expressed in the general form [74] as a function of the probability that a received packet is error-free. However, the estimation errors of the imperfect CSIT is generally assumed to be a Gaussian distribution. For correlated fading channels, both throughput and packet error rate are analyzed for an ARQ scheme based on a constant-power variable-rate adaptive M-quadrature amplitude modulation (M-QAM) [29]. The Markovian channel model is used to describe the time-varying multipath fading behavior. The impact of the imperfect CSIT on the performance of the system is considered.

Chapter 3

On the Ergodic Capacity of MIMO Rayleigh Fading Channels

The ergodic capacity of MIMO fading channels depends on the joint distribution of the eigenvalues of a Wishart matrix, and is quite complex in general. We obtain here three simple and tight bounds on the ergodic capacity of the MIMO Rayleigh fading channel. These simple bounds show explicitly the dependence of the ergodic capacity on the SNR and the numbers of transmit and receive antennas. They enable us to determine the optimum number of transmit antennas to be used for a given SNR and a given total number of antennas in the system.

3.1 Introduction

The information-theoretic analysis of MIMO channels has attracted a lot of research attention. One focus of such a study is the ergodic capacity, which is defined as the ensemble average of the mutual information (MI) over the statistical distribution of the channels. In particular, the ergodic capacity has been investigated in [7, 9, 77–79] for the Rayleigh fading case. However, the result in [7] is only for the case of a large

number of antennas, while the results in [9, 77–79] rely on the joint distribution of the eigenvalues of a Wishart matrix and are quite complex. Such complicated results do not allow one in general to study explicitly the dependence of the ergodic capacity on the system parameters. In particular, we are interested here in the optimum number of transmit antennas to be used, as a function of the SNR. Instead of using the joint distribution of the eigenvalues of a Wishart matrix, a new approach is proposed in this paper to derive upper and lower bounds on the ergodic capacity based on the results we obtained in [50]. Our approach to the ergodic capacity analysis greatly simplifies the computational procedure, and provides easy and accurate ways to deal with ergodic capacity-related calculations for MIMO Rayleigh fading channels. The bounds obtained here on the ergodic capacity are expressed in simple closed forms, and show explicitly the effects of the system parameters on the ergodic capacity. The bounds are valid for an arbitrary number of antennas, and they enable us to design an optimum antenna deployment strategy, i.e., to determine the optimum number of transmit antennas for a given SNR and a given total number of antennas in the system. For the Rician fading case, the ergodic capacity is considered in [44, 46, 49, 80, 81]. The results of [80] are based on the Gaussian approximation to the distribution of the MI, and are accurate only for a large number of antennas. The bounds in [46] are only tight in the high SNR regime for a large number of antennas. The results in [81] are only for dual MIMO systems. The bounds in [44, 49] involve complicated expressions that do not reduce to the simple closed forms when specialized to the Rayleigh fading case.

3.2 System Description

Consider a single user MIMO system with N transmit and M receive antennas. The M-dimensional received signal vector is mathematically represented as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{3.1}$$

where $\mathbf{H}$ is the $M \times N$ channel matrix with the mn-th entry, h_{mn}, being the channel gain between the n-th transmit antenna and the m-th receive antenna. The entries, h_{mn}, are i.i.d., complex, Gaussian random variables with mean zero and variance σ^2, i.e., $h_{mn} \sim \mathcal{CN}(0, \sigma^2)$. Vector $\mathbf{x}$ is an N-dimensional transmitted vector, and $\mathbf{n}$ is AWGN. We assume that $\mathbb{E}[\mathbf{n}\mathbf{n}^H] = N_0\mathbf{I}_M$ and the total transmitted energy is E_s, i.e., $\mathbb{E}[\mathbf{x}^H\mathbf{x}] = E_s$.

We consider the scenario where the receiver has perfect knowledge of the CSI, and the transmitter has no channel knowledge at all (neither CSI nor fading distribution). In this case, for any realization of $\mathbf{H}$, the MI, $\mathcal{I}(\mathbf{x}; \mathbf{y}|\mathbf{H})$, is maximized when the transmit signal is circularly symmetric, zero-mean, complex, Gaussian distributed [9]. Only the covariance matrix $\mathbb{E}[\mathbf{x}\mathbf{x}^H]$ of the capacity-achieving transmit signal depends on the fading distributions. For i.i.d Rayleigh fading channels, the capacity-achieving covariance matrix is $\mathbb{E}[\mathbf{x}\mathbf{x}^H] = (E_s/N)\mathbf{I}_N$ [9, Theorem 1]. Therefore, the MI is given by [8, eq (2.10)]

$$\begin{aligned}\mathcal{I} = \mathcal{I}(\mathbf{x}; \mathbf{y}|\mathbf{H}) &= \frac{1}{\ln 2} \ln\det(\mathbf{I}_N + \frac{E_s}{NN_0}\mathbf{H}^H\mathbf{H}) \\ &= \frac{1}{\ln 2} \ln\det(\mathbf{I}_N + \gamma\sigma^2\mathbf{Z}^H\mathbf{Z}) \quad \text{bps/Hz},\end{aligned} \tag{3.2}$$

where $\gamma = E_s/(NN_0)$ is the average SNR at each transmit antenna, $\mathbf{Z} = \sqrt{1/\sigma^2}\mathbf{H}$ and $\mathbf{Z} \sim \mathcal{CN}_{N,M}(\mathbf{0}, \mathbf{I}_N \otimes \mathbf{I}_M)$. Without loss of generality, we assume $M \geq N$. Let $\lambda_n, n = 1, 2, \cdots, N$ be the nonzero eigenvalues of $\mathbf{Z}^H\mathbf{Z}$. The mutual information in

(3.2) can be expressed as [9, Sec 3.2]

$$\mathcal{I} = \mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H}) = \frac{1}{\ln 2}\ln\left(\prod_{n=1}^{N}\left(1+\gamma\sigma^2\lambda_n\right)\right). \tag{3.3}$$

If $M < N$, all of our results in this chapter remain the same, except for interchanging M with N.

3.3 Trace Bounds

Since the MI in (3.2) is a random variable, the ergodic capacity depends on the distribution of the MI. Up till now, the MI distribution is studied using the joint distribution of the eigenvalues of $\mathbf{H}^H\mathbf{H}$, and the expressions for the MI distribution function are quite difficult to evaluate. In this section, we will use some new and simple approaches to investigate the ergodic capacity based on the distributions of the trace of the Wishart matrix.

3.3.1 Upper bound

Applying the arithmetic-and-geometric-means inequality: $(\prod_{n=1}^{N} x_n)^{1/N} \leq \sum_{n=1}^{N} x_n/N$ for all $x_n > 0$, to (3.3), we can upper bound the MI as

$$\mathcal{I} \leq \frac{1}{\ln 2}N\ln\left(1+\frac{\sigma^2\gamma}{N}\sum_{n=1}^{N}\lambda_n\right). \tag{3.4}$$

Since the trace of a matrix is equal to the sum of its eigenvalues: $\sum_{n=1}^{N}\lambda_n = \text{tr}(\mathbf{Z}^H\mathbf{Z})$, therefore, we can rewrite the above upper bound on the MI as

$$\mathcal{I} \leq \frac{1}{\ln 2}N\ln\left(1+\frac{\sigma^2\gamma}{N}\text{tr}(\mathbf{Z}^H\mathbf{Z})\right). \tag{3.5}$$

By taking the expectation of the upper bound over the distribution of $\text{tr}(\mathbf{Z}^H\mathbf{Z})$, the ergodic capacity is bounded by

$$\mathbb{E}[\mathcal{I}] \leq \mathbb{E}\left[\frac{N}{\ln 2}\ln(1+\frac{\gamma\sigma^2}{N}\text{tr}(\mathbf{Z}^H\mathbf{Z}))\right]. \tag{3.6}$$

Since the trace of $\mathbf{Z}^H\mathbf{Z}$ can be expressed as the sum of the magnitude square of all the elements of a Gaussian variate matrix, i.e., tr$(\mathbf{Z}^H\mathbf{Z}) = (\sigma^2)^{-1}||\mathbf{H}||_F^2 = (\sigma^2)^{-1}\sum_{m=1}^{M}\sum_{n=1}^{N}|h_{mn}|^2$, it can easily be seen that tr$(\mathbf{Z}^H\mathbf{Z})$ is central chi-square distributed with $2NM$ degrees of freedom. The underlying i.i.d Gaussian random variables are of mean zero and variance $1/2$. Defining $Z = \text{tr}(\mathbf{Z}^H\mathbf{Z})$, the pdf of Z is given by [6, eq (2.1.110)]

$$p_Z(z) = \frac{1}{\Gamma(NM)} z^{NM-1} e^{-z}, \quad z \geq 0. \tag{3.7}$$

Now, evaluating the expectation in (3.6) using the pdf in (3.7), we obtain an upper bound on the ergodic capacity given by

$$\begin{aligned}\mathbb{E}[\mathcal{I}] \leq \overline{\mathcal{I}}_{\text{tr\_U}} &\triangleq \frac{1}{\ln 2} N e^{N/(\sigma^2\gamma)} \\ &\times \sum_{j=0}^{NM-1} \left(\frac{N}{\sigma^2\gamma}\right)^{NM-1-j} \Gamma\left(-(NM-1-j), \frac{N}{\sigma^2\gamma}\right)\end{aligned} \tag{3.8}$$

in which, the term $\Gamma(\alpha, x)$ is the upper incomplete gamma function, i.e., $\Gamma(\alpha, x) = \int_x^\infty t^{\alpha-1}e^{-t}dt$. The details of the derivation of (3.8) are shown in Appendix A.

Reference [82] provides an alternative derivation, and [82, eq.(25)] is an alternative form for the result (3.8). The latter result, however, is computationally simpler. As will be seen in Section 3.5, this trace upper bound provides a tight bound with the same trend as the Monte Carlo results as well as the result computed from [9].

We next examine the behavior of (3.8) with respect to N to obtain insights on the dependence of the ergodic capacity on the antenna deployment. By applying a high-SNR assumption and using the fact that $\lim_{x\to 0}\Gamma(\alpha, x) = -x^\alpha/\alpha$, (3.8) can be simplified to

$$\overline{\mathcal{I}}_{\text{tr\_U}} \doteq \frac{N}{\ln 2} e^{\frac{N^2 N_0}{E_s\sigma^2}} \left[\sum_{k=1}^{NM-1} \frac{1}{k} + \text{E}_1\left(\frac{N^2 N_0}{E_s\sigma^2}\right)\right], \tag{3.9}$$

where $\text{E}_1(x)$ is the exponential integral function defined as $\text{E}_1(x) = \int_x^\infty e^{-t}/t dt, x > 0$ and the approximation symbol $\doteq$ denotes that the ratio of its two sides converges to

one as $E_s\sigma^2/N_0 \to \infty$. Given $N + M = N_A$, we examine the first derivative of (3.8) with respect to N at high SNR. Using (3.9), one can show that

$$\lim_{\frac{E_s\sigma^2}{N_0}\to\infty} \frac{d\overline{\mathcal{I}}_{\text{tr\_U}}}{dN} = \frac{1}{\ln 2}\left[\sum_{k=1}^{NM-1}\frac{1}{k} + \mathrm{E}_1\left(\frac{N^2N_0}{E_s\sigma^2}\right)\right] - \frac{1}{\ln 2}\left[2 + \frac{N(M-N)}{(NM-1)^2}\right] >> 0. \tag{3.10}$$

Thus, $\overline{\mathcal{I}}_{\text{tr\_U}}$ increases with increasing N at high SNR, and N should be as large as possible in order to maximize $\overline{\mathcal{I}}_{\text{tr\_U}}$, subject to the constraint that $N \leq M$.

For low SNR, by using the fact that $\lim_{x\to\infty}\Gamma(\alpha, x) = x^{\alpha-1}e^{-x}$, (3.8) can be simplified to

$$\overline{\mathcal{I}}_{\text{tr\_U}} \fallingdotseq \frac{N}{\ln 2}\left[\frac{E_s\sigma^2(NM-1)}{N_0N^2} + e^{\frac{N^2N_0}{E_s\sigma^2}}\mathrm{E}_1\left(\frac{N^2N_0}{E_s\sigma^2}\right)\right] \tag{3.11}$$

where the approximation symbol $\fallingdotseq$ denotes that the ratio of its two sides converges to one as $E_s\sigma^2/N_0 \to 0$. The first derivative of (3.8) with respect to N for low SNR can be shown to be

$$\lim_{\frac{E_s\sigma^2}{N_0}\to 0} \frac{d\overline{\mathcal{I}}_{\text{tr\_U}}}{dN} = \frac{E_s\sigma^2(1-N^2)}{N_0N^2\ln 2} \leq 0. \tag{3.12}$$

Thus, $\overline{\mathcal{I}}_{\text{tr\_U}}$ decreases as N increases at low SNR, and N should be as small as possible (i.e., $N = 1$) for maximizing $\overline{\mathcal{I}}_{\text{tr\_U}}$. In the derivation of (3.12), we use the fact that $\lim_{x\to\infty} e^x\mathrm{E}_1(x) = 0$ and $\lim_{x\to\infty} xe^x\mathrm{E}_1(x) = 1$. A simpler way to arrive at this conclusion is to note that a simple upper bound can be obtained by applying the inequality: $\ln x \leq x - 1, x > 0$ to (3.6). This gives

$$\mathbb{E}[\mathcal{I}] \leq \frac{E_s\sigma^2(N_A - N)}{N_0\ln 2}. \tag{3.13}$$

While it is tight only at low SNR, it clearly shows that $\mathbb{E}[\mathcal{I}]$ is maximized by setting $N = 1$ at low SNR.

3.3.2 Lower bounds

Expanding $\prod_{n=1}^{N}(1+\sigma^2\gamma\lambda_n)$, we obtain the inequality

$$\prod_{n=1}^{N}\left(1+\sigma^2\gamma\lambda_n\right) \geq 1+\sigma^2\gamma\sum_{n=1}^{N}\lambda_n+\left(\sigma^2\gamma\right)^N\prod_{n=1}^{N}\lambda_n$$
$$\geq 1+\sigma^2\gamma\text{tr}(\mathbf{Z}^H\mathbf{Z}). \tag{3.14}$$

By applying the inequality (3.14) to (3.3), the ergodic capacity can be lower bounded by

$$\mathbb{E}[\mathcal{I}] \geq \mathbb{E}\left[\frac{1}{\ln 2}\ln(1+\sigma^2\gamma\text{tr}(\mathbf{Z}^H\mathbf{Z}))\right]. \tag{3.15}$$

This lower bound is tight only at low SNR, because the inequality (3.14) ignores the terms containing second and higher powers of the SNR. The expectation in (3.15) can be evaluated in the same way as that in (3.6), giving

$$\mathbb{E}[\mathcal{I}] \geq \overline{\mathcal{I}}_{\text{tr\_L}} \triangleq \frac{1}{\ln 2}e^{1/(\sigma^2\gamma)}$$
$$\times\sum_{j=0}^{NM-1}\left(\frac{1}{\sigma^2\gamma}\right)^{NM-1-j}\Gamma\left(-(NM-1-j),\frac{1}{\sigma^2\gamma}\right) \tag{3.16}$$

Reference [82] provides an alternative derivation, and [82, eq.(26)] is an alternative form for the result (3.16). However, (3.16) is computationally simpler. The lower bound is only tight at low SNR whereas the upper bound is tight for all SNR when the number of antennas is small. However, with an increase in the number of antennas, the two bounds become looser. This is consistent with the results on the capacity in [45].

3.4 Determinant Bound

In this section, we derive a lower bound on the ergodic capacity in terms of the distribution of the determinant $\det(\mathbf{Z}^H\mathbf{Z})$. Expanding $\prod_{n=1}^{N}(1+\sigma^2\gamma\lambda_n)$, we obtain

the inequality

$$\prod_{n=1}^{N}\left(1+\sigma^2\gamma\lambda_n\right) \geq 1+\sigma^2\gamma\sum_{n=1}^{N}\lambda_n+\left(\sigma^2\gamma\right)^N\prod_{n=1}^{N}\lambda_n$$
$$\geq 1+\left(\sigma^2\gamma\right)^N\prod_{n=1}^{N}\lambda_n. \tag{3.17}$$

Since the determinant of a matrix is equal to the product of its eigenvalues: $\prod_{n=1}^{N}\lambda_n = \det(\mathbf{Z}^H\mathbf{Z})$, we can lower bound the MI as

$$\mathcal{I} \geq \frac{1}{\ln 2}\ln\left(1+(\sigma^2\gamma)^N\det(\mathbf{Z}^H\mathbf{Z})\right). \tag{3.18}$$

The complementary cumulative distribution function (CCDF) of the MI is then lower bounded as

$$P(\mathcal{I}>R) \geq P\left(\det(\mathbf{Z}^H\mathbf{Z}) > a^N(e^{R\ln 2}-1)\right), \tag{3.19}$$

where $a = 1/\sigma^2\gamma$. The closed-form lower bound on the CCDF of the MI for $N \geq 2$ is given by [50, eq(25)]

$$P(\mathcal{I}>R) \geq \sum_{r=0}^{M-N}\frac{(M-1-r)!}{r!(M-1)!}a^{rN}(e^{R\ln 2}-1)^r$$
$$\times\exp\left(-\frac{(M-N-r)!a^N(e^{R\ln 2}-1)}{(M-1-r)!}\right)$$
$$\times\prod_{n=2}^{N-1}\frac{(M-n-r)!}{(M-n)!}. \tag{3.20}$$

Denoting the PDF of the MI by $p_{\mathcal{I}}(t)$, we can compute its expectation as

$$\mathbb{E}[\mathcal{I}] = \int_0^\infty t p_{\mathcal{I}}(t)dt = \int_0^\infty\int_0^t p_{\mathcal{I}}(t)dRdt \tag{3.21}$$

Interchanging the order of integration, we have

$$\mathbb{E}[\mathcal{I}] = \int_0^\infty\int_R^\infty p_{\mathcal{I}}(t)dtdR = \int_0^\infty P(I>R)dR \tag{3.22}$$

Applying the lower bound on the CCDF of the MI in (3.20) to (3.22), we can obtain the determinant lower bound on ergodic capacity as follows. Let

$$\alpha = \frac{a^{rN}}{r!} \prod_{n=1}^{N-1} \frac{(M-n-r)!}{(M-n)!}, \quad \mu(r) = \frac{(M-N-r)!a^N}{(M-1-r)!},$$

and $x = \exp(R \ln 2) - 1$. The lower bound on the average MI can be evaluated by

$$\mathbb{E}[\mathcal{I}] \geq \sum_{r=0}^{M-N} \frac{\alpha}{\ln 2} \int_0^\infty x^r \exp\left(-\mu(r)x\right)(x+1)^{-1} dx. \tag{3.23}$$

By applying (A.6) in Appendix A to (3.23), after some manipulations, the above lower bound can be simplified into

$$\mathbb{E}[\mathcal{I}] \geq \overline{\mathcal{I}}_{\text{det}} \triangleq \sum_{r=0}^{M-N} \frac{a^{rN} e^{\mu(r)} \Gamma(-r, \mu(r))}{\ln 2} \prod_{n=1}^{N-1} \frac{(M-n-r)!}{(M-n)!}. \tag{3.24}$$

An alternative form for the result (3.24) is provided by [82, eq.(29)], but (3.24) is computationally simpler. Compared with the trace lower bound that is only tight for low SNRs, the determinant lower bound is only tight for high SNRs.

3.5 Simulation and Numerical Results

We have derived three bounds on the ergodic capacity of MIMO Rayleigh fading channels. In particular, they are one trace upper bound, one trace lower bound and one determinant lower bound. In this section, we compare them with the results from Monte Carlo simulations. Based on the observations concerning the ergodic capacity, an optimum antenna deployment strategy is further investigated.

3.5.1 Trace bounds and determinant bound

In this subsection, we compare the analytical results for the bounds on the ergodic capacity with the results obtained from Monte Carlo simulations as well as the result

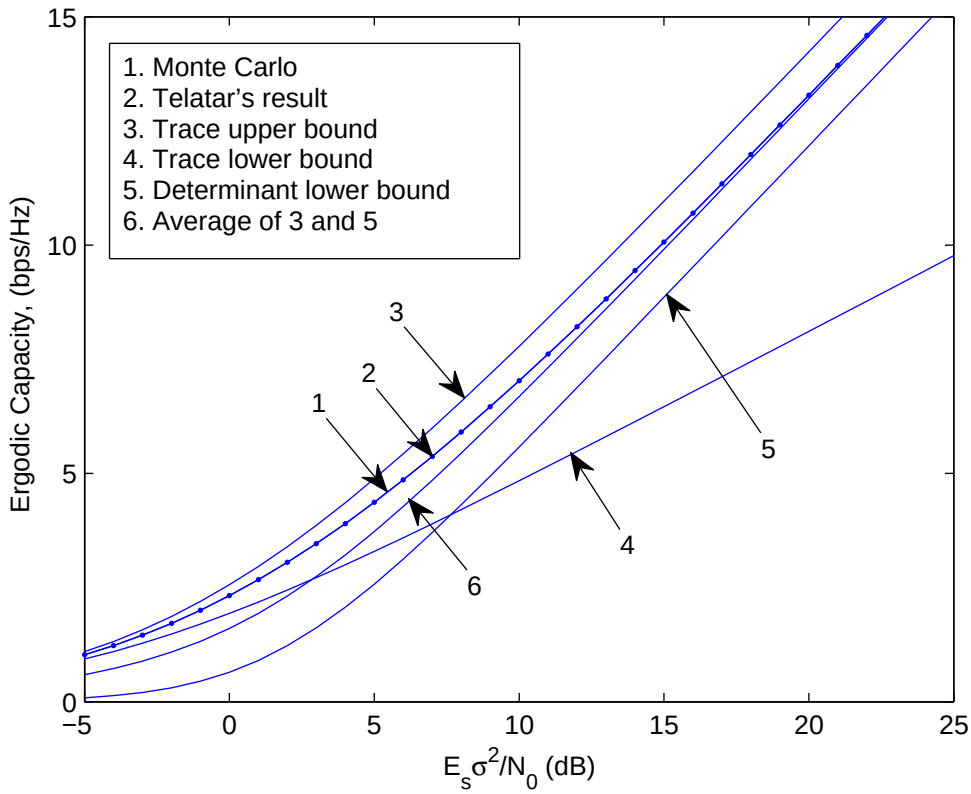

Figure 3.1: Bounds on the average MI of MIMO Rayleigh channels with $N = 2$ and $M = 3$.

obtained in [9] which is a function of Laguerre polynomials. The ergodic capacity in terms of Laguerre polynomials is given by [9, eq (8)]

$$\overline{\mathcal{I}} = \sum_{k=0}^{N-1} \frac{k!}{(k+M-N)!} \times \int_0^\infty \log(1+\frac{E_s\lambda}{N})[L_k^{M-N}(\lambda)]^2 \lambda^{M-N} e^{-\lambda} d\lambda, \tag{3.25}$$

where $L_k^{M-N}(x) = \frac{1}{k!} e^x x^{N-M} \frac{d^k}{dx^k}(e^{-x} x^{M-N+k})$ is the associated Laguerre polynomial of order k. It is easy to observe that it is complicated to evaluate (3.25) since the evaluation involves $N(N-1)/2$ differential operations and N integrations. Compared with the results in [9], the evaluation of either the trace upper bound (3.8) or the trace

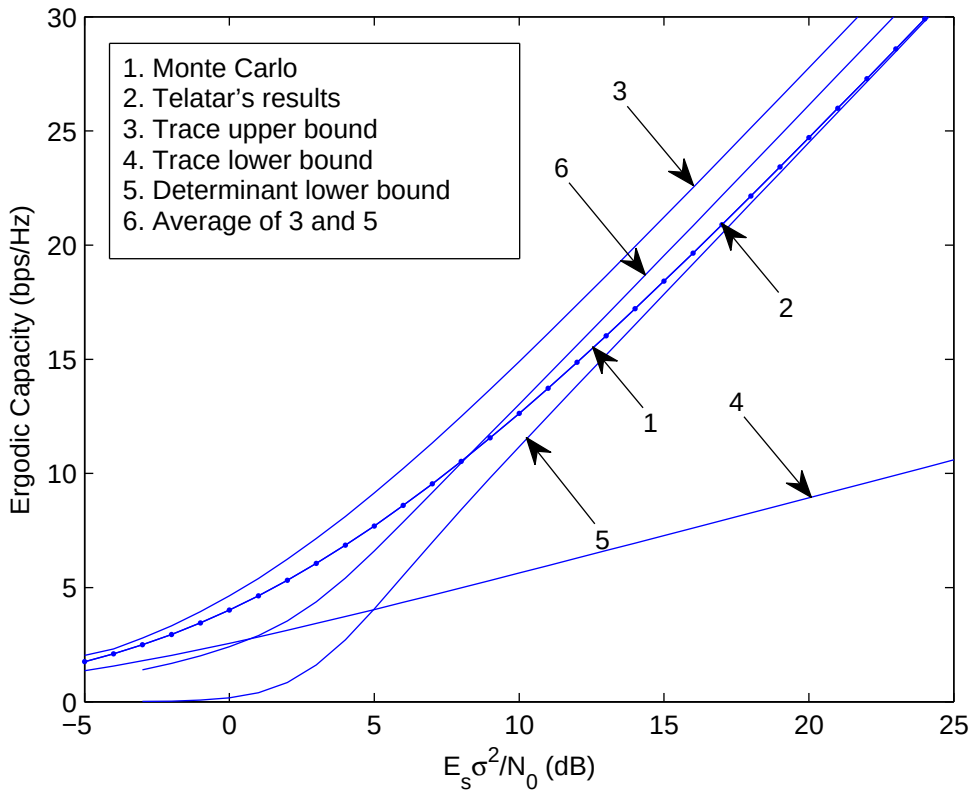

Figure 3.2: Bounds on the average MI of MIMO Rayleigh channels with $N = 4$ and $M = 5$.

lower bound (3.16) involves one integration operation. In Fig. 3.1 and Fig. 3.2, our three bounds and the results from Monte Carlo simulations are plotted. It is observed that the trace upper bound provides a tight bound with the same trend as the Monte Carlo results for all SNRs, whereas the trace lower bound is only tight at low SNR. With an increase in the number of antennas, the two bounds become a little looser. For the determinant lower bound, in the low SNR regime, it is looser than the trace lower bound. However, at high SNR, it provides a good prediction for the ergodic capacity. The tightness is more significant with an increasing number of antennas, which can be observed by comparing Fig. 3.1 and Fig. 3.2.

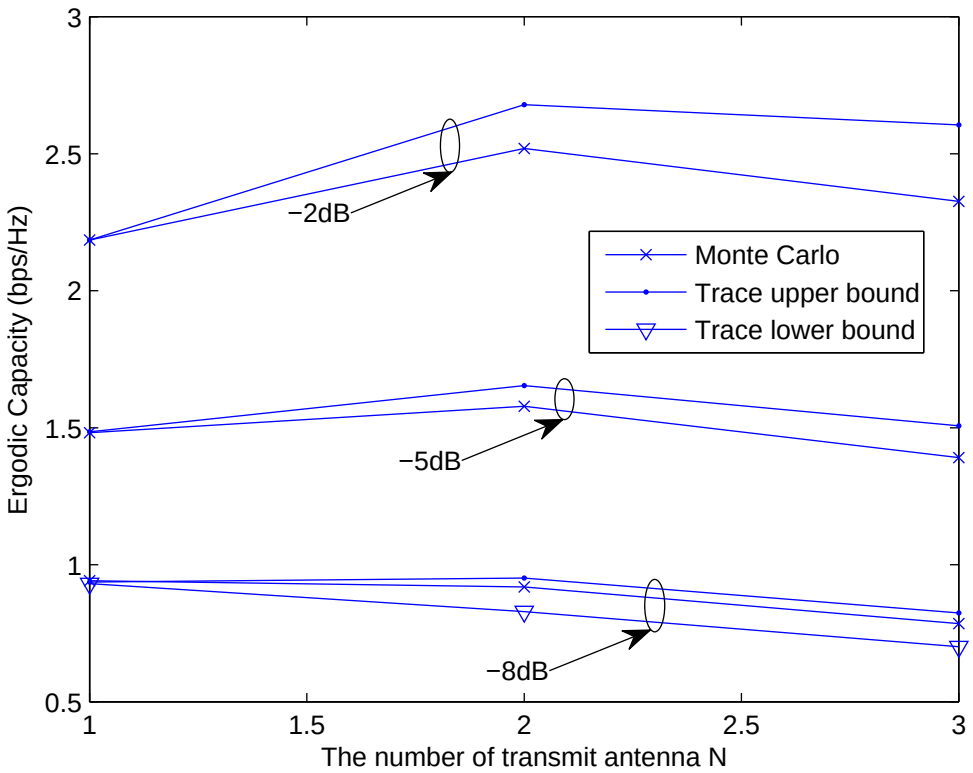

Figure 3.3: Optimum N for achieving the maximum ergodic capacity (N+M=7).

3.5.2 Optimum Antenna Deployment

For a system with a total number of antennas N_A i.e., $N + M = N_A$, and total transmit energy E_s, the optimum number of transmit antennas which maximizes the ergodic capacity will be investigated in this section. Since a simple closed-form expression for the ergodic capacity is not available, we will make use of the bounds on ergodic capacity to analyze numerically the relationship between the optimum number of transmit antennas and the SNR. For a given total transmit SNR and a fixed total number of antennas, using more transmit antennas will decrease the SNR per transmit channel. When the total transmit SNR is low, deploying more transmit antennas cannot help increase the channel capacity since the SNR per transmit-receive link is further

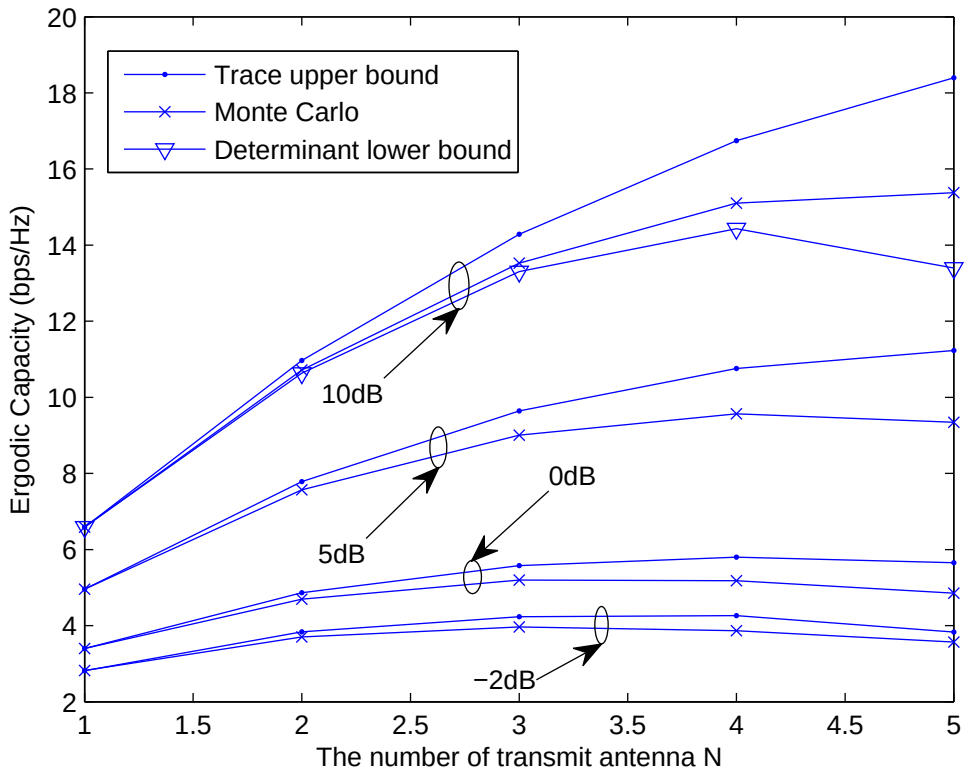

Figure 3.4: Optimum N for achieving the maximum ergodic capacity (N+M=11).

decreased. When the total transmit SNR is high, providing more transmit-receive links is beneficial for achieving a higher channel capacity because of the greater spatial diversity achieved. Therefore, an optimum number of transmit antennas exists, which depends on the total transmit SNR and the total number of antennas. In Fig. 3.3 and Fig. 3.4, the optimum number of transmit antennas can be easily observed for different transmit SNRs. In particular, in Fig. 3.4, at high SNR, i.e., 10dB, the optimum number of transmit antennas is the maximum possible number of transmit antennas, which is 5 in this case. However, at low SNR, i.e., -2dB, the optimum number of transmit antennas is only 2. These observations are consistent with the conclusions drawn from (3.9) and (3.11). In both cases, we can observe that the trace upper bound provides

good prediction at all SNRs, whereas the trace lower bound and the determinant lower bound can only work at low SNR and high SNR, respectively.

3.6 Conclusions

In this chapter, we derive bounds on the ergodic capacity of the MIMO Rayleigh fading channel by exploiting the properties and distributions of the trace and the determinant of a Wishart matrix. The expressions of the bounds are simple and easy to compute. The trace upper bound is tight for all SNRs, while the trace lower bound and the determinant lower bound are tight for low SNR and high SNR, respectively. Furthermore, for a system with a fixed total number of antennas at a certain SNR, increasing the number of transmit antennas cannot always guarantee increasing the ergodic capacity. At high SNR, a larger number of transmit antennas helps increase the ergodic capacity while at low SNR, a smaller number of transmit antennas works better.

Chapter 4

Power Control for MIMO Diversity Systems with Non-Identical Rayleigh Fading Channels

MIMO technology offers significant increases in data throughput and link reliability without additional bandwidth or transmit power in wireless communications. These advantages are well represented in two forms of gains from the information-theoretic perspective, namely, diversity and multiplexing. In particular, the diversity advantage is built upon the transmission of the same message over multiple, independently faded, spatial branches. It can be accomplished by using space-time block codes (STBC) or other spreading codes together with appropriate combining at the receiver. Tremendous amount of work has been done on the design and analysis of space-time diversity techniques, such as [83–85] and references therein. It is noted that in most previous works [9, 52, 85, 86], the MIMO channels are assumed to be i.i.d.

Recently, MIMO diversity schemes over i.n.d fading channels have attracted great attention because of their applications in cooperative communications and distributed antenna systems. In decode-and-forward (DF) cooperative communications systems

[87], a transmission from a source to a destination is facilitated with the help of a set of relays. In the second stage when those relays have decoded the transmitted signal, the subsequent transmission can be performed by employing space-time coding in a distributed manner, resulting in a non-identical MISO fading channel. In distributed antenna systems [88], multiple antennas that are distributed at different radio ports and connected through coaxial cables work together to simulcast signals. A non-identical MIMO channel is actually formed, which enhances signal quality, increases system capacity and improves coverage. In the aforementioned cases, the channels on different transmit-receive antenna pairs can be modeled as independent but not necessarily identically distributed fading channels. Most previous works on MIMO systems assume either perfect CSI [9, 45, 89, 90] or identical channel distribution [9]. The studies on the effects of i.n.d fading for MIMO diversity systems have focused on the bit error analysis [91–95]. In this chapter, we are interested in the performance limits and associated power allocation problems in a MIMO diversity system given that the non-identical fading statistics are available at the transmitter. It is well known that equal power allocation is optimal for traditional MIMO channels with identical fading distribution [9], assuming no instantaneous CSIT. However, it is no longer optimal for non-identical MIMO channels. In [41], the mutual information outage of a transmit diversity system with a single receive antenna over non-identical Rayleigh fading channels is studied. Therein, a heuristic power control scheme named *equal power allocation with channel selection* is proposed and is shown to be near optimal in minimizing the information outage probability.

For ergodic mutual information maximization in a i.n.d MIMO diversity system, to the best of our knowledge, no closed-form power control is available in the literature. Our first contribution here is therefore to obtain the optimal power allocation for achieving the maximum ergodic mutual information. For information outage

minimization, a water-filling based power allocation is proposed. The derivation is based on the Chernoff bound and is different from [41]. Moreover, it is generalized to the case with multiple receive antennas.

4.1 System Model

Consider a narrowband system with $N > 1$ transmit antennas and $M \geq 1$ receive antenna(s). The channel is Rayleigh fading with additive white Gaussian noise of double-sided power spectral density N_0. The entity h_{ij} of channel matrix $\mathbf{H}$ is the channel coefficient between the i-th receive antenna and the j-th transmit antenna, and $\{h_{ij}\}_{i=1,j=1}^{M,N}$ is a set of i.n.d zero-mean complex Gaussian variables, each with variance σ_{ij}^2. The information is transmitted over the multiple antenna system by using an orthogonal space-time block codes (OSTBC). The code matrix $\mathbf{C}$ can be constructed by using a generalized complex orthogonal design [8, Chapter 4]. The entries of $\mathbf{C}$ are linear combinations of s_1, s_2, $\cdots$, s_K and their conjugates, where s_k for $k = 1, \cdots, K$ are the signals to be transmitted and the energy of each symbol is normalized to be $|s_k|^2 = 1$. The entries c_{jt}, $j = 1, \cdots, N$ are transmitted simultaneously from transmit antennas 1, 2, $\cdots$, N at each time slot $t = 1, 2, \cdots, T$. The received signal can be mathematically represented in terms of the channel variances, the transmit power, the transmit signals and the noises

$$\mathbf{R} = \mathbf{HPC} + \mathbf{N}, \tag{4.1}$$

where $\mathbf{R}$ is the $M \times T$ matrix with the it-th entry, r_{it}, being the received signal from the ith receive antenna at time slot t, $\mathbf{N}$ is the $M \times T$ noise matrix with the entries n_{it} being the i.i.d, zero-mean complex Gaussian random variables with variance N_0 and $\mathbf{P}$ is a diagonal matrix with $\sqrt{p_j}, j = 1, \cdots, N$ as its eigenvalues. The p_j denote the power radiated from the j-th transmit antenna, and let it be subject to a normalized

total power constraint $\sum_{j=1}^{N} p_j = 1$.

To achieve the antenna diversity gain, the transmit information is spread across all the transmit antennas using the OSTBC while at the receiver, the received signal is decoded by the maximum-likelihood (ML) decoder and then combined by using a maximum ratio combiner. With the ML decoder, the different transmit symbols can be decoded separately [8, Chapter 4]. Define the $1 \times T$ vector $\mathbf{r}_i$, $i = 1, 2, \cdots M$ as the signal vector received at the ith receive antenna. Considering only the ith receive antenna, for each index k, the transmitted signal s_k can be decoded from the vector product of $\mathbf{r}_i^{'}$ and $\mathbf{\Omega}_i$, where the $1 \times T$ vector $\mathbf{r}_i^{'}$ and the $T \times 1$ vector $\mathbf{\Omega}_i$ are defined as follows. For the k-th symbol, the t-th element of the $1 \times T$ vector $\mathbf{r}_i^{'}$ is defined by [8, eq.(4.138)]

$$\mathbf{r}_i^{'}(t) = \begin{cases} \mathbf{r}_i^*(t), & \text{if } s_k^* \text{ or } -s_k^* \text{ exists in the } t\text{-th column of } \mathbf{C} \\ \mathbf{r}_i(t), & \text{otherwise.} \end{cases} \tag{4.2}$$

For the k-th symbol, the t-th element of the $T \times 1$ vector $\mathbf{\Omega}_i$ is defined by [8, eq.(4.139)]

$$\mathbf{\Omega}_i(t) = \begin{cases} h_{ij}\sqrt{p_j}, & \text{if } c_{jt} = s_k \\ h_{ij}^*\sqrt{p_j}, & \text{if } c_{jt} = s_k^* \\ -h_{ij}\sqrt{p_j}, & \text{if } c_{jt} = -s_k \\ -h_{ij}^*\sqrt{p_j}, & \text{if } c_{jt} = -s_k^*. \end{cases} \tag{4.3}$$

By using maximum ratio combining, the symbol s_k can be detected by

$$\sum_{i=1}^{M} \mathbf{r}_i^{'}\mathbf{\Omega}_i = \sum_{j=1}^{N}\sum_{i=1}^{M} s_k|\sqrt{p_j}h_{ij}|^2 + \sum_{i=1}^{M} N_i, \tag{4.4}$$

where N_i is an i.i.d. zero-mean complex Gaussian random variable with variance equal to $N_0 \sum_{j=1}^{N} p_j|h_{ij}|^2$. Therefore, the instantaneous *post-detection SNR* of the diversity system can then be expressed as

$$\gamma_e = \frac{1}{N_0}\sum_{i=1}^{M}\sum_{j=1}^{N} p_j|h_{ij}|^2. \tag{4.5}$$

The conditional mutual information for a given power vector $\mathbf{p} = (p_1, \ldots, p_N)$ and channel realization $\mathbf{H} = (\mathbf{h}_1, \ldots, \mathbf{h}_N)$ is given by [83]

$$\mathcal{I}(\mathbf{p}, \mathbf{H}) = \log_2 (1 + \gamma_e). \tag{4.6}$$

4.2 Ergodic Mutual Information and Power Allocation

4.2.1 Ergodic mutual information analysis

The ergodic mutual information of the considered system is given by

$$\overline{\mathcal{I}}(\mathbf{p}) = \mathbb{E}[\log_2(1 + \gamma_e)], \tag{4.7}$$

where the expectation $\mathbb{E}[\cdot]$ is taken over all realizations of $\mathbf{H}$. From (4.5) it is clear that γ_e is a weighted sum of NM independent and normalized exponential random variables with weights given by $p_j\sigma_{ij}^2/N_0$. Its PDF can thus be expressed as [96, (1)]

$$p(\gamma_e) = \sum_k \frac{-N_0 e^{-x/\lambda_k}}{(-\lambda_k)^{m_k}} \frac{g_k^{(m_k-1)}(0, x)}{(m_k - 1)!}\Bigg|_{x=N_0\gamma_e} \tag{4.8}$$

where $\{\lambda_k\}_{k=0}^{N'-1}$ represent the $N'(\leq\ NM)$ distinct values of $p_j\sigma_{ij}^2$, each with multiplicity m_k, as a result of reordering and grouping the NM values of $p_j\sigma_{ij}^2$; α_{kl} is defined as $\alpha_{kl} = 1 - \lambda_l/\lambda_k$; and $g_k^{(m_k-1)}(s, x)$ denotes the $(m_k - 1)$-th derivative of $g_k(s, x)$ given in (4.9) with respect to s.

$$g_k(s, x) = e^{-sx} \prod_{l \neq k} \frac{1}{(\alpha_{kl} - \lambda_l s)^{m_l}}. \tag{4.9}$$

In the case where $p_j\sigma_{ij}^2$'s are all distinct (i.e., i.n.d channels), (4.8) can be reduced to

$$p(\gamma_e) = N_0 \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{B_{ij}}{p_j\sigma_{ij}^2} \exp\left(-\frac{N_0\gamma_e}{p_j\sigma_{ij}^2}\right), \tag{4.10}$$

where

$$B_{ij} = \prod_{\{m,n\} \neq \{i,j\}} \frac{p_j\sigma_{ij}^2}{p_j\sigma_{ij}^2 - p_n\sigma_{mn}^2}. \tag{4.11}$$

Applying (4.10) to (4.7) and using [97, (4.331.2)], we obtain the closed-form expression of the ergodic mutual information over i.n.d channels as

$$\overline{\mathcal{I}}(\mathbf{p}) = \frac{1}{\ln 2} \sum_{i=1}^{M} \sum_{j=1}^{N} B_{ij} \exp\left(\frac{N_0}{p_j \sigma_{ij}^2}\right) E_1\left(\frac{N_0}{p_j \sigma_{ij}^2}\right), \tag{4.12}$$

where $E_1(\cdot)$ is the exponential integral function defined as $E_1(x) = \int_x^\infty e^{-t}/t dt$, for $x > 0$.

In the following, we study the power allocation $\mathbf{p}$ that maximizes the ergodic mutual information $\overline{\mathcal{I}}(\mathbf{p})$. The expression of B_{ij} in (4.11) makes it difficult to maximize $\overline{\mathcal{I}}(\mathbf{p})$ with respect to $\mathbf{p}$ directly. To make the problem more tractable, we only consider the MISO case where only one receive antenna is used. First, we consider the simplest case where there are only two transmit antennas $(N = 2)$. After that, we propose a suboptimal power allocation scheme for $N > 2$ transmit antennas.

4.2.2 Power Allocation for Two-Transmit One-Receive Antenna Systems

For the sake of brevity, in the rest of this section, we omit the receive antenna subindex i in both B_{ij} and σ_{ij}^2 as only $M = 1$ is considered. For a MISO diversity system with two transmit antennas, the problem of maximizing $\overline{\mathcal{I}}(\mathbf{p})$ is equivalent to maximizing the following function:

$$\max_{p_1} \left\{ B_1 e^{\frac{N_0}{p_1 \sigma_1^2}} E_1\left(\frac{N_0}{p_1 \sigma_1^2}\right) + B_2 e^{\frac{N_0}{(1-p_1)\sigma_2^2}} E_1\left(\frac{N_0}{(1-p_1)\sigma_2^2}\right) \right\}. \tag{4.13}$$

Next we prove that the second derivative of the objective function in (4.13) with respect to p_1 is non-positive for $0 \leq p_1 \leq 1$. Our proof shows that the optimization problem is convex. Hence, the optimization can be solved by letting its first derivative with respect to p_1 be zero.

Proof. We first derive the first derivative of objective function in (4.13) with respect to p_1. After that we show that its second derivative with respect to p_1 is non-positive for $0 \leq p_1 \leq 1$. Based on the identity: $E_1(x) = -E_i(-x) = -\int_{-\infty}^{-x} \frac{e^t}{t} dt, \quad x > 0$, let $A = B_1 \exp(N_0/(p_1\sigma_1^2))$, $B = -E_i(-N_0/(p_1\sigma_1^2))$, $C = B_2 \exp(N_0/((1-p_1)\sigma_2^2))$ and $D = -E_i(-N_0/((1-p_1)\sigma_2^2))$. Taking the first derivative of A, B, C and D with respect to p_1, respectively, we get

$$
\begin{aligned}
A' &= \frac{-\sigma_1^2\sigma_2^2 e^{N_0/(p_1\sigma_1^2)} p_1 - N_0 e^{N_0/(p_1\sigma_1^2)}(p_1\sigma_1^2 - (1-p_1)\sigma_2^2)}{p_1(p_1\sigma_1^2 - (1-p_1)\sigma_2^2)^2}, \\
B' &= e^{-N_0/(p_1\sigma_1^2)}/p_1, \\
C' &= \frac{e^{N_0/((1-p_1)\sigma_2^2)}((1-p_1)\sigma_1^2\sigma_2^2 + N_0((1-p_1)\sigma_2^2 - p_1\sigma_1^2))}{((1-p_1)\sigma_2^2 - p_1\sigma_1^2)^2(1-p_1)}, \\
D' &= -e^{-N_0/((1-p_1)\sigma_2^2)}/(1-p_1).
\end{aligned}
$$

Note that $\lim_{x\to 0} -xE_i(-x) = 0$ and $\lim_{N_0\to 0} E_1(N_0/(1-p_1)\sigma_2^2) - E_1(N_0/p_1\sigma_1^2) = \ln \frac{(1-p_1)\sigma_2^2}{p_1\sigma_1^2}$. When assuming the SNR is very large (i.e. $N_0 \to 0$), the first derivative of the objective function in (4.13) (i.e.$A'B + AB' + C'D + CD'$) becomes

$$
\frac{\sigma_1^2\sigma_2^2\left(\ln \frac{(1-p_1)\sigma_2^2}{p_1\sigma_1^2}\right)}{((1-p_1)\sigma_2^2 - p_1\sigma_1^2)^2} - \frac{\sigma_1^2 + \sigma_2^2}{(1-p_1)\sigma_2^2 - p_1\sigma_1^2}. \tag{4.14}
$$

By letting (4.14) be zero, after simple manipulations, equation (4.20) can be obtained.

Having the first derivative, we can obtain the second derivative as follows. Differentiating (4.14) with respect to p_1, and multiplying the result by a positive term $\{[(1-p_1)\sigma_2^2 - p_1\sigma_1^2]^2 p_1(1-p_1)\}^{-1}$, we obtain

$$
y = -\sigma_1^2\sigma_2^2 + \frac{2\sigma_1^2\sigma_2^2(\sigma_1^2+\sigma_2^2)p_1(1-p_1)}{(1-p_1)\sigma_2^2 - p_1\sigma_1^2} \ln \frac{(1-p_1)\sigma_2^2}{p_1\sigma_1^2} - p_1(1-p_1)(\sigma_1^2+\sigma_2^2)^2. \tag{4.15}
$$

Without loss of generality, by assuming $\xi = \sigma_1^2/\sigma_2^2 \geq 1$ and the sum of channel variances to be N_t (in this case $N_t = 2$), one has $\sigma_1^2 = 2\xi/(\xi+1)$ and $\sigma_2^2 = 2/(\xi+1)$. Hence, (4.15) can be expressed by a function of ξ and p_1 shown as

$$
y = \frac{-4\xi}{(1+\xi)^2} + \frac{4 \times 2\xi p_1(1-p_1)}{(\xi+1)(1-p_1-\xi p_1)} \ln \frac{1-p_1}{\xi p_1} - 4p_1(1-p_1). \tag{4.16}
$$

Now we need to show that y is non-positive for $0 \leq p_1 \leq 1$. In order to do so, we consider three different cases. First, in the case of $x = (1-p_1)/(\xi p_1) = 1$, the term $\frac{1}{1-p_1-\xi p_1} \ln \frac{1-p_1}{\xi p_1}$ as part of the second term in (4.16) can be rewritten as $\frac{\ln x}{\xi p_1 x - \xi p_1}$. By applying L'Hospital rule, it is easy to show that $y = 0$. In the second case of $x = (1-p_1)/(\xi p_1) > 1$, we substitute $p_1 = 1/(x\xi + 1)$ into (4.16) and get

$$y = \frac{-4\xi}{(\xi+1)^2} + \frac{4 \times 2\xi x \ln(x)}{(\xi+1)(x-1)(\xi x+1)} - \frac{4\xi x}{(\xi x+1)^2}. \tag{4.17}$$

Since $x > 1$, multiplying the two sides of (4.17) by a positive term $(1+\xi)(\xi x+1)(x-1)/(4\xi x)$, we get

$$y_1 = \frac{-(\xi x+1)(x-1)}{(\xi+1)x} - \frac{(x-1)(\xi+1)}{\xi x+1} + 2\ln x. \tag{4.18}$$

By differentiating (4.18) with respect to x and multiplying the result by positive term $x^2(1+\xi)(\xi x+1)^2$, y_1 comes to

$$y_2 = -(x-1)^2(x^2\xi^3+1). \tag{4.19}$$

By using the property of the quartic function in (4.19), we can show that $y_2 \leq 0$ for $\xi \geq 1$. Hence, (4.18) is a monotonic decreasing function of x and $y_1 = 0$ can be shown to be the global maximum at point $x = 1$. For the third case of $0 \leq x < 1$, the same method used here is applicable to show $y \leq 0$. □

Since the optimization problem has been shown to be convex, the optimization can be done by letting its first derivative with respect to p_1 be zero. Note that when there does not exist such a $p_1 \in [0,\ 1]$ that makes the first derivative zero, the objective function degenerates to a monotonic function of p_1. In other words, one of the two ends of the range of p_1 should be the optimum value. Consequently, one antenna should be turned off. Hence, two different cases regarding power allocation are analyzed.

In the first case, both antennas are active. Assume that the power on the first antenna is $0 < p_1 < 1$, while on the second antenna it is $p_2 = 1 - p_1$. Taking the first

derivative of objective function (4.13) with respect to p_1 and equating it to zero, after applying high SNR assumption ($N_0 \to 0$), we obtain

$$\ln \frac{(1-p_1)\sigma_2^2}{p_1\sigma_1^2} = \frac{(\sigma_1^2+\sigma_2^2)(\sigma_2^2 - p_1(\sigma_2^2+\sigma_1^2))}{\sigma_1^2\sigma_2^2}. \tag{4.20}$$

Though the solution to (4.20) cannot be obtained in closed form, we can still find the optimum p_1 by finding the intersection point of the two curves specified by the left and right sides of (4.20), respectively. Next, we prove that there always exists one tangent point and at most one intersection point for the two curves.

Proof. First we let $f_1(p_1) = \ln((1-p_1)\sigma_2^2/p_1\sigma_1^2)$ and $f_2(p_1) = (\sigma_1^2+\sigma_2^2)(\sigma_2^2 - p_1(\sigma_2^2+\sigma_1^2))/\sigma_1^2\sigma_2^2$, respectively. We can show that when $p_1 > 1/2$, the function $f_1(p_1)$ is convex while when $p_1 < 1/2$, function $f_1(p_1)$ is concave. Since $f_2(p_1)$ is a linear function, there are at most three values of p_1 which can make $f_1(p_1) = f_2(p_1)$. Since there is one tangent point between $f_1(p_1)$ and $f_2(p_1)$, two of the three values are always equal. This tangent point between the two curves lies at $p_1 = \sigma_2^2/(\sigma_1^2+\sigma_2^2)$ since

$$f_i'\left(p_1 = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right) = -\frac{(\sigma_1^2+\sigma_2^2)^2}{\sigma_1^2\sigma_2^2} \quad i = 1, 2, \tag{4.21}$$

and the equality $f_1(p_1) = f_2(p_1)$ holds when the value of p_1 satisfies $p_1 = \sigma_2^2/(\sigma_1^2+\sigma_2^2)$. However, the tangent point $p_1 = \sigma_2^2/(\sigma_1^2+\sigma_2^2)$ is not the valid solution since the equation $p_1\sigma_1^2 = p_2\sigma_2^2$ holds when p_1 satisfies $p_1 = \sigma_2^2/(\sigma_1^2+\sigma_2^2)$, which conflicts with the assumption made when obtaining (4.10). □

Therefore, the tangent point is $p_1 = \sigma_2^2/(\sigma_1^2+\sigma_2^2)$, but it is not the optimal solution, since it violates the assumption of distinct distribution, i.e., $p_1\sigma_1^2 \neq p_2\sigma_2^2$, made in (4.10). By inspection, we can see that $p_1 > 1/2$ when $\sigma_1^2 > \sigma_2^2$, and p_1 increases when the ratio $\xi_{12} = \sigma_1^2/\sigma_2^2$ increases. When ξ_{12} is large enough, there does not exist any intersection point anymore. In other words, a valid solution $0 < p_1 < 1$ to (4.20) does

not exist for highly unbalanced channels. This leads to the second case where only one antenna is active.

When only one antenna is active, the problem of maximizing $\overline{\mathcal{I}}(\mathbf{p})$ is equivalent to maximizing the function below:

$$\max_{j\in\{1,2\}} \left\{ \overline{\mathcal{I}}(p_j\sigma_j^2) = e^{\frac{N_0}{p_j\sigma_j^2}} E_1\left(\frac{N_0}{p_j\sigma_j^2}\right) \right\}. \tag{4.22}$$

Using the monotonic increasing property of $\overline{\mathcal{I}}(p_j\sigma_j^2)$, it is clear that the total power should be assigned to the antenna with larger channel variance.

Based on the discussion above, the asymptotic power allocation scheme at high SNR only depends on the channel ratio $\xi_{12} = \sigma_1^2/\sigma_2^2$. Without loss of generality, we assume $\sigma_1^2 \geq \sigma_2^2$. In Fig. 4.1, we plot the numerical solution of p_1 as a function of ξ_{12} to (4.20). It is seen that p_1 can be well approximated by

$$p_1 = f(\xi_{12}) = \begin{cases} 1 - \frac{1}{2}\exp\left(-\xi_{12}+1\right), & 1 \leq \xi_{12} \leq \xi_T \\ 1, & \xi_{12} > \xi_T. \end{cases} \tag{4.23}$$

where the ratio threshold ξ_T can be chosen larger than 10.

4.2.3 Power Allocation for Multiple-Transmit One-Receive Antenna Systems

For general multiple-transmit antenna systems, it is difficult to directly optimize (4.12) with respect to $\mathbf{p}$. Motivated by the results obtained for two-transmit antenna systems, we propose a simple power allocation scheme which can provide near-optimal performance. It is assumed without loss of generality that $\sigma_1^2 \geq \sigma_2^2 \cdots \geq \sigma_N^2$. The power on each antenna is assigned sequentially. In other words, p_1 is computed first, followed by p_2, then p_3, and so on, until p_N. At stage j for computing p_j, the other antennas $j+1, \ldots, N$ that are not assigned powers yet are grouped together as an antenna subset. Define $\sigma_{j+1}^{2(\mathrm{e})}$ as the equivalent channel gain of the antenna subset,

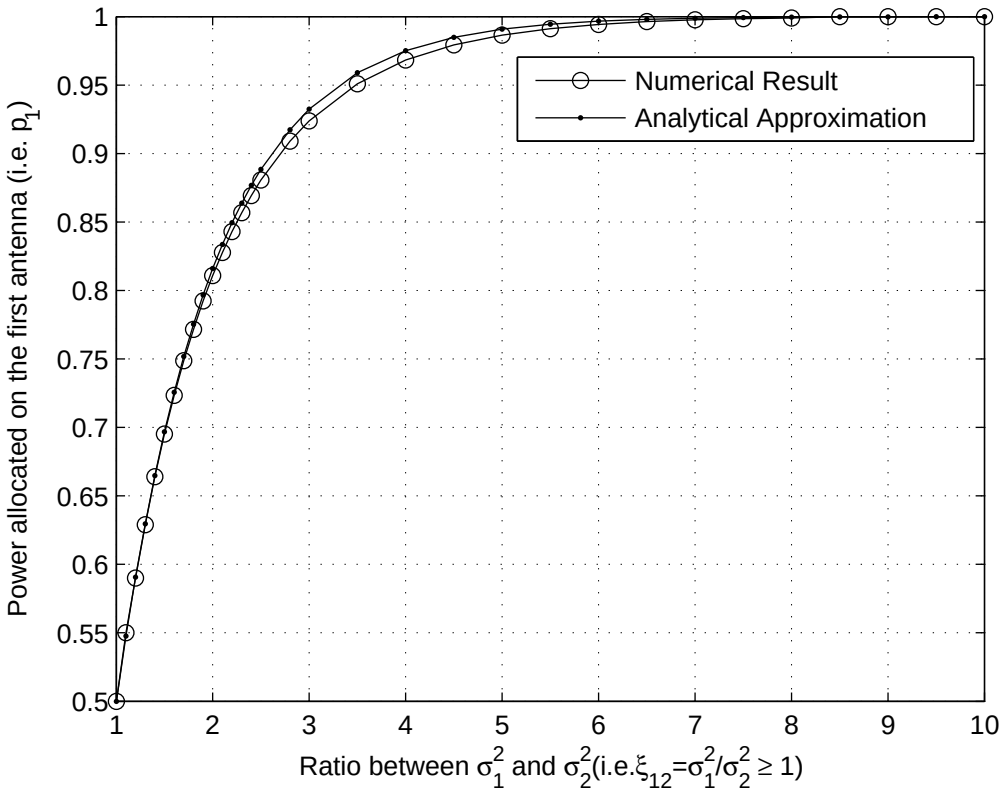

Figure 4.1: Power functions for a diversity system with two transmit antennas and one receive antenna

whose exact definition in terms of $\left\{\sigma_{j+1}^2, \cdots, \sigma_N^2\right\}$ is to be given shortly. Using (4.23), a proposed sub-optimal power allocation scheme is

$$p_j = \begin{cases} \left(1 - \sum_{k=1}^{j-1} p_k\right) f\left(\sigma_j^2/\sigma_{j+1}^{2(\mathrm{e})}\right), & \sigma_j^2 \geq \sigma_{j+1}^{2(\mathrm{e})} \\ \left(1 - \sum_{k=1}^{j-1} p_k\right) \left[1 - f\left(\sigma_{j+1}^{2(\mathrm{e})}/\sigma_j^2\right)\right], & \sigma_{j+1}^{2(\mathrm{e})} > \sigma_j^2 \end{cases}, \tag{4.24}$$

for $j = 1, \cdots, N$.

We now propose an efficient method to obtain $\sigma_{j+1}^{2(\mathrm{e})}$. First we compare each channel gain of $\{\sigma_{j+1}^2, \cdots, \sigma_N^2\}$ with σ_j^2 and form the set $\mathbf{S}_{j+1}$ in which each channel gain is larger than σ_j^2/ξ_T, i.e.,

$$\left\{\sigma_k^2 \in \mathbf{S}_{j+1} \left| \sigma_j^2/\sigma_k^2 \leq \xi_T, k \in [j+1, N] \right.\right\}. \tag{4.25}$$

Assume that there are $K_{j+1} \in [0, N-j]$ elements in $\mathbf{S}_{j+1}$. We simply regard those K_{j+1} associated antennas as one single antenna and discard the remaining antennas whose channel variances are small enough compared with σ_j^2. Then, we define the equivalent channel gain $\sigma_{j+1}^{2(\mathrm{e})}$ as the norm of the vector $[\sigma_{j+1}^2, \sigma_{j+2}^2, \cdots, \sigma_{j+K_{j+1}}^2]$, i.e.

$$\sigma_{j+1}^{2(\mathrm{e})} = \sqrt{\sum_{k=j+1}^{j+K_{j+1}} (\sigma_k^2)^2}. \tag{4.26}$$

We shall show in Section 4.4 that the proposed equivalent channel gain in (4.26) together with (4.24) provides a near capacity-achieving performance for both slightly unbalanced and highly unbalanced channels

4.3 Information Outage Probability and Power Allocation

Given the instantaneous mutual information $\mathcal{I}(\mathbf{p}, \mathbf{H})$ defined in (4.6) and an outage mutual information $\mathcal{I}_{out}$, the information outage probability is defined as

$$P_{out}(\mathbf{p}) = P\left(\mathcal{I}(\mathbf{p}, \mathbf{H}) < \mathcal{I}_{out}\right) = P(\gamma_e < \gamma_{out}), \tag{4.27}$$

where $\gamma_{out} = 2^{\mathcal{I}_{out}} - 1$. Hence, the outage probability is the same as the cumulative distribution function (CDF) of γ_e, which is expressed as [96, (32)]

$$P_{out}(\mathbf{p}) = 1 + \sum_k \left. \frac{e^{-x/\lambda_k} \hat{g}_k^{(m_k-1)}(0, x)}{(-\lambda_k)^{m_k}(m_k-1)!} \right|_{x=\gamma_{out}}. \tag{4.28}$$

Here $\hat{g}_k(s, x)$ is given by

$$\hat{g}_k(s, x) = -\lambda_k e^{-sx} \prod_l \frac{1}{(\alpha_{kl} - \lambda_l s)^{m_{kl}}}, \tag{4.29}$$

with $\alpha_{kk} \triangleq -1$, $m_{kl} = m_l$ for $l \neq k$, and $m_{kk} = 1$. In the case where $p_j \sigma_{ij}^2$'s are all distinct (i.e., i.n.d channels), the outage probability in (4.28) can be simplified to

$$P_{out}(\mathbf{p}) = \sum_{i=1}^{M} \sum_{j=1}^{N} B_{ij} \left(1 - \exp\left(-\frac{N_0 \gamma_{out}}{p_j \sigma_{ij}^2}\right)\right) \tag{4.30}$$

which is consistent with [41, (2)]. In [98], the authors derived the outage probability for DF cooperative communications. When assuming the source-relay link to be error free, we can verify that [98, (10a)] reduces to (4.30) with $M = 1$.

In the following, we derive a suboptimal power allocation scheme that can minimize an upper bound on the outage probability at any given $\mathcal{I}_{out}$. By applying the Chernoff bound, the outage probability in (4.27) can be upper-bounded by

$$P_{out}(\mathbf{p}) \leq \mathbb{E}[e^{u(\gamma_{out}-\gamma_e)}] = e^{u\gamma_{out}} \prod_{i=1}^{M}\prod_{j=1}^{N} \frac{1}{1+up_j\sigma_{ij}^2/N_0}, \tag{4.31}$$

where u is a non-negative constant that can be chosen to optimize the tightness of the bound. Nevertheless, we choose $u = NM/\gamma_{out}$, for simplicity. This bound can be minimized with respect to p_j's by maximizing the objective function

$$\varphi(p_j) = \prod_{i=1}^{M}\prod_{j=1}^{N} \left(1+up_j\sigma_{ij}^2/N_0\right). \tag{4.32}$$

By applying the inequality: $(1+\sqrt[M]{\prod_{i=1}^{M} x_i})^M \leq \prod_{i=1}^{M}(1+x_i)$ [99, (25)] to (4.32), it can be lower-bounded by

$$\varphi(p_j) \geq \prod_{j=1}^{N}\left(1+\frac{up_j}{N_0}\sqrt[M]{\prod_{i=1}^{M}\sigma_{ij}^2}\right)^M. \tag{4.33}$$

By taking the logarithm of (4.33), maximizing (4.33) is equivalent to minimizing

$$\begin{aligned} &\min -M\sum_{j=1}^{N}\log\left(1+\frac{up_j}{N_0}\sqrt[M]{\prod_{i=1}^{M}\sigma_{ij}^2}\right) \\ &\text{s.t.} \sum_{j=1}^{N} p_j = 1, \\ &p_j \geq 0 \end{aligned} \tag{4.34}$$

Using the Lagrange method, we define the Lagrangian as

$$\mathcal{L} = -M\sum_{j=1}^{N}\log\left(1+\frac{up_j}{N_0}\sqrt[M]{\prod_{i=1}^{M}\sigma_{ij}^2}\right) - \sum_{j=1}^{N} p_j\lambda_j + v\left(\sum_{j=1}^{N} p_j - 1\right) \tag{4.35}$$

where λ_j is the Lagrange multiplier associated with the jth inequality constraint, and v is the Lagrange multiplier associated with the equality constraint. Since $\mathcal{L}$ is a convex quadratic function of p_j, we can obtain the optimal p_j by taking the first order derivative of $\mathcal{L}$ with respect to p_j and letting it be zero. Hence we obtain the water-filling based suboptimal power allocation to be:

$$p_j = \left\{ \frac{1}{v} - \frac{N_0 \gamma_{out}}{NM \sqrt[M]{\prod_{i=1}^{M} \sigma_{ij}^2}} \right\}^+ , \tag{4.36}$$

where $\{a\}^+$ denotes $\max\{0, a\}$, and v is a constant determined by the constraint $\sum_{j=1}^{N} p_j = 1$. According to the properties of water-filling, at high transmit SNR, the power tends to be equally allocated among all transmit antennas, while at low SNR some of the antennas with the geometric mean of channel variances (i.e., $\sqrt[M]{\prod_{i=1}^{M} \sigma_{ij}^2}$) significantly lower than the others may be turned off. These conclusions with $M = 1$ are consistent with the heuristic power control scheme for MISO channels in [41].

4.4 Numerical Results

We consider the MIMO diversity systems over i.n.d channels. The i.n.d channels can be classified into highly or slightly unbalanced channels. As an example of the highly unbalanced channels, the distributed antenna system mentioned before is considered. Multiple directional antennas are placed in a distributed manner with sufficiently large spacing. Therefore, the channels seen by the receiver may experience significantly different propagation environments. Moreover, take the example of DF cooperative communications systems which have been given a simple description at the beginning of this chapter. In this scenario, the relay nodes can be placed in any location between the source and the destination. For a simply two-relay system, one relay node may

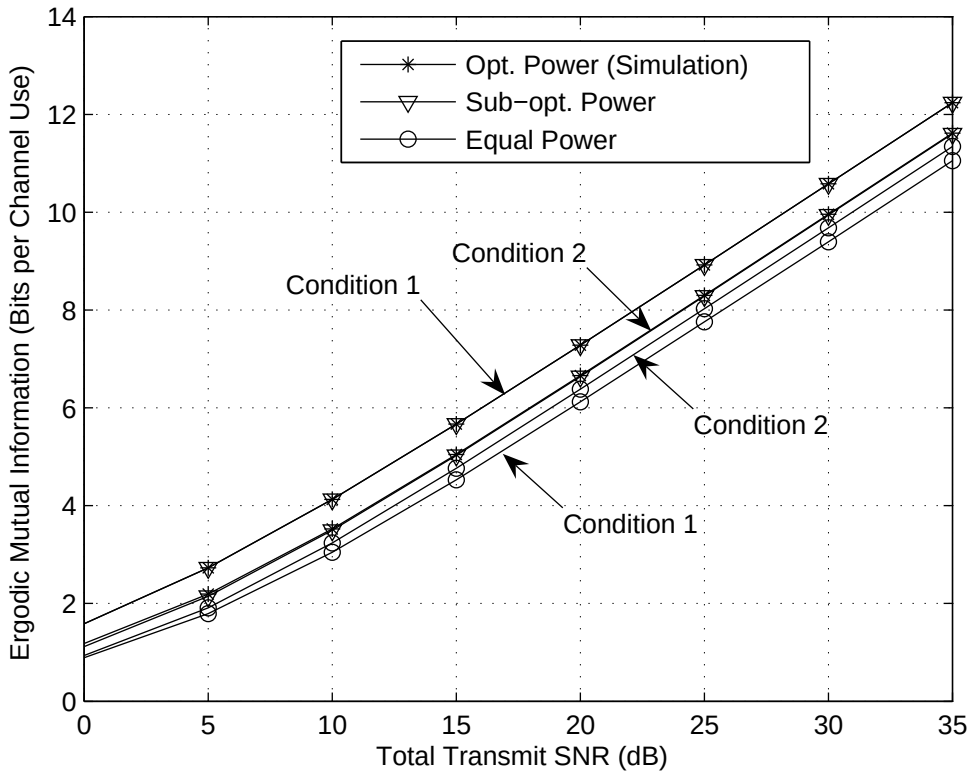

Figure 4.2: Ergodic mutual information with different power allocations

be close to the source, while the other may be midway between the source and the destination. Using the well-known path-loss model

$$PL = 10n \log_{10} d \tag{4.37}$$

where PL is the path loss in decibels, d is the distance between the transmitter and the receiver and n is the path loss exponent which is assumed to be 3 in the following discussion. Hence, the variance differences of the channels from the two relays seen by the destination is about 9-dB. In such a case, a pair of highly unbalanced channels is usually encountered. When the two relays are close to each other, the channels seen by relays are slightly unbalanced, which can be approximated by the i.i.d channel model. Hence, the i.n.d fading channel attracts greater attention than the i.i.d fading channel

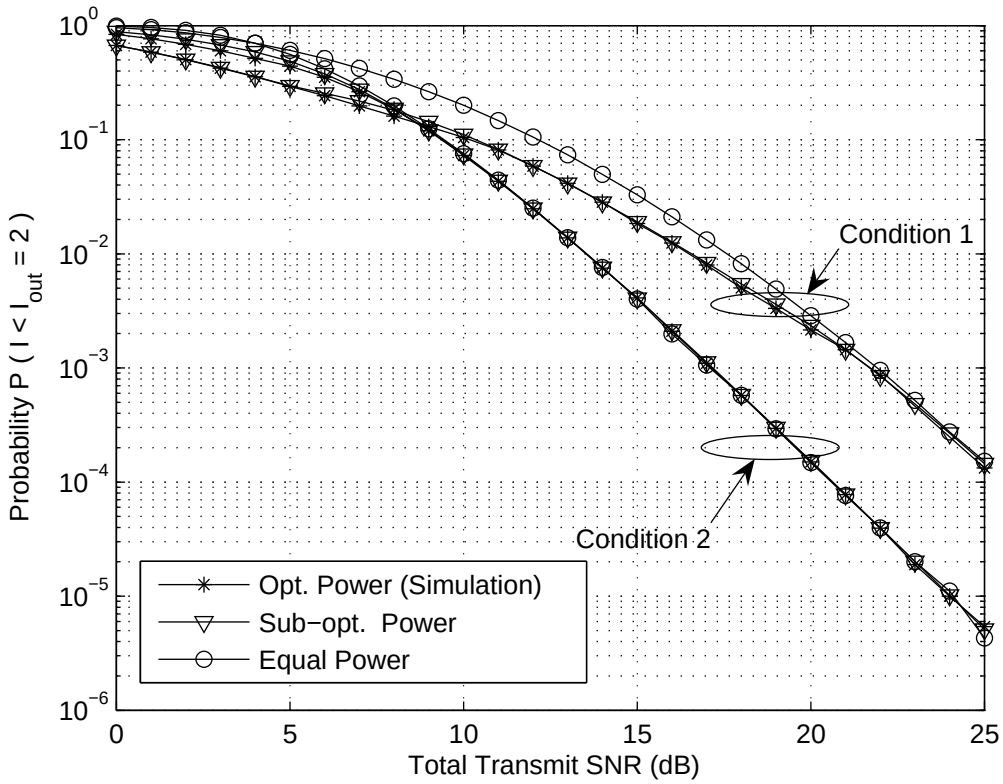

Figure 4.3: Outage probability with different power allocations.

since it occurs more naturally in practical MIMO systems

We first consider a system with three transmit antennas and one receive antenna. Both highly and slightly unbalanced channel conditions are discussed. For the highly unbalanced channel condition (condition 1), the parameters are $\sigma_{11}^2 = 300/111$, $\sigma_{12}^2 = 30/111$ and $\sigma_{13}^2 = 3/111$. The slightly unbalanced channel condition (condition 2) has parameters $\sigma_{11}^2 = 18/11$, $\sigma_{12}^2 = 9/11$ and $\sigma_{13}^2 = 6/11$. The sum of channel variances in both cases is equal to $N = 3$.

In Fig. 4.2, the ergodic mutual information using different power allocations is illustrated. The result of the optimal power allocation is obtained using a two-dimensional exhaustive search. It is seen that the proposed suboptimal

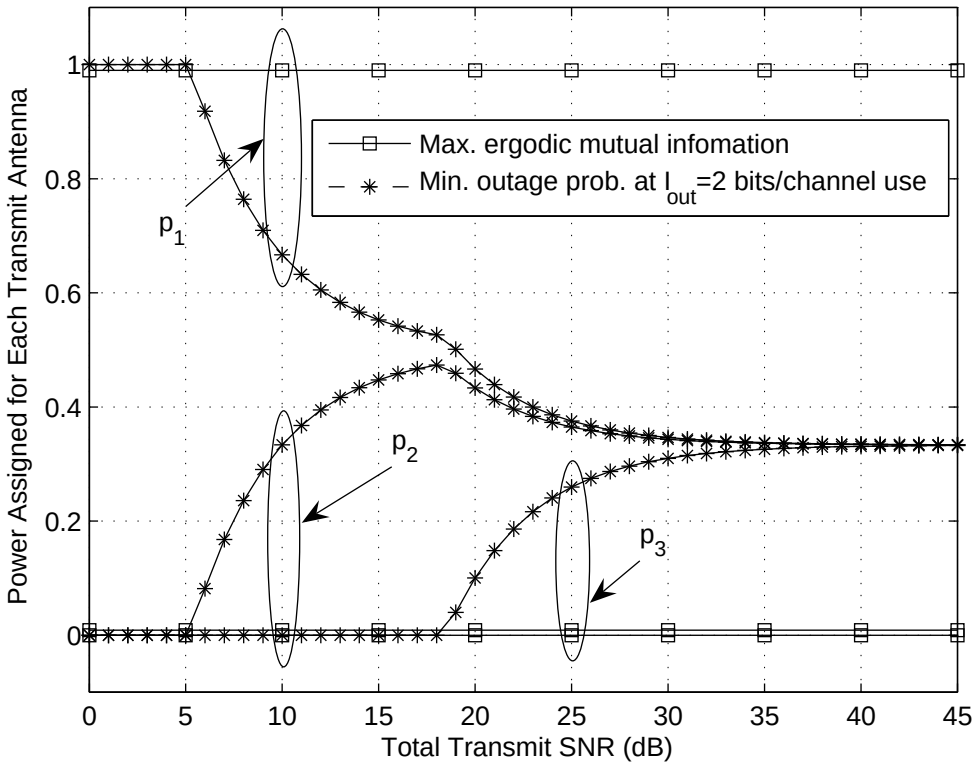

Figure 4.4: Power values using different criteria under channel condition 1.

scheme (4.24) performs almost the same as the optimal one and hence is near capacity-achieving. Also, 3-dB and 1-dB SNR gains are achieved over equal power allocation in the two channel conditions, respectively.

The outage probability for a given $\mathcal{I}_{out} = 2$ bits per channel use is shown in Fig. 4.3. We can see that the proposed power allocation (4.36) provides performance very close to the optimal scheme (via exhaustive search). The outage probability after using the proposed power allocation is smaller than that of equal power allocation. This improvement is more significant for highly unbalanced channels.

Furthermore, we compare in Figs. 4.4 and 4.5 the power values assigned to each antenna using the different criteria (4.24) and (4.36). It is clear that, for ergodic mutual

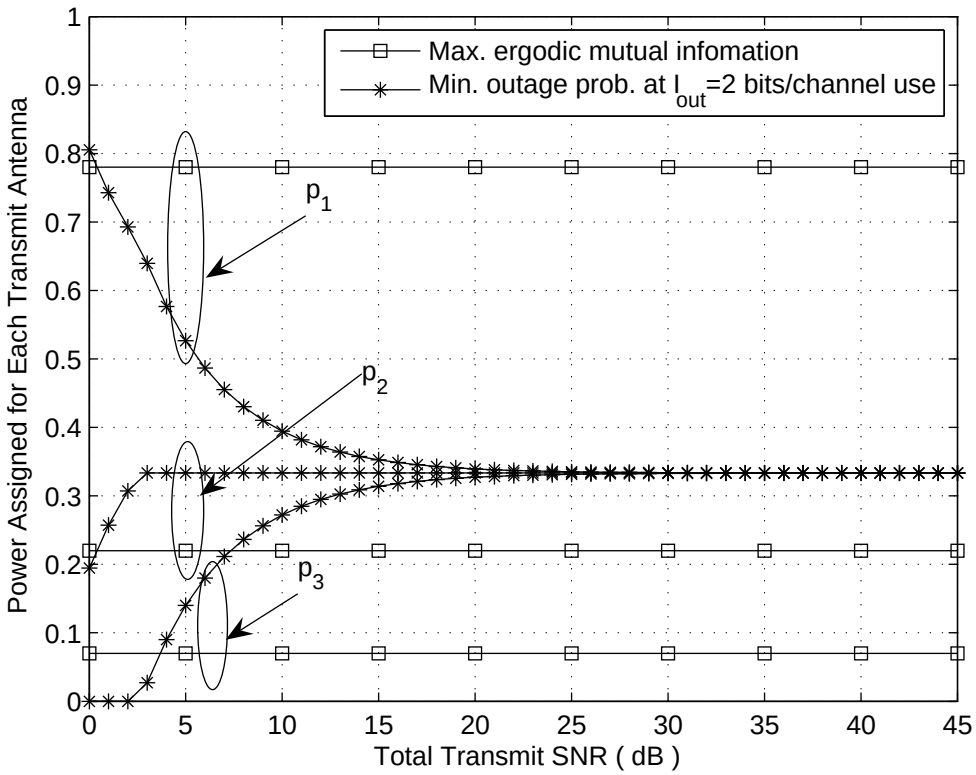

Figure 4.5: Power values using different criteria under channel condition 2.

information maximization, the power allocation is independent of the total transmit SNR and only depends on the channel variances. In particular, only one antenna is active in the highly unbalanced channel condition. For the slightly unbalanced channel condition, all the three antennas are active but more power is given to antennas with larger channel variances. For outage minimization, it follows the water-filling principle. When the total transmit SNR is high enough, all the antennas need to be active and the power tends to be equally allocated.

Next, we consider the outage probability in a system with three transmit antennas and two receive antennas using the proposed power allocation in (4.36). The channel variances are assumed to be $\sigma_{11}^2 = 300/111$, $\sigma_{12}^2 = 30/111$, $\sigma_{13}^2 = 3/111$, $\sigma_{21}^2 =$

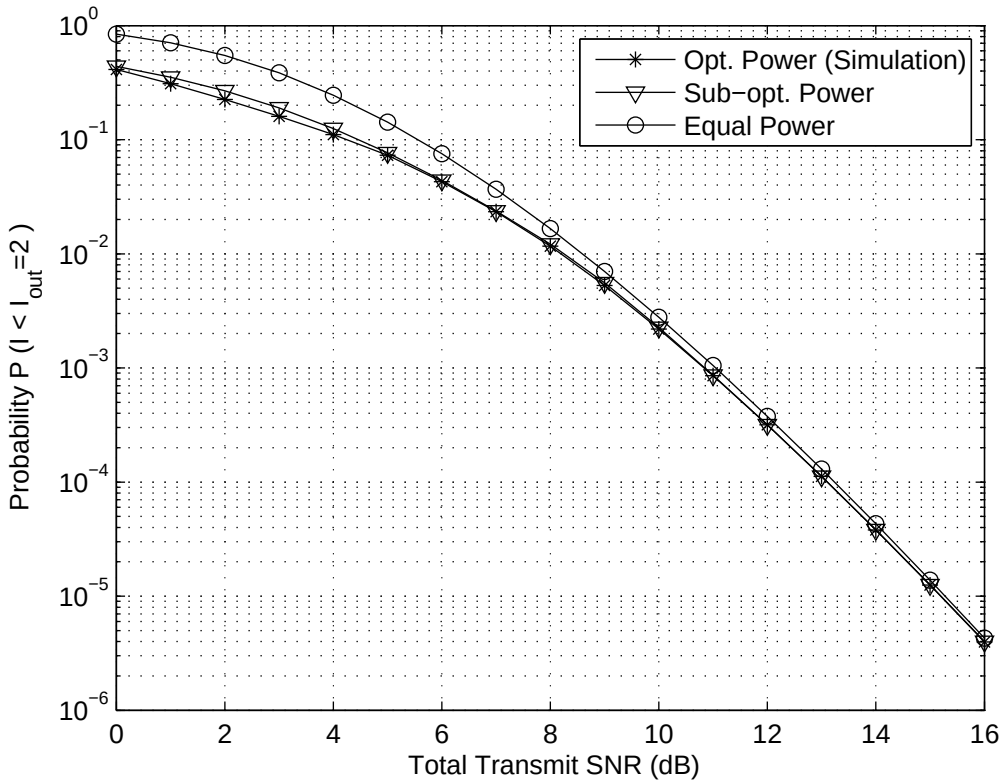

Figure 4.6: Outage probability with different power allocations.

$18/11$, $\sigma_{22}^2 = 9/11$ and $\sigma_{23}^2 = 6/11$, which sum to $NM = 6$. We can see from Fig. 4.6 that the proposed scheme in (4.36) provides a very close performance to optimal power allocation (obtained via multi-dimensional exhaustive search). It outperforms equal power allocation in low and moderate SNR regions.

4.5 An Application of Our Results

In this section, we consider an application of our results to the following communication scenario. For the non-identical fading channels, we consider distributed antenna deployment whose layout is sketched in Fig. 4.7. Here the three transmit ports $\{T1, T2, T3\}$ and receive port $R1$ are aligned on a straight line spaced

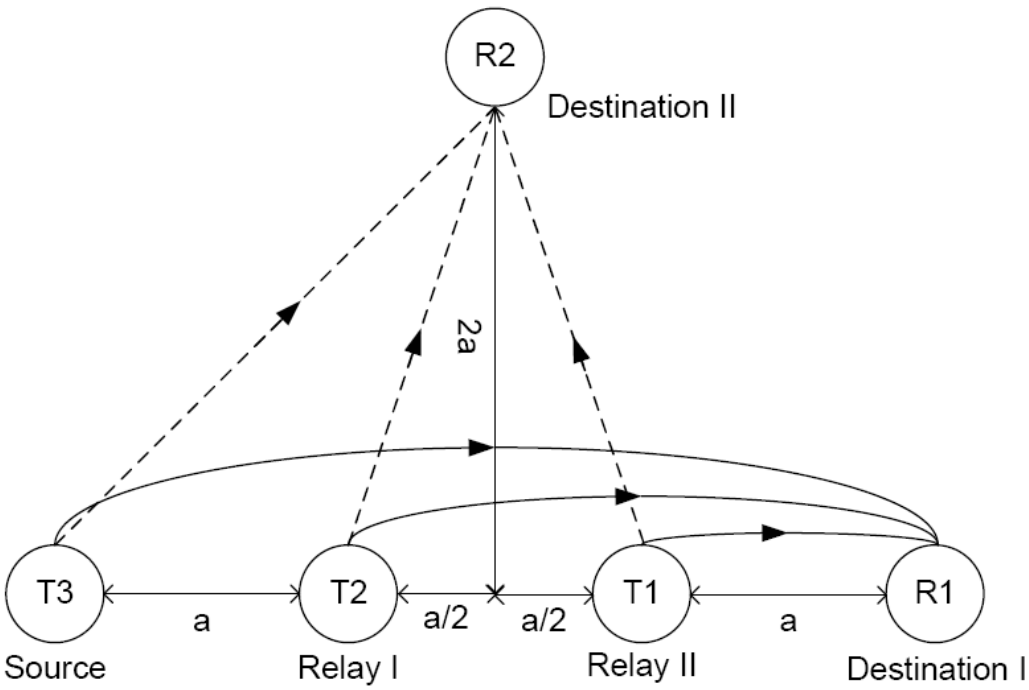

Figure 4.7: Wireless cooperative relay diversity system which can represent three different systems: system 1 (solid line), system 2 (dash line) and system 3 (solid line and dash line)

with equal distance, while receive port $R2$ is located at its perpendicular bisector. Suppose receiver $R1$ is active and receiver $R2$ is disabled. This scenario is referred to as system 1 and indicated by solid lines in the figure. By adopting the well-known path loss model in (4.37) and assuming the path loss exponent to be equal to 3, the channel variances can be obtained as $\sigma_1^2 = 648/251$, $\sigma_2^2 = 81/251$ and $\sigma_3^2 = 24/251$. Alternatively, in system 2, receiver $R1$ is disabled and receiver $R2$ is active, as denoted by dashed lines in Fig. 4.7. In this case, the channel variances become $\sigma_1^2 = 117/100$, $\sigma_2^2 = 117/100$ and $\sigma_3^2 = 66/100$. The sum of channel variances in both of the above systems is normalized to satisfy the constraint $NM = 3$ with $M = 1$. Either of the systems may be equivalent to a DF cooperating system with one source ($T3$), two

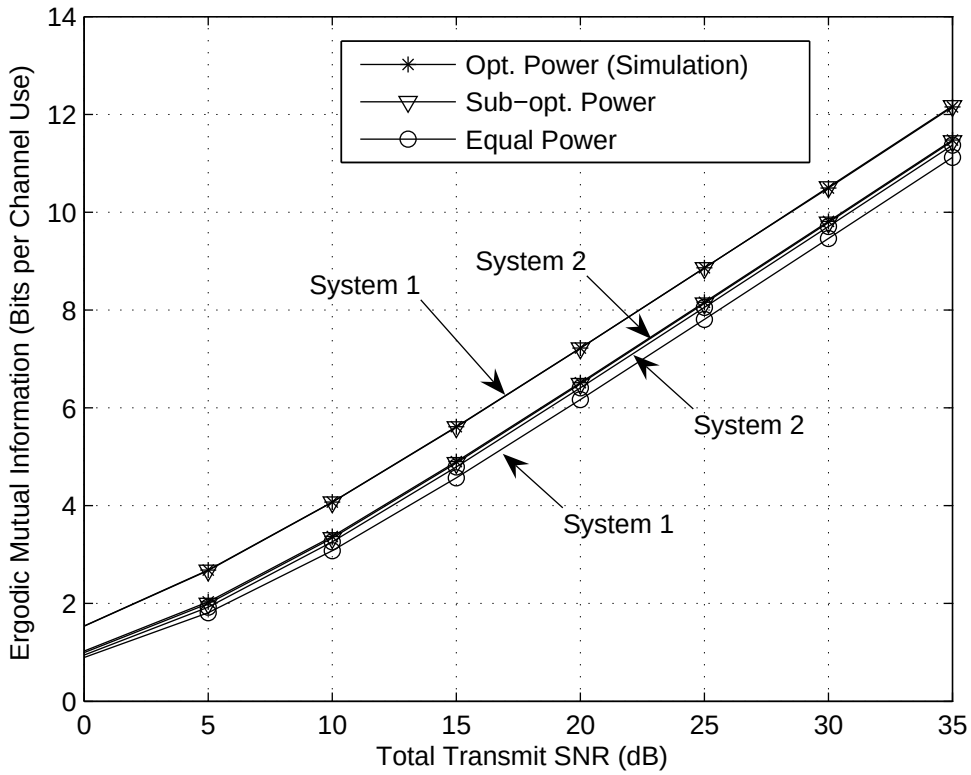

Figure 4.8: Ergodic mutual information using different power allocation in system 1 and 2 with $N = 3$ and $M = 1$.

relays ($T2$ and $T1$) and one destination ($R1$ or $R2$), by assuming error-free decoding at the relays. When both receivers are active, we obtain system 3, for which the channel gain parameters are obtained as $\sigma_{11}^2 = 648/251$, $\sigma_{12}^2 = 81/251$, $\sigma_{13}^2 = 24/251$, $\sigma_{21}^2 = 117/100$, $\sigma_{22}^2 = 117/100$ and $\sigma_{23}^2 = 66/100$, which sum to $NM = 6$.

To illustrate the results for ergodic mutual information, we consider only the 3-by-1 systems (system 1 and system 2). In Fig. 4.8, the ergodic mutual information using different power allocations is presented. The result of the optimal power allocation is obtained by using two-dimensional exhaustive search. It is seen that the proposed suboptimal scheme (4.24) performs almost the same as the optimal one and hence is near capacity-achieving. Also, 3-dB and 0.3-dB SNR gains are achieved over

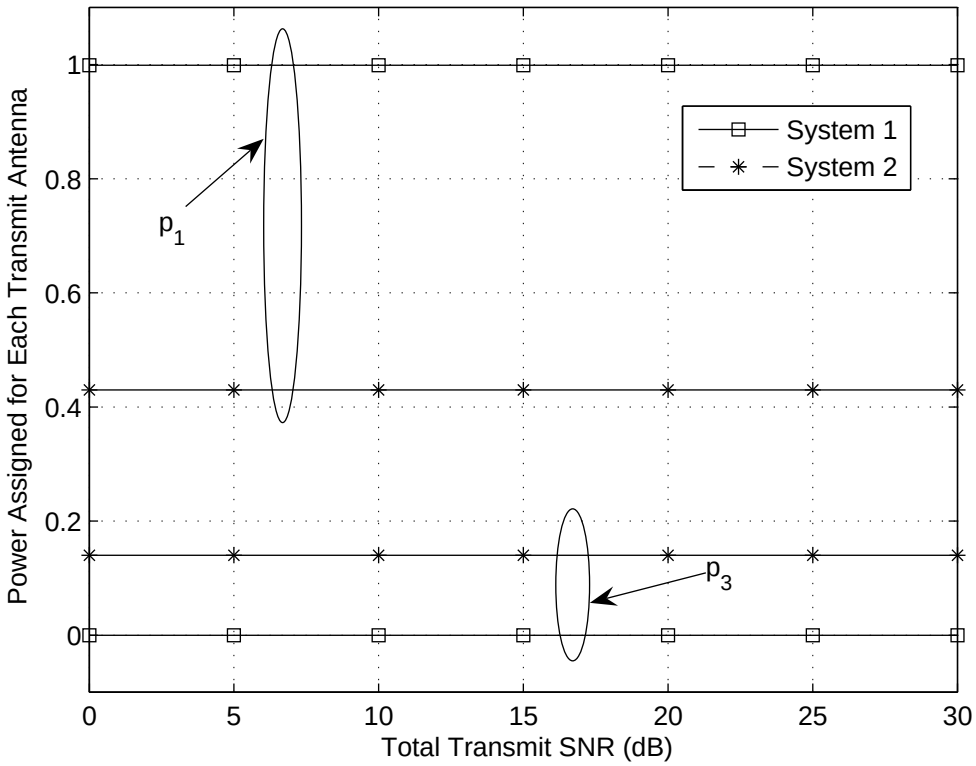

Figure 4.9: The power values assigned on relay II and source for maximizing the ergodic mutual information in system 1 and 2 with $N = 3$ and $M = 1$.

EPA in the two systems, respectively. Fig. 4.9 compares the power values assigned to $T1$ (p_1) and $T3$ (p_3) by using the criteria (4.24) (for simplicity p_2 is not shown but can be obtained straightforwardly). Note that the power values are all constant at different transmit SNR since the power allocation (4.24) only depends on the ratio of channel variances. The results show that for system 1 in the cooperative transmission scenario, only relay II is needed for forwarding signals without the cooperation from the source and relay I. This is expected as the channels are highly unbalanced. On the other hand, for the slightly unbalanced channels encountered in system 2, the source and the two relays all need to be active but more power is given to the node with larger channel gain.

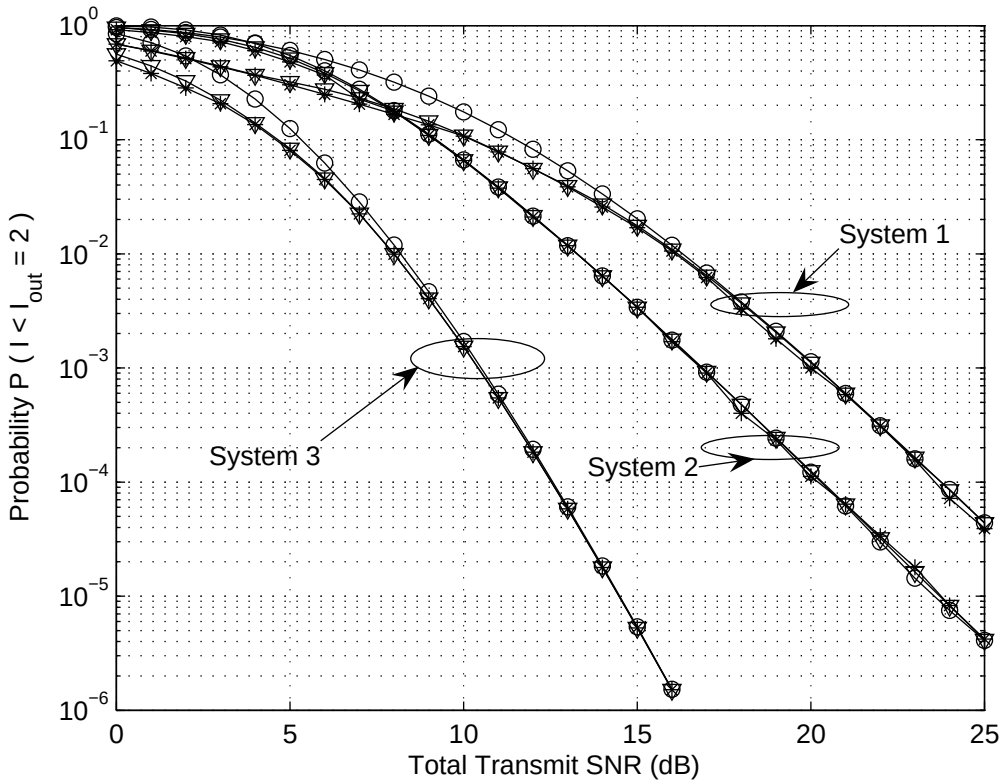

Figure 4.10: Outage probability with different power allocations in systems 1 and 2 with $N = 3$ and $M = 1$, and system 3 with $N = 3$ and $M = 2$.

The outage probability for a given $\mathcal{I}_{out} = 2$ bits per channel use is shown in Fig. 4.10. It is seen that the proposed power allocation (4.36) provides performance very close to the optimal scheme (via exhaustive search). The outage probability after using the proposed power allocation is smaller than that of EPA. This improvement is more significant in system 1 and system 3 in low and moderate SNR regions. As expected, the 3-by-2 system (system 3) achieves two times the diversity order of the other two systems.

In Fig. 4.11, the power values assigned to $T1$ and $T3$ by using the water-filling principle (4.36) are compared. An interesting finding is that for each given transmitter, the power value assigned to it in system 3 lies between the values assigned to it in

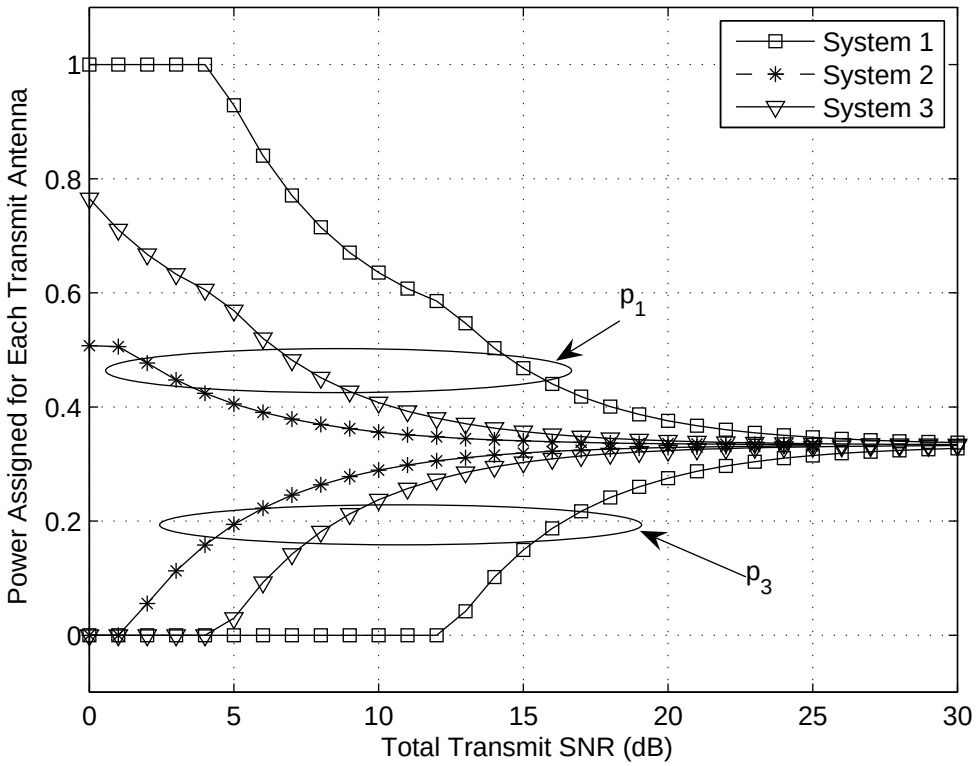

Figure 4.11: The power values assigned on relay II and source for minimizing the outage in systems 1 and 2 with $N = 3$ and $M = 1$, and system 3 with $N = 3$ and $M = 2$.

system 1 and system 2.

4.6 Conclusions

In this chapter, we analyzed the mutual information of MIMO diversity systems with i.n.d Rayleigh fading. Closed-form expressions for the ergodic mutual information and the outage probability over i.n.d channels are obtained with an arbitrary number of transmit and receive antennas. We then derived two near-optimal power allocation schemes for exploiting the non-identical channel statistics for ergodic mutual information maximization and information outage minimization, respectively.

4.6 Conclusions

With a single receive antenna, the power allocation scheme for maximizing the ergodic mutual information is particularly novel. It assigns more power to antennas with larger channel variances, and is independent of total transmit SNR. For minimizing the outage probability, we showed that the power allocation suitable for multiple receive antennas is dependent on the geometric mean of channel variances and follows the water-filling principle. Our analysis illustrates that the proposed power controls are beneficial in non-identical fading channels, especially when the channel variances are highly unbalanced.

Chapter 5

Performance of ARQ/HARQ Schemes With Imperfect CSIR Over Rayleigh Fading Channels

With imperfect CSI acquired by channel estimation at the receiver, the performance of basic ARQ and HARQ systems is studied as a function of the accuracy of channel estimation. The aim is to establish a link between data-link-layer performances and physical-layer parameters. The performance parameters we study in particular are the goodput, the APER and the drop rate, as a function of the channel estimation MSE and the factors which affect the MSE. Upper and lower bounds on the APER, the goodput as well as the drop rate are derived. These upper and lower bounds are close to one another, and therefore, enable the behavior of the exact performance parameters to be investigated. The precise dependence of the APER and the goodput on the channel estimation accuracy is quantified. The results show that the effect of the MSE on the system performance is nonlinear. For large MSE, the performance deteriorates very rapidly while at low values of the MSE, the performance improves gently toward that of the perfect CSI case. An attempt is made to study the inter-relationships among the

various relevant system performance parameters and the crucial dependence of these relationships on the CSI accuracy. The study here shows how accurate the CSI should be to achieve a specified performance level in ARQ schemes.

Differential BPSK (DBPSK) is a commonly used alternative modulation to BPSK, because it does not require additional pilot symbols for channel estimation for the acquisition of CSI. We extends the case of BPSK discussed in Section 5.3 to the case of DBPSK modulation in Section 5.5. The normalized fading correlation coefficient between two adjacent symbols is denoted by $\rho = R(1)/R(0)$. It is a measure of the fluctuation rate of the channel fading process. For the case of a static channel ($\rho = 1$), the exact closed-form expression of the performance will be derived. In addition to the case of a static channel, the effects of fading fluctuations in a time-varying channel ($\rho < 1$) on the performance are also examined. However, for the case of a slowly time-varying channel, the performance can only be studied via simulation in Section 5.5.3.

5.1 Introduction

ARQ error control strategies achieve high reliability by using an error-detecting code coupled with a packet retransmission scheme. There are three basic ARQ protocols: SW-ARQ, GBN-ARQ, and SR-ARQ [10, 11]. However, their throughputs fall rapidly with increasing channel error rate. Compared with the ARQ schemes, FEC schemes maintain a constant, high throughput regardless of the system reliability by using an error-correcting code. As a result, to achieve a higher reliability than an FEC system alone and a higher throughput than the system with ARQ only, a proper combination of FEC and ARQ is commonly used to further improve the robustness of the system. Such a combination of the two basic control schemes is referred to as a HARQ. In

the Type-I HARQ scheme, a coded packet is transmitted initially and, if the receiver fails to accept the packet, a retransmission request in the form of a NAK is fed back to the transmitter. Upon reception of this NAK, the transmitter re-sends the same coded packet again. If the receiver is capable of buffering previously received packets, these multiple packets can be combined to create a single packet till the resulting combined packet can be reliably decoded. The method of combining these multiple packets can be separated into two distinct types. In the code combining scheme, these repeated packets are concatenated to form a single packet at a lower code rate, which is often referred to as Chase combining [14]. In the diversity combining scheme, these repeated packets are combined into a single packet with more reliable constituent symbols by using symbol voting schemes [17] or by using symbol averaging schemes [16]. In the Type-II HARQ scheme, instead of re-sending the same packet, the transmitter tries to construct and sends additional parity bits when a NAK is received. This is also known as the incremental redundancy scheme [18].

In order to provide reliable communications over dynamic wireless channels, advanced transmission mechanisms based on HARQ are adopted. For example, in an adaptive modulation system combined with selection transmit diversity, HARQ is incorporated to provide the feedback channel for packet retransmission, modulation adaptation and transmit antenna selection [29, 76, 100]. In a cooperative diversity system, the relay is selected by checking the correctness of a CRC code of a packet, while at the same time an ARQ protocol is implemented at the destination for packet retransmissions [30]. In recent years, ARQ has been defined as an option at the medium access control (MAC) layer in WiMax standards, which has gained significant attention from both industry and academia. There are several existing research works on the performance analysis of ARQ-based WiMax networks [13, 101, 102].

The performance parameters in the data-link layer due to HARQ, such as the

APER, throughput, goodput and drop rate depend not only on the MAC protocol, but also on the physical-layer parameters. Much work has been done on the performance of HARQ schemes over fading channels [22–25]. Due to the large number and the complexity of the parameters as well as the protocols across the two layers, in previous works, by and large, perfect CSI in the physical-layer at the transmitter is assumed. Nevertheless, the CSI may be outdated or imperfect due to the feedback delays and the channel estimation errors both at the transmitter and the receiver. Since the CSI can be used to perform link adaption, transmit diversity selection [29] and relay selection [30], evaluating the effects of imperfect CSI on the system performance is important to provide insights on system operation and guidelines for designing effective system management schemes. The impact of using outdated CSI on the performance of a SR-ARQ system is considered in [29].

We will study the impact of imperfect CSIR on ARQ/HARQ schemes and demonstrate that the accuracy of the CSIR plays a crucial role in determining the performance in the data-link layer. We focus on establishing a link between network-layer and physical-layer performance parameters. We analyze the performance of three basic ARQ schemes as well as three Type-I HARQ schemes with diversity combining over a block fading channel with imperfect CSIR. The imperfect CSIR is acquired via MMSE channel estimation with the aid of pilot symbols. Three performance parameters: APER, goodput, and drop rate are investigated, respectively. We obtain closed-form upper and lower bounds on the APER, the goodput as well as the drop rate. Using numerical results, we compare the impact of the accuracy of the imperfect CSIR on basic ARQs and on Type-I HARQs. In practice, the number of transmissions is limited in Type-I HARQ, which can result in a drop rate of data packets without guaranteeing their error-free delivery. Hence, the impact of the accuracy of the imperfect CSIR on the transmission limits necessary for achieving

a certain requirement on the drop rate, is further illustrated. The study shows that at large channel estimation MSE, the performance deteriorates very rapidly while at low values of the MSE, the performance improves gently toward that of the perfect CSI case.

Differential phase shift keying (DPSK) is a commonly used alternative modulation to BPSK, because it does not require channel estimation for the acquisition of CSI. In ARQ schemes, information bits are grouped into packets for transmission. Therefore, the packet error probability (PEP) is an important performance measure of an ARQ scheme. In some works, the PEP is also known as the frame error probability. We study the PEP of ARQ schemes in Section 5.5 with DBPSK over a nonselective Rayleigh fading channel. For the case of a static channel, we derive the exact closed-form expressions as well as tight bounds for the PEP and the goodput. For the case of a slowly time-varying channel, the performance is studied via simulations.

5.2 System Description

When information is transmitted using an ARQ scheme, each block of m bits of information is first sent to an encoder. The code can usually correct some error patterns and simultaneously detect other error patterns. However, it is assumed in this work that the code is a linear block code, capable of purely detecting any error pattern of $d_{min} - 1$ or fewer errors which will result in a received vector that is not a code word. The term d_{min} denotes the minimum distance of a block code. After passing through a binary (n, m) systematic block encoder and being prefixed by N_H pilot bits, a packet of $n + N_H$ bits is produced. Each packet comprises m information bits for information transmission, $(n - m)$ parity-check bits for error detection, and N_H pilot bits for channel estimation. The rate of the error detecting code is $R = m/n$.

The effective rate of each packet is defined as $R_e = m/(n + N_H)$ which takes into account the redundancy introduced by the error detecting code and the pilot bits used for channel estimation. The number of pilot bits N_H is usually limited in practice by the channel bandwidth expansion factor $1/R_e$. To facilitate performance study at the data link layer, without channel coding, it is assumed that the N_H pilot bits are transmitted by binary phase shift keying (BPSK) modulation while the n data bits are modulated by either BPSK or quadrature phase shift keying (QPSK). The energy per pilot bit is E_p while the energy per information bit is E_b. Hence, at the output of the error detection encoder, the average energy per data bit is $E_a = E_b R$. The energy per data symbol is $E_s = E_b R$ for BPSK, and $E_s = 2E_b R$ for QPSK.

With L receive antennas and a single transmit antenna, the received pilot and data signal at the lth receive antenna over the kth symbol interval during the tth transmission of a packet, is given by

$$\begin{aligned} r_{tl}[k] &= \sqrt{E_p} m_p[k] h_{tl} + N_{tl}[k], \quad k = 1, \cdots, N_H, \\ z_{tl}[k] &= \sqrt{E_s} m_s[k] h_{tl} + N_{tl}[k], \quad k = N_H + 1, \cdots, N_H + n, \end{aligned} \tag{5.1}$$

where, terms $m_p[k]$ and $m_s[k]$ are the transmitted pilot and data symbols, respectively. For a block fading channel, the channel is assumed to be constant over a duration of $N_H + n$ symbols. The term h_{tl} is the fading gain experienced by the signal on the tth transmission of the packet and received on the lth receive antenna. Since the round-trip delay, which is defined as the time interval between the transmission of a packet and the receipt of an acknowledgment for that packet, is assumed to be larger than the coherence time of the channel, the block fading gain experienced by the retransmitted packet is independent of the gains experienced in previous transmission(s) of the same packet, and they are identically distributed. Therefore, the channel gains $\{h_{tl}\}_{t=1,l=1}^{J,L}$ are i.i.d, complex, Gaussian random variables with $\mathbb{E}[h_{tl}] = 0$ and $\mathbb{E}[|h_{tl}|^2] = 2\sigma^2$,

where J represents the number of transmissions of a packet. The terms $\{N_{tl}\}_{t=1,l=1}^{J,L}$ are due to channel additive, white, Gaussian noise, and are i.i.d, complex, Gaussian random variables with $\mathbb{E}[N_{tl}(k)] = 0$ and $\mathbb{E}[|N_{tl}(k)|^2] = N_0$.

The receiver, using the pilot signals received in each packet from each receive antenna, measures the CSI of the channel for each receive antenna in order to implement the optimum detection of data symbols [103]. The N_H pilot symbols on the tth transmission of the packet are used in a Wiener filter for generating the MMSE estimate $\hat{h}_{tl}$ of the channel gain for the lth receive antenna. Since the channel gain is complex Gaussian, the MMSE estimate $\hat{h}_{tl}$ is given by [104, eq.(2.1)]

$$\hat{h}_{tl} = \sum_{i=1}^{N_H} w_{tl}[i] r_{tl}[i], \tag{5.2}$$

where $w_{tl}[i] = 2\sigma^2\sqrt{E_p}(2N_H\sigma^2 E_p + N_0)^{-1}$ is the ith filter coefficient and is the same for all i, since the channel gain h_{tl} is constant for fixed t and l. The coefficient $w_{tl}[i]$ is also the same for all t and l, since $\{h_{tl}\}$ are i.i.d random variables. The channel estimator's MSE is [104, eq.(2.49)]

$$E\left[|h_{tl} - \hat{h}_{tl}|^2\right] = 2V^2 = \frac{2\sigma^2}{1 + N_H \frac{2\sigma^2 E_p}{N_0}}. \tag{5.3}$$

The estimate $\hat{h}_{tl}$ is a complex Gaussian random variable with mean zero and variance $2(\sigma^2 - V^2)$ [104, eq.(2.18)].

For each received packet, the channel gain is estimated using the pilot symbols. The signals received over the L receive antennas from the J transmissions are combined for optimal detection by using the optimum (maximal ratio) combiner (MRC) [105], which achieves both time diversity combining and spatial diversity combining. After detection, the received code vector is checked by the error detection code. When the received code vector has been detected to be in error, the receiver requests a retransmission. Define $e_d(J)$ as the error event that after combining the

signals received from J transmissions for optimal detection, the packet has been detected to be in error. The probability $P(e_d(J))$ is the probability of detectable error or the probability of retransmission which can be obtained by [11, (Example15-1)].

$$P(e_d(J)) = P(e_p(J)) - P(e_u(J)), \tag{5.4}$$

where $e_p(J)$ is the error event that the received packet, after combining JL received copies for optimal detection, contains one or more bit errors, and $e_u(J)$ represents the error event that the received packet contains an undetectable error pattern after combining JL received copies for optimal detection. The packet error probability $P(e_p(J))$ depends on the channel error statistics, whereas the probability of detectable error $P(e_d(J))$ and the probability of undetectable error $P(e_u(J))$ depend on both the channel error statistics and the choice of the (n, m) error detecting code.

There are two basic parameters that determine the performance of an ARQ protocol: reliability and throughput. In an ARQ scheme, a packet is erroneously accepted if, on any transmission attempt, it arrives at the receiver containing an undetectable error pattern. The APER P_{AE} is the percentage of packets accepted by the receiver that contain one or more bit errors. Clearly, for an ARQ system to be reliable, P_{AE} should be very small. Therefore, the reliability of an ARQ protocol is measured by its APER P_{AE}. Another measure of the performance of an ARQ system is its throughput, which is defined as the ratio of the average number of information bits successfully accepted by the receiver to the total number of bits that could be transmitted.

5.3 Basic ARQ with BPSK/QPSK in SIMO Systems with Imperfect CSIR

In this section, the three basic ARQ schemes are analyzed in a SIMO system without packet combining. Whenever a received packet is detected in error, that packet is discarded and replaced by a retransmitted copy. For the sake of brevity, in the rest of this section, we omit the transmission subindex t in both h_{tl} and $\hat{h}_{tl}$ as well as the index J in $P(e_d(J))$, $P(e_p(J))$, and $P(e_u(J))$, since the received Jth erroneous packet is discarded without packet combining. All three basic ARQ schemes achieve the same reliability; however, they have different throughputs. Next, we derive the throughput of each of the three basic ARQ schemes. For simplicity, we assume that the feedback channel is noiseless.

5.3.1 Bit Error Probability

Based on the optimal channel estimation receiver structure obtained in [103, eq.(7)], the conditional bit error probability (BEP) conditioned on the MMSE estimate of the channel gains $\{\hat{h}_l\}_{l=1}^{L}$, is given by [103, Appendix III]

$$P(e_b|\{\hat{h}_l\}) = \frac{1}{2}\text{erfc}\left(\frac{E_s \cos^2\alpha \sum\limits_{l=1}^{L} |\hat{h}_l|^2}{2E_s V^2 + N_0}\right)^{1/2}, \tag{5.5}$$

where, $\alpha = 0$ corresponds to the conditional BEP of BPSK modulation while $\alpha = \pi/4$ corresponds to that of QPSK. The term e_b denotes the event of bit error. Since $\{h_l\}$ is a set of i.i.d., complex, Gaussian random variables, the estimated channel gains $\hat{h}_l$ are also i.i.d., complex, Gaussian random variables with mean zero and variance $2(\sigma^2 - V^2)$. Therefore, the sum $g = \sum\limits_{l=1}^{L} |\hat{h}_l|^2$ is chi-square-distributed with $2L$ degrees

of freedom and its PDF is given by [6, eq.(2.1-137)]

$$p(g) = \frac{g^{L-1}}{2^L(\sigma^2 - V^2)^L \Gamma(L)} e^{-\frac{g}{2(\sigma^2 - V^2)}}. \tag{5.6}$$

Averaging the conditional BEP (5.5) over g using the pdf (5.6) gives the average BEP [103, eq.(8a)]

$$P(e_b) = \left(\frac{1-\mu}{2}\right)^L \sum_{l=0}^{L-1} \begin{pmatrix} L-1+l \\ l \end{pmatrix} \left(\frac{1+\mu}{2}\right)^l, \tag{5.7}$$

where,

$$\mu = \left(1 + \frac{2E_s V^2 + N_0}{2(\sigma^2 - V^2)E_s \cos^2\alpha}\right)^{-1/2}. \tag{5.8}$$

It is seen from (5.7) that the BEP depends on both E_s and E_p since $2V^2$ is a function of E_p. Taking into account the energy used for channel estimation, we define the effective energy per data bit as

$$E_b^{\text{eff}} = \frac{E_p N_H + E_b m}{n}. \tag{5.9}$$

Therefore, the effect of energy devoted to data delivery on the BEP can be illustrated as a function of the effective received SNR per data bit $\gamma_b^{\text{eff}} = 2E_b^{\text{eff}} L\sigma^2/N_0$.

5.3.2 Packet Error Probability

Conditioned on knowing the estimated channel gain $\hat{h}_l$, the channel is memoryless since the AWGN is independent from symbol to symbol. Hence, the conditional probability that a received packet contains at least one error bit, can be written as

$$P(e_p|\{\hat{h}_l\}) = 1 - \left(1 - P(e_b|\{\hat{h}_l\})\right)^n. \tag{5.10}$$

By averaging (5.10) over the chi-square random variable g, the packet error probability $P(e_p)$ can be obtained as

$$P(e_p) = \int_0^\infty \left(1 - \left(1 - P(e_b|\{\hat{h}_l\})\right)^n\right) p(g)dg. \tag{5.11}$$

By applying the Chernoff bound: $\text{erfc}(x) \leq e^{-x^2}$, and the identity: $\int_0^\infty x^n e^{-\mu x} dx = n!\mu^{-n-1}$ [97, eq.(3.351.3)] on (5.11), the latter can be upper bounded by

$$P(e_p) \leq 2\sum_{i=1}^{n} \binom{n}{i} \left(-\frac{1}{2}\right)^{i+1} \left(1 + \frac{2iE_s \cos^2\alpha(\sigma^2 - V^2)}{2E_s V^2 + N_0}\right)^{-L}. \tag{5.12}$$

In [106], when $L = 1$, an alternative expression for (5.12) without the binomial coefficient, is obtained as

$$P(e_p) \leq 1 - \sum_{l=0}^{n} \left(\frac{1}{2}\right)^{n-l} \prod_{j=0}^{l-1} \frac{(n-j)c}{2(b + (j+1)c)}, \tag{5.13}$$

where,

$$b = \frac{1}{2(\sigma^2 - V^2)}, \qquad c = \frac{E_s \cos^2\alpha}{2E_s V^2 + N_0}. \tag{5.14}$$

Compared with the Chernoff bound, a closer approximation of $\text{erfc}(x)$ without much loss in accuracy becomes useful when computing a closer approximation of $P(e_p)$. Applying the more accurate approximation below [107, eq.(31)]

$$\text{erfc}(x) \approx \frac{1}{6}e^{-x^2} + \frac{1}{3}e^{-4x^2/3} \tag{5.15}$$

to (5.5), we have

$$P(e_b|\{\hat{h}_l\}) \approx \frac{1}{12} e^{-\frac{E_s \cos^2\alpha g}{2E_s V^2 + N_0}} \left(1 + 2e^{-\frac{E_s \cos^2\alpha g}{3(2E_s V^2 + N_0)}}\right). \tag{5.16}$$

By substituting (5.16) for $P(e_b|\{\hat{h}_l\})$ in (5.11) and making use of [97, eq.(3.351.3)], the packet error probability $P(e_p)$ can be more accurately approximated by

$$P(e_p) \approx -\sum_{i=1}^{n} \binom{n}{i} \left(\frac{-1}{12}\right)^i \sum_{j=0}^{i} \binom{i}{j} 2^j \left(1 + \frac{c}{b}\left(\frac{j}{3} + i\right)\right)^{-L}. \tag{5.17}$$

5.3.3 Undetectable Error Rate

Next, we evaluate the probability of undetectable error $P(e_u)$. This probability is often quite difficult to determine since it requires a complete characterization of the output

of the FEC decoder. A bound or an approximation, however, is usually sufficient [11]. For (n, m) linear codes, except for some short linear codes, the weight distributions for many codes are still unknown. Consequently, it is considerably difficult to compute their $P(e_u)$, but it is fairly easy to derive upper and lower bounds on $P(e_u)$ for the ensemble of all (n, m) linear codes. Conditioned on knowing the estimated channel gain $\hat{h}_l$, the composite channel including binary modulator and hard-decision detector can be modeled as a binary symmetric channel (BSC) [6, Chapter 7]. Hence, the upper bound on the conditional undetectable error probability can be evaluated by [10, eq.(3.42)]

$$P(e_u|\{\hat{h}_l\}) \leq 2^{-n(1-R)} \left[1 - \left(1 - P(e_b|\{\hat{h}_l\})\right)^n\right], \tag{5.18}$$

and the lower bound can be evaluated by [108, eq.(12)]

$$P(e_u|\{\hat{h}_l\}) \geq \sum_{w=n-m+1}^{n} \begin{pmatrix} n \\ w \end{pmatrix} \left(2^{m+w-n} - 1\right) P(e_b|\{\hat{h}_l\})^w \left(1 - 2P(e_b|\{\hat{h}_l\})\right)^{n-w}. \tag{5.19}$$

Taking the mean of (5.18) over g, the average undetectable error probability can be upper-bounded as

$$P(e_u) \leq \int_0^\infty 2^{-n(1-R)} \left(1 - \left(1 - P(e_b|\{\hat{h}_l\})\right)^n\right) p(g)dg. \tag{5.20}$$

The last integral can be evaluated like that in (5.11), giving

$$P(e_u) \leq 2^{-n(1-R)+1} \sum_{i=1}^{n} \begin{pmatrix} n \\ i \end{pmatrix} \left(-\frac{1}{2}\right)^{i+1} \left(1 + \frac{ic}{b}\right)^{-L}. \tag{5.21}$$

The ratio between $P(e_u)$ and $P(e_p)$ can then be upper bounded by the ratio of (5.20) to (5.11):

$$\frac{P(e_u)}{P(e_p)} \leq 2^{-n(1-R)}. \tag{5.22}$$

We thus have $P(e_u) << P(e_p)$ when the number of parity-check bits $n-m$ is large enough to make $2^{-n(1-R)} << 1$, say, $(n-m) > 5$, and therefore $P(e_d) \approx P(e_p)$ is inferred from (5.4). This assumption is applied in the rest of the chapter. The lower bound on the average undetectable error probability can be obtained by taking the mean of (5.19) over g as

$$P(e_u) \geq \int_0^\infty \sum_{w=n-m+1}^{n} \binom{n}{w} (2^{m+w-n}-1) P(e_b|\{\hat{h}_l\})^w \left(1-2P(e_b|\{\hat{h}_l\})\right)^{n-w} p(g)dg. \tag{5.23}$$

By making use of the binomial theorem, the last integral becomes

$$P(e_u) \geq \sum_{w=n-m+1}^{n} \binom{n}{w} (2^{m+w-n}-1) \sum_{i=0}^{n-w} \binom{n-w}{i} (-2)^i \int_0^\infty P(e_b|\{\hat{h}_l\})^{i+w} p(g)dg. \tag{5.24}$$

By substituting (5.16) for $P(e_b|\{\hat{h}_l\})$ in (5.24) and using [97, eq.(3.351.3)], the right-hand side of the last integral can be evaluated, giving

$$P(e_u) \geq \sum_{w=n-m+1}^{n} \binom{n}{w} \sum_{i=0}^{n-w} \binom{n-w}{i} \sum_{j=0}^{i+w} \binom{i+w}{j} (-1)^i \frac{2^{m-n+j}-2^{j-w}}{6^{i+w}} \left(1+\frac{c(i+w+j/3)}{b}\right)^{-L} \tag{5.25}$$

5.3.4 Selective-repeat ARQ scheme

If buffering at both the transmitter and the receiver is allowed, an SR-ARQ protocol can be implemented. The throughput of the SR-ARQ scheme follows from [11, eq.(15-8)] and [12, Sec. III] as

$$\eta^{sr} = R_e \left(1 - P(e_d)\right). \tag{5.26}$$

By applying (5.22) on (5.26), the throughput is then approximated by

$$\eta^{sr} \approx R_e \left(1 - P(e_p)\right). \tag{5.27}$$

One can only evaluate a lower bound to right-hand side of (5.27) by using the upper bound on $P(e_p)$ in (5.12).

Another useful system parameter, the APER, which shows the reliability of the ARQ system, is given by [11, eq.(15-2)]

$$P_{AE} = \frac{P(e_u)}{1 - P(e_d)}. \tag{5.28}$$

Using $P(e_d) \approx P(e_p)$ and substituting (5.21) and (5.12) into (5.28), P_{AE} is upper-bounded by

$$P_{AE} \leq \frac{2^{-n(1-R)} Z}{1 - Z}, \tag{5.29}$$

where,

$$Z = 2 \sum_{i=1}^{n} \binom{n}{i} \left(-\frac{1}{2}\right)^{i+1} \left(1 + \frac{ic}{b}\right)^{-L}. \tag{5.30}$$

The lower bound to the right-hand side of (5.28) can be evaluated by using the lower bound on $P(e_u)$ in (5.25) and the approximation of $P(e_p)$ in (5.17).

The throughput is meaningful only when considered in conjunction with the reliability. Therefore, the goodput η_g is defined as the ratio of the expected number of information bits correctly received per unit of time to the total number of bits that can be transmitted per unit of time [12,13]. The goodput of an SR-ARQ scheme, which shows the proportion of the throughput consisting of correct packets, can be expressed as

$$\eta_g^{sr} = (1 - P_{AE})\eta^{sr}. \tag{5.31}$$

By substituting (5.28) and (5.26) into (5.31), the goodput can be obtained as

$$\eta_g^{sr} = \frac{m}{n + N_H}(1 - P(e_p)). \tag{5.32}$$

The goodput can be considered as a lower bound on the throughput since the η_g^{sr} in (5.32) is the same as the η^{sr} in (5.26) when neglecting the term $P(e_u)$. For a sufficiently

small $P(e_u)$, the goodput approaches the throughput. Substituting (5.12) into (5.32), we obtain the lower bound on η_g^{sr} as

$$\eta_g^{sr} \geq \frac{m}{n+N_H}(1-Z). \tag{5.33}$$

Consider the special case when $L = 1$, (5.29) and (5.33) can be simplified to [106, eq.(17)]

$$P_{AE} \leq 2^{-n(1-R)} \left[\left(\sum_{l=0}^{n} \left(\frac{1}{2}\right)^{n-l} \prod_{j=0}^{l-1} \frac{(n-j)c}{2(b+(j+1)c)} \right)^{-1} - 1 \right] \tag{5.34}$$

and [106, eq.(21)]

$$\eta_g^{sr} \geq \frac{m}{n+N_H} \sum_{l=0}^{n} \left(\frac{1}{2}\right)^{n-l} \prod_{j=0}^{l-1} \frac{(n-j)c}{2(b+(j+1)c)}. \tag{5.35}$$

respectively.

5.3.5 Stop-and-wait ARQ scheme

Let D be the idle time from the end of transmission of one packet to the beginning of transmission of the next. Let τ be the bit transmission rate which is defined as the number of bits transmitted per unit of time. In one round-trip delay time, which is defined as the time interval between the transmission of a packet and the receipt of an acknowledgment for that packet, the transmitter can transmit a total of $1 + D\tau/(n + N_H)$ packets if it does not stay idle. By evaluating the average number of packets that the transmitter could have transmitted during the interval from the beginning of transmission of one packet to the receipt of a positive acknowledgment for that packet, the throughput of a SW-ARQ system follows [10, eq. (22.6)] as

$$\eta^{sw} = \frac{m(1-P(e_d))}{n+N_H+D\tau} \approx \frac{m(1-P(e_p))}{n+N_H+D\tau}. \tag{5.36}$$

By making use of (5.22) and (5.12), (5.36) can be limited by its lower bound as

$$\eta^{sw} \geq \eta_L^{sw} = \frac{m(1-Z)}{n+N_H+D\tau}. \tag{5.37}$$

Since the APER merely depends on the channel error statistics and the choice of the error detecting code, the SW-ARQ scheme has the same reliability as the SR-ARQ scheme. Therefore, the goodput of a SW-ARQ scheme can be derived and its lower bound is further obtained in the same way as in the SR-ARQ scheme, i.e.,

$$\eta_g^{sw} = \frac{m\,(1 - P(e_p))}{n + N_H + D\tau} \geq \frac{m(1 - Z)}{n + N_H + D\tau}. \tag{5.38}$$

We see that even if the data transmission rate is low and the round-trip delay is short, the goodput can never achieve the maximum value of R_e since the inaccurate channel estimation will deteriorate the performance.

5.3.6 Go-back-N ARQ scheme

The term 'go-back-N' (GBN) derives from the fact that when a transmitter receives a retransmission request, it must go back into its buffer some N packets and restart transmission from there. When a NAK is received, the transmitter resends that packet and the $N - 1$ subsequent packets that were transmitted earlier. The parameter N depends on the bit transmission rate τ and on the round-trip delay $D + (n + N_H)/\tau$, and is therefore evaluated as $N = 1 + D\tau/(n + N_H)$. Hence, the throughput of a GBN-ARQ scheme is given by [10, eq.(22.5)]

$$\eta^{gbn} = \frac{(1 - P(e_d))m}{n + N_H + P(e_d)D\tau} \approx \frac{(1 - P(e_p))m}{n + N_H + P(e_p)D\tau}. \tag{5.39}$$

The lower bound can be obtained in the same way as in (5.26) and (5.36) and expressed as

$$\eta^{gbn} \geq \eta_L^{gbn} = \frac{(1 - Z)m}{n + N_H + P(e_p)D\tau}. \tag{5.40}$$

Making use of the APER of GBN-ARQ which is the same as that of SR-ARQ shown in (5.28), the goodput and its lower bound can be obtained as follows:

$$\eta_g^{gbn} = \frac{(1 - P(e_p))m}{n + N_H + P(e_d)D\tau} \geq \frac{m(1 - Z)}{n + N_H + ZD\tau}. \tag{5.41}$$

5.3.7 Power Allocation between Pilot and Data Bits

Each packet is sent with a fixed total energy E_T. When more energy is devoted to channel estimation, the estimates of channel gains are more accurate, leading to a smaller error probability. However, this reduces the energy available for data transmission and leads to a higher error probability. For this reason, there must exist an optimum fraction ε of the total energy E_T that should be devoted to channel estimation so as to maximize the lower bound on the goodput η_g^{sr} in (5.33).

All signalling messages are assumed to be significantly shorter than the user data packets, and therefore transmitted with negligible overall energy consumption. For a given total packet energy E_T, the amount of total energy assigned to pilot symbols is $\varepsilon E_T = N_H E_p$ while the remainder of total energy devoted to data transmission equals $(1-\varepsilon)E_T = nE_s$. The number of pilot symbols N_H which is limited by the allowable channel bandwidth, is assumed to be fixed for each packet. Thus an optimum ε will lead to an optimum E_p and similarly an optimum E_s. Since N_H is fixed, (5.32) shows that maximizing the goodput η_g^{sr} amounts to minimizing the packet error rate $P(e_p)$. By considering the special case when $L = 1$ and using the lower bound (5.35), the maximization problem now comes to

$$\varepsilon^* = \arg\max_{0\le\varepsilon\le 1}\left\{ f\left(\frac{c}{b}\right) = \sum_{l=0}^{n}\left(\frac{1}{2}\right)^{n-l}\prod_{j=0}^{l-1}\frac{(n-j)c}{2(b+(j+1)c)}\right\}, \tag{5.42}$$

where b and c given in (5.14) can be rewritten as

$$b = \frac{1}{2\sigma^2 - \frac{2N_0\sigma^2}{N_0+2\varepsilon E_T\sigma^2}} = \frac{1+\varepsilon\gamma}{2\varepsilon\sigma^2\gamma} \tag{5.43}$$

and

$$c = \frac{(1+\varepsilon\gamma)(1-\varepsilon)E_T}{2(1-\varepsilon)E_T\sigma^2 + n(N_0+2\varepsilon E_T\sigma^2)}. \tag{5.44}$$

Here, $\gamma = 2E_T\sigma^2/N_0$ is defined as the total transmit SNR. The objective function $f\left(\frac{c}{b}\right)$ defined in (5.42) is a monotonically increasing function of the variable c/b. Therefore,

the optimization problem can be reduced to

$$\varepsilon^* = \arg\max_{0\le\varepsilon\le 1}\left\{\frac{c}{b}\right\} = \arg\max_{0\le\varepsilon\le 1}\left\{\frac{\varepsilon(1-\varepsilon)\gamma^2}{\gamma(1-\varepsilon)+n(1+\varepsilon\gamma)}\right\}. \tag{5.45}$$

Setting the derivative of c/b with respective to ε equal to zero, and solving the resulting quadratic equation, we obtain the optimal ε^* as

$$\varepsilon^* = \frac{n+\gamma-\sqrt{n^2+n\gamma+\gamma^2 n+\gamma n^2}}{\gamma-n\gamma}. \tag{5.46}$$

The optimal ε^*, which is in the range $[0,1]$, satisfies an equality shown in the following proposition which indicates the upper and the lower limits on the optimum amount of energy devoted to channel estimation.

Proposition 5.1. For a given n, the optimal value ε^* satisfies the following inequality for any γ.

$$\frac{1-\sqrt{n}}{1-n} \le \varepsilon^* \le 0.5 \tag{5.47}$$

Proof. The optimal value ε^* is a monotonically decreasing function of γ, which can be shown as follows. By changing to the variable $x=\frac{\gamma}{n}$, (5.46) becomes

$$\varepsilon^* = \frac{1+x-\sqrt{1+x+nx^2+nx}}{x-nx}. \tag{5.48}$$

Note that n is also a function of x. By taking the first derivative of ε^* with respective to x and simplifying, the derivative can be arranged to be

$$\frac{d\varepsilon^*}{dx} = \frac{(n-1)\left(2\sqrt{1+x+nx^2+nx}-(nx+x+2)\right)}{2\sqrt{1+x+nx^2+nx}\,(x-nx)^2}.$$

By applying the inequality of arithmetic and geometric means, the inequality $d\varepsilon^*/dx \le 0$ is true because of the relationship

$$2\sqrt{1+x+nx^2+nx} = 2\sqrt{(1+x)(nx+1)} \le 2+nx+x.$$

Using the monotonically decreasing property of ε^*, the upper and lower limits of ε^* can be seen to be

$$\lim_{x\to\infty} \varepsilon^*(x) \leq \varepsilon^* \leq \lim_{x\to 0} \varepsilon^*(x). \tag{5.49}$$

The lower limit can be shown to be

$$\lim_{x\to\infty} \frac{1 + x^{-1} - \sqrt{x^{-2} + x^{-1} + nx^{-1} + n}}{1-n} = \frac{1-\sqrt{n}}{1-n}.$$

By applying L'Hospital rule, the upper limit is obtained as

$$\lim_{x\to 0} \frac{1 - 0.5(1+n+2nx)(1+nx^2+nx+x)^{-\frac{1}{2}}}{1-n} = 0.5$$

□

The upper limit on ε^* in (5.47) indicates that the optimum amount of energy devoted to channel estimation is always at most half of the total energy. For the lower limit, at least a fraction of total energy (i.e. $\frac{1-\sqrt{n}}{1-n}E_T$) is necessary to be devoted to channel estimation. A small amount of energy is assigned to channel estimation when a long code (i.e. large n) is used. This is because the lower limit in (5.47) can be reduced to $1/\sqrt{n}$ when n is large.

5.3.8 Numerical Results for Basic ARQ Schemes

Since the bounds on goodput and APER only depend on the values of n and m and they are not influenced by specific code structures, different values of n and m are chosen to demonstrate the performances based on (5.34) and (5.35). The fixed total energy E_T for each packet is $(m + N_H)E_a$, where E_a represents the average energy per bit. The normalized channel estimation MSE V^2/σ^2 depends on the energy $\varepsilon^* E_T$ devoted to pilot symbols rather than the value of N_H.

In Fig. 5.1 and Fig. 5.2, the performance bounds are plotted with the number of information bits $m = 210$ and with equal power allocation $E_p = E_b^B = E_a$

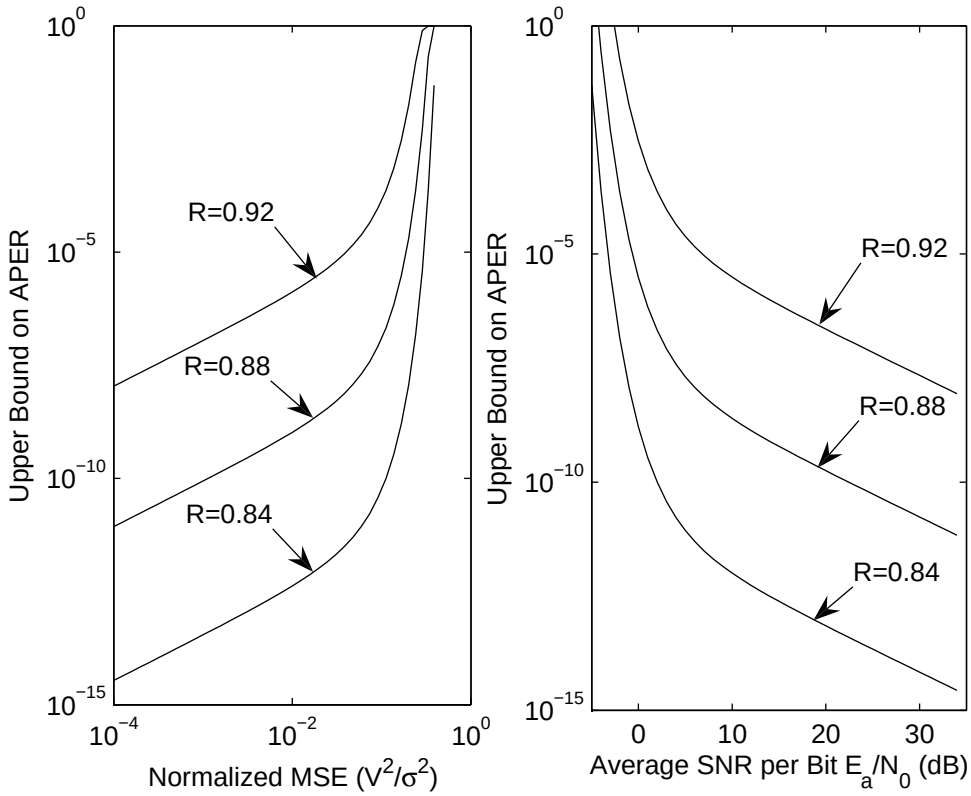

Figure 5.1: The APER versus the normalized MSE and E_a/N_0.

(i.e. $\varepsilon = N_H/(n + N_H)$). For the fair comparison, we consider the cases where the transmit energy (mE_b) is the same, but with the different values of $R = 0.92$, $R = 0.88$, and $R = 0.84$. The number of pilot bits N_H is usually limited in practice by the channel bandwidth expansion factor $1/R_e$. Assume that the channel bandwidth expansion factor is $1/R_e = 1.11$, $1/R_e = 1.16$, and $1/R_e = 1.21$ respectively for $R = 0.92$, $R = 0.88$, and $R = 0.84$. This bandwidth expansion factor leads to the choice here of $N_H = 5$ pilot symbols, which is the maximum value of N_H for the given m and R. Fig. 5.1 indicates that there exists a critical value for both V^2/σ^2 and average SNR per bit E_a/N_0 that separates two different trends in the APER curves. In particular, when V^2/σ^2 is above a critical value of around 10^{-2}, the APER

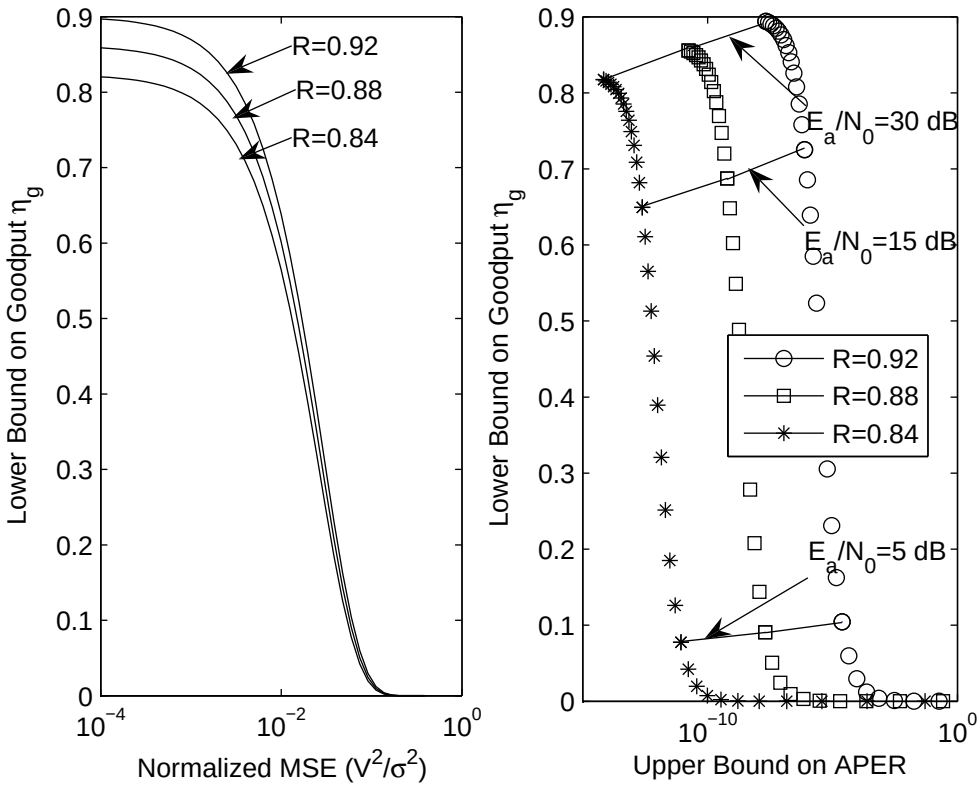

Figure 5.2: The goodput versus the normalized MSE and the APER.

deteriorates very fast. Below this critical value, the APER decreases more gently. The corresponding critical value of SNR E_a/N_0 is around 10 dB. Additionally, it can be seen that in order to achieve a certain APER requirement, greater channel estimation accuracy or higher E_a/N_0 is needed for a code with fewer parity-check bits. Fig. 5.2 shows how the parameter V^2/σ^2 and the APER affect the goodput performance. The values of normalized MSE above a critical value of around 10^{-2} degrade the goodput performance very rapidly. The goodput can be improved by more accurate channel estimation. However, decreasing the normalized MSE below a value of about 10^{-3} leads to diminishing increments in the goodput. Furthermore, in order to achieve a certain goodput requirement, greater channel estimation accuracy is needed for a

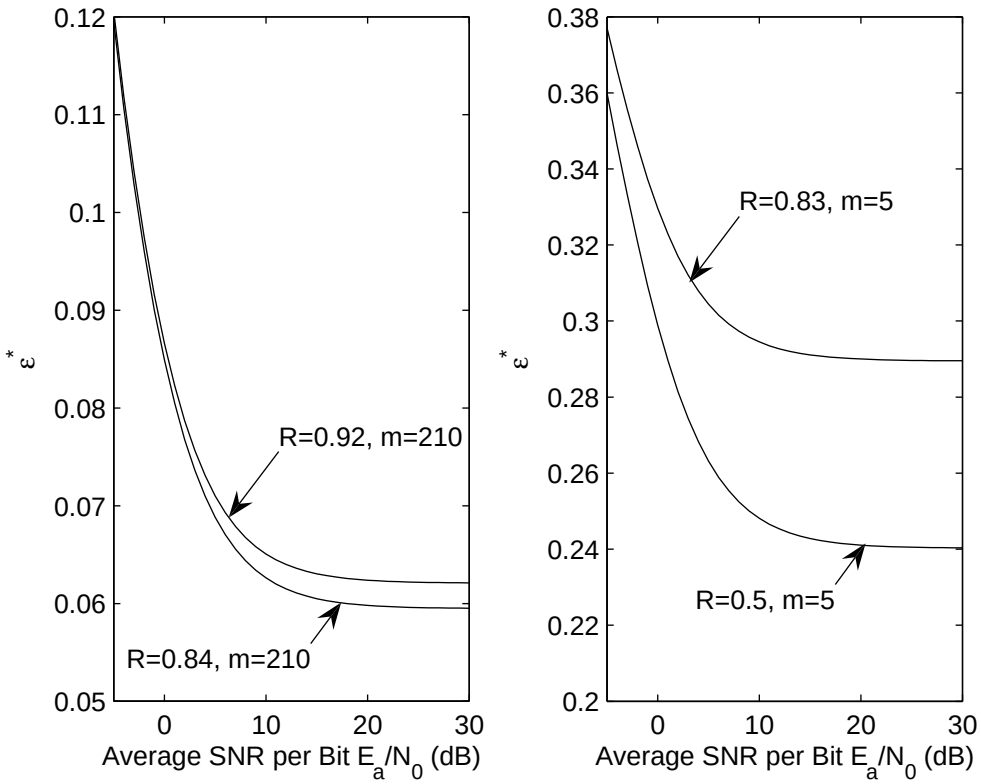

Figure 5.3: The value of ε^* versus average SNR per bit E_a/N_0

very high rate code with more parity-check bits. This is because using more accurate channel estimation compensates for the higher packet error rate induced by a longer packet. The goodput versus the APER in Fig. 5.2 shows that, for a lower-rate code, a smaller APER is required to achieve desired goodput performance. For a fixed SNR E_a/N_0, higher goodput is obtained by using a higher-rate code but at the expense of worse APER. This proposition indicates that the optimum amount of energy devoted to channel estimation is always at most half of the total energy. On the other hand, at least a fraction of total energy (i.e.$\frac{1-\sqrt{n}}{1-n}E_T$) is necessary to be devoted to channel estimation. A small amount of energy is assigned to channel estimation when a long code (i.e. large n) is used. This is because the left-hand side of (5.47) can be reduced

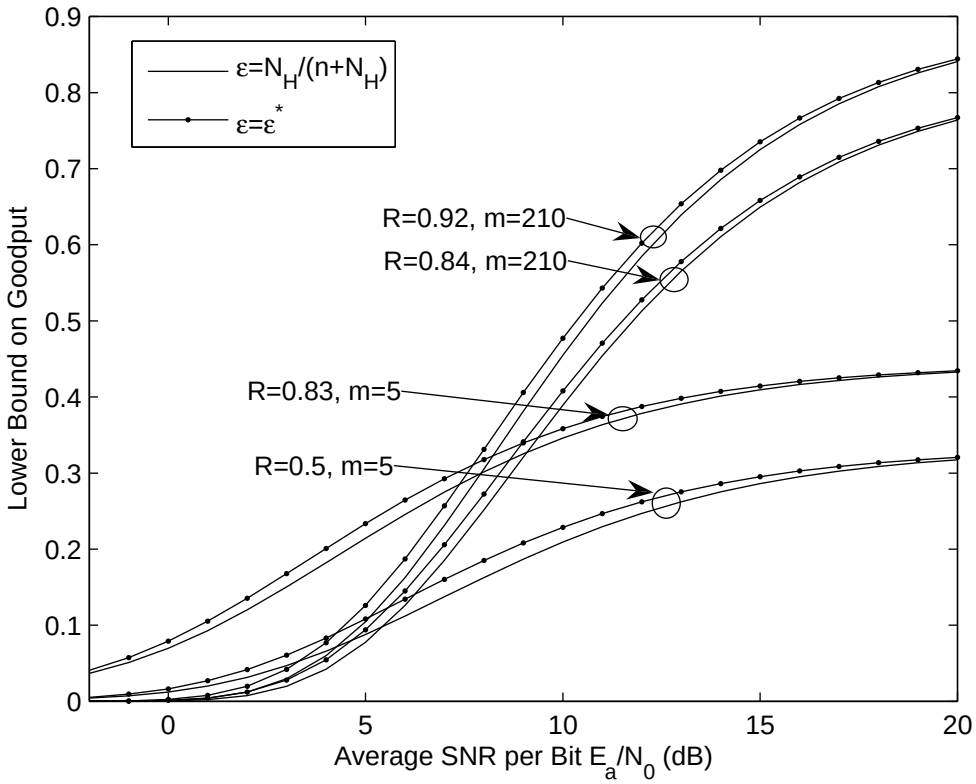

Figure 5.4: The lower bound on the goodput achieved by different values of ε

to $1/\sqrt{n}$ when n is large. The above discussion of channel estimation accuracy leads to the consideration of optimum power allocation between channel estimation and data transmission to achieve the maximum goodput. The power allocation and the goodput improvement achieved are shown in Fig. 5.3 and Fig. 5.4, respectively. The curves in Fig. 5.3 indicate that ε^* is a decreasing function of the total transmit SNR γ but it converges to $\frac{1-\sqrt{n}}{1-n}$ and 0.5 in the high and low SNR regions, respectively. For different codes, an improvement of about 0.5 dB to 1 dB can be observed in Fig. 5.4 for a given goodput. Less improvement is achieved when a longer code is used. This is due to that there is no significant difference between the value of ε^* and ε when the length of the code is long. For a code with $m = 210$ and $R = 0.84$, a 0.5 dB improvement is

obtained.

5.4 Type-I HARQ with BPSK/QPSK in SIMO Systems with Imperfect CSIR

In this section, the three Type-I HARQ schemes are analyzed in a SIMO system. For each received packet, the channel gain is estimated using pilot symbols. The receiver retains the packets detected in error in previous transmissions for combining with the repeated copy in the current transmission in an optimum manner. The signals received over the L receive antennas from J transmissions are combined for optimal detection by using the optimum MRC [105], which achieves both time diversity combining and spatial diversity combining. The maximum number of allowed transmissions for each packet is denoted by $J_m + 1$. A packet will be dropped when it is not accepted successfully in $J_m + 1$ transmissions.

We first evaluate the packet error probability $P(e_p(J))$ at the Jth transmission. Based on the optimal channel estimation receiver structure obtained in [103, eq.(7)], the conditional bit error probability after combining JL signal copies from the L receive antennas over J transmissions, is given by [103, Appendix III]

$$P(e_b(J)|\{\hat{h}_{tl}\}) = \frac{1}{2}\text{erfc}\left[\frac{E_s \cos^2\alpha \sum_{t=1}^{J}\sum_{l=1}^{L}|\hat{h}_{tl}|^2}{2E_s V^2 + N_0}\right]^{1/2}. \tag{5.50}$$

Since $\{h_{tl}\}$ is a set of i.i.d., complex, Gaussian random variables, the estimated channel gains $\hat{h}_{tl}$ are also i.i.d., complex, Gaussian random variables, each with mean zero and variance $2\sigma^2$. Therefore, the sum $g(J) = \sum_{t=1}^{J}\sum_{l=1}^{L}|\hat{h}_{tl}|^2$ is chi-square-distributed with

$2JL$ degrees of freedom and its pdf is given by [6, eq.(2.1-137)]

$$p(g) = \frac{g^{JL-1}}{(2\sigma^2 - 2V^2)^{JL}\Gamma(JL)} e^{-g/(2\sigma^2-2V^2)}. \tag{5.51}$$

The term $P(e_p(J))$ is the probability that the packet contains one or more bit errors at the Jth transmission. It can be evaluated by

$$P(e_p(J)) = \int_0^\infty \left(1 - \left(1 - P(e_b(J)|\{\hat{h}_{tl}\})\right)^n\right) p(g)dg. \tag{5.52}$$

By applying the Chernoff bound: $\mathrm{erfc}(x) < e^{-x^2}$, and [97, eq.(3.351.3)] on (5.52), it can be upper bounded by

$$P(e_p(J)) < 2\sum_{i=1}^{n} \binom{n}{i} \left(-\frac{1}{2}\right)^{i+1} \left(1 + \frac{ic}{b}\right)^{-JL}. \tag{5.53}$$

Using the approximation of the erfc(x) in (5.15), (5.52) can be evaluated like that in (5.17), giving

$$P\big(e_p(J)\big) \approx -\sum_{i=1}^{n} \binom{n}{i} \left(\frac{-1}{12}\right)^i \sum_{j=0}^{i} \binom{i}{j} 2^j \left(1 + \frac{c}{b}\left(\frac{j}{3} + i\right)\right)^{-JL}. \tag{5.54}$$

By substituting (5.50) into the bounds in (5.18) and (5.19) and taking the average over the chi-square random variable $g(J)$, the term $P(e_u(J))$ can be respectively upper and lower bounded by

$$P(e_u(J)) \leq 2^{-n(1-R)+1} \sum_{i=1}^{n} \binom{n}{i} \left(-\frac{1}{2}\right)^{i+1} \left(1 + \frac{ic}{b}\right)^{-JL}, \tag{5.55}$$

and

$$P(e_u(J)) \geq \sum_{w=n-m+1}^{n} \binom{n}{w} \sum_{i=0}^{n-w} \binom{n-w}{i} \sum_{j=0}^{i+w} \binom{i+w}{j} (-1)^i \frac{2^{m-n+j} - 2^{j-w}}{6^{i+w}} \times \left(1 + \frac{c(i+w+j/3)}{b}\right)^{-JL} \tag{5.56}$$

Define $e_r(J)$ to be the event that a packet is not accepted successfully in J transmissions (including the original transmission). The probability that a packet is to be retransmitted the Jth time (excluding the original transmission) can be expressed as

$$P(e_r(J)) = P\left(e_d(1), e_d(2), \cdots, e_d(J)\right).$$

Based on the discussion for (5.22), we have that $P(e_d(J)) \approx P(e_p(J))$ when $(n-m)$ is large enough. The evaluation of the probability $P(e_r(J))$ is complicated since the statistical dependence among different receptions is induced by combining all received packets. A lower and an upper bound on $P(e_r(J))$ can be obtained as [28, eq.(35)]

$$\prod_{t=1}^{J} P(e_p(t)) \leq P(e_r(J)) \leq P(e_p(J)). \tag{5.57}$$

5.4.1 Selective-repeat based Type-I HARQ scheme

The average number of transmissions it takes before a packet is accepted is given by [11, eq.(15-3)]

$$\begin{aligned} T_r^{sr} &= \sum_{J=1}^{J_m+1} J\left[P\big(e_r(J-1)\big) - P\big(e_r(J)\big)\right] \\ &= \sum_{J=0}^{J_m} P\big(e_r(J)\big) - (J_m+1)P\big(e_r(J_m+1)\big). \end{aligned} \tag{5.58}$$

where $P\big(e_r(0)\big) = 1$ and $P\big(e_r(J-1)\big) - P\big(e_r(J)\big)$ is the probability that a packet is accepted at its Jth transmission. Therefore, the throughput of the SR-HARQ system with the allowed maximum number of retransmission J_m, follows from [11, eq.(15-8)] as

$$\eta^{sr}(J_m) = \frac{m}{n+N_H}\frac{1}{T_r^{sr}}. \tag{5.59}$$

The upper bound on T_r^{sr} can be obtained as [28, eq.(26)]

$$T_r^{sr} < 1 + \sum_{J=1}^{J_m} P\big(e_r(J)\big), \tag{5.60}$$

which is reasonable since the term $P(e_r(J_m+1))$ can be small enough to be ignored when L or J_m is large enough. By using the bounds (5.57), (5.60) and (5.53) in (5.59), a lower bound on throughput for SR-HARQ scheme is obtained as

$$\eta^{sr}(J_m) > \frac{m}{(n+N_H)\left(1+\sum_{J=1}^{J_m} Z(J)\right)}, \tag{5.61}$$

where

$$Z(J) = 2\sum_{i=1}^{n} \begin{pmatrix} n \\ i \end{pmatrix} \left(\frac{-1}{2}\right)^{i+1} \left(1+\frac{ic}{b}\right)^{-JL}. \tag{5.62}$$

By applying (5.57) to (5.58), (5.59) can be upper-bounded by

$$\eta^{sr}(J_m) < \frac{R_e}{1+\sum_{J=1}^{J_m}\prod_{t=1}^{J} P\big(e_p(t)\big) - (J_m+1)P\big(e_p(J_m+1)\big)}. \tag{5.63}$$

The right-hand side of (5.63) can be approximated by substituting (5.54) for $P\big(e_p(t)\big)$.

The term $e_u(J)$ denotes the event that a packet is erroneously accepted if, after packet combining at the Jth transmission attempt, it contains an undetectable error pattern. The packet can be accepted on the first transmission, the second (after a retransmission request), etc. Therefore, the APER of the HARQ system with the allowed maximum number of retransmissions J_m, can be computed by summing the probabilities of the various events that lead to the acceptance of an erroneous packet as [11, eq.(15-2)], [12].

$$P_{AE}(J_m) = P\big(e_u(1)\big) + \sum_{J=2}^{J_m+1} P\big(e_d(1),\cdots,e_d(J-1),e_u(J)\big). \tag{5.64}$$

By applying (5.57) to (5.64), the term $P_{AE}(J_m)$ can be upper and lower bounded by

$$P_{AE}(J_m) \le P\big(e_u(1)\big) + \sum_{J=2}^{J_m+1} P\big(e_u(J)\big) = \sum_{J=1}^{J_m+1} P\big(e_u(J)\big), \tag{5.65}$$

and

$$P_{AE}(J_m) \ge P\big(e_u(1)\big) + \sum_{J=2}^{J_m+1}\prod_{t=1}^{J-1} P\big(e_d(t)\big)P\big(e_u(J)\big). \tag{5.66}$$

By substituting the bounds given in (5.55) for $P(e_u(J))$ in (5.65), an upper bound on $P_{AE}(J_m)$ can be obtained as

$$P_{AE}(J_m) \leq \sum_{J=1}^{J_m+1} 2^{-n(1-R)} Z(J). \tag{5.67}$$

An approximation on the right-hand side of (5.66) can be evaluated by substituting (5.56) for $P\big(e_u(J)\big)$ and (5.54) for $P\big(e_d(t)\big)$ since $P\big(e_d(J)\big) \approx P\big(e_p(J)\big)$ when $n-m$ is sufficiently large.

5.4.2 Stop-and-wait based Type-I HARQ scheme

During the interval from the beginning of transmission of one packet to the receipt of a positive acknowledgment for that packet, the average number of packets (including the idling effect which refers to the idle time spent waiting for an acknowledgement for each transmitted packet) that the transmitter could have transmitted is [10, eq.(22.6)]

$$\begin{aligned} T_r^{sw} &= \sum_{J=1}^{J_m+1} J\left(1+\frac{D\tau}{n+N_H}\right)\left[P\big(e_r(J-1)\big)-P\big(e_r(J)\big)\right] \\ &= \left(1+\frac{D\tau}{n+N_H}\right)\left[\sum_{J=0}^{J_m} P\big(e_r(J)\big)-(J_m+1)P\big(e_r(J_m+1)\big)\right]. \end{aligned}$$

Therefore, the throughput of an SW-HARQ scheme is [10, eq.(22.6)]

$$\eta^{sw}(J_m) = \frac{R_e}{T_r^{sw}}. \tag{5.68}$$

By using the same derivation as that for (5.60), the upper bound on T_r^{sw} can be obtained as

$$T_r^{sw} < \left(1+\frac{D\tau}{n+N_H}\right)\sum_{J=0}^{J_m} P\big(e_r(J)\big). \tag{5.69}$$

Therefore, the lower bound on the throughput of an SW-HARQ scheme becomes

$$\eta^{sw}(J_m) > \frac{m}{(n+N_H+D\tau)\left(1+\sum\limits_{J=1}^{J_m} Z(J)\right)}. \tag{5.70}$$

By using the same approach as that for (5.63), (5.68) can be upper-bounded by

$$\eta^{sw}(J_m) < \frac{R_e}{\left(1+\frac{D\tau}{n+N_H}\right)\left(1+\sum_{J=1}^{J_m}\prod_{t=1}^{J} P\big(e_p(t)\big) - (J_m+1)P\big(e_p(J_m+1)\big)\right)}. \tag{5.71}$$

The right-hand side of (5.71) can be approximated by using the approximation in (5.54) for $P\big(e_p(t)\big)$.

5.4.3 Go-back-N based Type-I HARQ scheme

When a NAK is received, the transmitter resends that packet and the $N-1$ subsequent packets that were transmitted earlier. Therefore, for a packet to be successfully accepted by the receiver, the average number of retransmissions (including the original transmission) required is [10, eq. (22.5)]

$$\begin{aligned} T_r^{gbn} =& \sum_{J=1}^{J_m+1}\left[1+(J-1)\left(1+\frac{D\tau}{n+N_H}\right)\right]\Big(P\big(e_r(J-1)\big) - P\big(e_r(J)\big)\Big) \\ =& 1 - P\big(e_r(J_m+1)\big) + \left(1+\frac{D\tau}{n+N_H}\right)\left(\sum_{J=1}^{J_m} P\big(e_r(J)\big) - J_m P\big(e_r(J_m+1)\big)\right) \\ =& 1 + \left(1+\frac{D\tau}{n+N_H}\right)\sum_{J=1}^{J_m} P\big(e_r(J)\big) - P\big(e_r(J_m+1)\big)\left[J_m\left(1+\frac{D\tau}{n+N_H}\right)+1\right]. \end{aligned} \tag{5.72}$$

Since the drop rate $P\big(e_r(J_m+1)\big)$, which is defined to be the probability that a packet is drop when it is not accepted successfully in J_m+1 transmissions, must be small enough to satisfy the quality of service requirement, by ignoring the effect of the term $P\big(e_r(J_m+1)\big)$, we have

$$T_r^{gbn} < 1 + \left(1+\frac{D\tau}{n+N_H}\right)\sum_{J=1}^{J_m} P\big(e_r(J)\big). \tag{5.73}$$

Using the bounds (5.57) and (5.53), we obtain the lower bound on throughput for GBN-HARQ scheme to be

$$\eta^{gbn}(J_m) > \frac{m}{n + N_H + (n + N_H + D\tau) \sum_{J=1}^{J_m} Z(J)}. \tag{5.74}$$

By using the same approach as that for (5.63), the upper bound on throughput can be shown to be

$$\eta^{gbn}(J_m) < \frac{R_e}{1 + \left(1 + \frac{D\tau}{n+N_H}\right) \left[\sum_{J=1}^{J_m} \prod_{t=1}^{J} P\big(e_p(t)\big) - J_m P\big(e_p(J_m+1)\big)\right] - P\big(e_p(J_m+1)\big)}. \tag{5.75}$$

The right-hand side of (5.75) can be approximated by using the approximation in (5.54) for $P\big(e_p(J)\big)$.

5.4.4 Numerical Results for Type-I HARQ

Since the bounds on goodput and APER depend on the values of n and m and they are not influenced by specific code structures, different values of n and m are chosen to demonstrate the performances. A $(50, 45)$ linear block code is adopted in the section. For packet ARQ schemes, the round-trip delay $D + (n + N_H)/\tau$ is assumed to be equal to the amount of time for transmitting N packets. Thus, the term $D\tau$ can be evaluated from the equation: $N = 1 + D\tau/(n + N_H)$. The average received SNR per symbol per diversity channel is $\gamma_s = 2E_s\sigma^2/N_0$, while the average received SNR per data bit is γ_b where $\gamma_b = L\gamma_s$ for BPSK modulation, and $\gamma_b = L\gamma_s/2$ for QPSK. By taking into account the energy devoted to channel estimation, the effective received SNR per data bit γ_b^{eff} is defined to be $2E_b^{\text{eff}}L\sigma^2/N_0$, from (5.9). Regarding GBN, the energy consumption for retransmission is not taken into consideration in the effective energy per data bit [10] as it is impossible to obtain the exact number

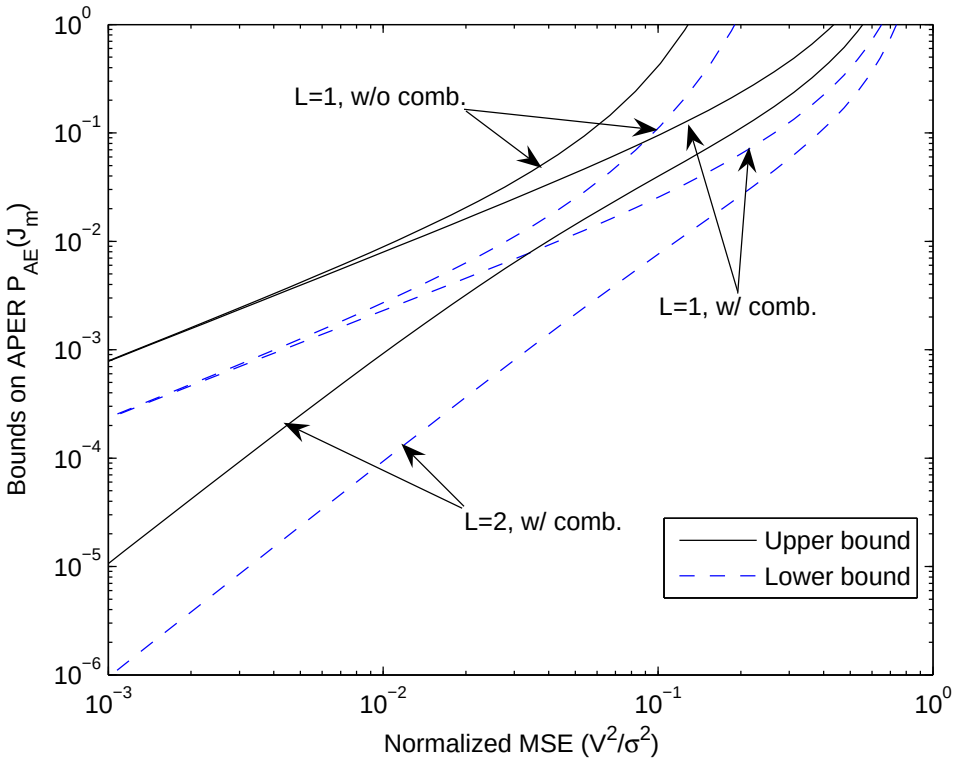

Figure 5.5: The APER versus the NMSE for basic ARQ without packet combining (w/o comb.) and Type-I HARQ with packet combining (w/ comb.)

of retransmissions of each packet for energy consumption calculation. However, the effect of the retransmission has been taken into consideration in the throughput performance. The normalized MSE (NMSE) of channel estimation V^2/σ^2 depends on the energy devoted to pilot symbols rather than on the value of N_H. The number of pilot symbols which is limited by the allowable channel bandwidth, is assumed to be $N_H = 5$. With equal power allocation, it is assumed that the energy per pilot bit equals the energy per data bit, i.e. $E_p = E_b$. The optimum power allocation between E_p and E_b of pure ARQ schemes for achieving a maximum goodput has been presented in [106].

In Fig. 5.5 and Fig. 5.6, the dependence of the reliability of packet ARQ schemes

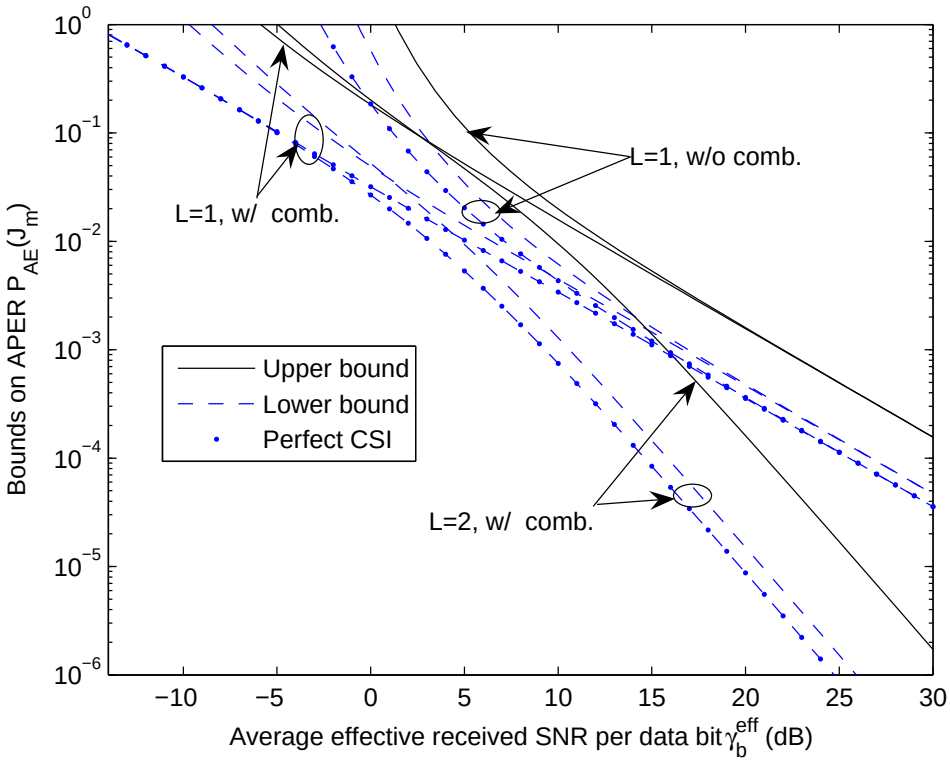

Figure 5.6: The APER versus the effective received SNR per bit for basic ARQ (w/o comb.) and Type-I HARQ (w/ comb.)

on the accuracy of the channel estimation and the effective received SNR per data bit is studied. The upper and lower bounds on APER of both basic ARQ schemes and HARQ schemes with BPSK modulation are illustrated. For basic ARQ schemes, the transmitter is allowed to request retransmissions till the packet is accepted successfully. Therefore, the maximum number of transmissions can theoretically be infinite. For fair comparison between basic ARQ schemes and HARQ schemes, the maximum transmission limit $J_m + 1$ of HARQ is assumed to be infinite. The two figures indicate that the upper and lower bounds are close to one another and follow the similar trend. Therefore, these bounds enable the behavior of the exact performance parameters to be investigated. Seen from the two figures, there exists a critical value for both V^2/σ^2

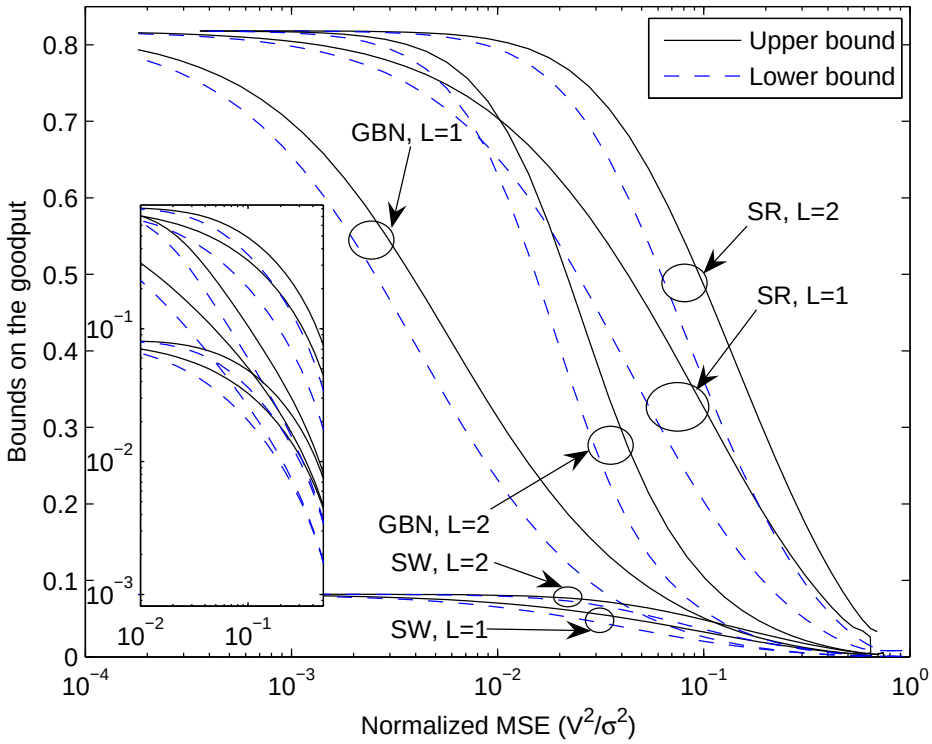

Figure 5.7: The goodput versus the NMSE for SR, GBN, and SW based Type-I HARQ with $N = 10$.

and γ_b^{eff} that separates two different trends in the APER curves. In particular, as can be seen from Fig. 5.5, the APER of the basic ARQ scheme deteriorates very fast at large NMSE, say, larger than around 10^{-2}. For HARQ scheme with first order spatial diversity $L = 1$, the APER is less sensitive to the MSE, hence it deteriorates fast only at higher values of NMSE, say, around 10^{-1}. With second order spatial diversity $L = 2$, the APER continues to be less sensitive. A similar change in the slop of the APER as a function of γ_b^{eff} can be observed in Fig. 5.6. Hence, the spatial diversity achieved by using multiple receive antennas as well as the time diversity achieved by using packet combining are capable of maintaining a significant, low APER over a wide range of NMSE of the channel estimation. When the channel gain

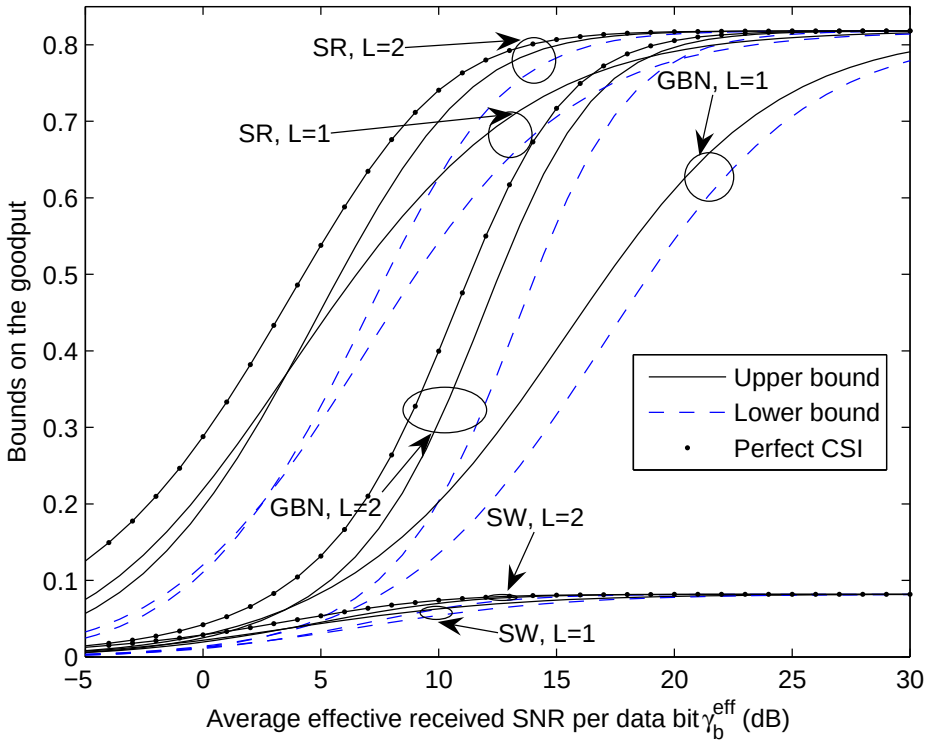

Figure 5.8: The goodput versus the effective received SNR per bit for Type-I HARQ schemes with $N = 10$.

is not estimated accurately, with the same order spatial diversity, HARQ achieves a smaller APER than basic ARQ does because of its inherent capability of achieving time diversity, whereas, it performs close to the basic ARQ when the channel is estimated accurately. In other words, with a larger NMSE, the time diversity of the HARQ is more beneficial in terms of decreasing the APER. With the spatial diversity achievable by multiple antennas in HARQ schemes, when the CSIR is estimated accurately, the APER can be decreased more dramatically by using more receive antennas, however, the improvement is diminished when V^2/σ^2 increases. This is because the effect of diversity combining using multiple antennas cannot be achieved when the CSIR is poorly estimated.

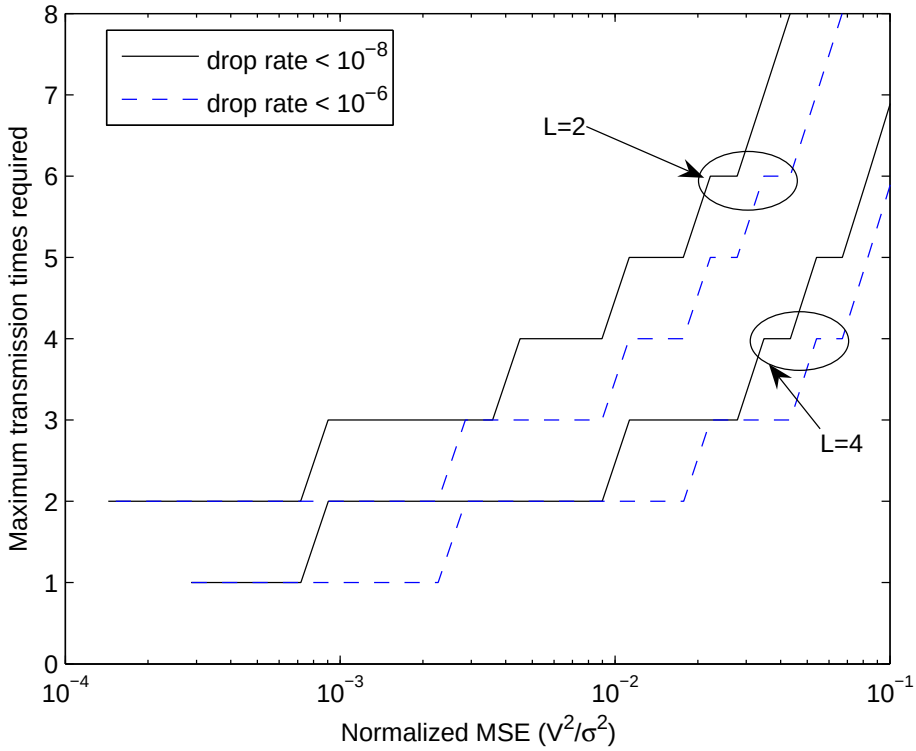

Figure 5.9: The number of necessary transmissions for achieving a drop rate less than 10^{-8} and 10^{-6} for Type-I HARQ versus the NMSE.

Fig. 5.7 and Fig. 5.8 show how the accuracy of the channel estimation NMSE V^2/σ^2 and the effective SNR γ_b^{eff} affect the goodput performance. The closed-form expressions of the bounds on the goodput are nonlinear functions of the NMSE and the effective SNR. They enable us to study numerically the dependence of the goodput on the NMSE and the effective SNR. At the large MSE, the performance deteriorates very rapidly with increasing values of MSE, while at low values of the MSE, the performance improves gently with decreasing values of MSE, i.e., decreasing the NMSE leads to diminishing increments in the goodput. A similar observation with regard to the effective SNR can be seen from Fig. 5.8. Moreover, to achieve a specified goodput performance requirement, SR-HARQ requires less energy devoted

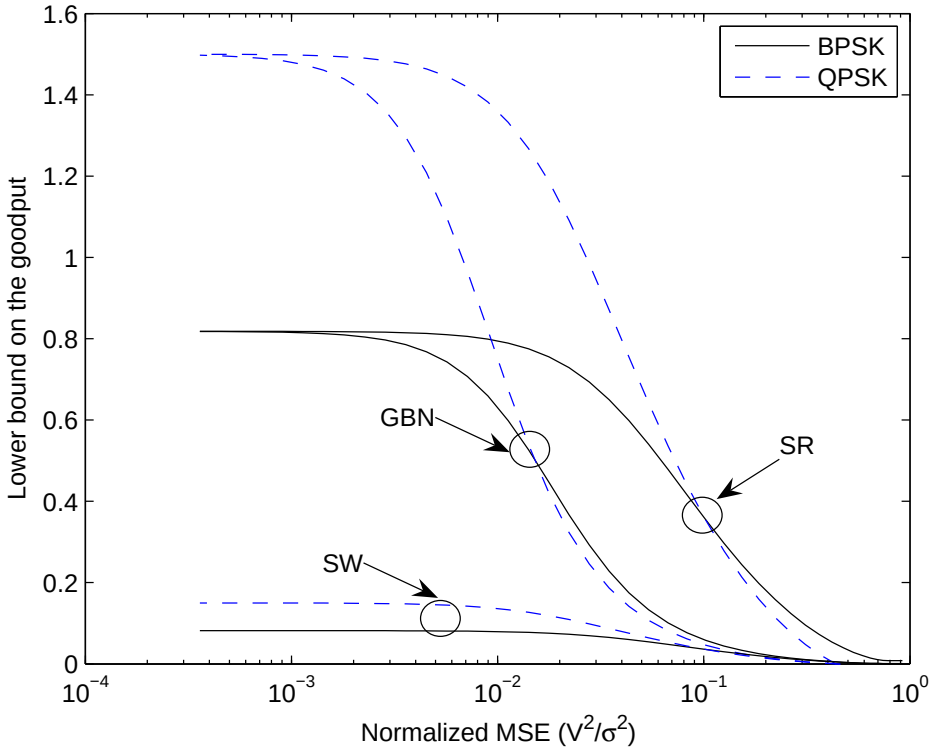

Figure 5.10: The comparison of lower bounds on goodput of Type-I HARQ schemes with BPSK and QPSK when $L = 2$.

to channel estimation than GBN-HARQ does, since its goodput is less sensitive to the accuracy of the channel estimation. However, the implementation of SR-HARQ is more complicated than that of GBN-HARQ.

In practical packet ARQ systems, the number of necessary transmissions is usually limited. To achieve a specified requirement on the drop rate, the number of necessary transmissions is shown in Fig. 5.9 versus the accuracy of the channel estimation. For example, to achieve a drop rate less than 10^{-8}, with $L = 2$, the number of necessary transmissions increases dramatically when the value of NMSE is larger than around 0.7. Using more receive antennas or alleviating the drop rate requirement can reduce the number of necessary transmissions.

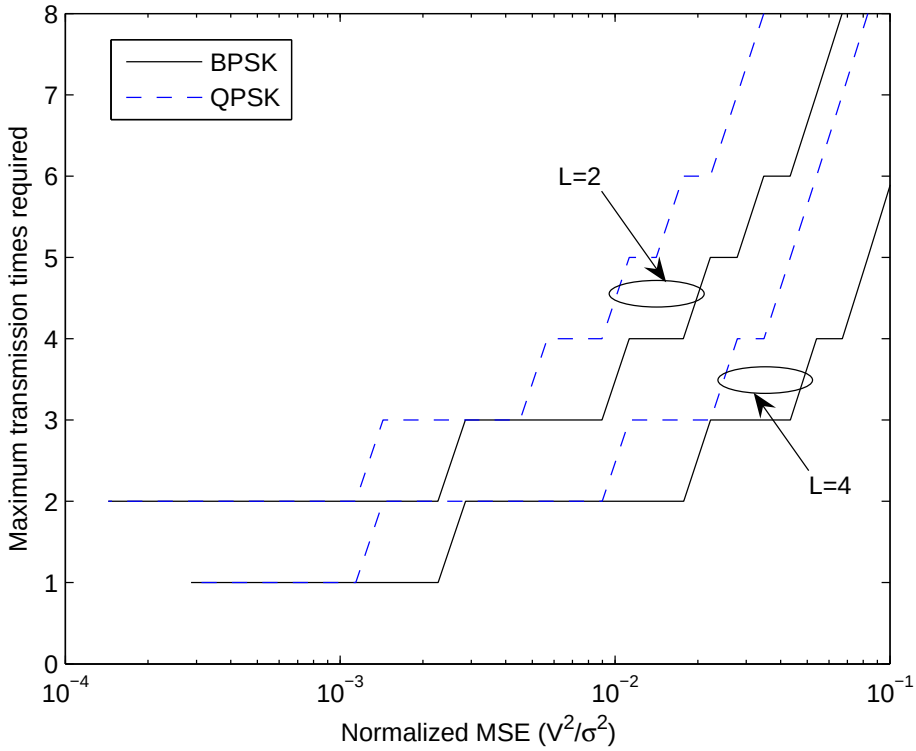

Figure 5.11: The comparison of the necessary number of transmissions of Type-I HARQ with BPSK and QPSK for achieving a drop rate less than 10^{-6}.

For illustrating the performance loss due to imperfect CSIR, the performance parameters: APER and goodput, with perfect CSIR versus the effective SNR are plotted in Fig. 5.6 and Fig. 5.8, respectively. We assume that perfect CSIR is provided by a genie-aided receiver without using any pilot symbols. Thus, the effective received SNR is obtained from (5.9) by making $E_p = 0$. The bounds on the APER and the goodput are computed from the closed-form expressions by setting $V^2 = 0$. Compared with the case of perfect CSIR, the packet ARQ schemes with imperfect CSIR have a performance loss which is due to the greater energy consumed for channel estimation resulting in a higher γ_b^{eff}, as well as the non-zero channel estimation error V^2/σ^2. As the effective SNR γ_b^{eff} increases, the performance gap decreases because of the

decreased value of V^2/σ^2. At the high γ_b^{eff} region, for the APER, the performance loss is constant and irreducible, and is due to the higher γ_b^{eff} resulting from the energy devoted to the channel estimation. However, for the goodput, the performance loss can be reduced to zero at the high γ_b^{eff} region. This is because the goodput is approaching a limiting value with the increasing γ_b^{eff}.

In Figs. 5.5–5.9, all the performance curves are for the case of BPSK modulation. For comparison, the performance curves of HARQ schemes with QPSK modulation are also shown in Figs. 5.10–5.11. The expressions of the APER for QPSK can be obtained by substituting $\alpha = \pi/4$ and $E_s = 2E_bR$ into (5.54), (5.65) and (5.66). One can easily show that the results can be obtained by simply replacing n by $n/2$ in (5.61), (5.70), and (5.74), respectively. As expected, at low values of V^2/σ^2, packet ARQ schemes with QPSK modulation provide a higher thoughput than those with BPSK modulation. However, the imperfect CSIR with a large value of V^2/σ^2, say, larger than around 10^{-1}, leads to a lower goodput for QPSK modulation than for BPSK modulation. To achieve a specified requirement on the drop rate, a larger number of necessary transmissions is needed for QPSK modulation than for BPSK modulation.

With imperfect CSIR, we have studied the effect of channel estimation accuracy on the performance of ARQ and HARQ systems over block Rayleigh fading channels. Our work provides a systematic approach whereby the link-layer performance parameters can be evaluated in terms of the parameters at the lowest physical-layer. While the closed-form expressions of the bounds on the APER, goodput and drop rate are nonlinear functions of the NMSE, they enable the system designer to study numerically the dependence of the link-layer performance parameters on the NMSE and the effective SNR, for any given (n, m) linear block code and any modulation format for transmitting the code bits. A key physical-layer parameter that plays an implicit but crucial role in the analysis here is the channel bandwidth. The bandwidth,

together with the code rate, determines the allowable number of pilot symbols per packet, which in turn determines the required SNR for achieving the desired channel estimation NMSE that leads to the target link-layer performance level.

5.5 Basic ARQ with BDPSK in SIMO Systems

Most of the existing ARQ schemes assume that the receiver has perfect CSI, and utilize the CSI for coherent detection [10, 11, 13, 22–24, 76, 100]. However, it is rather costly or even infeasible to obtain accurate channel estimates especially in rapid fading environments. In the differential transmission scheme, the information bits are differentially encoded and transmitted while at the receiver the signals are differentially decoded by using the previous received symbol. DPSK is a commonly used alternative modulation to BPSK, because it does not require additional pilot symbols for channel estimation for the acquisition of CSI used. But its performance is degraded because of only one previous received signal is used for the detection of the current signal. Therefore, it is necessary to study the performance of ARQ schemes with DPSK modulation. The PEP performance of an ARQ scheme for packet transmission depends on the modulation format and the channel characteristics, among other factors. For the case of a static channel, we derive the exact closed-form expressions as well as tight bounds for the PEP and the goodput. For the case of a slowly time-varying channel, the performance can only be studied via simulations.

We have considered in [106] the impact of imperfect receiver CSI on the performance of FEC-ARQ schemes over a static, nonselective, Rayleigh fading channel with BPSK modulation. There, we have shown a crucial dependence of the performance on the accuracy of the receiver CSI. DBPSK is a commonly used alternative modulation to BPSK, because it does not require channel estimation for the

acquisition of CSI. This section thus extends the work of [106] to the case of DBPSK modulation. In addition to the case of a static channel, the effects of fading fluctuations in a time-varying channel on the performance are also examined.

In ARQ schemes, information bits are grouped into packets for transmission. Therefore, the PEP is an important performance measure of an ARQ scheme. In some works, the PEP is also known as the frame error probability. We study the PEP of ARQ schemes here with DBPSK over a nonselective Rayleigh fading channel. For the case of a static channel, we derive the exact closed-form expressions as well as tight bounds for the PEP and the goodput. For the case of a slowly time-varying channel, the performance is studied via simulations.

When information is transmitted using an ARQ scheme, each information stream with m bits is encoded by a (n, m) CRC code into a packet of n bits for transmission. Without the use of pilot bits ($N_H = 0$), all the n bits of each packet are transmitted by DBPSK modulation. At the receiver, the received packet is then demodulated and the detected bits are checked by the error detection code. When the receiver detects an error in a received packet, a NAK is sent to the transmitter. The feedback channel is assumed to be error-free. The round-trip delay is assumed to be larger than the coherence time of the channel. Hence, the fading gain experienced by the retransmitted packet is independent of the gains experienced in previous transmission(s), and they are identically distributed.

We consider a SIMO system with L receive antennas and a single transmit antenna. The received signal at the lth receive antenna over the kth symbol interval, is given by

$$r_l(k) = \sqrt{E_s} m(k) h_l(k) + n_l(k), \quad k = 1, \cdots, n, \tag{5.76}$$

where, E_s is the energy per symbol and the term $m(k)$ is the transmitted data symbol. The channel gain for the lth antenna $\{h_l(k)\}_k$ is a sequence of complex

Gaussian random variables (CGRVs) with $\mathbb{E}[h_l(k)] = 0$ and autocorrelation function $\mathbb{E}[h_l(k)h_l^*(k+t)] = 2R(t)$. The channel gains $\{h_l(k)\}$ and $\{h_j(k)\}$ are independent of each other for $l \neq j$, and are identically distributed. The AWGN for the lth receive antenna $\{n_l(k)\}_k$ is a sequence of CGRVs with $\mathbb{E}[n_l(k)] = 0$ and $\mathbb{E}[|n_l(k)|^2] = N_0$. Terms $\{n_l(k)\}$ and $\{n_j(k)\}$ are independent of each other for $l \neq j$, and are identically distributed.

5.5.1 Packet Error Probability

The normalized fading correlation coefficient between two adjacent symbols is denoted by $\rho = R(1)/R(0)$. It is a measure of the fluctuation rate of the channel fading process. For the case of a static channel ($\rho = 1$), the exact closed-form expression and a tight upper bound for the PEP $P(e_p)$ will be derived in this section. However, for the case of a slowly time-varying channel ($\rho < 1$), the PEP can only be studied via simulation in Section 5.5.3.

For the case of a static channel ($\rho = 1$), channel gain $h_l(k)$ remains constant over the duration of a packet for each receive antenna. For the sake of brevity, in the rest of this section, we omit the symbol index k in $h_l(k)$ as the static channel is considered. Conditioned on fixed values of the gains $\{h_l\}_{l=1}^{L}$, the conditional BEP of DBPSK is given by [109, eq.(9A.15)]

$$P(e_b|\{h_l\}) = \frac{1}{2} + \frac{1}{2^{2L-1}} \sum_{l=1}^{L} \binom{2L-1}{L-l} \left(Q_l(\alpha,\beta) - Q_l(\beta,\alpha)\right), \tag{5.77}$$

where, $Q_m(\alpha,\beta)$ is mth order Marcum Q-function with

$$\alpha = 0, \text{ and } \beta = \left(2\sum_{l=1}^{L} \frac{E_s|h_l|^2}{N_0}\right)^{1/2}.$$

The instantaneous, received, effective SNR is defined to be

$$\gamma_s = \sum_{l=1}^{L} \frac{E_s |h_l|^2}{N_0}, \tag{5.78}$$

which can be derived from (5.9) by letting $N_H = 0$ because there are no pilot symbols for channel estimation. Furthermore, it can be easily shown that

$$\frac{1}{2^{2L-1}} \sum_{l=1}^{L} \begin{pmatrix} 2L-1 \\ L-l \end{pmatrix} = \frac{1}{2}. \tag{5.79}$$

Using the identity $Q_m(\beta, 0) = 1$ [109, eq.(4.75)] as well as the identity [109, eq.(4.73)]

$$Q_m(0, \beta) = \sum_{n=0}^{m-1} \exp\left(-\frac{\beta^2}{2}\right) \frac{(\beta^2/2)^n}{n!},$$

after simple manipulation, we can simplify the conditional BEP of DBPSK in (5.77) into

$$P(e_b|\{h_l\}) = \frac{1}{2^{2L-1}} \sum_{l=1}^{L} \begin{pmatrix} 2L-1 \\ L-l \end{pmatrix} \sum_{n=0}^{l-1} e^{-\gamma_s} \frac{\gamma_s^n}{n!}. \tag{5.80}$$

Conditioned on knowing the channel gain h_l, the channel is memoryless since the AWGN is independent from symbol to symbol. The occurrences of bit errors are independent of one another. Therefore, the conditional packet error probability $P(e_p|\{h_l\})$ that a received packet contains at least one error bit, can be written as

$$P(e_p|\{h_l\}) = 1 - (1 - P(e_b|\{h_l\}))^n. \tag{5.81}$$

Because of the dependence of the conditional BEP in (5.80) on the instantaneous received effective SNR γ_s, the conditional PEP also depends on the chi-square-distributed random variable γ_s, which has a PDF [6, eq.(2-1-110)]

$$p(\gamma_s) = \frac{1}{\bar{\gamma}_s^L \Gamma(L)} \gamma_s^{L-1} e^{-\gamma_s/\bar{\gamma}_s}, \tag{5.82}$$

where, the term $\bar{\gamma}_s = 2E_sR(0)/N_0$ is defined to be the mean received SNR per symbol per channel. By averaging (5.81) over the PDF of γ_s, the average PEP is given by

$$P(e_p) = \int_0^\infty P(e_p|\{h_l\})p(\gamma_s)d\gamma_s. \tag{5.83}$$

We consider the following two special cases to get a closed form for the integral in (5.83) to evaluate the PEP .

PEP for Single Receive Antenna: Case (a)

Considering the single-input-single-output (SISO) system of $L = 1$, the conditional BEP in (5.80) is reduced to

$$P(e_b|h_1) = \frac{1}{2}e^{-\gamma_s}, \tag{5.84}$$

which agrees with [6, eq.(12.1-15)], and $\gamma_s = E_s|h_1|^2/N_0$. The PDF of γ_s is obtained by substituting $L = 1$ in (5.82) as

$$p(\gamma_s) = \frac{1}{\bar{\gamma}_s}e^{-\gamma_s/\bar{\gamma}_s}. \tag{5.85}$$

Thus, the PEP for a SISO system is given by

$$P^{(1)}(e_p) = 1 - \int_0^\infty \left(1 - \frac{1}{2}e^{-\gamma_s}\right)^n p(\gamma_s)d\gamma_s \tag{5.86}$$

By using integration by parts with $u = (1-\frac{1}{2}e^{-\gamma_s})^n$ and $dv = de^{-\gamma_s/\bar{\gamma}_s}$, (5.86) becomes

$$P^{(1)}(e_p) = 1 - 2^{-n} - \frac{n}{2}\int_0^\infty\left(1 - \frac{1}{2}e^{-\gamma_s}\right)^{n-1} e^{-\gamma_s(1+\bar{\gamma}_s^{-1})}d\gamma_s.$$

By continuing the integration by parts till the last integral, $P^{(1)}(e_p)$ further comes to

$$\begin{aligned} P^{(1)}(e_p) = 1 - 2^{-n} - \frac{n2^{1-n}}{2(1+\bar{\gamma}_s^{-1})} - \frac{n(n-1)2^{2-n}}{2(1+\bar{\gamma}_s^{-1})2(2+\bar{\gamma}_s^{-1})} \\ - \cdots - \frac{n(n-1)\cdots 1}{2(1+\bar{\gamma}_s^{-1})2(2+\bar{\gamma}_s^{-1})\cdots 2(n+\bar{\gamma}_s^{-1})}. \end{aligned} \tag{5.87}$$

After some manipulation, we obtain the PEP in closed form as

$$P^{(1)}(e_p) = 1 - 2^{-n}\left(1 + \sum_{k=1}^{n}\prod_{j=1}^{k}\frac{n-j+1}{j+1/\bar{\gamma}_s}\right). \tag{5.88}$$

PEP for Multiple Receive Antennas: Case (b)

For the SIMO system with $L > 1$, we can only derive an upper bound in closed form for the PEP. The conditional BEP $P(e_b|\{h_l\})$ in (5.80) is smaller than 1, i.e., $P(e_b|\{h_l\}) < 1$. When x is much smaller than 1, we have the approximation [97, eq.(1.511)]:

$$\ln(1-x) \approx -x. \tag{5.89}$$

Multiplying both sides of (5.89) by a factor n, we have

$$\ln(1-x)^n = n\ln(1-x) \approx -nx, \tag{5.90}$$

which gives the approximation that

$$(1-x)^n \approx e^{-nx}. \tag{5.91}$$

Using the result: $(1-x)^n \approx e^{-nx}$, for $0 < x << 1$, we can approximate the average PEP in (5.83) as

$$P(e_p) \approx 1 - \int_0^\infty e^{-nP(e_b|\{h_l\})} p(\gamma_s) d\gamma_s. \tag{5.92}$$

Since e^x is a convex function in x, applying the Jensen's inequality, we have

$$\int_0^\infty e^{-nP(e_b|\{h_l\})} p(\gamma_s) d\gamma_s \geq e^{-n\int_0^\infty P(e_b|\{h_l\})p(\gamma_s)d\gamma_s}. \tag{5.93}$$

The term $\int_0^\infty P(e_b|\{h_l\})p(\gamma_s)d\gamma_s$ in (5.93) is the average BEP $P(e_b)$. By substituting (5.80) and making use of the result [97, eq.(3.351-3)]

$$\int_0^\infty x^n e^{-\mu x} dx = n!\mu^{-n-1},$$

the average BEP can be obtained as

$$P(e_b) = \frac{(1+\bar{\gamma}_s)^{-L}}{2^{2L-1}\Gamma(L)} \sum_{l=1}^{L} \begin{pmatrix} 2L-1 \\ L-l \end{pmatrix} \sum_{n=0}^{l-1} \frac{1}{n!} \frac{(n+L-1)!\bar{\gamma}_s^n}{(1+\bar{\gamma}_s^n} \tag{5.94}$$

which agrees with [6, eq.(14.4-26)]. Substituting (5.94) and (5.93) in (5.92), we obtain an approximate upper bound on the average PEP in closed form as

$$P(e_p) \leq P^{(L)}(e_p) = 1 - e^{-nP(e_b)}. \tag{5.95}$$

5.5.2 Goodput Analysis of ARQ Schemes

In this section, the performance of three basic ARQ schemes are analyzed in a SIMO system. Whenever a received packet is detected in error, that packet is discarded and replaced by a retransmitted copy. The APER showing the reliability of the ARQ system is given by (5.28). The probability of detectable error or the probability of retransmission $P(e_d)$ can be obtained by omitting the index J in (5.4), since the received Jth erroneous packet is discarded without packet combining. In Section 5.5.1, the PEP $P(e_p)$ has been evaluated for different cases. Next, we evaluate the probability of undetectable error $P(e_u)$. An exact determination of the $P(e_u)$ of a CRC code requires knowledge of the weight distribution of the code. As this is usually unavailable, the probability of undetectable error can be upper bounded by [10, Theorem 4.11]

$$P(e_u) \leq 2^{-n(1-R)}. \tag{5.96}$$

When the number of parity-check bits $n-m$ is large enough so that we have $P(e_u) \approx 0$. Therefore $P(e_d) \approx P(e_p)$. By substituting (5.96) and $P(e_d) \approx P(e_p)$ into (5.28), the P_{AE} is then upper bounded by

$$P_{AE} \leq \frac{2^{-n(1-R)}}{1 - P(e_p)}. \tag{5.97}$$

By substituting (5.88) and (5.95) for $P(e_p)$ in (5.97), we have

$$P_{AE} \leq \begin{cases} 2^{-n(1-R)}/(1 - P^{(1)}(e_p)), & L = 1, \quad (a) \\ 2^{-n(1-R)}/(1 - P^{(L)}(e_p)), & L > 1, \quad (b) \end{cases} \tag{5.98}$$

As another useful system parameter, the throughput, is investigated for the three basic ARQ schemes in the following.

SR-ARQ Schemes

An infinite receiver buffer is assumed to be available to store successful packets following a packet which is detected in error. The throughput of an SR-ARQ scheme is given by replacing R_e by R in (5.26). The rate of the error detecting code is $R = m/n$. The goodput defined in (5.31) shows the proportion of the throughput consisting of correct packets and can be obtained by setting $N_H = 0$ in (5.32). By substituting (5.88) and (5.95) for $P(e_p)$ in (5.32) with $N_H = 0$, the goodput can be lower bounded in closed form as

$$\eta_g^{sr} \geq \begin{cases} R(1 - P^{(1)}(e_p)), & L = 1, \quad (a) \\ R(1 - P^{(L)}(e_p)), & L > 1, \quad (b) \end{cases} \tag{5.99}$$

SW-ARQ Schemes

Let D be the idle time from the end of transmission of one packet to the beginning of transmission of the next. Let τ be the bit transmission rate which is defined as the number of bits transmitted per unit of time. In one round-trip delay time, which is defined as the time interval between the transmission of a packet and the receipt of an acknowledgment for that packet, the transmitter can transmit a total of $N = 1 + D\tau/n$ packets if it does not stay idle. By evaluating the average number of packets that the transmitter could have transmitted during the interval from the beginning of transmission of one packet to the receipt of a positive acknowledgment for that packet, the throughput of an SW-ARQ system can be given by setting $N_H = 0$ in (5.36). Since the APER merely depends on the channel error statistics and the choice of the error detecting code, the SW-ARQ scheme has the same reliability as the SR-ARQ scheme.

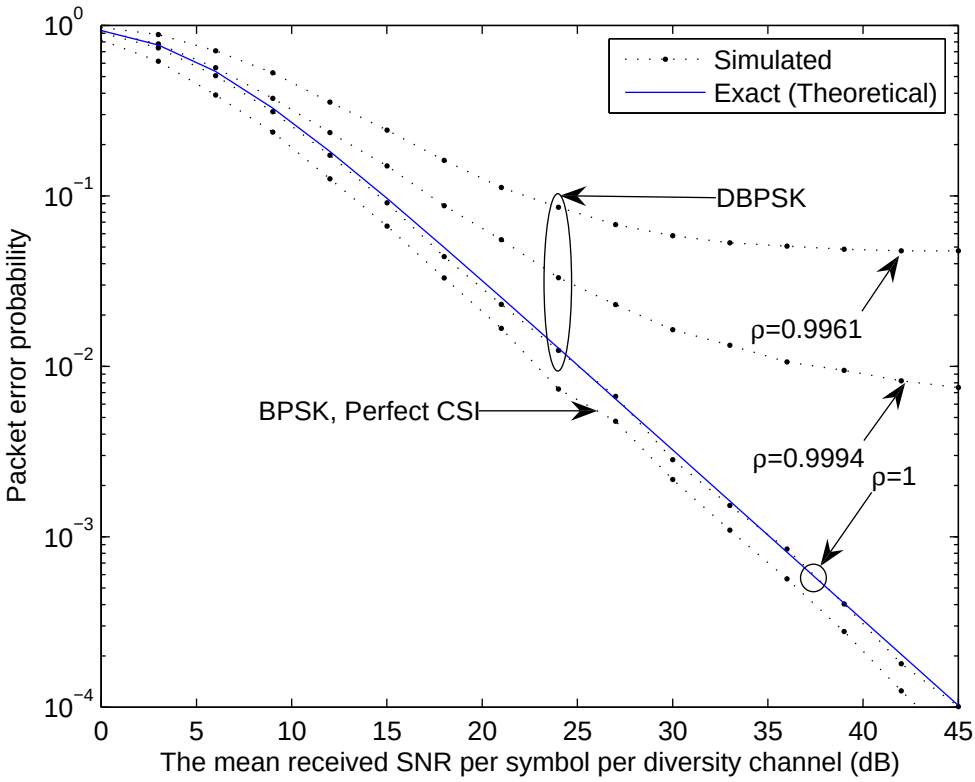

Figure 5.12: The PEP versus the average SNR $\bar{\gamma}_s$ with a single receive antenna.

Therefore, the goodput of a SW-ARQ scheme can be given by setting $N_H = 0$ in (5.38). By substituting (5.88) and (5.95) for $P(e_p)$ in (5.38) with $N_H = 0$ and replacing $D\tau$ by $(N-1)n$, the goodput of SW-ARQ schemes can be lower bounded by

$$\eta_g^{sw} \geq \begin{cases} R(1 - P^{(1)}(e_p))/N, & L = 1, \quad (a) \\ R(1 - P^{(L)}(e_p))/N, & L > 1, \quad (b) \end{cases} \tag{5.100}$$

GBN-ARQ Schemes

The term'go-back-N' derives from the fact that when a transmitter receives a NACK, it must go back into its buffer some N packets and restart transmission from there. In other words, the transmitter resends that packet and the $N-1$ subsequent packets

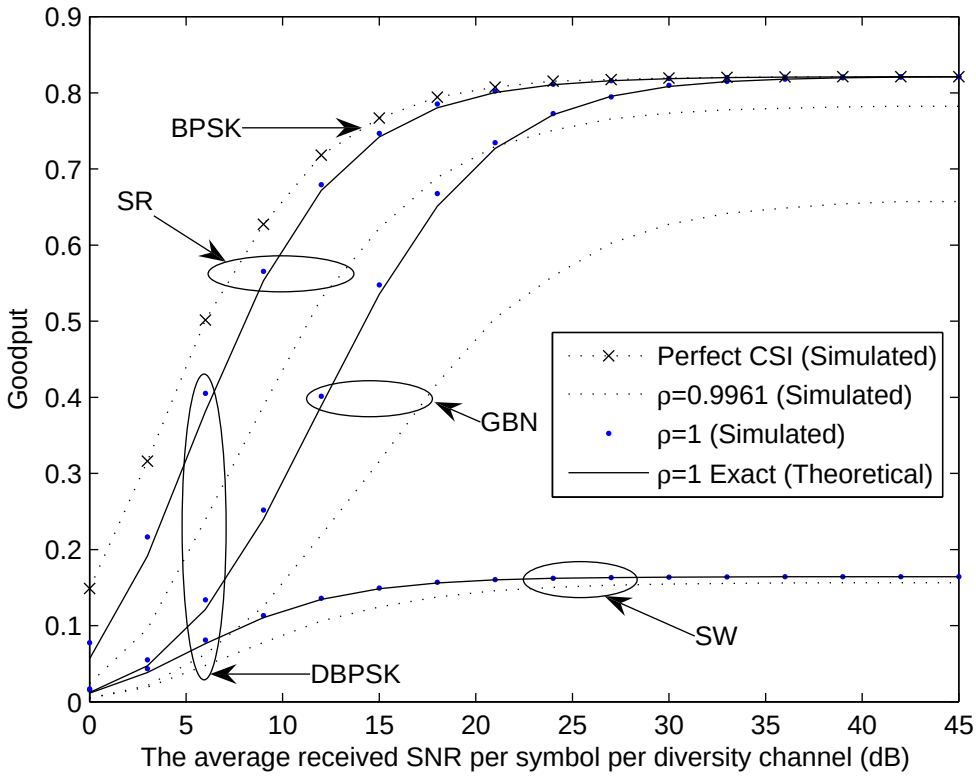

Figure 5.13: The goodput versus the average SNR $\bar{\gamma}_s$ with a single receive antenna.

that were transmitted earlier before the NACK is received. The parameter N depends on transmission rate τ and the round-trip delay $D + n/\tau$, which is evaluated by $N = 1 + D\tau/n$. Therefore, the throughput of a GBN-ARQ scheme is given by setting $N_H = 0$ in (5.39). Making use of the APER of GBN-ARQ which is the same as that of the SR-ARQ shown in (5.97), the goodput can be obtained to be (5.41) with $N_H = 0$. Since $P(e_d) \leq P(e_p)$ according to (5.4), substituting $P(e_p)$ for $P(e_d)$ in (5.41) with $N_H = 0$ gives an upper bound on the goodput:

$$\eta_g^{gbn} \leq \frac{(1 - P(e_p))m}{n + P(e_p)D\tau}. \tag{5.101}$$

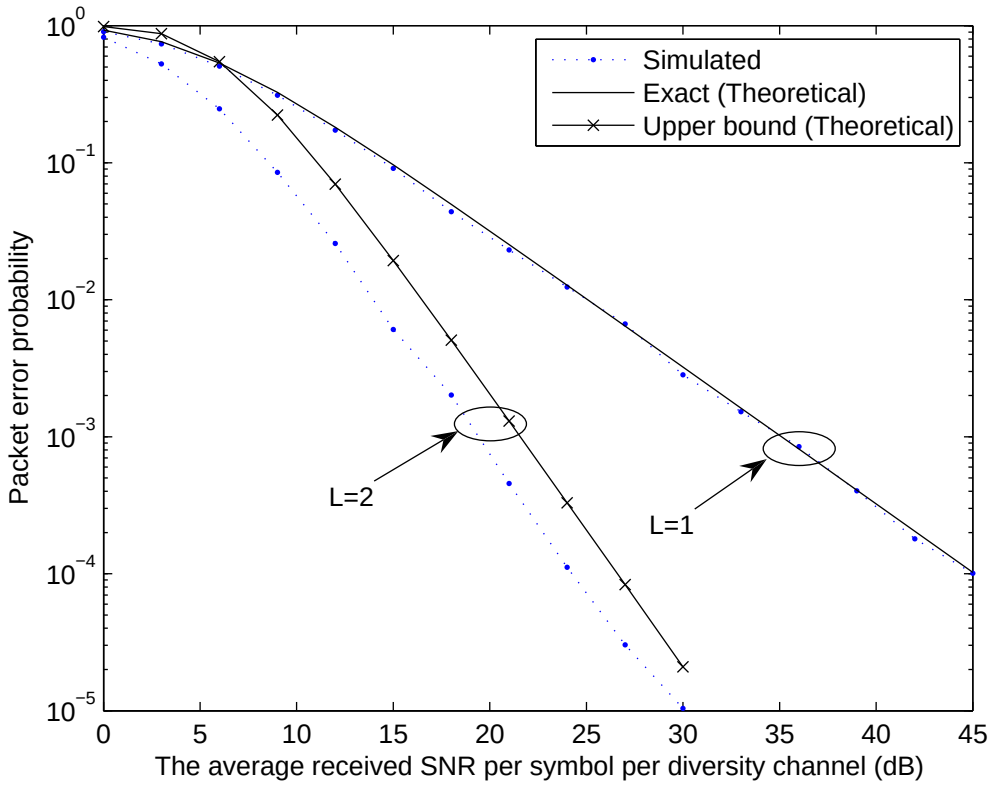

Figure 5.14: The PEP versus the average SNR $\bar{\gamma}_s$ when $\rho = 1$.

By substituting (5.88) and (5.95) for $P(e_p)$ in (5.101) and replacing $D\tau$ by $(N-1)n$, the goodput of SW-ARQ schemes can be lower bounded by

$$\eta_g^{gbn} \geq \begin{cases} R\frac{1-P^{(1)}(e_p)}{1+(N-1)P^{(1)}(e_p)}, & L=1, \quad (a) \\ R\frac{1-P^{(L)}(e_p)}{1+(N-1)P^{(L)}(e_p)}, & L>1, \quad (b) \end{cases} \tag{5.102}$$

5.5.3 Simulation and Numerical Results

In this section, we present simulation and numerical results for the performance of different ARQ schemes over Rayleigh fading channels. The autocorrelation coefficient of the fading process is given according to Jake's model by $2R(t) = 2R(0)J_0(2\pi f_d t T_s)$, where f_d is the maximum Doppler frequency and T_s is the symbol period. Therefore, we have the parameter $\rho = R(1)/R(0) = J_0(2\pi f_d T_s)$. We consider

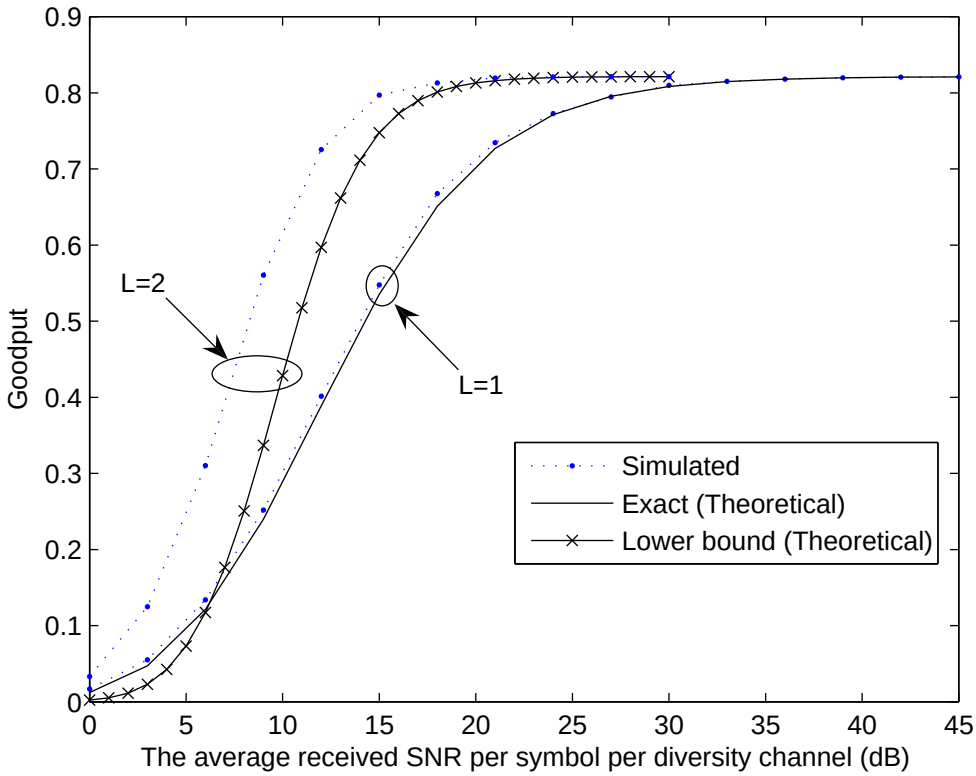

Figure 5.15: The goodput of GBN-ARQ versus the average SNR $\bar{\gamma}_s$ when $\rho = 1$.

three cases of the fading channels with different values of f_d. For static fading, we set $f_d = 0$ which gives $\rho = 1$. For time-varying channels, we use two values of f_d for $f_d T_s$. Hence, the term $f_d T_s$ is set to be 2×10^{-2} and 0.8×10^{-2} which gives $\rho = 0.9961$ and $\rho = 0.9994$, respectively. A $(28, 23)$ CRC code with the generator polynomial $g(x) = 1 + x^2 + x^5$ is used for error detection. It is assumed that the round-trip delay is the time during which $N = 5$ packets can be transmitted.

In Figs. 5.12 and 5.13, the PEP and goodput are shown for both static channels and time varying channels. For static channels, the PEP is computed based on (5.88) while the goodput is computed by (5.99.(a)), (5.100.(a)) and (5.102.(a)). The correctness of these numerical results is verified by the simulation results. For the

performance comparison with the results of DBPSK, the case of BPSK considered in [106] are also included in Figs. 5.12 and 5.13. For fair comparison, the performance of the basic ARQ scheme with BPSK is considered by assuming the perfect CSI provided by a genie without using any pilot symbols. It can be observed that SR-ARQ with DBPSK has a performance loss of around 2 dB in both PEP and goodput compared with that of ideal coherent BPSK. The reason is that, for time-varying channels, it is expected that an error-rate floor for PEP exists for $\rho < 1$, since it is well known that an inherent irreducible BEP floor exists for time-varying channels [110]. Furthermore, the irreducible error-rate floor of PEP results in a significantly reduced goodput due to the dependence of goodput on the PEP.

In Figs. 5.14 and 5.15, the exact results are plotted according to (5.88) and (5.100.(a)) while the performance bounds are computed based on (5.95) and (5.100.(b)), respectively. The tightness of the bounds and the correctness of the exact results are verified by simulation results. The two figures show that the spatial diversity achieved by multiple receive antennas is capable of achieving a significantly lower PEP and higher goodput over a wide range of SNR. In this section, we study the PEP of ARQ schemes with DBPSK over a nonselective Rayleigh fading channel. For the case of a static channel, we derive the exact closed-form expressions as well as tight bounds for the PEP and the goodput. The correctness and tightness of these closed-form expressions are verified by the simulation results. Furthermore, we observe that SR-ARQ with DBPSK has a performance loss of around 2 dB in both PEP and goodput performance, compared with that of ideal coherent BPSK. For the case of a slowly time-varying channel, the performance is studied via simulations. It is observed that there is an inherent irreducible error-rate floor for PEP which significantly reduces the goodput.

5.6 Conclusions

In this chapter, with imperfect CSIR, we have studied the effect of channel estimation accuracy on the performance of ARQ and HARQ systems over Rayleigh fading channels. Our work provides a systematic approach whereby the link-layer performance parameters can be evaluated in terms of the parameters at the lowest physical-layer. The results enable the system designer to study numerically the dependence of the link-layer performance parameters on the NMSE and the effective SNR, for any given (n, m) linear block code and any modulation format for transmitting the code bits.

Chapter 6

Goodput-Optimal Rate Adaptation with Imperfect CSIT and CSIR

Channel-adaptive rate allocation is a promising technique which makes use of the wireless fading channel in a more efficient manner. The transmission rates are adapted with respect to the channel conditions. Therefore, the CSI plays a crucial role in determining the performance of the systems. Rate adaptation schemes require channel state information CSI, which can be acquired at the receiver by inserting pilot symbols in the transmit signals. We develop here a rate adaptation scheme that takes account of both channel estimation and prediction errors. The transmitter adapts the transmission rate relative to the predicted channel state and a utilization factor. In turn, this utilization factor is optimized so as to achieve a maximum goodput.

6.1 Introduction

Channel-adaptive rate allocation is a promising technique which makes use of the wireless fading channel in a more efficient manner. The transmission rates are adapted with respect to the channel conditions. Compared with fixed rate transmission which

results in insufficient utilization of the full channel capacity, adaptive rate allocation can improve the system throughput significantly. When perfect CSIT is available, the transmission rate can be adapted relative to the channel state to maximize spectral efficiency [19, 111]. However, it is too optimistic in practice to assume the availability of perfect CSIT, since CSIT may be outdated or imperfect due to the feedback delays or the imperfect CSIR. The effect of the imperfect CSIT has attracted much attention since the CSIT plays a crucial role in determining the rates to adapt. By assuming perfect CSIR, the effect of imperfect CSIT due to the feedback delay has been investigated for adaptive transmitter designs in [32, 35, 112, 113] and the references therein. Considering a more practical case with imperfect CSIT and imperfect CSIR, the authors in [114] design an adaptive M-QAM pilot symbol assisted modulation (PSAM) to meet target performance and maximize spectral efficiency.

In contrast to the work in [114], in which the modulation is restricted to M-QAM, a continuous rate adaptation scheme following the approach proposed in [35] will be considered in this Chapter. Instead of assuming perfect CSIR [35], we consider the imperfect CSIT due to the channel prediction errors and the imperfect CSIR due to the channel estimation errors. While a strictly causal channel predictor is employed to predict the CSIT for adapting transmission rates, a noncausal channel estimator estimates the CSIR for coherent demodulation. The PSAM scheme is applied at the transmitter to facilitate the channel prediction and estimation at the receiver. Based on the predicted channel state and a utilization factor, the transmitter allocates the optimal rates. In turn, the utilization factor, which takes into account both channel estimation and prediction errors, is to be optimized in order to achieve maximum goodput. As a performance measure, goodput is the amount of data delivered to the receiver correctly per unit time, considering both the throughput and the reliability [26].

6.2 System Model

The adaptive system under consideration is described as follows. At the receiver, based on known pilot symbols, the channel estimator and channel predictor extract the pilot signal to estimate and predict the channel periodically. According to the predicted channel, the transmitter selects a transmission rate with constant transmission power E_s and fixed symbol rate T.

For signal transmission over a frequency nonselective Rayleigh fading channel with AWGN, the received signal in the kth symbol interval $kT \leq t < (k+1)T$ is given by

$$r[k] = s[k]h[k] + n[k], \tag{6.1}$$

where $s[k]$ is a signal symbol with constant energy $|s[k]|^2 = E_s$. The channel $\{h[k]\}_k$ is a sequence of zero-mean, circularly symmetric, complex, Gaussian random variables with covariance function

$$\mathbb{E}\left[h[k]h^*[k-j]\right] = 2R[j] = 2R[0]J_0(2\pi j f_d T), \tag{6.2}$$

where Jakes' model is used and $J_0(\cdot)$ represents the zero-order Bessel function of the first kind and f_d is the Doppler frequency. The noise $\{n[k]\}_k$ is a sequence of independent, identically distributed, zero-mean, complex Gaussian random variables with double sided power spectral density N_0.

6.3 PSAM Scheme with Channel Prediction and Channel Estimation

In a PSAM system, known pilot symbols are inserted periodically into the data sequence. It is assumed that each frame contains N symbols, with the first symbol

being the pilot. The parameters i and j represent the frame index and symbol index in each frame, respectively. The time index k can therefore be expressed as $k = iN + j$, where $j = 0$ represents the pilot symbol in each frame. The channel estimator and channel predictor extract the pilot signal to estimate and predict the channel periodically.

6.3.1 Channel Estimation

Without loss of generality, it is assumed that the estimated channel state of the current frame is obtained from the nearest $L_p + L_e + 1$ modulation-free pilot symbols. The L_p pilot symbols are transmitted in frames prior the current frame while the L_e symbols are transmitted after the current frame. After dividing the received signal by the known pilot symbols, the channel fading at the pilot symbol times can be expressed as

$$\mathbf{z}=\left[\frac{r[(i-L_p)N]}{s[(i-L_p)N]}\cdots\frac{r[iN]}{s[iN]}\quad\frac{r[(i+1)N]}{s[(i+1)N]}\cdots\frac{r[(i+L_e)N]}{s[(i+L_e)N]}\right]^T.$$

The function of channel estimation is to accurately recover the true fading gain $h[iN + j]$ of the jth data symbol in the ith frame from the pilot channel observation $\mathbf{z}$. Therefore the maximum *a posteriori* (MAP) estimation of the true fading gain denoted by $\hat{h}[iN + j]$ is given by [104, (2.1)]

$$\hat{h}[iN + j] = \mathbf{w}_o^T\mathbf{z}. \tag{6.3}$$

The corresponding optimum interpolation coefficient $\mathbf{w}_o$ can be obtained by using [115, (19)],

$$\begin{aligned}\mathbf{w}_o &= [w_o\,[(i-L_p)N]\cdots w_o\,[iN]\cdots w_o\,[(i+L_e)N]]^T\\ &= \left(E\left[\mathbf{z}\mathbf{z}^H\right]\right)^{-1}E\left[h^*[iN+j]\mathbf{z}\right]\\ &= \left[\mathbf{R}+\frac{N_0}{E_s}\mathbf{I}\right]^{-1}\mathbf{v},\end{aligned} \tag{6.4}$$

where, $\mathbf{R}$ is given by

$$\mathbf{R}=\begin{bmatrix} R[0] & R[N] & \cdots & R[(L_p+L_e)N] \\ R[N] & R[0] & \cdots & R[(L_p+L_e-1)N] \\ \vdots & \vdots & \vdots & \vdots \\ R[(L_p+L_e)N] & \cdots & \cdots & R[0] \end{bmatrix}$$

and $\mathbf{v}$ can be written as

$$\mathbf{v} = [R[L_pN+j]\cdots R[j]\cdots R[L_eN-j]]^T.$$

The MMSE of estimated channel can be expressed as [115, (20)]

$$\begin{aligned} 2\sigma_{\hat{h}}^2 &= E\left[|h[iN+j]-\hat{h}[iN+j]|^2\right] = E\left[|e[iN+j]|^2\right] \\ &= 2R[0] - \mathbf{v}^T\left[\mathbf{R}+\frac{N_0}{E_s}\mathbf{I}\right]^{-1}\mathbf{v}. \end{aligned} \tag{6.5}$$

Since the channel of interest is a Gaussian random process, $\hat{h}[iN+j]$ is Gaussian random variable with mean zero and variance $2(R[0]-\sigma_{\hat{h}}^2)$ [104, (2.18)]. Given $\hat{h}[iN+j]$, the gain $h[iN+j]$ is a conditional Gaussian random variable with mean $\hat{h}[iN+j]$ and variance $2\sigma_{\hat{h}}^2$. The channel estimation error $e[iN+j]$ is a zero-mean Gaussian random variable with variance $2\sigma_{\hat{h}}^2$ and is uncorrelated with $\hat{h}[iN+j]$.

6.3.2 Channel Prediction

Due to the feedback delay which takes account of both actually transmission delay and the processing time, channel prediction is needed so that the transmission rates can be adjusted properly. If the length of the linear MMSE channel predictor is L_p' and the feedback delay is a integer of NT (i.e. DNT), the predicted channel $\tilde{h}[iN+j]$ can then be obtained from a number (L_p') of observations of channel fading before the current transmission

$$\mathbf{z}' = \left[\frac{r[(i-D-L_p'+1)N]}{s[(i-D-L_p'+1)N]}\cdots\frac{r[(i-D)N]}{s[(i-D)N]}\right]^T.$$

Let $\tilde{h}[iN+j]$ be the predicted channel, the MMSE of $\tilde{h}[iN+j]$ is

$$\begin{aligned} 2\sigma_{\tilde{h}}^2 &= E\left[|h[iN+j]-\tilde{h}[iN+j]|^2\right] = E\left[|\varepsilon[iN+j]|^2\right] \\ &= 2R[0] - \mathbf{v}'^T\left[\mathbf{R}' + \frac{N_0}{E_s}\mathbf{I}'\right]^{-1}\mathbf{v}', \end{aligned} \tag{6.6}$$

where

$$\begin{aligned} \mathbf{v}' &= E\left[h^*[iN+j]\mathbf{z}'\right] \\ &= 2\left[R[(D+L_p'-1)N+j]\cdots R[DN+j]\right]^T, \end{aligned}$$

and $\mathbf{R}'$ can be evaluated from

$$\begin{aligned} \mathbf{R}' &= E\left[\mathbf{z}'\mathbf{z}'^H\right] - \frac{N_0}{E_s}\mathbf{I}' \\ &= \begin{bmatrix} R[0] & R[N] & \cdots & R\left[(L_p'-1)N\right] \\ R[N] & R[0] & \cdots & R\left[(L_p'-2)N\right] \\ \vdots & \vdots & \vdots & \vdots \\ R\left[(L_p'-1)N\right] & \cdots & \cdots & R[0] \end{bmatrix}. \end{aligned}$$

Similar to the estimated channel $\hat{h}[iN+j]$, the predicted channel $\tilde{h}[iN+j]$ is a Gaussian random variable with mean zero and variance $2(R[0]-\sigma_{\tilde{h}}^2)$. With a given $\tilde{h}[iN+j]$, the channel $h[iN+j]$ is Gaussian distributed with mean $\tilde{h}[iN+j]$ and variance $2\sigma_{\tilde{h}}^2$. The channel prediction error $\varepsilon[iN+j]$ is a zero mean Gaussian random variable with variance $2\sigma_{\tilde{h}}^2$ and is uncorrelated with $\tilde{h}[iN+j]$.

6.3.3 The Relationship Between Channel Estimation and Prediction

The following equations between the true channel and the estimation/prediction error

$$h[iN+j] = \hat{h}[iN+j] + e[iN+j]$$

$$h[iN+j] = \tilde{h}[iN+j] + \varepsilon[iN+j]$$

can be used to express the relationship between the estimated and predicted channels shown below

$$\hat{h}[iN+j] = \tilde{h}[iN+j] + \varepsilon[iN+j] - e[iN+j]. \tag{6.7}$$

The term $\tilde{h}[iN+j]$ and $\varepsilon[iN+j]$ are uncorrelated because of the orthogonality principle, however, $\tilde{h}[iN+j]$ is correlated with $e[iN+j]$. Before deriving the conditional statistical distribution of $\hat{h}[iN+j]$ conditioned on $\tilde{h}[iN+j]$, we give Lemma 6.1 first [116], which will be used to derive the conditional distribution of $\hat{h}[iN+j]$ later.

Lemma 6.1. If X and Y are jointly Gaussian circularly symmetric complex random variables with means $\mathbb{E}[X]=0$ and $\mathbb{E}[Y]=0$ and variances σ_X^2 and σ_Y^2, the covariance $\mu_{XY} = \mathbb{E}[XY^*]$ is shown to be a real number. Conditioned on X, Y is a conditional Gaussian random variable with conditional mean

$$\mathbb{E}[X|Y] = \rho_{XY}\frac{\sigma_X}{\sigma_Y}Y,$$

and conditional variance $\mathrm{Var}[X|Y] = \sigma_X^2(1-\rho_{XY}^2)$. The normalized covariance ρ_{XY} is defined to be

$$\rho_{XY} = \frac{\mu_{XY}}{\sigma_X\sigma_Y}.$$

Now we are ready to derive the conditional statistical distribution of $\hat{h}[iN+j]$ conditioned on $\tilde{h}[iN+j]$. Since $\tilde{h}[iN+j]$ and $\varepsilon[iN+j]$ are uncorrelated, in addition $\tilde{h}[iN+j]$ and $\hat{h}[iN+j]$ are jointly Gaussian random variables, according to Lemma 6.1, the conditional mean can be expressed as

$$\begin{aligned} m_{\hat{h}|\tilde{h}} &= \mathbb{E}[\hat{h}[iN+j]|\tilde{h}[iN+j]] \\ &= \tilde{h}[iN+j] + 0 - \mathbb{E}[e[iN+j]|\tilde{h}[iN+j]] \\ &= \left(1 - \frac{\mathbb{E}[e[iN+j]\tilde{h}^*[iN+j]]}{2(R[0]-\sigma_{\tilde{h}}^2)}\right)\tilde{h}[iN+j], \end{aligned} \tag{6.8}$$

and the conditional variance can be written as

$$\begin{aligned}2\sigma^2_{\hat{h}|\tilde{h}} &= \mathbb{E}[|\hat{h}[iN+j] - m_{\hat{h}|\tilde{h}}|^2|\tilde{h}[iN+j]] \\ &= 0 + \mathbb{E}[|\varepsilon[iN+j]|^2] + \mathrm{Var}[e[iN+j]|\tilde{h}[iN+j]] \\ &= 2\sigma^2_{\tilde{h}} + 2\sigma^2_{\hat{h}}\left(1 - \frac{(\mathbb{E}[e[iN+j]\tilde{h}^*[iN+j]])^2}{4(R[0]-\sigma^2_{\tilde{h}})\sigma^2_{\hat{h}}}\right). \end{aligned} \tag{6.9}$$

Therefore, with a given $\tilde{h}[iN+j]$, the amplitude of $\hat{h}[iN+j]$ has a Rician distribution with PDF in the form of [6]

$$p\left(|\hat{h}||\tilde{h}\right) = \frac{2|\hat{h}|}{\sigma^2_{\hat{h}|\tilde{h}}} e^{-(|\hat{h}|^2+|m_{\hat{h}|\tilde{h}}|^2)/(2\sigma^2_{\hat{h}|\tilde{h}})} I_0\left(\frac{2|\hat{h}||m_{\hat{h}|\tilde{h}}|}{\sigma^2_{\hat{h}|\tilde{h}}}\right). \tag{6.10}$$

6.4 Goodput-Optimal Rate Allocation

With a given estimated channel $\hat{h}[iN+j]$, the decision variable for $s[iN+j]$ is

$$d[iN+j] = \frac{s[iN+j](\hat{h}[iN+j] + e[iN+j]) + n[iN+j]}{\hat{h}[iN+j]}. \tag{6.11}$$

The post-detection SNR can then be expressed as the ratio of energy between signal and noise component shown below

$$\gamma_d[iN+j] = \frac{E_s|\hat{h}|^2}{2E_s\sigma^2_{\hat{h}} + N_0}, \tag{6.12}$$

Given the transmitted signal $s[iN+j]$, (6.11) is effectively an AWGN channel with SNR $\gamma_d[iN+j]$. In the rest of this section, the time indexes of $s[iN+j]$, $\hat{h}[iN+j]$ and $\tilde{h}[iN+j]$ are omitted for notational brevity. Hence, given an estimated channel knowledge $\hat{h}$, a lower bound on the mutual information is given by [117, eq.(2)]

$$\mathcal{C}_{low} = \log\left(1 + \frac{E_s|\hat{h}|^2}{2E_s\sigma^2_{\hat{h}} + N_0}\right). \tag{6.13}$$

In the rate allocation scheme, the transmitter makes use of the predicted channel and allocates a rate $\mathcal{R}$ as

$$\mathcal{R} = \log\left(1 + \frac{E_s|\tilde{h}|^2\lambda}{N_0}\right) \tag{6.14}$$

where λ is defined to be a utilization factor of the predicted channel $\tilde{h}$. When $\lambda = 1$, the predicted channel is used as if it was the true channel and the transmitter does not take advantage of the utilization factor of predicted channel at all. The goodput metric can be defined as [35]

$$\eta = \mathcal{R}P\left(\mathcal{C}_{low} > \mathcal{R}\right) \tag{6.15}$$

and can be used as a cost function to optimize the rate adaptation scheme. The probability $P\left(\mathcal{C}_{low} > \mathcal{R}\right)$ represents the probability that the rate $\mathcal{R}$ can be successfully delivered and is expressed by

$$P\left(\mathcal{C}_{low} > \mathcal{R}\right) = P\left(|\hat{h}| > \sqrt{\frac{(2E_s\sigma_{\hat{h}}^2 + N_0)\lambda}{N_0}}|\tilde{h}|\right). \tag{6.16}$$

The goodput in (6.15) can be expressed as

$$\begin{aligned}\eta &= \mathcal{R}P\left(|\hat{h}| > \sqrt{\frac{(2E_s\sigma_{\hat{h}}^2 + N_0)\lambda}{N_0}}|\tilde{h}|\right)\\ &= \log\left(1 + \frac{E_s|\tilde{h}|^2\lambda}{N_0}\right) Q_1\left(\frac{\sqrt{2}|m_{\hat{h}|\tilde{h}}|}{2\sigma_{\hat{h}|\tilde{h}}}, \frac{\sqrt{2A\lambda|\tilde{h}|^2}}{2\sigma_{\hat{h}|\tilde{h}}}\right)\end{aligned} \tag{6.17}$$

where $A = (2E_s\sigma_{\hat{h}}^2 + N_0)/N_0$ and $Q_1(a, b)$ is the first-order Marcum Q function. The goodput optimization problem is to maximize the product of the transmission rate and the outage probability, which can be formulated as

$$\max_{\lambda} \mathcal{R}P\left(\mathcal{C}_{low} > \mathcal{R}\right). \tag{6.18}$$

6.4.1 Optimal Solution λ_o^*

By taking the first derivative of (6.17) and making use of the integral expression for the Marcum Q function [6, (2.1-121)], the optimal utilization factor λ_o^* can be obtained

from the following equation

$$\frac{E_s|\tilde{h}|^2}{N_0+E_s|\tilde{h}|^2\lambda}Q_1\left(\sqrt{2}K_{11},\sqrt{2A\lambda}K_{12}\right)-\log\left(1+\frac{E_s|\tilde{h}|^2\lambda}{N_0}\right)$$
$$\times AK_{12}^2e^{-(K_{11}^2+A\lambda K_{12}^2)}I_0\left(2\sqrt{A\lambda}K_{13}\right)=0, \tag{6.19}$$

where

$$K_{11}=\frac{|m_{\hat{h}|\tilde{h}}|}{2\sigma_{\hat{h}|\tilde{h}}};\quad K_{12}=\frac{|\tilde{h}|}{2\sigma_{\hat{h}|\tilde{h}}};\quad K_{13}=\frac{|\tilde{h}||m_{\hat{h}|\tilde{h}}|}{2\sigma^2_{\hat{h}|\tilde{h}}},$$

and $I_0(\cdot)$ is the zeroth order modified Bessel function of the first kind.

6.4.2 Approximation of λ_o^*

The optimum λ_o calculated from (6.19) depends on many parameters and is complicated to obtain. In this subsection, an approximation of λ_o^* will be derived. The following expression appearing in (6.8) can be written as

$$\frac{\mathbb{E}[e[iN+j]\tilde{h}^*[iN+j]]}{2(R[0]-\sigma^2_{\tilde{h}})}=\frac{2\sigma^2_{\hat{h}}-\mathbb{E}[e\varepsilon^*]}{2(R[0]-\sigma^2_{\tilde{h}})}, \tag{6.20}$$

which can be upper and lower bounded by [114, (14)]

$$\frac{\sigma^2_{\hat{h}}-\sigma_{\hat{h}}\sigma_{\tilde{h}}}{R[0]-\sigma^2_{\tilde{h}}}\leq\frac{2\sigma^2_{\hat{h}}-\mathbb{E}[e\varepsilon^*]}{2(R[0]-\sigma^2_{\tilde{h}})}\leq\frac{\sigma^2_{\hat{h}}+\sigma_{\hat{h}}\sigma_{\tilde{h}}}{R[0]-\sigma^2_{\tilde{h}}}. \tag{6.21}$$

Since $2\sigma^2_{\hat{h}}$ and $2\sigma^2_{\tilde{h}}$ are very small and much smaller than $2R[0]$ in practical systems, the right-hand side of (6.20) is upper and lower bounded by a small value, which can be inferred from (6.21). Hence the value of (6.20) can be assumed to be zero. The conditional mean in (6.8) is therefore $m_{\hat{h}|\tilde{h}}\approx\tilde{h}$. Similarly, one of the terms in (6.9) which is bounded by

$$\frac{(\sigma_{\hat{h}}-\sigma_{\tilde{h}})^2}{R[0]-\sigma^2_{\tilde{h}}}\leq\frac{(\mathbb{E}[e[iN+j]\tilde{h}^*[iN+j]])^2}{4(R[0]-\sigma^2_{\tilde{h}})\sigma^2_{\hat{h}}}\leq\frac{(\sigma_{\hat{h}}+\sigma_{\tilde{h}})^2}{R[0]-\sigma^2_{\tilde{h}}}$$

can also be assumed to be zero. Hence, the conditional variance in (6.9) is $2\sigma^2_{\hat{h}|\tilde{h}}\approx 2(\sigma^2_{\tilde{h}}+\sigma^2_{\hat{h}})$. By applying $m_{\hat{h}|\tilde{h}}\approx\tilde{h}$ and $2\sigma^2_{\hat{h}|\tilde{h}}\approx 2(\sigma^2_{\tilde{h}}+\sigma^2_{\hat{h}})$ on (6.19), it can be observed

that $K_{11}^2 = K_{12}^2 = K_{13} = K$. Term $K = |\tilde{h}|^2/(2\sigma_{\tilde{h}}^2 + 2\sigma_{\hat{h}}^2)$ is the Rician factor of the Rician distribution shown in (6.10). The K actually represents the ratio of the energy of predicted channel to the energy of prediction and estimation error. Hence, K can be used as a quality indicator of the feedback channel. An approximation of λ_o^* can be obtained by numerically solving (6.19) with K_{11}, K_{12}, K_{13} replaced by $\sqrt{K}$, i.e.,

$$\frac{E_s|\tilde{h}|^2}{N_0 + E_s|\tilde{h}|^2\lambda} Q_1\left(\sqrt{2K}, \sqrt{2A\lambda K}\right) - \log\left(1 + \frac{E_s|\tilde{h}|^2\lambda}{N_0}\right)$$
$$\times AKe^{-(K+A\lambda K)} I_0\left(2\sqrt{A\lambda K}\right) = 0 \tag{6.22}$$

Therefore, the goodput-optimal rate allocation can be described by

$$\mathcal{R} = \log\left(1 + \frac{E_s|\tilde{h}|^2\lambda_o^*}{N_0}\right).$$

Remark: The MMSEs of the channel prediction and estimation $2\sigma_{\tilde{h}}^2$ and $2\sigma_{\hat{h}}^2$ do not depend on frame index i. It is shown in [114] that the MMSE depends only on E_s/N_0, f_dT and N when the number of filter coefficients approaches infinity.

6.5 Numerical Results

We consider a rate adaptive system which employs the predicted channel $\tilde{h}$ and the utilization factor λ_o^*. The channel is assumed to have a Jakes' Doppler spectrum. For a given estimator or a predictor, the MMSEs of estimation and prediction can be evaluated by (6.5) and (6.6), respectively. The values of $2\sigma_{\tilde{h}}^2$ and $2\sigma_{\hat{h}}^2$ have incorporated the effects of Doppler frequency f_d and pilot spacing N. For the sake of simplicity, $2R[0]$ is assumed to be 1 in this section so that $2\sigma_{\tilde{h}}^2$ and $2\sigma_{\hat{h}}^2$ become the normalized MMSEs which will be denoted as $2\sigma_{\varepsilon}^2$ and $2\sigma_{e}^2$ in the following figures, respectively.

In Fig. 6.1, the optimal utilization factor approximated by (6.22) is plotted versus Rician factor K for different transmit SNR E_s/N_0. The sum of $2\sigma_e^2$ and $2\sigma_\varepsilon^2$ is assumed

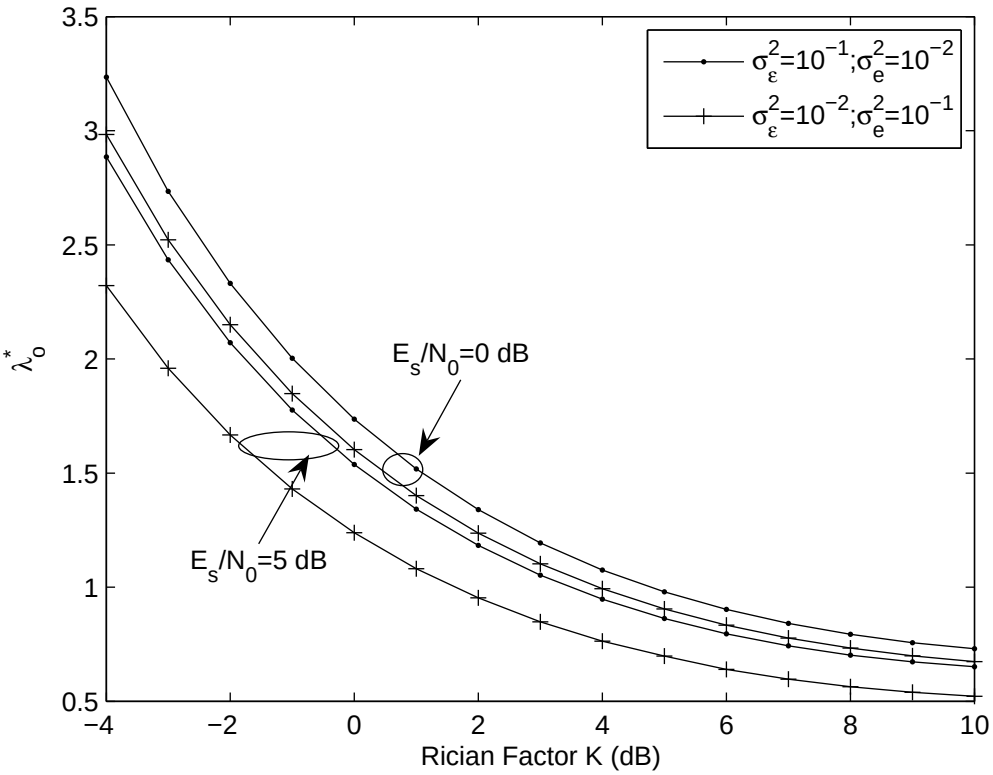

Figure 6.1: The optimal utilization factor versus Rician factor K for different transmit SNR with different accuracy of channel estimation and prediction.

to be constant. Two cases with different values of $2\sigma_e^2$ and $2\sigma_\varepsilon^2$ are considered in Fig. 6.1. For the same E_s/N_0 and at a fixed K value, a larger $2\sigma_\varepsilon^2$ results in a larger λ_o^*. This suggests that a worse channel prediction enables a larger λ_o^* providing a larger error margin to attain the maximum goodput. Utilization factor λ_o^* is a decreasing function of K. This is because a high K, (i.e., a high $|\tilde{h}|^2$) increases the outage probability, which can be inferred from (6.16). The significantly increased outage probability causes the maximum goodput to be achieved at a rate allocation with $\lambda_o^* < 1$ having a smaller error margin. For the same reason, a higher transmit SNR E_s/N_0 gives a smaller value of λ_o^*. As K continues to grow, the optimum utilization factor converges to 1. This is consistent with the intuition that the CSIT feedback of high quality can be regarded

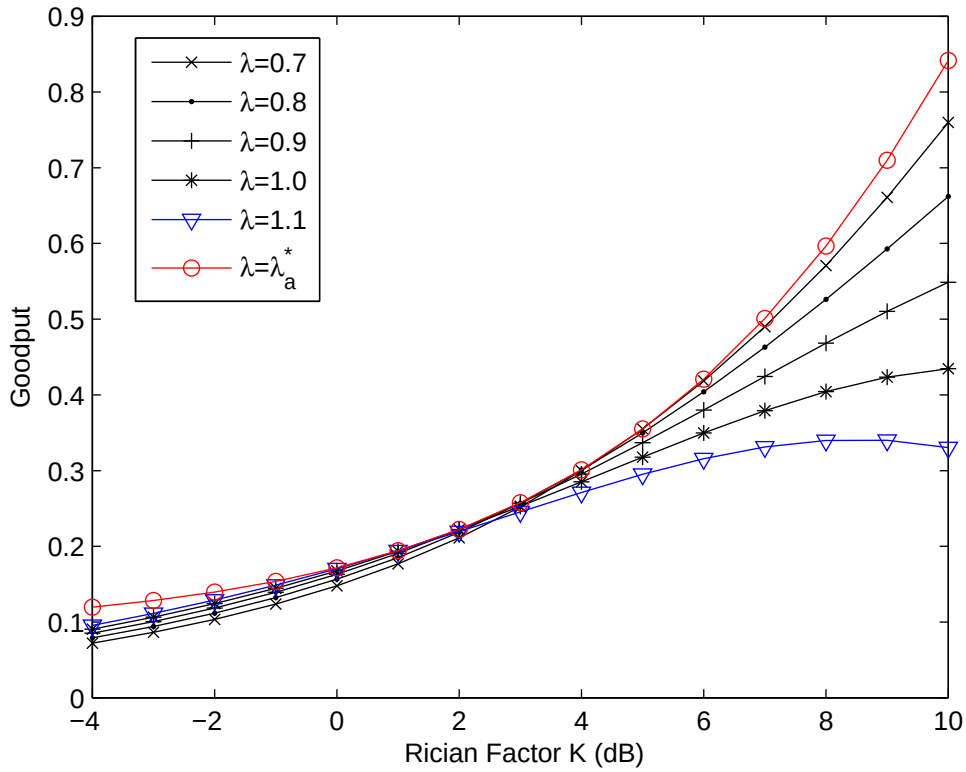

Figure 6.2: The goodput achieved by different utilization factors versus Rician factor K with $2\sigma_\varepsilon^2 = 10^{-2}$, $2\sigma_e^2 = 10^{-1}$ and $E_s/N_0 = 5$ dB.

as perfect CSIT and the transmission rate can be allocated accordingly. However, the convergence of the utilization factor is so slow that the transmission rate is reduced with a utilization factor of less than 1.

The goodput for different utilization factors as a function of Rician factor K is plotted in Fig. 6.2. The transmit SNR is assumed to be 5 dB with $2\sigma_\varepsilon^2 = 10^{-2}$ and $2\sigma_e^2 = 10^{-1}$. The λ_o^* obtained from (6.22) gives a maximum goodput while other fixed utilization factors provide the lower goodput. A fixed utilization factor larger than 1 makes the goodput diverge from the maximum value with the increasing K. This indicates that excessive rate allocation when K is large leads to a reduced goodput.

Fig. 6.3 depicts the goodput achieved by optimal rate allocation as a function of

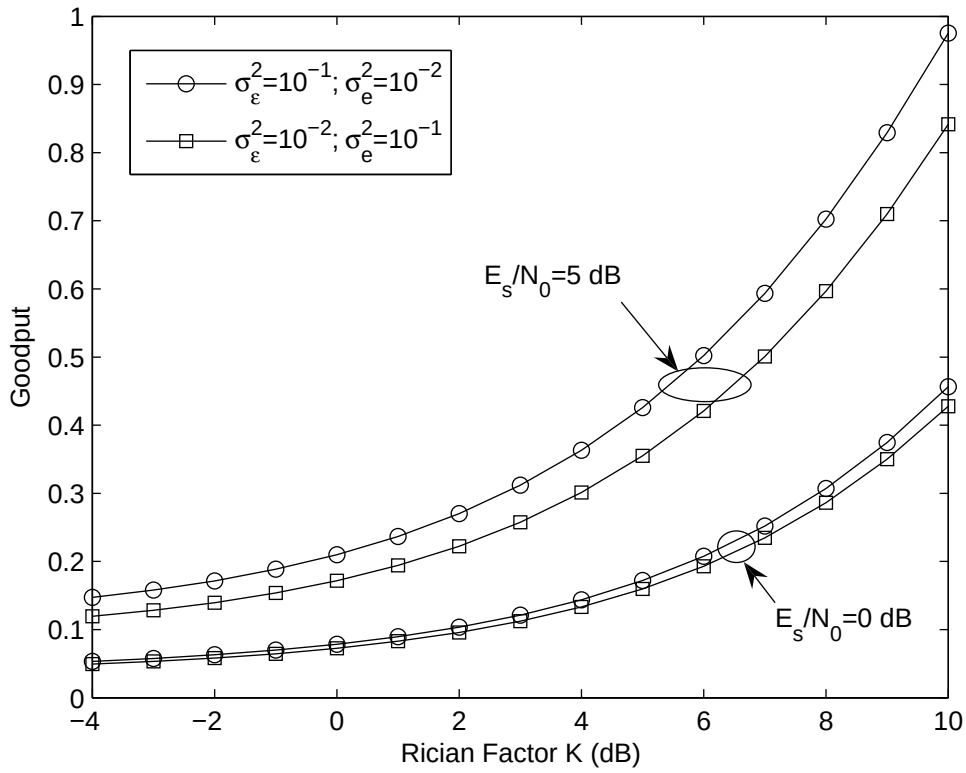

Figure 6.3: The goodput achieved by λ_o^* versus K for different transmit SNR and different accuracy of channel estimation and prediction.

K for different transmit SNR. For the same E_s/N_0 and at a fixed K value, a more accurate channel estimation (i.e a smaller $2\sigma_e^2$) can increase the goodput despite the fact that a poor prediction is made. The impact of prediction error $2\sigma_\varepsilon^2$ on the goodput is not so significant as the estimation error $2\sigma_e^2$.

6.6 Conclusions

In this Chapter, a rate adaptive transmission scheme with pilot-symbol-assisted estimation and prediction of fading channels is studied. The effect of channel estimation and prediction errors on rate allocation which helps achieve maximum

goodput is investigated. It is shown that with a high quality indicator K of feedback channel, the margin for the error of rate allocation needs to be reduced in order that the goodput is maximized. With a constant value of the quality indicator, a smaller estimation error (i.e. a better channel estimation) enables a higher rate allocation to reach the maximum goodput. Furthermore, a more accurate channel estimation can increase the goodput in spite of a poor channel prediction. The simple bounds on the first-order Marcum Q function obtained in [118] provide a great potential to further simplify the solution given by (6.19).

Chapter 7

Conclusions and Future Work

7.1 Conclusions

Many of the current and emerging wireless communication systems make use of diversity in one form or another to combat the effects of multipath fading. Diversity combining consists of receiving redundantly the same information-bearing signal over two or more fading channels, then combining these multiple replicas at the receiver in order to increase the overall received SNR. Depending on the domain where replicas of the same information-bearing signal are obtained, diversity techniques can be categorized into space diversity and time diversity. A conventional approach to achieving space diversity is to employ a MIMO communication system with multiple transmit and receive antennas. For achieving time diversity by combining packets transmitted in different time slots, ARQ/HARQ provide a great opportunity for obtaining additional copies of the packets by requesting a retransmission. In this dissertation, we studied the performance and transmission strategies of MIMO systems and ARQ/HARQ systems.

The ergodic capacity of a MIMO system in a fading environment is an important information theoretic performance measure. For the case of no CSIT but perfect CSIR,

we obtained three simple and tight bounds on the ergodic capacity of the i.i.d MIMO Rayleigh fading channels in closed-form. These simple bounds show explicitly the dependence of the ergodic capacity on SNR and the numbers of transmit and receive antennas. Since the CSIT is not available and the channel is assumed to be i.i.d, it is well known that the optimal transmission strategy for achieving the maximum ergodic capacity is to transmit independent, circularly symmetric, complex Gaussian distributed signals on all the antennas with equal power allocation. For the special case when the number of transmit antennas is equal to the number of receive antennas, previous results have shown that the ergodic capacity increases linearly with the number of transmit antennas. However, the simple bounds obtained in this dissertation enable us to determine the optimum number of transmit antennas to be used for a given SNR and a given total number of antennas in the system. We concluded that increasing the number of transmit antennas cannot always guarantee increasing the ergodic capacity. When the total transmit SNR is low, deploying more transmit antennas cannot increase the channel capacity since the SNR per transmit-receive link is too low when further decreased. When the total transmit SNR is high, providing more transmit-receive links is beneficial for achieving a higher channel capacity because of the greater spatial diversity achieved. We then considered a more practical MIMO system over i.n.d fading channels with the partial CSIT available at the transmitter. The partial CSIT is referred as the covariance feedback, i.e., the distribution of the channel is known at the transmitter to be a Gaussian distribution with mean zero and a certain nonzero covariance matrix. In practice, the covariance matrix could be computed at the receiver via long-term time averaging of the channel realizations and reliably transmitted to the transmitter through a low data rate feedback channel. The i.n.d fading channel attracts greater attention than the i.i.d fading channel since it occurs more naturally in practical MIMO systems. In order to enjoy the space diversity, the

antenna spacing needs to be sufficiently large to minimize the correlation between channels. The large spatial channel separation implies that the channels would more likely encounter different propagation environments. Therefore, we derived bounds on ergodic capacity and outage probability over i.n.d Rayleigh fading channels with the partial CSIT. By studying the simple and closed-form bounds, we derived two simple and near-optimal power-allocation schemes for maximizing the ergodic capacity and minimizing the outage probability, respectively. For ergodic capacity maximization at high SNR, the power control only depends on the ratios between the channel variances. In particular, for highly unbalanced channels, some antennas associated with lower channel variances may be turned off. For slightly unbalanced channels, all antennas may be active but more power is given to the antenna associated with larger channel variances. For minimizing the outage probability, the power allocation depends on the geometric mean of channel variances and follows the "water-filling" principle. According to the properties of this principle, at a high SNR, the power tends to be equally allocated among all transmit antennas whereas at low SNR, some of the transmit antennas with geometric mean of channel variances significantly lower than the others may be turned off.

In ARQ/HARQ systems, when the received packets have been detected in error by an error detection code, an additional copy of the packet is requested by the receiver. This retransmission mechanism provides the opportunity for packets to be transmitted in different time slots for achieving the time diversity. Instead of discarding the previously received packets, the receiver can combine the received packets in an optimal manner to obtain a more reliable packet. We considered such an ARQ/HARQ system with time diversity combining when the CSIR is imperfect and the maximum number of retransmissions is limited. In practice, only finite delays and buffer sizes can be accommodated, therefore, it is more meaningful to bound the maximum number

of retransmissions to a small number. Since the CSIR may be imperfect due to the channel estimation errors at the receiver, evaluating the effects of the imperfect CSIR on the system performance is important to provide insights on system operation and guidelines for designing effective system management schemes. Therefore, the performance of ARQ/HARQ systems are evaluated as a function of the accuracy of channel estimation, with the imperfect CSIR acquired by channel estimation at the receiver. A link between data-link-layer performances and physical-layer parameters is therefore established. The data-link layer performance parameters we studied in particular are the goodput, the APER and the drop rate, as a function of the channel estimation MSE and the factors which affect the MSE. Upper and lower bounds on the APER, the goodput as well as the drop rate were derived. These upper and lower bounds are close to one another, and therefore, enable the behavior of the exact performance parameters to be investigated. The precise dependence of the APER and the goodput on the channel estimation accuracy was quantified. An attempt was made to study the inter-relationships among the various relevant system performance parameters and the crucial dependence of these relationships on the CSIR accuracy. In particular, when the channel gain is not estimated with sufficient accuracy, for a given order of receive space diversity, HARQ achieves a smaller APER than basic ARQ does because of its inherent capability of achieving time diversity, whereas, it performs as close as the basic ARQ does when the channel is estimated accurately. In other words, with a lager value of MSE, the time diversity of the HARQ is more beneficial in terms of decreasing the APER. When receive space diversity is available in HARQ schemes and the CSIR is estimated accurately, the APER can be decreased more dramatically by using more receive antennas.

For enhancing the throughput, adaptive transmission strategies have been adopted to match the modulation and coding to time-varying channel conditions to achieve

higher spectral efficiency. As a MIMO system provides high reliability, an ARQ/HARQ system is an alternative way to mitigate channel fading to ensure a higher probability in the acceptance of retransmitted packets. Therefore, with regard to maximizing the throughput, in addition to providing a more reliable transmission, ARQ/HARQ schemes with adaptive transmissions are extensively adopted. There are two aspects to adaptive transmissions: channel quality estimation and rate selection. Channel quality estimation involves measuring the time-varying state of the wireless channel for the purpose of generating predictions of future quality. Rate selection involves using the channel quality predictions to select an appropriate rate. Among the factors that influence the effectiveness of rate adaptation, of particular importance is the accuracy of the channel quality estimates. We therefore developed a continuous-rate adaptation scheme that takes account of both channel estimation and prediction errors. While a strictly causal channel predictor is employed to predict the channel state for the transmitter to adapt its rates, a noncausal channel estimator estimates the channel for the receiver to perform coherent demodulation. The goodput, defined as the amount of data delivered to the receiver correctly per time unit, considering both the throughput and the reliability, was considered as the performance metric to be optimized. The transmission rate determined by the transmitter is adapted relative to the predicted channel state and a utilization factor so as to achieve a maximum goodput. It is shown that both excessive and insufficient rate adaptation lead to a reduced goodput. The utilization factor depends on the quality of the channel prediction.

7.2 Future Work

7.2.1 Effects of Imperfect CSIR on MIMO Systems

In Chapters 3 and 4 of this dissertation, it is assumed that the CSIR is perfectly known at the corresponding receivers. We have obtained bounds on the ergodic capacity and outage probability with perfect CSIR in [39, 40, 119]. However, the perfect CSIR is difficult to obtain in MIMO systems due to the increased number of channel parameters to be estimated at the receive. Therefore, the ergodic mutual information with imperfect CSIR is an important problem to investigate. To the best of our knowledge, deriving a closed-form expression for the ergodic mutual information with imperfect CSIR is still an open problem. Further more, it is important and meaningful to investigate MIMO systems with imperfect CSI on their applications in cooperative communications and distributed antenna systems.

7.2.2 Transmission Strategies in MIMO Systems with Imperfect CSIR and Outdated CSIT

In Chapters 3 and 4, it is assumed that either no CSIT or the partial CSIT is available. We have derived two sub-optimal power control schemes with only partial CSIT for optimizing the ergodic capacity and outage probability, respectively [39,40]. However, the CSIT may be outdated or imperfect due to the feedback delay or the imperfect CSIR that is fed back. The effect of the imperfect CSIR and outdated CSIT on the ergodic capacity and outage probability is a meaningful problem to study. Since the channel state may change to a new one from the time the receiver estimates the channel state to the time the transmitter receives the CSIR that is fed back, adaptive transmission without taking into account the feedback delay may reduce the performance. Thus,

transmission strategies for systems with outdated CSIT should be studied to reduce the capacity loss due to the feedback delay.

7.2.3 Extension of HARQ with Diversity Combining to Code Combining

In Chapter 5, the effect of the imperfect CSIR on the performance of the ARQ/HARQ systems with diversity combining was investigated [106,120]. The diversity combining combines the symbols in the received packets with a code rate of R in an optimal manner i.e., using MRC, and decodes the packet with the same code rate R. The diversity combining is effective for simple fading channels, but not necessarily effective in the presence of jamming [14]. Code combining treats the J received packets as a packet of a rate R/J. The decoder may be a rate-R/J soft or hard decoder, but also has the capability of weighting the reliability of each received packet. Since code combining is designed to work in a very noisy (jamming) environment, its performance should be studied with imperfect CSIR to see if it still works with a large MSE.

7.2.4 Adaptive Transmission in HARQ Schemes with Imperfect CSIT/CSIR

For enhancing the throughput, adaptive transmission strategies have been adopted to match the modulation and coding to time-varying channel conditions to achieve higher spectral efficiency. In Chapter 6, a continuous-rate adaptation scheme was developed with imperfect CSIR and imperfect CSIT for maximizing the goodput performance [121]. The adaptive transmission rate can be achieved by either changing the modulation scheme directly, or by sending more redundancy bits to reduce the

code rate indirectly. In Type-I HARQ systems with code combining, the individual transmitted packets are encoded at some code rate R. If the receiver has J packets that have caused retransmission requests, these packets are concatenated to form a single packet encoded at rate R/J. As J increase, the decoder eventually acquires sufficient power to reliably decode the packets under existing channel conditions. In Type-II HARQ, the transmitter responds to retransmission requests by sending additional parity bits to the receiver. The receiver appends these bits to the received packets to reduce the code rate allowing for increased error correction capability. With a rate adaptation scheme with imperfect CSI, a study on the maximum number of retransmissions in Type-I HARQ and the FEC codes used in Type-II HARQ is useful to design a HARQ combined with the rate adaptation, which can achieve high spectral efficiency and high reliability.

Bibliography

[1] D. Brennan, "Linear Diversity Combining Techniques," in *Proceedings of the Institute of Radio Engineers (IRE)*, Jun 1959, pp. 1075–1102.

[2] W. C. Jakes, *Microwave Mobile Communication*. Piscataway, NJ: IEEE Press, 1994.

[3] G. L. Stuber, *Principles of Mobile Communications*. Norwell, MA: Kluwer Academic Publishers, 1996.

[4] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Upper Saddle River, NJ: PTR Prentice-Hall, 1996.

[5] H. L. V. Trees, *Detection, Estimation, and Modulation Theory*. New York: John Wiley & Sons, Inc., 2001.

[6] J. G. Proakis, *Digital Communications*. New York: McGraw Hill, 2001.

[7] G. J. Foschini and M. J. Gans, "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998.

[8] H. Jafarkhani, *Space-Time Coding: Theory and Practice*. Reading, MA: Cambridge University Press, 2005.

[9] E. Telatar, "Capacity of Multi-antenna Gaussian channles," *European Transaction on Telecommunications*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[10] S. Lin and D. J. Costello, *Error control coding*. Englewood Cliffs, New Jersey: Prentice-Hall, 2004.

[11] S. B. Wicker, *Error Control System for Digital Communication and Storage*, 1st ed. Upper Saddle River, New Jersey: Prentice-Hall, 1995.

[12] F. Babich, "Performance of hybrid ARQ schemes for the fading channel," *IEEE Transactions on Communnications*, vol. 50, no. 12, pp. 1882–1885, Dec. 2002.

[13] D. Qiao, S. Choi, and K. G. Shin, "Goodput Analysis and Link Adaptation for IEEE 802.11a Wireless LANs ," *IEEE Tansactions on Mobile Computing*, vol. 1, no. 4, pp. 278–292, Dec. 2002.

[14] D. Chase, "Code Combining–A Maximum-Likelihood Decoding Approach for Combining an Arbitrary Number of Noisy Packets," *IEEE Transaction on Communications*, vol. Com-33, no. 5, pp. 385–393, May 1985.

[15] S. B. Wicker and M. D. Bartz, "Type-ii Hybrid-ARQ Protocols Using Punctured MDS Codes," *IEEE Transaction on communications*, Apr 1994.

[16] B. A. Harvey and S. B. Wicker, "Packet Combining Systems Based on the Viterbi Decoder," *IEEE Transaction on communications*, vol. 42, no. 2, pp. 1544–1557, February 1994.

[17] S. B. Wicker, "Adaptive Rate Error Control Through the Use of Diversity Combining and Majority-Logic Decoding in a Hybrid-ARQ Protocol," *IEEE Transactions on Communications*, vol. 39, no. 3, pp. 380–385, March 1991.

[18] D. M. Mandelbaum, "An Adaptive-Feedback Coding Scheme Using Incremental Redundancy," *IEEE Transactions on Information Theory*, vol. 20, no. 3, pp. 388–389, May 1974.

[19] A. J. Goldsmith and S.-G. Chua, "Variable-Rate Variable-Power MQAM for Fading Channels," *IEEE Transactions on Communications*, vol. 45, no. 10, pp. 1218–1230, Oct 1997.

[20] S. A. Jafar and A. Goldsmith, "On optimality of beamforming for multiple antenna systems with imperfect feedback," in *Proceedings, IEEE International Symposium on Information Theory (ISIT'01)*, Washington, DC, June 2001, p. 321.

[21] A. L. Moustakas and S. H. Simon, "Optimizing Multiple-Input Single-Output (MISO) Communication Systems With General Gaussian channels: Nontrivial Covariance and Nonzero Mean," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2770–2780, October 2003.

[22] C. Pimentel and R. L. Siqueira, "Analysis of the Go-Back-N Protocol on Finite-State Markov Rician Fading Channels," *IEEE Transactions on Wireless Communications*, vol. 57, no. 4, pp. 2627–2632, July 2008.

[23] A. Steiner and S. Shamai, "Multi-Layer Broadcasting Hybrid-ARQ Strategies for Block Fading Channels," *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2640–2650, July 2008.

[24] Y. Zhou and J. Wang, "Optimum subpacket transmission for hybrid ARQ systems," *IEEE Transaction on Communications*, vol. 54, no. 5, pp. 934–942, May 2006.

[25] J. F. T. Cheng, "Coding Performance of Hybrid ARQ Schemes," *IEEE Transactions on Communications*, vol. 54, no. 6, pp. 1017–1029, Jun. 2006.

[26] F. Babich, "Performance of Hybrid ARQ Schemes For the Fading Channel," in *Proc. IEEE International Conference on Communications (ICC'01)*, Helsinki, Finland, Jun.

[27] M. Zorzi, R. R. Rao, and L. B. Milstein, "ARQ Error Control for Fading Mobile Radio Channels," *IEEE Transaction on Vehicular Technology*, vol. 46, no. 2, pp. 445–455, May 1997.

[28] S. Kallel, "Analysis of Memory and Incremental Redundancy ARQ Schemes Over a Nonstationary Channel," *IEEE Transactions on Communications*, vol. 40, no. 9, pp. 1474–1480, Sep. 1992.

[29] G. Femenias, "SR ARQ for Adaptive Modulation Systems Combined With Selection Transmit Diversity," *IEEE Transactions on Communications*, vol. 53, no. 6, pp. 998–1006, June 2005.

[30] V. Mahinthan, H. Rutagemwa, J. W. Mark, and X. S. Shen, "Cross-Layer Performance Study of Cooperative Diversity System With ARQ," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 2, pp. 705–719, Feb. 2009.

[31] A. J. Goldsmith and S.-G. Chua, "Adaptive Coded Modulation for Fading Channels," *IEEE Transactions on Communications*, vol. 46, no. 5, pp. 595–602, May 1998.

[32] D. L. Goeckel, "Adaptive coding for time-varying channels using outdated fading estimates," *IEEE Transactions on Communications*, vol. 47, no. 6, pp. 844–855, Jun 1999.

[33] K. J. Hlole, H. holm, and G. E. Oien, "Adaptive Multidemensional Coded Modulation over Flat Fading Channels," *IEEE Journal On Selected Areas In Communications*, vol. 18, no. 7, pp. 1153–1158, July 2000.

[34] M. B. Pursley and J. M. Shea, "Adaptive Nonuniform Phase-Shift-Key Modulation for Multimedia Traffic in Wireless Networks," *IEEE Journal On Selected Areas In Communications*, vol. 18, no. 8, pp. 1394–1407, Aug 2000.

[35] J. Zhang and J. S. Lehnert, "Throughput-Optimal Precoding and Rate Allocation for MISO Systems With Noisy Feedback Channels," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2139–2155, May 2008.

[36] S. Falahati and A. Svensson, "Hybird Type-II ARQ Schemes with Adaptive Modulation Systems for Wireless Channels," in *Proc. IEEE Vehicular Technology Conference (VTC'99)*, Amsterdam, Netherlands, September 1999, pp. 2691–2695.

[37] E. A. Jorswieck and H. Boche, "Optimal Transmission Strategies and Impact of Correlation in Multiantenna Systems with Different Types of Channel State Information," *IEEE Transactions on Signal Processing*, vol. 52, no. 12, pp. 3440–3453, December 2004.

[38] A. L. Moustakas and S. H. Simon, "Optimizing MIMO Antenna Systems With Channel Covariance Feedback," *IEEE Transactions on Selected Areas in Communications*, vol. 21, no. 3.

[39] L. Cao, M. Tao, and P. Y. Kam, "Capacity Analysis and Power Allocation over Non-identical MISO Rayleigh Fading Channels," in *Proc. IEEE International Conference on Communications (ICC'08)*, Beijing, China, May 2008, pp. 4659–4663.

[40] ——, "Power Control for MIMO Diversity Systems with Non-Identical Rayleigh Fading," *IEEE Transactions on Vehicular Technology*, vol. 58, pp. 998–1003, Feb. 2009.

[41] J. Luo, R. S. Blum, L. Cimini, L. Greenstein, and A. Haimovich, "Power Allocation in a Transmit Diversity System with Mean Channel Gain Information," *IEEE Communications Letters*, vol. 9, no. 7, pp. 616– 618, July 2005.

[42] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[43] O. Oyman, R. U. Nabar, H. Bolcskei, and A. J. Paulraj, "Tight Lower Bounds on the Ergodic Capacity of Rayleigh Fading MIMO Channels," in *Proceedings, IEEE Global Communications Conference (GLOBECOM'02)*, November 2002, pp. 1172–1176.

[44] X. W. Cui, Q. T. Zhang, and Z. M. Feng, "Generic Procedure for Tightly Bounding the Capacity of MIMO Correlated Rician Fading Channels," *IEEE Transactions on Communications*, no. 5, pp. 890–898, May 2005.

[45] Y. Zhu, Y. Xin, and P. Y. Kam, "On the Outage Performance of Rician Fading Relay Channels," in *Proceedings, IEEE Military Communications Conference (MILCOM'06)*, Washington D.C., October 2006, pp. 1–7.

[46] S. Jin, X. Gao, and X. You, "On the Ergodic Capacity of Rank-1 Ricean-Fading MIMO Channels," *IEEE Transactions On Information Theory*, vol. 53, no. 2, pp. 502–517, February 2007.

[47] O. Oyman, R. U. Nabar, H. Bolcskei, and A. J. Paulraj, "Characterizing the Statistical Proerties of Mutual Information in MIMO Channels," *IEEE Transactions on Signal Processing*, vol. 51, no. 11, pp. 2784–2795, Nov 2003.

[48] Z. Wang and G. B. Giannakis, "Outage Mutual Information of Space-Time MIMO Channels," *IEEE Transactions on Information Theory*, vol. 50, no. 4, pp. 657–661, April 2004.

[49] S. K. Jayaweera and H. V. Poor, "On the Capacity of Multiple-Antenna Systems in Rician Fading," *IEEE Transactions On Wireless Communications*, vol. 4, no. 3, pp. 1102–1111, May 2005.

[50] Y. Zhu, P.-Y. Kam, and Y. Xin, "On the Mutual Information Distribution of MIMO Rician Fading Channels," *IEEE Transactions on Communications*, vol. 57, no. 5, pp. 1453–1462, May 2009.

[51] E. Visotsky and U. Madhow, "Space-Time Transmit Precoding With Imperfect Feedback," *IEEE Transactions on Information Theory*, vol. 47, no. 6, pp. 2632–2639, September 2001.

[52] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO Capacity Through correlated Channels in the Presence of Correlated Interferers and Noise: A Not So Larege N Analysis," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2545– 2561, October 2003.

[53] A. Narula, M. D. Trott, and G. W. Wornell, "Performance Limits of Coded Diversity Methods for Transmitter Antenna Arrays," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2418–2433, November 1999.

[54] E. Abbe, E. Telatar, and L. Zheng, "The Algebra of MIMO Channels," in *Allerton Annual Conference on Communication, Control and Computing*, October 2005.

[55] A. Narula, M. D. Lopez, M. D. Trott, and G. W. Wornell, "Efficient Use of Side Information in Multiple-Antenna Data Transmission Over Fading Channels," *IEEE Journal On Selected Areas In Communications*, vol. 16, no. 8, pp. 1423–1436, October 1998.

[56] S. A. Jafar and A. Goldsmith, "Transmitter Optimization and Optimality of Beamforming for Multiple Antenna Systems," *IEEE Transactions on Wirless Communications*, vol. 3, no. 4, pp. 1165–1174, July 2004.

[57] K. Brayer, "Error Control Techniques Using Binary Symbol Burst Codes," *IEEE Transactions on Communications*, vol. COm, no. 16, pp. 199–214, April 1968.

[58] A. Drukarev and J. Daniel J. Costello, "Hybird ARQ Error Control Using Sequential Decoding," *IEEE Transactions on Information Theory*, vol. IT-29,, no. 4, pp. 521–535, July 1983.

[59] H. Yamamoto and K. Itoh, "Viterbi Decoding Algorithm for Convolutional Codes with Repeat Request," *IEEE Transactions on Information Theory*, vol. IT-26,, no. 5, pp. 540–547, September 1980.

[60] M. D. Rice and S. B. Wicker, "Modified Majority Logic Decoding of Cyclic Codes in Hybrid-ARQ Systems," *IEEE Transaction on communications*, vol. 40, no. 9, pp. 1413–1417, September 1992.

[61] S. B. Wicker, "High-Reliability Data Transfer Over the Land Mobile Radio Channel Using Interleaved Hybrid-ARQ Error Controls," *IEEE Transactions On Vehicular Technology*, vol. 39, no. 1, pp. 48–55, February 1990.

[62] ——, "Reed-Solomon Error Control Coding for Rayleigh Fading Channels with Feedback," *IEEE Transactions On Vehicular Technology*, vol. 41, no. 2, pp. 124–133, May 1992.

[63] J. J. Metzner, "Improvements in Block-Retransmission Schemes," *IEEE Transactions on Communications*, vol. COM, no. 27, pp. 525–532, February 1979.

[64] S. Kallel, "Generalized Type II Hybrid ARQ Scheme Using Punctured Convolutional Coding," *IEEE Transactions on Communications*, vol. 38, no. 11, pp. 1938–1946, Nov. 1990.

[65] ——, "Analysis of A Type-II Hybrid ARQ Scheme with Code Combining," *IEEE Transactions on Communications*, vol. 38, no. 8, pp. 1133–1137, Aug. 1990.

[66] ——, "Efficient Hybrid ARQ Protocols with Adaptive Forward Error Correction," *IEEE Transactions on Communications*, vol. 42, no. 2, pp. 281–289, February 1994.

[67] Q. Yang and V. K. Bhaargava, "Delay and Coding Gain Analysis of A Trun-cated Type-II hybrid ARQ Protocol," *IEEE Transaction on Vehicular Technology*, vol. 42, no. 1, pp. 22–32, Feb. 1993.

[68] Q. Zhang and S. A. Kassam, "Hybrid ARQ with Selective Combining for Fading Channels," *IEEE Journal On Selected Areas In Communications*, vol. 17, no. 5, pp. 867–879, May 1999.

[69] E. Malkamaki and H. leib, "Performance of Truncated Type-II Hybrid ARQ Schemes with Noisy Feedback over Block Fading Channels," *IEEE Transactions on Communications*, no. 9, pp. 1477–1487, September 2000.

[70] S. Kim and C. Un, "Throughput Analysis for Two ARQ Schemes Using Combined Transition Matrix," *IEEE Transactions On Communications*, vol. 40, no. 11, pp. 1679–1683, Nov 1992.

[71] Y. J. Cho and C. K. Un, "Performance Analysis of ARQ Error Controls under Markovian Block Error Pattern," *IEEE Transactions On Communications*, vol. 42, no. 2, pp. 2051–2061, Feb 1994.

[72] M. Zorzi and R. R. Rao, "Throughput analysis of ARQ error controls Go-Back-N protocol in Markov channels with unreliable feedback," in *Proc. IEEE International Conference on Communications (ICC'95)*, Seattle, WA, USA, June.

[73] W. Turin, "Throughput Analysis of the Go-Back-N Protocol in Fading Radio Channels," *IEEE Journal On Selected Areas In Communications*, vol. 17, no. 5, pp. 881–887, May 1999.

[74] J. Yun, W. Jeong, and M. Kavehrad, "Throughput analysis of selective repeat ARQ combined with adaptive modulation for fading channels," in *Proc. IEEE Military Communications Conference, (MILCOM'02)*, Oct 2002, pp. 710–714.

[75] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-Layer Combining of Adaptive Modulation and Coding With Truncated ARQ Over Wireless Links," *IEEE Transactions On Wireless Communications*, vol. 3, no. 5, pp. 1746–1755, September 2004.

[76] J. Yun and M. Kavehrad, "Markov Error Structure for Throughput Analysis of Adaptive Modulation Systems Combined with ARQ over Correlated Fading Channels," *IEEE Transaction on Vehicular Technology*, vol. 54, no. 1, pp. 235–245, Jan. 2005.

[77] T. L. Marzetta and B. M. Hochwald, "Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading," *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 139–157, Jan 1999.

[78] M. Chiani, M. Z. Win, and A. Zanella, "On the Capacity of Spatially Correlated MIMO Rayleigh-Fading Channels," *IEEE Transactions On Information Theory*, vol. 49, no. 10, pp. 2363–2371, October 2003.

[79] M. Kang and M.-S. Alouini, "Capacity of Correlated MIMO Rayleigh Channels," *IEEE Transactions on Wireless Communications*, vol. 5, no. 1, pp. 143–155, Jan 2006.

[80] M. R. McKay and I. B. Collings, "General Capacity Bounds for Spatially Correlated Rician MIMO Channels," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3121–3145, Sep 2005.

[81] M. Matthaiou, Y. Kopsinis, D. I. Laurenson, and A. M. Sayeed, "Ergodic Capacity Upper Bound for Dual MIMO Ricean Systems: Simplified Derivation and Asymptotic Tightness," *IEEE Transactions On Communications*, vol. 57, no. 12, pp. 3589–3596, December 2009.

[82] Y. Zhu, P. Y. Kam, and Y. Xin, "A New Approach to the Capacity Distribution of MIMO Rayleigh Fading Channels," in *Proceedings, IEEE Global Communications Conference (GLOBECOM'06)*, San Francisco, CA, November 2006, pp. 1–5.

[83] S. Sandhu and A. Paulraj, "Space-time Block Codes: A Capacity Perspective," *IEEE Wireless Communications Letters*, vol. 4, no. 12, pp. 384–386, Dec. 2000.

[84] G. Ganesan and P. Stoica, "Space-time block codes: A maximum SNR approach," *IEEE Transactions on Infomation Theory*, vol. 47, no. 4, pp. 1650–1656, May 2001.

[85] H. Zhang and T. A. Gulliver, "Capacity and error probability analysis for orthogonal space-time block codes over fading channels," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 808–819, March 2005.

[86] C. Gao, A. Gaimovich, and D. Lao, "Bit Error Probability for Space-Time Block Code with Coherent and Differential Detection," in *Proc. IEEE VTC'02*, Vancouver BC, Canada, Sept. 2002, pp. 410–414.

[87] J. N. Laneman and G. W. Wornell, "Distribued Space-Time-Coded Protocols for Exploiting Cooperative Diversity in Wireless Networks," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2415– 2425, Oct. 2003.

[88] M. V. Clark, T. M. W. III, L. J. Greenstein, J. R. A, V. Erceg, and R. S. Roman, "Distributed versus centralized antenna arrays in broadband wireless networks," in *Proc. IEEE VTC'01 Spring*, Rhodes, Greece, May 2001, pp. 33–37.

[89] J. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior," *IEEE Transactions on Information Theory*, vol. 15, no. 12, pp. 3062–3080, Dec. 2004.

[90] G. Caire, G. Taricco, and E. Biglieri, "Optimal Power Control over Fading Channels," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1468–1489, July 1999.

[91] J. He and P. Y. Kam, "On the Performance of Orthogonal Space-Time Block Codes over Independent, Nonidentical Rayleigh/Ricean Fading Channels," in *Proc. IEEE Globalcom'06*, San Francisco, CA, Nov. 2006, pp. 1–5.

[92] H. Zhao, Y. Gong, Y. L. Guan, C. L. Law, and Y. Tang, "Space-Time Block Codes in Nakagami Fading Channels with Non-identical m-distributions," in *Proc. IEEE WCNC'07*, Kowloon, Hongkong, Mar. 2007, pp. 536–540.

[93] M. Tao and P. Y. Kam, "Analysis of differential orthogonal space-time block codes over semi-identical MIMO fading channels," *IEEE Transactions on Communications*, vol. 55, no. 2, pp. 282–291, Feb. 2007.

[94] H. Fu and P. Y. Kam, "Performance of Optimum and Suboptimum Combining Diversity Reception for Binary and Quadrature DPSK Over Independent,Nonidentical Rayleigh Fading Channels," *IEEE Transactions on Communications*, vol. 55, no. 5, pp. 887–894, May 2007.

[95] L. Cao, M. Tao, and P. Y. Kam, "Closed-Form Performance of MFSK Signal with Diversity Reception over Non-identical Fading Channels," in *IEEE Wireless Communications and Networking Conference (IEEEE WCNC'07)*, HongKong, China, March 2007, pp. 740–745.

[96] D. Raphaeli, "Distribution of noncentral indefinite quadratic forms in complex normal variables," *IEEE Trans. on Info. Theory*, vol. 42, 1996.

[97] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. U.S.A.: Academic Press Limited, 1994.

[98] N. C. Beaulieu and J. Hu, "A closed-form expression for the outage probability of decode-and-forward relaying in dissimilar rayleigh fading channels," *IEEE Communications Letters*, vol. 10, no. 12, pp. 813–815, Dec. 2006.

[99] M.-K. Byun and B. G. Lee, "New Bounds of Pairwise Error Probability for Space-Time Codes in Rayleigh Fading Channels," *IEEE Transactions on Communications*, no. 8, pp. 1484–1493, Aug 2007.

[100] H. C. Yang and S. Sasankan, "Analysis of Channel-Adaptive Packet Transmission Over Fading Channels With Transmit Buffer Management," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 1, pp. 404–413, Jan. 2008.

[101] W. Wang, Z. Guo, X. S. Shen, C. Chen, and J. Cai, "Dynamic Bandwidth Allocation for Qos Probisioning in IEEE 802.16 Networks with ARQ-SA," *IEEE Transactions on Wireless Communications*, vol. 7, no. 9, pp. 3477–3487, Sep. 2008.

[102] C. Wang, T. Lin, and K. W. Chang, "On Throughput Performance of Channel Inequality in IEEE 802.11 WLANs," *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, pp. 4425–4431, Nov. 2008.

[103] P. Y. Kam, "Optimal Detection of Digital Data Over the Nonselective Rayleigh Fading Channel with Diversity Reception," *IEEE Transactions on Communications*, vol. 39, no. 2, pp. 214–219, Feb. 1991.

[104] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, New Jersey: Prentice-Hall, 2002.

[105] P. Y. Kam, "Adaptive Diversity Reception Over a Slow Nonselective Fading Channel," *IEEE Transactions on Communications*, vol. Com-35, no. 5, pp. 572–574, May 1987.

[106] L. Cao, P. Y. Kam, and M. Tao, "Impact of Imperfect Channel State Information on ARQ Schemes over Rayleigh Fading Channels," in *Proceedings, IEEE ICC'09*, Dresden, Germany, Jun 2009, pp. 1–5.

[107] N. Kim, Y. Lee, and H. Park, "Performance Analysis of MIMO System with Linear MMSE Receiver," *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, pp. 4474–4478, Nov. 2008.

[108] K. A. S. Abdel-Ghaffar, "A Lower Bound on the Undetected Error Probability and Strictly Optimal Codes," *IEEE Transactions on Information Theory*, vol. 43, no. 5, pp. 1489–1502, Sep. 1997.

[109] M. K. Simon and M. S. Alouini, *Digital Communications Over Fading Channels*. New York: Wiley, 2000.

[110] P. Y. Kam, "Bit Error Probabilities of MDPSK Over the Nonselective Rayleigh Fading Channel with Diversity Reception," *IEEE Transaction on Communications*, vol. 39, no. 2, pp. 220–224, Feb. 1991.

[111] V. K. N. Lau and M. D. Macleod, "Variable-rate trellis coded QAM for flat-fading channels," *IEEE Transactions on Communications*, vol. 49, no. 9, pp. 1550–1556, Sep 2001.

[112] S. Falahati, A. Svensson, T. Ekman, and M. Sternad, "Adaptive Modulation Systems for Predicted Wireless Channels," *IEEE Transaction on Communications*, vol. 52, no. 2, pp. 307–316, Feb. 2004.

[113] Y. Ma, D. Zhang, A. Leith, and Z. Wang, "Error Performance of Transmit Beamforming with Delayed and Limited Feedback," *IEEE Transaction on Wireless Communications*, vol. 8, no. 3, pp. 1164–1170, March 2009.

[114] X. Cai and G. B. Giannakis, "Adaptive PSAM Accounting for Channel Estimation and Prediction Errors," *IEEE Transactions on Wireless Communications*, vol. 4, no. 1, pp. 246–256, Jan. 2005.

[115] J. K. Cavers, "An Analysis of Pilot Symbol Assisted Modulation for Rayleigh Fading Channels," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 4, pp. 686–693, Nov. 1991.

[116] A. Papoulis, *Probability Random Variables and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.

[117] T. Yoo and A. Goldsmith, "Capacity and Power Allocation for Fading MIMO Channels With Channel Estimation Error," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[118] P. Y. Kam and R. Li, "Computing and Bounding the First-Order Marcum Q-Function: A Geometric Approach," *IEEE Transactions On Communications*, vol. 56, no. 7, pp. 1101–1110, July 2008.

[119] L. Cao and P. Y. Kam, "Optimal Antenna Deployment for Capacity Maximization in a MIMO Rayleigh Fading Channel," in *Proc. IEEE Vehicular Technology Conference (VTC'10)*, Ottawa, Canada, September 2010, pp. 1–5.

[120] ——, "On the Performance of Packet ARQ Schemes in Rayleigh Fading: The Role of Receiver Channel State Information and Its Accuracy," *IEEE Transaction on Vehicular Technology*, 2010.

[121] ——, "Goodput-Optimal Rate Adaptation with Imperfect Channel State Information," in *Proceedings, IEEE Vehicular Technology Conference (VTC'09)*, Anchorage, Alaska, USA, September 2009, pp. 1–5.

Appendix A

Proof of the Inequality (3.8)

Substituting (3.7) into (3.6), we obtain the upper bound on the ergodic capacity to be

$$\mathbb{E}[\mathcal{I}] \leq \frac{N}{\Gamma(NM)\ln 2}\int_0^\infty \ln(1+\gamma\sigma^2/Nz)z^{NM-1}e^{-z}dz \tag{A.1}$$

Letting $c = \gamma\sigma^2/N$ and $b = NM - 1$, the above integral can be rewritten as

$$\mathbb{E}[\mathcal{I}] \leq \frac{N}{\Gamma(NM)\ln 2}\int_0^\infty \ln(1+cz)z^{b}e^{-z}dz. \tag{A.2}$$

Denote the integral term in the above equality as

$$F(b) = \int_0^\infty \ln(1+cz)z^{b}e^{-z}dz. \tag{A.3}$$

Making use of integration by parts, we have

$$\begin{aligned} F(b) &= -\int_0^\infty \ln(1+cz)z^{b}de^{-z} \\ &= -\ln(1+cz)z^{b}e^{-z}\,|_0^\infty + \int_0^\infty \frac{cz^{b}e^{-z}}{1+cz}dz + bF(b-1) \end{aligned} \tag{A.4}$$

Due to the following two limits:

$$\lim_{z\to\infty}\frac{\ln(1+cz)z^{b}}{e^{z}} = 0$$

and

$$\lim_{z\to 0}\frac{\ln(1+cz)z^{b}}{e^{z}} = 0,$$

term $F(b)$ in (A.4) can be written as

$$F(b) = \int_0^\infty \frac{c}{1+cz} z^b e^{-z} dz + bF(b-1). \tag{A.5}$$

Making use of [97, eq.(366.10)]:

$$\int_0^\infty \frac{x^{v-1}e^{-\mu x}}{x+\beta} dx = \beta^{v-1} e^{\beta\mu} \Gamma(v)\Gamma(1-v, \beta\mu), \tag{A.6}$$

when $|\arg\beta| < \pi$, Re$[\mu] > 0$, and Re$[v] > 0$, we can evaluate the integral in (A.4) to be

$$\int_0^\infty \frac{c}{1+cz} z^b e^{-z} dz = (1/c)^b e^{1/c} \Gamma(b+1)\Gamma(-b, 1/c) = A(b), \tag{A.7}$$

in which, term $\Gamma(\alpha)$ is the gamma function defined as $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt, \alpha > 0$. Making use of (A.5) recursively, the quantity $F(b)$ can be expressed as

$$\begin{aligned} F(b) = A(b) &+ \sum_{j=1}^{b-1} A(b-j) \prod_{i=1}^{j} (b+1-i) \\ &+ \prod_{i=1}^{b} (b+1-i) \int_0^\infty \ln(1+cz) e^{-z} dz. \end{aligned} \tag{A.8}$$

After some simple manipulation, term $F(b)$ can be simplified to be

$$F(b) = \sum_{j=0}^{b} A(b-j) \prod_{i=1}^{j} (b+1-i) \tag{A.9}$$

where $\prod_{i=1}^{0}(b+1-i)$ is defined to be 1. By applying (A.9) to (A.2), the average mutual information $\mathbb{E}[\mathcal{I}]$ can be upper bounded as

$$\begin{aligned} \mathbb{E}[\mathcal{I}] \leq &\frac{Ne^{N/(\sigma^2\gamma)}}{\Gamma(NM)\ln 2} \sum_{j=0}^{NM-1} \left(\frac{N}{\sigma^2\gamma}\right)^{NM-1-j} \prod_{i=1}^{j} (NM-i) \\ &\times \Gamma(NM-j)\Gamma(-(NM-1-j), N/(\sigma^2\gamma)) \end{aligned} \tag{A.10}$$

Since $\prod_{i=1}^{j}(NM-i)\Gamma(NM-j) = \Gamma(NM)$, the above bound can be further reduced to

$$\begin{aligned} \mathbb{E}[\mathcal{I}] \leq \overline{\mathcal{I}}_{\text{tr\_U}} &\triangleq \frac{1}{\ln 2} N e^{N/(\sigma^2\gamma)} \\ &\times \sum_{j=0}^{NM-1} \left(\frac{N}{\sigma^2\gamma}\right)^{NM-1-j} \Gamma(-(NM-1-j), N/\sigma^2\gamma). \end{aligned}$$

Appendix B

Proof of the equation (5.12)

By using the Chernoff bound: $\mathrm{erfc(x)} < e^{-x^2}$, an upper bound can be obtained as

$$P_e \leq 1 - \int_0^\infty \left(1 - \frac{1}{2}e^{-c|\hat{h}|^2}\right)^n e^{-b|\hat{h}|^2} d(b|\hat{h}|^2) = 1 - Z.$$

Using integration by parts with $u = (1 - \frac{1}{2}e^{-c|\hat{h}|^2})^n$ and $dv = de^{-b|\hat{h}|^2}$, $Z = -[uv - \int vdu]_0^\infty$ can be expressed as

$$Z = \left(\frac{1}{2}\right)^n + \frac{nc}{2}\int_0^\infty (1 - \frac{1}{2}e^{-c|\hat{h}|^2})^{n-1} e^{-|\hat{h}|^2(b+c)} d|\hat{h}|^2 \tag{B.1}$$

Continuing the integration by parts with $u = (1 - \frac{1}{2}e^{-c|\hat{h}|^2})^{n-1}$ and $dv = e^{-|\hat{h}|^2(b+c)}d|\hat{h}|^2$, and performing the similar process till the last integral, term Z comes to

$$\begin{aligned} Z = &\left(\frac{1}{2}\right)^n + \frac{nc}{2(b+c)}\left(\frac{1}{2}\right)^{n-1} + \frac{nc(n-1)c}{2(b+c)2(b+2c)}\left(\frac{1}{2}\right)^{n-2} \\ &+ \cdots - \frac{nc(n-1)c\cdots c}{2(b+c)2(b+2c)\cdots 2(b+nc)}. \end{aligned} \tag{B.2}$$

Hence, we can get

$$Z = \sum_{l=0}^{n} \left(\frac{1}{2}\right)^{n-l} \prod_{j=0}^{l-1} \frac{(n-j)c}{2(b+(j+1)c)}. \tag{B.3}$$

Note that when $l = 0$, term $\prod_{j=0}^{l-1} \frac{(n-j)c}{2(b+(j+1)c)}$ is defined to be 1.

List of Publications

1. Le Cao and Pooi Yuen Kam, "On the Performance of Packet ARQ Schemes in Rayleigh Fading: The Role of Receiver Channel State Information and Its Accuracy," submitted to *IEEE Transaction on Vehicular Technology*, vol. 60, no. 2, pp. 704–709, March 2011

2. Le Cao and Pooi Yuen Kam, "Optimal Antenna Deployment for Capacity Maximization in a MIMO Rayleigh Fading Channel" in *Proc. IEEE Vehicular Technology Conference (VTC'10)*, pp. 1–5, Ottawa, Canada, September, 2010.

3. Le Cao and Pooi Yuen Kam, "Goodput-Optimal Rate Adaptation with Imperfect Channel State Information" in *Proc. IEEE Vehicular Technology Conference (VTC'09)*, pp. 1–5, Anchorage, Alaska, USA, September, 2009.

4. Le Cao, Pooi Yuen Kam, and Meixia Tao, "Impact of Imperfect Channel Estimation Error on Performance of ARQ Schemes over Rayleigh Fading Channels," in *Proc. IEEE International Conference on Communications (ICC'09)*, pp. 1–5, Dresden, Germany, June, 2009.

5. Le Cao, Meixia Tao, and Pooi Yuen Kam, "Power Control for MIMO Diversity Systems with Non-identical Rayleigh Fading," in *IEEE Transaction on Vehicular Technology*, vol. 58, no. 2, pp. 998-1003, February 2009.

6. Le Cao, Meixia Tao, and Pooi Yuen Kam, "Capacity Analysis and Power Allocation over Non-identical MISO Rayleigh Fading Channels," in *Proc. IEEE International Conference on Communications (ICC'08)*, pp. 4659–4663, Beijing, China, May 2008.

7. Le Cao, Meixia Tao, and Pooi Yuen Kam, "Closed-form Performance of MFSK Signals with Diversity Reception over Non-identical Fading Channels," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC'07)*, pp. 740–750, Hong Kong, March 2007.