Transmission Resource Allocation in Multi-Antenna Wireless Communication Systems with Channel Uncertainty

Xiangyun Zhou

B.E. (Hons 1)(The Australian National University, Canberra)

June 2010

A thesis submitted for the degree of Doctor of Philosophy
of The Australian National University

School of Engineering
College of Engineering and Computer Science
The Australian National University

Declaration

The contents of this thesis are the results of original research and have not been submitted for a higher degree to any other university or institution.

Much of the work in this thesis has been published or has been submitted for publication as journal papers or conference proceedings. These papers are:

Journal articles

J1. Xiangyun Zhou, Parastoo Sadeghi, Tharaka A. Lamahewa, and Salman Durrani, "Optimizing Antenna Configuration for MIMO Systems with Imperfect Channel Estimation", *IEEE Trans. Wireless. Commun.*, vol. 8, no. 3, pp. 1177-1181, Mar. 2009.

J2. Xiangyun Zhou, Parastoo Sadeghi, Tharaka A. Lamahewa, and Salman Durrani, "Design Guidelines for Training-based MIMO Systems with Feedback", *IEEE Trans. Signal Processing*, vol. 57, no. 10, pp. 4014-4026, Oct. 2009.

J3. Xiangyun Zhou, Tharaka A. Lamahewa, Parastoo Sadeghi, and Salman Durrani, "Two-way Training: Optimal Power Allocation for Pilot and Data Transmission", *IEEE Trans. Wireless Commun.*, vol 9, no. 2, pp. 564-569, Feb. 2010.

J4. Xiangyun Zhou and Matthew R. McKay, "Secure Transmission with Artificial Noise over Fading Channels: Achievable Rate and Optimal Power Allocation", submitted to *IEEE Trans. Veh. Technol.*, revised in May 2010.

Conference papers

C1. Xiangyun Zhou, Tharaka A. Lamahewa, Parastoo Sadeghi, and Salman Durrani, "Capacity of MIMO Systems: Impact of Spatial Correlation with Channel Estimation Errors", in *Proc. IEEE Int. Conf. on Commun. Syst. (ICCS)*, Guangzhou, China, Nov. 2008, pp. 817-822.

C2. Xiangyun Zhou and Matthew R. McKay, "Physical Layer Security with Artificial Noise: Secrecy Capacity and Optimal Power Allocation", in *Proc. Int. Conf. on Signal Processing and Commun. Syst. (ICSPCS)*, Omaha, NE, Sept. 2009, pp. 1-5.

C3. Xiangyun Zhou, Tharaka A. Lamahewa, Parastoo Sadeghi, and Salman Durrani, "Optimizing Training-based Transmission for Correlated MIMO Systems with Hybrid Feedback", in *Proc. IEEE Global Commun. Conf. (Globecom)*, Honolulu, HI, Nov. 2009, pp. 1-6.

C4. Xiangyun Zhou, Parastoo Sadeghi, and Tharaka A. Lamahewa, "Optimizing Training-based MIMO Systems: How Much Time is Needed for Actual Transmission?", in *Proc. IEEE Veh. Tech. Conf. (VTC-Spring)*, Taipei, Taiwan, May 2010, pp. 1-5.

The following publications are also the results from my Ph.D. study but not included in this thesis:

Journal articles

J5. Tharaka A. Lamahewa, Parastoo Sadeghi, and Xiangyun Zhou "On Lower Bounding the Information Capacity of Amplify and Forward Wireless Relay Channels with Channel Estimation Errors", submitted to *IEEE Trans. Wireless. Commun.*, May 2010.

Conference papers

C5. Xiangyun Zhou, Tharaka A. Lamahewa, Parastoo Sadeghi, and Salman Durrani, "Designing PSAM Schemes: How Optimal are SISO Pilot Parameters for Spatially Correlated SIMO?", in *Proc. IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Cannes, France, Sept. 2008, pp. 1-6.

C6. Xiangyun Zhou, Tharaka A. Lamahewa, and Parastoo Sadeghi "Kalman Filter-based Channel Estimation for Amplify and Forward Relay Communications", in *Proc. IEEE Asilomar Conf. on Signals, Syst. and Computers (ACSSC)*, Pacific Grove, CA, Nov. 2009, pp. 1-5.

The research work presented in this thesis has been performed jointly with Dr. Parastoo Sadeghi (The Australian National University), Dr. Tharaka A. Lamahewa (The Australian National University), Dr. Salman Durrani (The Australian National University), and Prof. Matthew R. McKay (Hong Kong University of

Science and Technology). The substantial majority of this work was my own.

Xiangyun Zhou
School of Engineering,
The Australian National University,
Canberra,
ACT 0200,
Australia.

Acknowledgements

The work presented in this thesis would not have been possible without the support of a number of individuals and organizations and they are gratefully acknowledged below:

- I would like to express my sincere thanks to my supervisors and friends Drs Parastoo Sadeghi, Tharaka Lamahewa, Salman Durrani and Prof. Rod Kennedy for their guidance, support and encouragement throughout my PhD studies. They are always willing to have in-depth discussion whenever I approach them for help. Apart from my research studies, they have also given me great support in obtaining teaching and supervision experiences.

- I would like to thank Prof. Matt McKay from the Hong Kong University of Science and Technology (HKUST), Profs Jeff Andrews and Robert Heath from the University of Texas at Austin (UT) for kindly welcoming me to visit their research groups. During the one-week visit at HKUST, Matt introduced the problem of information-theoretic security to me, which has led to part of the work presented in this thesis. The two-month visit at UT was also very valuable. The discussions with Jeff stimulated many interesting ideas in the area of wireless random networks.

- I also would like to thank Prof. Meixia Tao from Shanghai Jiao Tong University, Prof. Pingzhi Fan from Southwest Jiao Tong University and Prof. Youxi Tang from University of Electronic Science and Technology for kindly hosting my visit at these three Chinese universities.

- It is my great pleasure to study in the wireless signal processing (WSP) group at the Research School of Information Sciences and Engineering. I would like to thank everyone for making WSP group a friendly and relaxing research environment. I certainly enjoy the weekly badminton matches with my supervisors and fellow students. Special thanks go to Ms. Elspeth Davies, our student administrator, for her assistance.

- Thanks must go to the Australian Government and Australian National University for providing me with the PhD scholarship and stipend, and the Australian Research Council Discovery Project Grants (DP0773898 and DP0984950) for supporting my conference attendances and university visits.

- I thank my parents for providing me the opportunity of studying abroad and their great financial support. My every little success would not have been possible without their love and encouragement. Last but not least, I would like to give my special thanks to my girlfriend Iris for her constant support, encouragement and understanding.

Abstract

In this thesis we investigate the design of transmission resource allocation in current and future wireless communication systems. We focus on systems with multiple antennas and characterize their performance from an information-theoretic viewpoint. The goal of this work is to provide practical transmission and resource allocation strategies taking into account imperfections in estimating the wireless channel, as well as the broadcast nature of the wireless channel.

In the first part of the thesis, we consider training-based transmission schemes in which pilot symbols are inserted into data blocks to facilitate channel estimation. We consider one-way training-based systems with and without feedback, as well as two-way training-based systems. Two-way training enables both the transmitter and the receiver to obtain the channel state information (CSI) through reverse training and forward training, respectively. In all considered cases, we derive efficient strategies for transmit time and/or energy allocation among the pilot and data symbols. These strategies usually have analytical closed-form expressions and can achieve near optimal capacity performance evidenced by extensive numerical analysis.

In one-way training-based systems without feedback, we consider both spatially independent and correlated channels. For spatially independent channels, we provide analytical bounds on the optimal training length and study the optimal antenna configuration that maximizes an ergodic capacity lower bound. For spatially correlated channels, we provide simple pilot and data transmission strategies that are robust under least-favorable channel correlation conditions.

In one-way training-based systems with feedback, we study channel gain feedback (CGF), channel covariance feedback (CCF) and hybrid feedback. For spatially independent channels with CGF, we show that the solutions to the optimal training length and energy coincide with those for systems without feedback. For spatially correlated channels with CCF, we propose a simple transmission scheme, taking into account the fact that the optimal training length is at most as large as the number of transmit antennas. We then provide a solution to the optimal energy

allocation between pilot and data transmissions, which does not depend on the channel spatial correlation under a mild condition. Our derived resource allocation strategies in CGF and CCF systems are extended to hybrid CCF-CGF systems.

In two-way training-based systems, we provide analytical solutions to the transmit power distribution among the different training phases and the data transmission phase. These solutions are shown to have near optimal symbol error rate (SER) and capacity performance. We find that the use of two-way training can provide noticeable performance improvement over reverse training only when the system is operating at moderate to high signal-to-noise ratio (SNR) and using high-order modulations. While this improvement from two-way training is insignificant at low SNR or low-order modulations.

In the second part of the thesis, we consider transmission resource allocation in security-constrained systems. Due to the broadcast nature of the wireless medium, security is a fundamental issue in wireless communications. To guarantee secure communication in the presence of eavesdroppers, we consider a multi-antenna transmission strategy which sends both an information signal to the intended receiver and a noise-like signal isotropically to confuse the eavesdroppers. We study the optimal transmit power allocation between the information signal and the artificial noise. In particular, we show that equal power allocation is a near optimal strategy for non-colluding eavesdroppers, while more power should be used to generate the artificial noise for colluding eavesdroppers. In the presence of channel estimation errors, we find that it is better to create more artificial noise than to increase the information signal strength.

List of Acronyms

AWGN	additive white Gaussian noise
CCF	channel covariance feedback
CGF	channel gain feedback
CSI	channel state information
EVD	eigenvalue decomposition
GSM	global system for mobile communications
LMMSE	linear minimum mean square error
MIMO	multiple-input multiple-output
MISO	multiple-input single-output
OFDM	orthogonal frequency-division multiplexing
PSAM	pilot-symbol-assisted modulation
PSK	phase shift keying
QAM	quadrature amplitude modulation
QoS	quality of service
QPSK	quadrature phase shift keying
SER	symbol error rate
SIMO	single-input multiple-output
SINR	signal to interference and noise ratio
SNR	signal-to-noise ratio
TDD	time-division duplex
UCA	uniform circular array
WLAN	wireless local area network
WMAN	wireless metropolitan area network
ZMCSCG	zero-mean circularly symmetric complex Gaussian

Notations

$\boldsymbol{A}^\dagger$	complex conjugate transpose of matrix $\boldsymbol{A}$
$\boldsymbol{a}^\dagger$	complex conjugate transpose of vector $\boldsymbol{a}$
$\boldsymbol{A}^T$	transpose of matrix $\boldsymbol{A}$
$\boldsymbol{a}^T$	transpose of vector $\boldsymbol{a}$
$\boldsymbol{A}^*$	complex conjugate of matrix $\boldsymbol{A}$
$\boldsymbol{a}^*$	complex conjugate of vector $\boldsymbol{a}$
$\|\boldsymbol{a}\|$	Euclidian norm of vector $\boldsymbol{a}$
$\|\boldsymbol{A}\|$	determinant of matrix $\boldsymbol{A}$
$\|a\|$	absolute value of scalar a
$\mathrm{tr}\{\boldsymbol{A}\}$	trace of matrix $\boldsymbol{A}$
$\boldsymbol{I}_n$	$n \times n$ identity matrix
$E\{\cdot\}$	mathematical expectation
$\Gamma(\cdot)$	the Gamma function
$\mathrm{E}_n(\cdot)$	the generalized exponential integral
$\mathrm{B}(\cdot,\cdot)$	the Beta function
${}_2\mathrm{F}_1(\cdot)$	the Gauss hypergeometric function

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and Background

Over the past thirty years, the wireless communication technology has undergone three generations trying to meet the increasing demand for high-speed and reliable communications. Error-resilient high-rate information transfer has proven to be very challenging in the time-varying channel conditions due to the user mobility. The existing wireless networks have gradually become saturated by conventional mobile telephone traffic. To address the ever-increasing demand, the telecommunication industry is constantly looking into new network architectures and more efficient use of transmission resource.

The main impairment in wireless communication systems is the wireless channel, a key feature of which is multipath fading [1]. The transmitted signal is reflected and diffracted by the obstacles in the wireless environment, resulting in multiple copies of the same signal superimposed at the receiver. Due to the movement in the wireless environment and the change in locations of the transmit-receive pair, the multipath phenomenon causes rapid change in both the amplitude and the phase of the received signal. Early designs of wireless communication systems aimed to mitigate the multipath fading effect, while recent designs utilize the multipath fading to improve the system performance. The technique that makes use of the channel fading effect is named diversity, which allows the reception of a number of different replica of the transmitted signal over time (temporal diversity), space (spatial diversity), frequency (frequency diversity) or antenna polarization (polarization diversity) [2–4]. The system performance measures used to assess the diversity schemes are usually in terms of the information capacity or the detection error probability. The former is an important theoretical metric characterizing the maximum data rate that the wireless channel can support, while the latter is a prac-

tical metric characterizing the quality of service (QoS) of a particular transceiver design.

In this thesis, we focus on the information-theoretic performance of spatial diversity systems with multiple transmit and/or receive antennas. The idea of spatial diversity using multiple transmit and receiver antennas was first discussed in [3]. The spatial diversity technique combines multiple copies of the signal received at different antennas in an optimal way to improve the information capacity with no extra expense on transmit power or bandwidth. This surprising result has stimulated enormous amount of research in the area of multi-antenna systems. Pioneering works on multi-antenna systems demonstrated a linear improvement in the capacity with the number of antennas [5, 6]. Recent literature surveys on multi-antenna techniques can be found in [7, 8]. Multi-antenna technology has been incorporated into current wireless communication standards, including wireless local area networks (WLAN) [9–11], wireless metropolitan area networks (WMAN) [11–13], the third generation cellular networks [14,15] and even in a second generation terrestrial digital video broadcasting system (DVB-T2) [16].

In multi-antenna systems, there are many degrees of freedom in transmission resource optimization, such as the structure of data transmission and the power allocation among different transmit antennas. Therefore, the main focus of this thesis is on the transmission resource allocation to realize the enormous capacity gains with multi-antenna transmission. In particular, we study the problem of resource allocation in the following two scenarios:

- systems with channel estimation errors
- systems with physical-layer security constraints

In order to achieve the maximum capacity gain offered by multi-antenna systems, the channel state information (CSI) is required at the transmitter and/or the receiver. In practical systems, however, the states of the wireless channel are unknown and hence, need to be estimated. Therefore, an accurate channel estimation is crucial in achieving high system performance. Both blind and pilot-assisted channel estimation methods were proposed in the literature, *e.g.*, [17–23]. The use of pilot signals, which is known to the receiver, simplifies the channel estimation problem. However, a portion of the transmission resource, such as time and energy, needs to be allocated to pilot transmission. As a result, less resource is left for transmitting the actual data. Hence, the optimal resource allocation between pilot and data becomes a very interesting and important design problem.

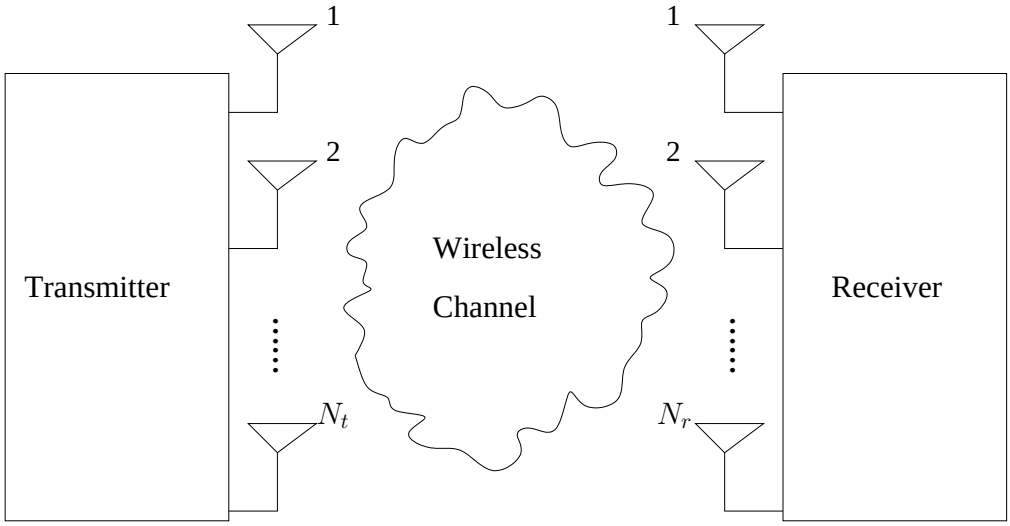

Figure 1.1: An example of a multiple-input multiple-output (MIMO) communication system over a wireless channel. The transmitter has N_t antennas and the receiver has N_r antennas.

Security is a fundamental problem in wireless communications due to the broadcast nature of the wireless medium. This problem deals with the transmission of a confidential message to the intended receiver in the presence of malicious users (also called eavesdroppers). With multiple transmit antennas, the transmitter is able to send the confidential message to the intended receiver and at the same time generating a noise-like signal to confuse the eavesdroppers [24]. As the total transmit power is limited, the ratio of power allocated to the transmission of the confidential message and the artificial noise is a crucial design parameter for maximizing the achievable data rate while ensuring security.

In the reminder of this chapter, we first provide background information on the capacity of multi-antenna systems without channel estimation errors and security constraints. Then we discuss new design problems by considering imperfect channel estimation or security constraints, which are to be investigated in this thesis.

1.1.1 Capacity of Multi-Antenna Systems

Fig. 1.1 shows an example of a multiple-input multiple-output (MIMO) communication system. Assuming perfect symbol timing synchronization and matched filter sampling [20], the discrete-time input-output relationship is given by

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}, \tag{1.1}$$

where

- N_t is the number of antennas at the transmitter
- N_r is the number of antennas at the receiver
- $\boldsymbol{y}$ is the $N_r \times 1$ received symbol vector
- $\boldsymbol{x}$ is the $N_t \times 1$ transmitted symbol vector
- $\boldsymbol{H}$ is the $N_r \times N_t$ channel gain matrix with zero-mean complex Gaussian entries. The $(i,j)^{\text{th}}$ entry, denoted by $[\boldsymbol{H}]_{ij}$, represents the channel gain from the j^{th} transmit antenna to the i^{th} receive antenna
- $\boldsymbol{n}$ is the $N_r \times 1$ additive white Gaussian noise (AWGN) vector.

The signal model in (1.1) is suitable for narrowband communication systems where the channel is flat fading [1], *i.e.*, the signal bandwidth is smaller than the coherence bandwidth of the wireless channel.

From an information-theoretic viewpoint, the information capacity is an fundamental system measure which depends on various parameters, such as the antenna configuration and the availability of CSI. An overview of results on the capacity limits of MIMO channels can be found in [25]. Most existing results in the literature, which focused on coherent detection, assumed perfect CSI at the receiver. With the perfect receiver CSI assumption, the ergodic capacity of a generic MIMO system described in (1.1) is given by [6]

$$C \;=\; E_{\boldsymbol{H}}\Big\{ \log_2 \Big| \boldsymbol{I}_{N_t} + \frac{1}{\sigma_{\boldsymbol{n}}^2} \boldsymbol{H}^{\dagger} \boldsymbol{H} \boldsymbol{Q} \Big| \Big\}, \tag{1.2}$$

where $\boldsymbol{Q} = E\{\boldsymbol{x}\boldsymbol{x}^{\dagger}\}$ is the input (data) covariance matrix and $\sigma_{\boldsymbol{n}}^2$ is the AWGN noise power at the receiver. The expectation takes into account all possible realizations of the MIMO channel. Hence, the ergodic capacity is a long-term performance measure of a communication system.

The studies on MIMO systems can be categorized based on the availability of CSI at the transmitter side as:

- *Systems with no transmitter CSI*: the transmitter has no knowledge about the channel.
- *Systems with statistical transmitter CSI*: the transmitter knows the spatial correlation between the MIMO channels, but not the instantaneous channel gains.

- *Systems with full transmitter CSI*: the transmitter knows the instantaneous channel gains and the channel spatial correlation.

In the following, we discuss the existing results on each of the above categorizes with the perfect receiver CSI assumption.

No transmitter CSI with spatially independent and identically distributed (i.i.d.) channels: The case of spatially i.i.d. channels can be justified by having sufficient scattering around the transmitter and the receiver as well as sufficient spacing between antenna elements. For spatially i.i.d. channels, the results in [5, 6] showed that it is optimal to evenly distribute the transmit power among the N_t antennas and transmit statistically independent Gaussian input (data) signals, when no CSI is available at the transmitter. Mathematically, the transmitter chooses $\boldsymbol{Q} = \frac{\mathcal{P}_d}{N_t}\boldsymbol{I}_{N_t}$, where $\mathcal{P}_d$ is the total data transmit power. More importantly, their results demonstrated a linear improvement in the capacity with the number of antennas, without increasing transmission resource. The ergodic capacity of a MIMO system with the CSI perfectly known at the receiver is given by

$$C \;=\; E_{\boldsymbol{H}}\Big\{ \log_2 \Big| \boldsymbol{I}_{N_t} + \frac{\mathcal{P}_d}{\sigma_{\boldsymbol{n}}^2 N_t}\boldsymbol{H}^{\dagger}\boldsymbol{H} \Big| \Big\}. \tag{1.3}$$

No transmitter CSI with spatially correlated channels: The spatial correlation between channels arises due to practical constraints on the antenna configuration, the distance between the transmitter and the receiver, as well as insufficient scattering environment [26–28]. Without transmitter CSI, the transmitter may still evenly distribute the transmit power among the N_t transmit antennas resulting in the same ergodic capacity expression as in the case of i.i.d. channels given in (1.3).

Furthermore, the impact of channel spatial correlation on the capacity was investigated in [29, 30]. Their result showed that the ergodic capacity decreases as the channel spatial correlation among either the transmit antennas or the receive antennas increases. Since the transmitter does not have any knowledge about the channel correlation, it is difficult to find the optimal transmission strategy that maximizes the capacity. An alternative approach was used in [31] to find a *robust* transmission strategy that maximizes the capacity for the least-favorable channel correlation. Their result showed that evenly distributing the transmit power among the N_t antennas is indeed the most robust transmission strategy.

Statistical transmitter CSI with spatially correlated channels[1]: The case where the channel spatial correlation is known at the transmitter is often referred to as systems with covariance feedback. For this type of systems, the transmit power can be optimally distributed among the N_t antennas for maximum capacity. Most existing studies on covariance feedback systems considered transmitter-side correlation [32–34, 36–38]. We define the covariance matrix of the channels as

$$\boldsymbol{R_H} = \frac{E\{\boldsymbol{H}^\dagger \boldsymbol{H}\}}{N_r}. \tag{1.4}$$

Therefore, $\boldsymbol{H} = \boldsymbol{H}_0 \boldsymbol{R}_{\boldsymbol{H}}^{1/2}$, where $\boldsymbol{H}_0$ has i.i.d. complex Gaussian entries. Let the eigenvalue decomposition (EVD) of the channel covariance matrix be $\boldsymbol{R_H} = \boldsymbol{U}\boldsymbol{G}\boldsymbol{U}^\dagger$. $\boldsymbol{U}$ is a unitary matrix whose columns represent the eigenvectors of $\boldsymbol{R_H}$. $\boldsymbol{G}$ is a diagonal matrix whose entries represent the eigenvalues of $\boldsymbol{R_H}$ given by g_i, $\forall\, i = 1, ..., N_t$. It was shown in [32,36] that the optimal transmit power allocation is given by $\boldsymbol{Q} = \boldsymbol{U}\widehat{\boldsymbol{Q}}\boldsymbol{U}^\dagger$, where $\widehat{\boldsymbol{Q}}$ is a diagonal matrix whose entries are $\widehat{q}_i$, $\forall\, i = 1, ..., N_t$. That is to say, $\boldsymbol{Q}$ has the same eigenvectors as $\boldsymbol{R_H}$. Then, the ergodic capacity in (1.2) reduces to

$$\begin{aligned} C &= E_{\boldsymbol{H}}\Big\{ \log_2 \Big| \boldsymbol{I}_{N_r} + \frac{1}{\sigma_{\boldsymbol{n}}^2} \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^\dagger \Big| \Big\} \\ &= E_{\boldsymbol{H}_0}\Big\{ \log_2 \Big| \boldsymbol{I}_{N_r} + \frac{1}{\sigma_{\boldsymbol{n}}^2} \boldsymbol{H}_0 \boldsymbol{G}^{1/2} \widehat{\boldsymbol{Q}} \boldsymbol{G}^{1/2} \boldsymbol{H}_0{}^\dagger \Big| \Big\} \\ &= E_{\boldsymbol{z}}\Big\{ \log_2 \Big| \boldsymbol{I}_{N_r} + \frac{1}{\sigma_{\boldsymbol{n}}^2} \sum_{i=1}^{N_t} g_i \widehat{q}_i \boldsymbol{z}_i \boldsymbol{z}_i{}^\dagger \Big| \Big\}, \end{aligned} \tag{1.5}$$

where $\boldsymbol{z}_i$ is the i^{th} column of $\boldsymbol{H}_0$.

To achieve maximum capacity, $\widehat{q}_i$ needs to be optimized subject to a total power constraint given by $\sum_{i=1}^{N_t} \widehat{q}_i = \mathcal{P}_d$. Intuitively, more power should be allocated into stronger (statistical) eigen-channels, that is to say, $\widehat{q}_i$ is an increasing function of g_i. Unfortunately, there is no closed-form solution for $\widehat{q}_i$ and only numerical methods are available [38]. In the special case where only one of $\widehat{q}_i$ is non-zero, the transmission is referred to as beamforming. The use of beamforming implies $\widehat{q}_i = \mathcal{P}_d$ if g_i is the largest eigenvalue of $\boldsymbol{R_H}$. The condition under which beamforming transmission is the optimal transmission strategy was found in [33, 34, 36, 37]. The general message is that beamforming is optimal when the channel spatial correlation

[1]The most commonly considered statistical transmitter CSI is the channel spatial correlation, which is discussed in this thesis. Apart from spatial correlation, the mean value of the channel gains were also studied in the literature, usually referred to as mean feedback. The results on capacity of multi-antenna systems with mean feedback can be found in [32–35].

is sufficiently large or the signal-to-noise ratio (SNR) is sufficiently low. When both the transmitter-side and receiver-side correlations are considered, the results in [39] showed that the optimal eigenvectors of $\boldsymbol{Q}$ are still given by the eigenvectors of the transmitter-side covariance matrix. However, the optimal eigenvalues of $\boldsymbol{Q}$ depend on both the transmitter-side and receiver-side correlations. The optimality of beamforming transmission was also studied in [39].

Furthermore, the impact of channel spatial correlation on the capacity was investigated in [30]. With covariance feedback, it was shown that the ergodic capacity is an increasing function of the channel spatial correlation. This is in contrast with the result for systems with no transmitter CSI.

Full transmitter CSI with either i.i.d. or correlated channels: When each realization of the channel gains is known at the transmitter, the data transmission can be adaptive according to the channel realization and hence, the capacity can be further improved. Regardless of the spatial correlation between the channels, the transmit power allocation among the N_t antennas follows a water-filling solution [6]. Let the EVD of $\boldsymbol{H}^\dagger\boldsymbol{H}$ be $\boldsymbol{H}^\dagger\boldsymbol{H} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\dagger$, where the (diagonal) elements of $\boldsymbol{\Lambda}$ are given by $\chi_i, \forall\, i = 1, ..., N_t$. The water-filling solution for the transmit power allocation gives $\boldsymbol{Q} = \boldsymbol{V}\widehat{\boldsymbol{Q}}\boldsymbol{V}^\dagger$ where the entries of the diagonal matrix $\widehat{\boldsymbol{Q}}$ are given by

$$\widehat{q_i} \;=\; [\eta - \sigma_{\boldsymbol{n}}^2\chi_i^{-1}]^+, \tag{1.6}$$

where $[a]^+ = \max\{0, a\}$ and η represents the water level determined by the total power constraint $\sum_{i=1}^{N_t}\widehat{q_i} = \mathcal{P}_d$. With the optimal transmit power allocation, the ergodic capacity in (1.2) reduces to

$$\begin{aligned} C &= E_{\boldsymbol{\Lambda}}\Big\{\log_2\Big|\boldsymbol{I}_{N_t} + \frac{1}{\sigma_{\boldsymbol{n}}^2}\boldsymbol{\Lambda}\widehat{\boldsymbol{Q}}\Big|\Big\} \\ &= E_{\boldsymbol{\chi}}\Big\{\sum_{i=1}^{N_t}\log_2\Big(1 + \frac{1}{\sigma_{\boldsymbol{n}}^2}\chi_i\widehat{q_i}\Big)\Big\} \\ &= E_{\boldsymbol{\chi}}\Big\{\sum_{i=1}^{m}\log_2\Big(\frac{\eta\chi_i}{\sigma_{\boldsymbol{n}}^2}\Big)\Big\}, \end{aligned} \tag{1.7}$$

where m denotes the number of non-zero $\widehat{q_i}$ and the expectation in (1.7) is taken over the m largest χ_i. Compared with the capacity of systems with no transmitter CSI, the capacity improvement from water-filling is significant at low SNR while the improvement reduces as SNR increases. At asymptotically high SNR, the water-filling solution converges to the equal power distribution, *i.e.*, $\widehat{q_i} = \mathcal{P}_d/N_t$ or

$\boldsymbol{Q} = \frac{\mathcal{P}_d}{N_t}\boldsymbol{I}_{N_t}$, which is the optimal solution for systems with no transmitter CSI.

Taking into account the time-varying nature of the fading channels, a temporal water-filling with a long-term average power constraint was introduced in [40,41] in addition to the spatial water-filling discussed above. Intuitively, less power should be used when the channel is weak and more power should be used when the channel is strong. Therefore, the capacity can be further improved by applying the temporal water-filling.

For correlated channels, the optimality of beamforming, which only uses the strongest eigen-channel for transmission, was studied in [42]. Furthermore, the impact of channel spatial correlation on the capacity was investigated in [30]. With full transmitter CSI, the ergodic capacity is a decreasing function of the channel spatial correlation. This agrees with the result for systems with no transmitter CSI, while it is in contrast with that for systems with covariance feedback.

The capacity results from the literature discussed so far assume perfect receiver CSI. In the next section, we focus on practical transmission schemes and identify new design problems with imperfect receiver CSI.

1.1.2 Training-Based Systems

In practical communication systems with coherent detection, the CSI needs to be estimated at the receiver and hence, the CSI is never perfectly known due to the noise and time variations in the fading channels. Pilot-symbol-assisted modulation (PSAM) has been widely used in many practical communication systems [43], *e.g.*, in Global System for Mobile Communications (GSM) [44], to assist estimation of unknown channel parameters. In PSAM schemes, training symbols or pilots are inserted into data blocks periodically to facilitate channel estimation at the receiver [43]. Therefore, the channel gains can be modeled as

$$\boldsymbol{H} = \widehat{\boldsymbol{H}} + \widetilde{\boldsymbol{H}}, \tag{1.8}$$

where $\widehat{\boldsymbol{H}}$ denotes the estimate of $\boldsymbol{H}$ and $\widetilde{\boldsymbol{H}}$ denotes the estimation error. The choice of channel estimation method depends on the performance criteria, *e.g.*, mean square error or likelihood. In this thesis, we consider the minimum mean square error (MMSE) estimator as the optimal estimator.

A survey on various PSAM schemes can be found in [45]. Generally speaking, there are two types of PSAM schemes, namely time-division multiplexing and superimposed transmission, as shown in Fig. 1.2. The pilot and data symbols are time multiplexed in the transmission blocks in the former type, while the pilot and

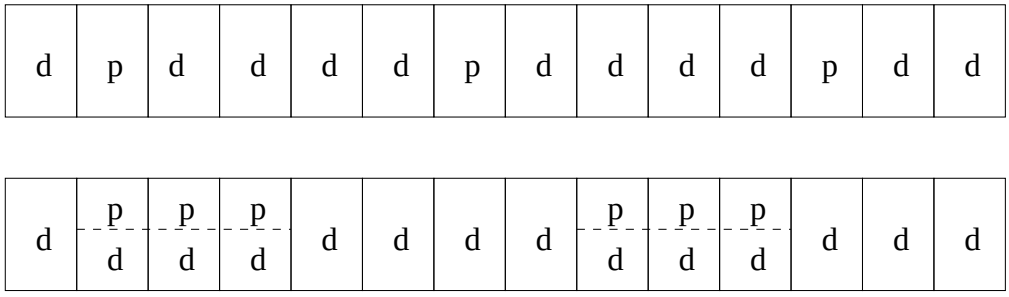

Figure 1.2: Examples of pilot-symbol-assisted modulation (PSAM). Each rectangular block represents a transmitted symbol. d denotes a data symbol and p denotes a pilot symbol. The time-division multiplexing scheme is shown on the top and the superimposed transmission scheme is shown at the bottom.

data symbols are transmitted simultaneously in the latter type. This thesis focuses on the time-division multiplexing schemes.[2]

Similar to the previous section, we can categorize the studies on training-based MIMO systems based on the availability of CSI at the transmitter side as:

- *One-way training-based systems with no feedback*: the transmitter has no knowledge about the channel, *i.e.*, no transmitter CSI.
- *One-way training-based systems with feedback*: there exists a feedback link from the receiver to the transmitter to provide the following two forms of transmitter CSI:
  - *Channel covariance feedback (CCF)*: the transmitter knows the spatial correlation between the MIMO channels, *i.e.*, statistical transmitter CSI.
  - *Channel gain feedback (CGF)*: the transmitter knows the estimated channel gains. Note that when the channels are spatially correlated, the channel correlation can be readily obtained from CGF, which effectively results in a hybrid CCF-CGF system, *i.e.*, full transmitter CSI.
- *Two-way training-based systems*: the channel gains are estimated by both the transmitter and the receiver with two-way pilot transmissions.

In the following, our discussions are focused on traditional one-way training-based systems to provide background knowledge for Chapters 2 and 3. On the other hand, two-way training was recently proposed in [49] as an alternative way of obtaining transmitter CSI, instead of using explicit feedback. In this scheme,

[2]The design of superimposed transmission schemes was discussed in [22, 46–48].

the transmitter acquires the outgoing CSI using the pilots sent from the receiver (*i.e.*, reverse training) and then the receiver estimates the effective CSI using the pilots sent from the transmitter (*i.e.*, forward training). We will discuss two-way training-based systems in detail in Chapter 4.

Ergodic capacity with imperfect receiver CSI: With channel estimation errors at the receiver, the capacity results discussed in Section 1.1.1 are not achievable. The degradation in the capacity due to imperfect receiver CSI was analyzed in [50–52]. The exact ergodic capacity with imperfect receiver CSI is still unknown. However, analytical bounds were derived in the literature [50, 53]. In particular, a lower bound and an upper bound on the ergodic capacity were derived for spatially i.i.d. MIMO channels in [53], and they were shown to be reasonably tight. The ergodic capacity lower bound is given by

$$C \;=\; E_{\widehat{\boldsymbol{H}}}\Big\{ \log_2 \Big| \boldsymbol{I}_{N_t} + \frac{1}{\sigma_{\boldsymbol{n}}^2 + \sigma_{\widetilde{\boldsymbol{H}}}^2 \mathcal{P}_d} \widehat{\boldsymbol{H}}^{\dagger} \widehat{\boldsymbol{H}} \boldsymbol{Q} \Big| \Big\}, \tag{1.9}$$

where $\sigma_{\widetilde{\boldsymbol{H}}}^2$ is the variance of the channel estimation error $\widetilde{\boldsymbol{H}}$. One can see from (1.9) that the effect of channel estimation errors is amplified by the data transmit power $\mathcal{P}_d$. Therefore, for a fixed quality of channel estimation, the ergodic capacity saturates as the data transmit power increases [53]. This is in contrast with the perfect receiver CSI case where the ergodic capacity always increases with the data transmit power.

The optimal data transmission strategy for spatially i.i.d. MIMO channels was discussed in [53–56]. With no transmitter CSI, it is optimal to evenly distribute the transmit power among N_t antennas, *i.e.*, $\boldsymbol{Q} = \mathcal{P}_d/N_t$. With full transmitter CSI[3], the optimal strategy follows a spatial water-filling solution according to $\widehat{\boldsymbol{H}}^{\dagger}\widehat{\boldsymbol{H}}$ (normalized by $\sigma_{\boldsymbol{n}}^2 + \sigma_{\widetilde{\boldsymbol{H}}}^2 \mathcal{P}_d$) instead of $\boldsymbol{H}^{\dagger}\boldsymbol{H}$. Taking into account the time-varying nature of the fading channels, another temporal water-filling can be applied for such MIMO systems with a long-term power constraint. However, it was shown in [53] that the temporal water-filling does not give noticeable capacity gains when the spatial water-filling is already used.

For spatially correlated MIMO channels, the derivation of capacity bounds with imperfect receiver CSI remains an open problem. Some initial result was reported in [57] in which a suboptimal channel estimation method was chosen to obtain a lower bound on the ergodic capacity. Another attempt can be found in [58] in which a relaxed ergodic capacity lower bound was found for a special class of rank-

[3]Full transmitter CSI means the transmitter has the same channel knowledge as the receiver, that is the channel estimates $\widehat{\boldsymbol{H}}$.

deficient channel covariance matrices. However, it is unclear how tight the relaxed capacity bound is. This thesis aims to derive a reasonably tight lower bound on the ergodic capacity for correlated MIMO channels with the optimal channel estimation method.

Design of pilot structure: The quality of the channel estimation is determined by the structure and power of pilot transmission. The pilot structure becomes important for MIMO systems in which there is a large number of channel gains to be estimated. By optimally designing the pilot structure, the channel estimation error can be minimized, which in turn, maximizes the ergodic capacity. For spatially i.i.d. channels, the optimal pilot signals are orthogonal with respect to time among the transmit antennas and the total energy of the pilots transmitted on each antenna is equal to each other [21, 54, 59].

For spatially correlated MIMO channels, the design of pilot structure depends on the availability of statistical transmitter CSI. For systems with no transmitter CSI, it is difficult to find the optimal pilot structure for the unknown channel spatial correlation. In this thesis, we look for a robust design that minimizes the channel estimation error for the least-favorable channel correlation. For systems with covariance feedback, the optimal pilot structure is to train along the eigenvectors of the channel covariance matrix with training power being water-filled according to the eigenvalues of the channel covariance matrix [60–63]. It was shown in [62, 63] that the use of the optimal pilot structure, compared to the orthogonal pilot structure, significantly reduces the channel estimation error at low SNR for correlated MIMO channels. While at asymptotically high SNR, the water-filling solution converges to the equal power distribution, which makes the optimal pilot signals having an orthogonal structure.

Transmission resource allocation: A crucial and nontrivial problem in PSAM systems is the transmission resource allocation between the pilot and data symbols. Since pilots carry no useful information, the insertion of pilots reduces the information capacity as less resource is allocated to the actual data transmission. Trade-off analysis is required on the distribution of transmission resource, such as the transmit time and energy to pilots and data. On one hand, allocating more resource to pilot transmission results in more accurate channel estimation which helps the detection during data transmission. On the other hand, when more resource is allocated to pilots, less resource is left for data transmission which directly affects the information capacity. Therefore, it is important to find the optimal resource allocation in order to achieve maximum capacity.

In the time-division multiplexing PSAM scheme, only a portion of the total

transmit time is used for data transmission and the rest of the time is used for training. Denote the proportion of time for data transmission as ψ. Taking into account the training overhead, the ergodic capacity lower bound for spatially i.i.d. channels is given by

$$\overline{C} \;=\; \psi E_{\widehat{\boldsymbol{H}}}\Big\{ \log_2 \Big| \boldsymbol{I}_{N_t} + \frac{1}{\sigma_{\boldsymbol{n}}^2 + \sigma_{\widetilde{\boldsymbol{H}}}^2 \mathcal{P}_d} \widehat{\boldsymbol{H}}^{\dagger} \widehat{\boldsymbol{H}} \boldsymbol{Q} \Big| \Big\}. \tag{1.10}$$

From the results in [53] we know that the ergodic capacity lower bound in (1.10) is reasonably tight. Hence, it can be used as the objective function to study the transmission resource allocation problem.

Existing studies on the optimal training length and energy allocation mainly focused on spatially i.i.d. channels with no transmitter CSI [54,64–67]. The common trends found in these works are as follows. Firstly, the optimal training length increases with the number of transmit antennas, while it is independent of the number of receive antennas. Secondly, the frequency of pilot transmission increases with the Doppler spread of the fading channel. Thirdly, the ratio of total energy allocated to pilot transmission decreases as the average operating SNR increases. The specific design guidelines depend on the system model used. For example, the authors in [54] considered a block-fading MIMO channel model which assumes that the channel gains stay constant over a transmission block and change to independent realizations in the next block. For the block-fading channel model, they showed that the optimal training length equals the number of transmit antennas and the optimal ratio of energy allocated to training was found in a simple closed form.

However, optimal solutions for the training length and energy allocation are generally unknown for both spatially i.i.d. and correlated MIMO systems with statistical or full transmitter CSI. For i.i.d. channels with full transmitter CSI, the water-filling solution of the data transmission results in a complicated expression for the ergodic capacity lower bound, which makes the resource allocation problem difficult to solve. For correlated channels, the lack of a tight ergodic capacity bound is the key hurdle in studying the resource allocation problem from an information-theoretic viewpoint. In this thesis, we address these issues and provide design guidelines on transmission resource allocation with various forms of transmitter CSI.

1.1.3 Security-Constrained Systems

Security is a fundamental problem in wireless communications due to the broadcast nature of the wireless medium. Traditionally, secure communication is achieved

by using cryptographic technologies such as encryption. On the other hand, the studies from an information-theoretic viewpoint have found conditions for reliable secure communication without using traditional cryptographic technologies that are based on unproven mathematical assumptions such as the assumed infeasibility of factoring large integers and finding discrete logarithms over large finite fields [68]. The notion of information-theoretic security provides a fundamental measure of the security level of a system at the physical layer, which is independent of the computational power of the eavesdropper. The technique to realize such security is particularly important in the initial stage of exchanging the secret key in a practical secure communication system. Furthermore, this physical-layer approach can be used jointly with encryption to enhance security in future wireless communication systems.

In this thesis, we carry out an information-theoretic study on the transmit power allocation for physical-layer security-constrained systems with a specific multi-antenna transmission scheme. In the following, we first introduce the notion of secrecy capacity and the transmission scheme. Then, we briefly discuss the problem of transmit power allocation.

Secrecy Capacity: In the pioneering works on information-theoretic security, Wyner introduced the wiretap channel for single point-to-point communication [69], which was extended to broadcast channels by Csiszár and Körner [70]. In particular, a widely-used lower bound on the secrecy capacity of a wireless link in the presence of an eavesdropper is given as [70]

$$C_S = [C_1 - C_2]^+, \tag{1.11}$$

where C_1 denotes the information capacity of the channel between the transmitter and the intended receiver, C_2 denotes the information capacity of the channel between the transmitter and the eavesdropper, and $[a]^+ = \max\{0, a\}$ ensures the secrecy capacity lower bound to be non-negative. From (1.11) we see that a positive secrecy capacity can be achieved if the intended receiver has a better channel than the eavesdropper.

Recently, information-theoretic security with multi-antenna transmission has drawn considerable attention. Many works have been devoted to deriving and analyzing the secrecy capacity with various antenna configurations and channel conditions. With multiple antennas at the transmitter, the optimal input structure (for Gaussian codes) that achieves the secrecy capacity of Gaussian channels was found to be in the form of beamforming transmission [71, 72]. The secrecy capacity of Gaussian channels with multiple antennas at both the transmitter and

the receiver was obtained in [73, 74]. These results were also extended to multiple-access [75, 76] and broadcast channels [77, 78]. One of the main assumptions in the above-mentioned works is that both the intended receiver's channel and the eavesdropper's channel are known at the transmitter. Clearly the assumption of knowing eavesdropper's channel is not always possible, especially for fading channels. The ergodic secrecy capacity without knowing the eavesdropper's channel was studied in [79, 80]. The authors in [79] considered the single antenna case and proposed an on-off power transmission with variable-rate allocation scheme, which was shown to approach the optimal performance at asymptotically high SNR. The authors in [80] extended the ergodic secrecy capacity result to systems with multiple antennas and developed capacity bounds in the large antenna limit.

Transmission of Artificial Noise: Various physical-layer signal processing techniques were proposed to achieve secure communication even if the receiver's channel is worse than the eavesdropper's channel. One of the main techniques is the use of interference or artificial noise to confuse the eavesdropper. With two base stations connected by a high capacity backbone, one base station can simultaneously transmit an interfering signal to secure the uplink communication for the other base station [81, 82]. In the scenario where the transmitter has a helping interferer or a relay node, the secrecy level can also be increased by having the interferer [83] or relay [84] to send codewords independent of the source message at an appropriate rate. When multiple cooperative nodes are available to help the transmitter, the secrecy rate maximizing weights of the signal transmitted from cooperative nodes were derived for both decode-and-forward [85] and amplify-and-forward [86] protocols.

When multiple antennas are available at the transmitter, it is possible to simultaneously transmit both the information bearing signal and the artificial noise to achieve secrecy in a fading environment [24]. The artificial noise acts as a mask for the information transmission. This technique was studied under the scenario where the transmitter knows the intended receiver's channel but not the eavesdropper's channel. In this case, as shown in Fig. 1.3, the information signal is transmitted into the intended receiver's channel, while the artificial noise is radiated isotropically into all channel directions other than the intended receiver's channel. Hence, the artificial noise is always received by the eavesdropper but not by the intended receiver.

Optimal Power Allocation: We focus on this multi-antenna transmission technique with a constraint on the total transmit power. Since the limited amount of transmit power needs to be used for both the transmission of the information

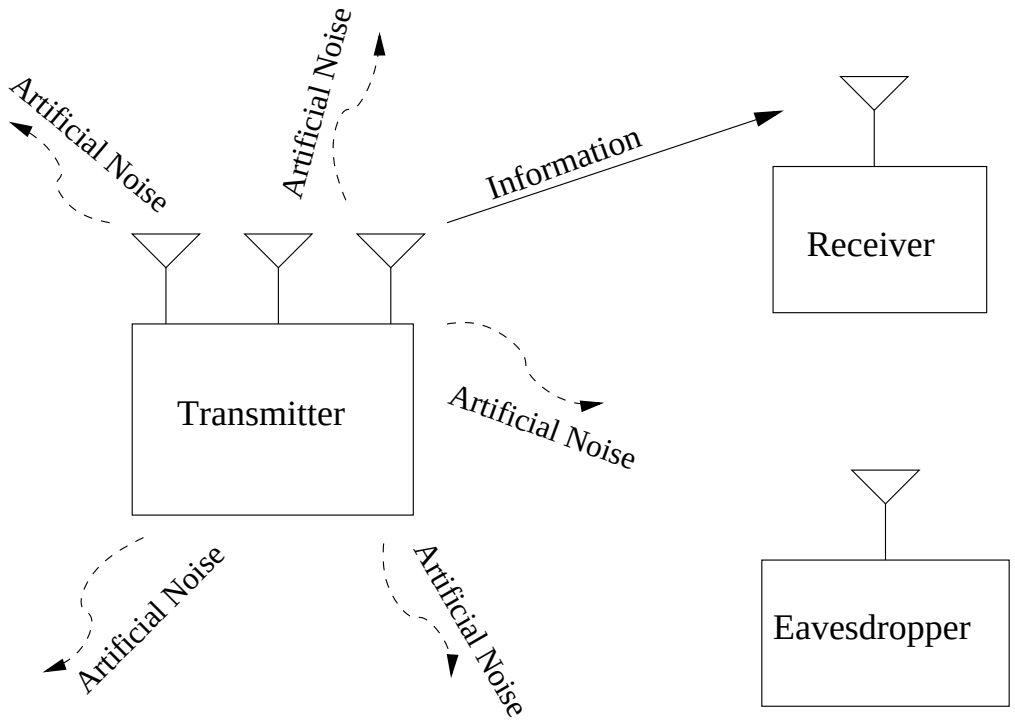

Figure 1.3: Multi-antenna transmission in the presence of the eavesdroppers. The transmitter is able to send the information signal to the intended received, at the same time generating the artificial noise isotropically (except in the direction of the intended receiver) to confuse the eavesdropper.

signal and the artificial noise, the ratio of power allocation between the two becomes an important design parameter. Denote the ratio of the total transmit power allocated to the information signal as ϕ. The secrecy capacity lower bound is clearly a function of ϕ given by

$$C_S(\phi) = [C_1(\phi) - C_2(\phi)]^+, \tag{1.12}$$

where $C_1(\phi)$ and $C_2(\phi)$ are both increasing functions of ϕ. When the transmitter introduces the artificial noise, *i.e.*, ϕ decreases from 1, both $C_1(\phi)$ and $C_2(\phi)$ decrease as less power is used to transmit the information signal. The decrease in $C_2(\phi)$ may be faster than that in $C_1(\phi)$ due to the reception of the artificial noise at the eavesdropper. In this thesis we aim to find the optimal value of the power allocation ratio ϕ to maximize the secrecy capacity lower bound C_S.

1.2 Overview and Contribution of Thesis

The main focus of this thesis is on the transmission resource allocation of multi-antenna wireless communication systems. The ergodic capacity or its bounds are used as the figure of merit. For training-based transmission systems, we study the optimal allocation of the transmit time and energy between pilot symbols and data symbols. For security-constrained systems, we consider the multi-antenna transmission with artificial noise and study the optimal transmit power allocation between the information signal and the artificial noise.

1.2.1 Questions to be Answered

The following open questions are answered in the thesis:

- What is the optimal transmit and receive antenna configuration taking channel estimation errors into account?
- How does the channel spatial correlation affect the design of pilot and data transmission?
- How does the statistical transmitter CSI change the optimal transmission resource allocation?
- How does the full transmitter CSI change the optimal transmission resource allocation?
- What is the optimal transmit energy allocation between pilots and data in two-way training-based systems?
- For multi-antenna transmission in the presence of eavesdroppers, what is the optimal transmit power allocation between the information signal and the artificial noise?
- How does the channel estimation error affect the optimal power allocation between the information signal and the artificial noise?

1.2.2 Thesis Contributions and Organization

Chapter 2 - One-Way Training-Based Systems with No Feedback

Chapter 2 considers and revisits the traditional one-way training-based systems where pilot symbols are periodically inserted into data symbols for channel estimation at the receiver. The transmitter does not have any CSI. We obtain

new practical design solutions to transmit time allocation between pilots and data, transmit energy allocation between pilots and data, as well as the optimal number of transmit and receive antennas. An ergodic capacity lower bound is used as the figure of merit in solving the optimization problems. The new contributions in this chapter are:

- For fixed power transmission, we show that all available time should be used for transmission if the training and data lengths can be jointly optimized. We also provide analytical bounds on the optimal training and data lengths.
- For fixed power transmission with a fixed training length, we derived a threshold SNR above which it is optimal to use all available time for transmission.
- When an extra antenna is available to be placed on either end of the system, we find that it should be placed at the receiver at moderate to high SNR when $N_t \geq N_r$. When $N_t < N_r$, we show that a critical SNR value needs to be considered, below which the extra antenna should be placed at the receiver.
- When it is only practical to change the number of transmit antennas, adding extra antennas generally improves the capacity at high SNR and large channel coherence interval, provided that $N_t < N_r$. More importantly, the optimal number of transmit antennas can exceed the number of receive antennas, and the capacity improvement by optimal antenna configuration is significant when the channel coherence interval is large. We also show that optimizing antenna configuration can be more beneficial than power optimization over pilot and data symbols particularly at large channel coherence intervals.
- We derive a lower and upper bound on the ergodic capacity for spatially correlated MIMO systems. Our numerical results show that the derived bounds are reasonably tight.
- We show that the capacity increases with channel correlation at low SNR, but decreases with correlation at high SNR. We also derive a robust transmission scheme which achieves the best channel estimation and capacity performance for the least-favorable channel correlation.

The results in this chapter have been presented in the following publications which are listed again for ease of reference:

J1. Xiangyun Zhou, Parastoo Sadeghi, Tharaka A. Lamahewa, and Salman Durrani, "Optimizing Antenna Configuration for MIMO Systems with Imperfect

Channel Estimation", *IEEE Trans. Wireless. Commun.*, vol. 8, no. 3, pp. 1177-1181, Mar. 2009.

C1. Xiangyun Zhou, Tharaka A. Lamahewa, Parastoo Sadeghi, and Salman Durrani, "Capacity of MIMO Systems: Impact of Spatial Correlation with Channel Estimation Errors", in *Proc. IEEE Int. Conf. on Commun. Syst. (ICCS)*, Guangzhou, China, Nov. 2008, pp. 817-822.

C4. Xiangyun Zhou, Parastoo Sadeghi, and Tharaka A. Lamahewa, "Optimizing Training-based MIMO Systems: How Much Time is Needed for Actual Transmission?", in *Proc. IEEE Veh. Tech. Conf. (VTC-Spring)*, Taipei, Taiwan, May 2010, pp. 1-5.

Chapter 3 - One-Way Training-Based Systems with Feedback

Chapter 3 considerers the traditional one-way training-based systems as in Chapter 2. The major difference between the two chapters is the availability of transmitter CSI. In Chapter 3, both CCF and CGF, as well as hybrid CCF-CGF are considered. We obtain practical solutions to transmit time and energy allocation between the pilots and data. An ergodic capacity lower bound is used as the figure of merit in solving the optimization problems. The new contributions in this chapter are:

- For systems with i.i.d. channels and estimated channel gains available at the transmitter at the start of the data transmission phase (*i.e.*, no feedback delay), we prove that the solutions to the optimal energy allocation to pilot and data transmission, as well as the optimal training length coincide with the solutions for systems with no transmitter CSI.

- For systems with i.i.d. channels and estimated channel gains available at the transmitter after the data transmission phase starts (due to feedback delay), our numerical results show that evenly distributing the power over the entire data transmission (regardless of the delay time) gives near optimal performance at practical SNR values. As a result, the solutions to the optimal energy allocation to pilot and data transmission, as well as the optimal training length for the delayless feedback system stay nearly optimal for the delayed system regardless of the delay time.

- For systems with correlated channels and statistical transmitter CSI, we propose a simple transmission scheme, taking into account the fact that the

optimal training length is at most as large as the number of transmit antennas. Our numerical results show that this scheme is very close to optimal and optimizing training length can result in a significant capacity improvement for correlated channels.

- Using the proposed transmission scheme mentioned above, we find the solution to the optimal energy allocation to pilot and data transmission, which does not depend on the channel spatial correlation under a mild condition on the channel coherence interval or SNR. Therefore, the proposed transmission and resource allocation schemes give near optimal performance while having very low computational complexity.

- For systems with correlated channels and full transmitter CSI, we prove that the optimal data transmission is given by a water-filling solution according to the estimated channel gains, rotated and truncated into the trained eigendirections.

- For systems with correlated channels and full transmitter CSI, we show that the optimal training length is at most as large as the number of transmit antennas. We consider a closed-form solution of optimal energy allocation between pilot and data and numerically show that this solution achieves near optimal performance.

The results in this chapter have been presented in the following publications which are listed again for ease of reference:

J2. Xiangyun Zhou, Parastoo Sadeghi, Tharaka A. Lamahewa, and Salman Durrani, "Design Guidelines for Training-based MIMO Systems with Feedback", *IEEE Trans. Signal Processing*, vol. 57, no. 10, pp. 4014-4026, Oct. 2009.

C3. Xiangyun Zhou, Tharaka A. Lamahewa, Parastoo Sadeghi, and Salman Durrani, "Optimizing Training-based Transmission for Correlated MIMO Systems with Hybrid Feedback", in *Proc. IEEE Global Commun. Conf. (Globecom)*, Honolulu, HI, Nov. 2009, pp. 1-6.

Chapter 4 - Two-Way Training-Based Systems

Chapter 4 considers two-way training which allows the transmitter CSI and the receiver CSI to be obtained by pilot transmissions from both the receiver (reverse training) and the transmitter (forward training), respectively. We obtain practical solutions to the transmit power distribution on the reverse training, forward

training and data transmission. Both the symbol error rate (SER) and an ergodic capacity approximation are used as the performance measure. The new contributions in this chapter are:

- We derive the linear minimum mean square error (LMMSE) channel estimation for the forward training and find that the forward channel estimation error is almost independent of the number of transmit antennas.
- An average SNR lower bound is used to obtain closed-form solutions to the optimal power allocation at high SNR in three different scenarios. The derived power allocation strategies are independent of the number of transmit antennas.
- We carry out Monte-Carlo simulations to obtain both SER and ergodic capacity performance. For both performance metrics, we verify the near optimality of the derived power allocation strategies over a wide range of SNR values.
- We derive the optimal power allocation for systems with reverse training only. Our numerical results show that two-way training provides no or marginal performance gain over reverse training only, at low SNR or when low-order modulations are used.

Part of the results in this chapter has been presented in the following publication which is listed again for ease of reference:

J3. Xiangyun Zhou, Tharaka A. Lamahewa, Parastoo Sadeghi, and Salman Durrani, "Two-way Training: Optimal Power Allocation for Pilot and Data Transmission", *IEEE Trans. Wireless Commun.*, vol 9, no. 2, pp. 564-569, Feb. 2010.

Chapter 5 - Physical-Layer Security-Constrained Systems

Chapter 5 considers multi-antenna transmission in the presence of eavesdroppers. The transmitter sends the information bearing signal to the intended receiver as well as the artificial noise to confuse the eavesdroppers. The eavesdroppers' channels are unknown and hence, the artificial noise is radiated isotropically except in the direction of the intended receiver. We obtain the optimal transmit power allocation between the information signal and artificial noise and study its sensitivity to channel estimation errors. The new contributions in this chapter are:

- We obtain a closed-form expression for the average secrecy capacity lower bound, which is an achievable data rate that can be guaranteed for secure communication without knowing the noise level at the eavesdroppers.

- We derive the optimal power allocation between transmission of the information signal and the artificial noise. In particular, equal power allocation is shown to be a near optimal strategy in the case of non-colluding eavesdroppers. When the number of colluding eavesdroppers increases, more power should be used to generate artificial noise.
- We also derive an upper bound on the critical SNR above which the achievable secrecy rate is positive and this bound is shown to be tight at low SNR.
- In the presence of channel estimation errors, we find that it is wise to create more artificial noise to confuse the eavesdroppers than to increase the signal strength for the intended receiver.

The results in this chapter have been presented in the following publications which are listed again for ease of reference:

J4. Xiangyun Zhou and Matthew R. McKay, "Secure Transmission with Artificial Noise over Fading Channels: Achievable Rate and Optimal Power Allocation", submitted to *IEEE Trans. Veh. Technol.*, revised in May 2010.

C2. Xiangyun Zhou and Matthew R. McKay, "Physical Layer Security with Artificial Noise: Secrecy Capacity and Optimal Power Allocation", in *Proc. Int. Conf. on Signal Processing and Commun. Syst. (ICSPCS)*, Omaha, NE, Sept. 2009, pp. 1-5.

Finally, Chapter 6 gives a summary of results presented and suggestions for future research work.

Chapter 2

One-Way Training-Based Systems with No Feedback

2.1 Introduction

In resource constrained communications, it is important to optimally allocate the limited amount of resource to achieve maximum data rates. In training-based systems where training symbols or pilots are periodically inserted into data symbols, the total available transmission time and energy should be optimally distributed among training and data symbols. When multiple antennas are available to be used at both the transmitter and the receiver, the optimal antenna configuration (*i.e.*, the number of transmit and receive antennas) is another important design parameter. In addition, the spatial correlation between the channels in multi-antenna systems can significantly affect the system design and performance. In the past few years, numerous results on optimizing resource allocation with certain system model assumptions have been obtained. However, there is still some gap in the knowledge. This chapter aims to fill this gap and studies the optimal resource allocation in training-based systems with no transmitter CSI by addressing the following questions:

Q1. What is the general optimal transmit energy and time allocation between training and data transmission?

Q2. What is the optimal antenna configuration taking into account the channel estimation errors?

Q3. How does the channel spatial correlation affect the optimal transmission resource allocation and capacity performance?

For constantly time-varying channels, the optimal training length and training interval were investigated in [64, 66, 87, 88]. For block fading channels with a fixed coherence block length, the training length for MIMO systems was optimized in [21, 54]. A common assumption in these works is that all available time is used for transmission. In Section 2.4, we revisit this common assumption used in aforementioned works by first relaxing it and then investigating its optimality in training-based transmissions. This is particularly important for many practical systems in which the transmit power for training symbols and data symbols is fixed to the same level. When the training and data lengths are allowed to be jointly optimized, we show that all available time should always be used for transmission. We also provide analytical bounds on the optimal training and data lengths, although the exact optimal values can only be found numerically. These bounds significantly reduce the ranges of the possible values of the optimal training and data lengths.

The optimal number of transmit and receive antennas has been studied in [29, 89–91] assuming perfect CSI at the receiver. In particular, the authors in [29] studied the situation where one extra antenna is available to be allocated at either end of a MIMO system. Their results show that one should always allocate the extra antenna to the side with less number of antennas at high SNR. In Section 2.5, we study the same problem taking channel estimation errors into account and show that the optimal antenna configuration in the perfect CSI case may not be applicable to systems with imperfect CSI. Taking into account the channel estimation errors, the authors in [54] found that the capacity decreases as the number of transmit antennas increases beyond the number of receive antennas at sufficiently high SNR. However, we show that this asymptotic result does not hold at a practical range of high SNR, *e.g.*, from 20 dB to 30 dB.

Most studies on the effect of channel spatial correlation on the ergodic capacity do not consider channel estimation errors and assume perfect CSI at the receiver [29, 30, 92]. For multiple-input single-output (MISO) systems, the authors in [30] showed that the ergodic capacity decreases as correlation increases when no CSI is available at the transmitter. Similar studies on MIMO systems found that the channel spatial correlations among both the transmit antennas and the receive antennas can reduce the capacity at high SNR [29] or for large number of transmit and receive antennas [92]. In Section 2.6, we extend previous studies of systems with perfect CSI and analyze the impact of channel spatial correlation on the ergodic capacity of MIMO systems with channel estimation errors. In particular, we show that the ergodic capacity increases with channel correlation at sufficiently low

SNR, which is in contrast to the existing results in [30] for the perfect CSI case.

2.2 System Model

We consider a MIMO flat-fading channel model with input-output relationship given by

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}, \tag{2.1}$$

where $\boldsymbol{y}$ is the $N_r \times 1$ received symbol vector, $\boldsymbol{x}$ is the $N_t \times 1$ transmitted symbol vector, $\boldsymbol{H}$ is the $N_r \times N_t$ channel gain matrix, and $\boldsymbol{n}$ is the $N_r \times 1$ noise vector having zero-mean circularly symmetric complex Gaussian (ZMCSCG) entries with unit variance. The entries of $\boldsymbol{H}$ are also ZMCSCG with unit variance.

2.2.1 Channel Spatial Correlation

The correlation between different entries of the channel gain matrix depends on the scattering environment around the transmitter and the receiver as well as the antenna spacing. To study the effects of channel correlation, we consider a typical downlink transmission scenario where there may be insufficient scattering around the base station transmitter, resulting in possible spatial correlations among the channel gains. The spatial correlation is characterized by the covariance matrix $\boldsymbol{R_H} = E\{\boldsymbol{H}^\dagger\boldsymbol{H}\}/N_r$. Therefore, $\boldsymbol{H} = \boldsymbol{H}_0\boldsymbol{R}_{\boldsymbol{H}}^{1/2}$, where $\boldsymbol{H}_0$ has i.i.d. ZMCSCG entries with unit variance. We assume that $\boldsymbol{R_H}$ is a positive definite matrix and denote the eigenvalues of $\boldsymbol{R_H}$ by $\boldsymbol{g} = [g_1\ g_2\ \ldots\ g_{N_t}]^T$.

Furthermore, we use the concept of majorization to characterize the degree of channel spatial correlation [30, 92]. A vector $\boldsymbol{a} = [a_1\ a_2\ ...\ a_n]^T$ is said to be majorized by another vector $\boldsymbol{b} = [b_1\ b_2\ ...\ b_n]^T$ if

$$\sum_{i=1}^{k} a_i \leq \sum_{i=1}^{k} b_i, \quad k = 1, ..., n-1, \quad \text{and} \quad \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i, \tag{2.2}$$

where the elements in both vectors are sorted in descending order [93]. We denote the relationship as $\boldsymbol{a} \prec \boldsymbol{b}$. Any real-valued function Φ, defined on a vector subspace, is said to be Schur-convex, if $\boldsymbol{a} \prec \boldsymbol{b}$ implies $\Phi(\boldsymbol{a}) \leq \Phi(\boldsymbol{b})$ [93]. Similarly Φ is Schur-concave, if $\boldsymbol{a} \prec \boldsymbol{b}$ implies $\Phi(\boldsymbol{a}) \geq \Phi(\boldsymbol{b})$. Following [30], we have the following definition:

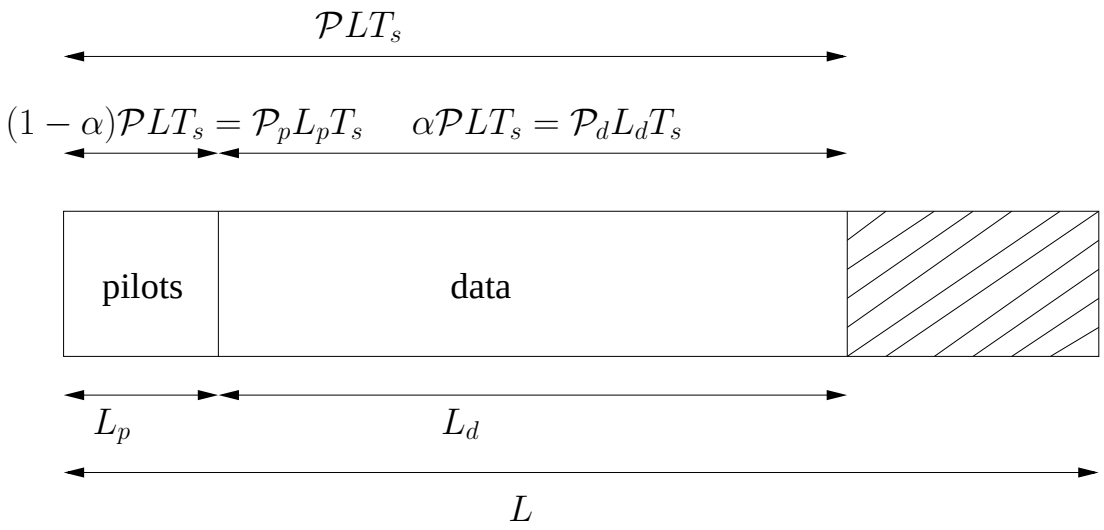

Figure 2.1: An example of a transmission block of L symbols in a non-feedback system. Temporal energy allocations are shown at the top and the length of each sub-block is shown at the bottom.

Definition 2.1 *Let* $\boldsymbol{a}$ *contains the eigenvalues of the channel covariance matrix* $\boldsymbol{R_a}$*, and* $\boldsymbol{b}$ *contains the eigenvalues of the channel covariance matrix* $\boldsymbol{R_b}$*. The elements in both vectors are sorted in descending order. Then* $\boldsymbol{R_a}$ *is less correlated than* $\boldsymbol{R_b}$ *if and only if* $\boldsymbol{a} \prec \boldsymbol{b}$*.*

2.2.2 Transmission Scheme

We assume that the channel gains remain constant over one coherence block of L symbol periods and change to independent realizations in the next block. This is an appropriate channel model for time-division multiple access or frequency-hopping systems [54]. Fig. 2.1 shows an example of a transmission block of L symbol periods in a PSAM scheme. At the beginning of each transmission block, each transmit antenna sends L_p training symbols followed by L_d data symbols, with $L_p + L_d \leq L$. No training or data symbol is transmitted during the remaining time slots if any.

The total transmission energy per block is given by $\mathcal{P}LT_s$ as shown in Fig. 2.1, where $\mathcal{P}$ is the average transmit power and T_s is the symbol duration. Note that the transmission energy $\mathcal{P}LT_s$ is only distributed over $L_p + L_d$ symbol periods, which can be shorter than the duration of a transmission block. We define the PSAM energy factor, α, as the ratio of the total energy allocated to the data transmission. We also denote the power per pilot and data symbol by $\mathcal{P}_p$ and $\mathcal{P}_d$, respectively. Therefore, $\mathcal{P}_p = \mathrm{tr}\{\boldsymbol{x}\boldsymbol{x}^\dagger\}$ during the pilot transmission and $\mathcal{P}_d = \mathrm{tr}\{\boldsymbol{Q}\}$ during the data transmission, where we have defined the data covariance matrix as $\boldsymbol{Q} =$

$E\{\boldsymbol{x}\boldsymbol{x}^\dagger\}$. Due to the normalization of the receiver noise power and the channel gains, we also refer to $\mathcal{P}$, $\mathcal{P}_p$ and $\mathcal{P}_d$ as average SNR, pilot SNR and data SNR, respectively. Therefore, we have the following relationships.

$$\begin{aligned} \mathcal{P}LT_s &= \mathcal{P}_p L_p T_s + \mathcal{P}_d L_d T_s, \\ \mathcal{P}_p &= (1-\alpha)\frac{\mathcal{P}L}{L_p}, \text{ and } \mathcal{P}_d = \alpha\frac{\mathcal{P}L}{L_d}. \end{aligned} \tag{2.3}$$

Depending on the degrees of freedom in the system design and the cost of power amplifiers, one can either have fixed power transmission or variable power transmission. In the case of fixed power transmission, the pilot SNR and data SNR are the same, *i.e.*, $\mathcal{P}_p = \mathcal{P}_d = \mathcal{P}_t$, where $\mathcal{P}_t = \mathcal{P}L/(L_p + L_d)$ is referred to as the actual transmit power in the fixed power transmission scheme. On the other hand, $\mathcal{P}_p$ and $\mathcal{P}_d$ are allowed to be different in the variable power transmission scheme.

2.2.3 Channel Estimation

In each transmission block, the receiver performs channel estimation during the pilot transmission. Assuming the channel spatial correlation is known at the receiver, the channel gain $\boldsymbol{H}$ can be estimated using the LMMSE estimator [94]. The assumption of knowing the channel correlation is reasonable since the correlation is a long-term statistic which can be accurately measured by the receiver. We denote the channel estimate and estimation error as $\widehat{\boldsymbol{H}} = \widehat{\boldsymbol{H}}_0 \boldsymbol{R}_{\widehat{\boldsymbol{H}}}^{1/2}$ and $\widetilde{\boldsymbol{H}} = \widetilde{\boldsymbol{H}}_0 \boldsymbol{R}_{\widetilde{\boldsymbol{H}}}^{1/2}$, respectively, where $\widehat{\boldsymbol{H}}_0$ and $\widetilde{\boldsymbol{H}}_0$ have i.i.d. ZMCSCG entries with unit variance. $\widehat{\boldsymbol{H}}$ is given as [62]

$$\widehat{\boldsymbol{H}} = \boldsymbol{Y}(\boldsymbol{X}_p{}^\dagger \boldsymbol{R}_{\boldsymbol{H}} \boldsymbol{X}_p + \boldsymbol{I}_{L_p})^{-1} \boldsymbol{X}_p{}^\dagger \boldsymbol{R}_{\boldsymbol{H}}, \tag{2.4}$$

where $\boldsymbol{Y}$ is the $N_t \times L_p$ matrix combining the L_p received symbol vectors during the pilot transmission and $\boldsymbol{X}_p$ is the $N_t \times L_p$ pilot matrix. The covariance matrix of the estimation error is given by [62]

$$\boldsymbol{R}_{\widetilde{\boldsymbol{H}}} = \frac{E\{\widetilde{\boldsymbol{H}}^\dagger \widetilde{\boldsymbol{H}}\}}{N_r} = (\boldsymbol{R}_{\boldsymbol{H}}^{-1} + \boldsymbol{X}_p \boldsymbol{X}_p{}^\dagger)^{-1}. \tag{2.5}$$

From the orthogonality property of LMMSE estimator, we have

$$\boldsymbol{R}_{\widehat{\boldsymbol{H}}} = \frac{E\{\widehat{\boldsymbol{H}}^\dagger \widehat{\boldsymbol{H}}\}}{N_r} = \boldsymbol{R}_{\boldsymbol{H}} - \boldsymbol{R}_{\widetilde{\boldsymbol{H}}}. \tag{2.6}$$

2.3 Ergodic Capacity Bounds

From an information-theoretic viewpoint, we investigate the optimal resource allocation using the ergodic capacity as the objective function. The exact capacity expression under imperfect receiver CSI is still unavailable. We consider one lower bound and one upper bound on the ergodic capacity for systems using LMMSE channel estimation. In particular, the authors in [53] derived a lower bound and an upper bound for spatially independent channels. Here we generalize their results to include spatially correlated channels as follows.

A lower bound on the ergodic capacity per channel use is given by [53]

$$C_{\mathrm{LB}} = E_{\widehat{\boldsymbol{H}}}\Big\{ \log_2 \Big| \boldsymbol{I}_{N_t} + \widehat{\boldsymbol{H}}^{\dagger}(\boldsymbol{I}_{N_r} + \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{H}}\boldsymbol{x}})^{-1}\widehat{\boldsymbol{H}}\boldsymbol{Q}\Big|\Big\}, \tag{2.7}$$

where

$$\begin{aligned}
\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{H}}\boldsymbol{x}} &= E\{\widetilde{\boldsymbol{H}}\boldsymbol{x}\boldsymbol{x}^{\dagger}\widetilde{\boldsymbol{H}}^{\dagger}\} \\
&= E\{\widetilde{\boldsymbol{H}}_0\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}^{1/2}\boldsymbol{x}\boldsymbol{x}^{\dagger}(\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}^{1/2})^{\dagger}\widetilde{\boldsymbol{H}}_0^{\dagger}\} \\
&= E\Big\{ \mathrm{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}^{1/2}\boldsymbol{x}\boldsymbol{x}^{\dagger}(\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}^{1/2})^{\dagger}\}\Big\}\boldsymbol{I}_{N_r} \\
&= \mathrm{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\boldsymbol{Q}\}\boldsymbol{I}_{N_r},
\end{aligned}$$

where we have used $E\{\widetilde{\boldsymbol{H}}_0\boldsymbol{Z}\widetilde{\boldsymbol{H}}_0^{\dagger}\} = E\{\mathrm{tr}\{\boldsymbol{Z}\}\}\boldsymbol{I}_{N_r}$, given that $\widetilde{\boldsymbol{H}}_0$ has i.i.d. entries with unit variance and is independent of $\boldsymbol{Z}$. Therefore, the ergodic capacity lower bound per channel use in (2.7) can be rewritten as

$$C_{\mathrm{LB}} = E_{\widehat{\boldsymbol{H}}}\Big\{ \log_2 \Big| \boldsymbol{I}_{N_t} + (1 + \mathrm{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\boldsymbol{Q}\})^{-1}\widehat{\boldsymbol{H}}^{\dagger}\widehat{\boldsymbol{H}}\boldsymbol{Q}\Big|\Big\}. \tag{2.8}$$

An upper bound on the ergodic capacity per channel use is given by [53]

$$C_{\mathrm{UB}} = E_{\widehat{\boldsymbol{H}}}\Big\{ \log_2 \Big|\pi e\boldsymbol{\Sigma}_{\boldsymbol{y}|\widehat{\boldsymbol{H}}}\Big|\Big\} - E_{\boldsymbol{x}}\Big\{ \log_2 \Big|\pi e(\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{H}}\boldsymbol{x}|\boldsymbol{x}} + \boldsymbol{I}_{N_r})\Big|\Big\},$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{y}|\widehat{\boldsymbol{H}}} = E\{\boldsymbol{y}\boldsymbol{y}^{\dagger}|\widehat{\boldsymbol{H}}\} = \widehat{\boldsymbol{H}}\boldsymbol{Q}\widehat{\boldsymbol{H}}^{\dagger} + \mathrm{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\boldsymbol{Q}\}\boldsymbol{I}_{N_r} + \boldsymbol{I}_{N_r},$$

and

$$\begin{aligned}
\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{H}}\boldsymbol{x}|\boldsymbol{x}} &= E\{\widetilde{\boldsymbol{H}}\boldsymbol{x}\boldsymbol{x}^{\dagger}\widetilde{\boldsymbol{H}}^{\dagger}|\boldsymbol{x}\} \\
&= E\{\widetilde{\boldsymbol{H}}_0\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}^{1/2}\boldsymbol{x}\boldsymbol{x}^{\dagger}(\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}^{1/2})^{\dagger}\widetilde{\boldsymbol{H}}_0^{\dagger}|\boldsymbol{x}\}
\end{aligned}$$

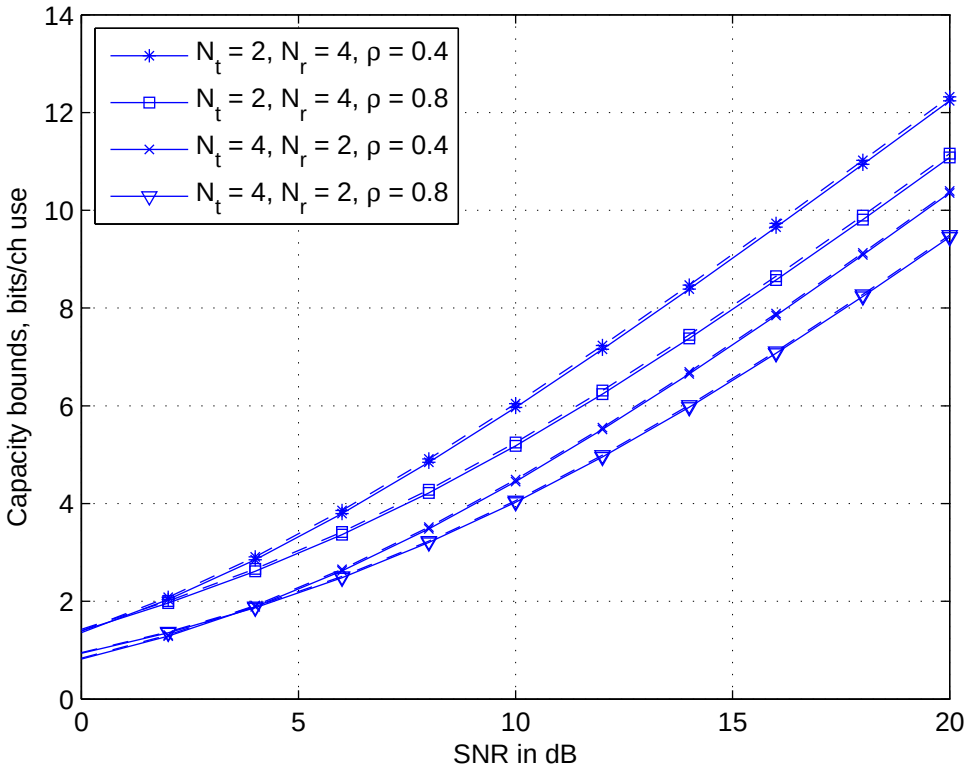

Figure 2.2: Capacity bounds in (2.8) and (2.9) versus SNR $\mathcal{P}$ for different correlation factors and antenna sizes. $L_p = N_t$, $L_d = L - L_p$ and $\mathcal{P}_p = \mathcal{P}_d = \mathcal{P}$ (*i.e.*, fixed power transmission) is used. Dashed lines indicate the upper bound and solid lines indicate the lower bound.

$$\begin{aligned} &= \operatorname{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}^{1/2}\boldsymbol{x}\boldsymbol{x}^{\dagger}(\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}^{1/2})^{\dagger}\}\boldsymbol{I}_{N_r} \\ &= \boldsymbol{x}^{\dagger}\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\boldsymbol{x}\boldsymbol{I}_{N_r}. \end{aligned}$$

Therefore, the ergodic capacity upper bound per channel use can be written as

$$\begin{aligned} C_{\mathrm{UB}} &= E_{\widehat{\boldsymbol{H}}}\left\{\log_2\left|\boldsymbol{I}_{N_t}+(1+\operatorname{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\boldsymbol{Q}\})^{-1}\widehat{\boldsymbol{H}}^{\dagger}\widehat{\boldsymbol{H}}\boldsymbol{Q}\right|\right\}+N_r E_{\boldsymbol{x}}\left\{\log_2\frac{1+\operatorname{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\boldsymbol{Q}\}}{1+\boldsymbol{x}^{\dagger}\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\boldsymbol{x}}\right\} \\ &= C_{\mathrm{LB}} + C_{\mathrm{gap}}, \end{aligned} \tag{2.9}$$

where C_{gap} is the difference between the upper bound and the lower bound, which indicates the maximum error of the bounds.

The authors in [53] studied the tightness of the bounds for spatially independent channels for fixed channel estimation errors. They observed that the gap between bounds is small for Gaussian inputs. Similarly, we numerically study the tightness of the above two bounds for correlated channels with Gaussian inputs

and LMMSE channel estimation. For numerical analysis, we choose the channel covariance matrix to be in the form of $[\boldsymbol{R}_{\boldsymbol{H}}]_{ij} = \rho^{|i-j|}$, where ρ is referred to as the spatial correlation factor [62]. We investigate the gap between the capacity bounds in (2.8) and (2.9) under different channel correlation conditions and antenna sizes. Examples are shown in Fig. 2.2 in which the upper bounds are plotted using dashed lines and the lower bounds are plotted using the solid lines. In general, we see that the gaps between the bounds are insignificant for any channel correlation factors. Therefore, the capacity lower bound per channel use in (2.8) is accurate enough to be used in our analysis assuming Gaussian inputs.

Taking the training overhead into account, the average capacity lower bound is therefore given by

$$\begin{aligned}\overline{C}_{\mathrm{LB}} &= \frac{L_d}{L} C_{\mathrm{LB}} \\ &= \frac{L_d}{L} E_{\widehat{\boldsymbol{H}}}\Big\{\log_2\Big|\boldsymbol{I}_{N_t} + (1+\mathrm{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\boldsymbol{Q}\})^{-1}\widehat{\boldsymbol{H}}^{\dagger}\widehat{\boldsymbol{H}}\boldsymbol{Q}\Big|\Big\}. \end{aligned} \tag{2.10}$$

We will use "capacity lower bound" and "capacity" interchangeably throughout the rest of this chapter. The average capacity lower bound in (2.10) will be used as the figure of merit to study the optimal transmission and resource allocation strategies in the following sections.

2.4 Optimal Transmission Scheme with Independent Channels

For PSAM transmission, the four important design parameters are the structure of pilot transmission, structure of data transmission, transmit energy allocation between pilots and data as well as transmit time allocation between pilots and data. The pilot structure is characterized by the pilot matrix $\boldsymbol{X}_p$. The data structure is characterized by the input data covariance matrix $\boldsymbol{Q}$. The transmit energy allocation is characterized by the PSAM energy factor α. And the transmit time allocation is characterized by the training length L_p and the data length L_d. Table 2.1 summarizes these design parameters.

We first consider the channels are spatially i.i.d. and will study the impact of channel correlation in Section 2.6. For i.i.d. channels, the data transmission utilizes all the channels with equal probability. Hence, it is reasonable to have at least as many measurements as the number of channels for channel estimation, which implies that $L_p \geq N_t$. For non-feedback systems with spatially i.i.d. channels, the

Table 2.1: Design parameters in PSAM transmission

Design parameter	Notation	Description
pilot structure	$\boldsymbol{X}_p$	transmitted pilot matrix ($N_t \times L_p$)
data structure	$\boldsymbol{Q}$	input data covariance matrix
PSAM energy factor	α	ratio of energy allocated to data
training length	L_p	number of training symbols in a block
data length	L_d	number of data symbols in a block

optimal pilot structure is orthogonal transmission, *i.e.*, $\boldsymbol{X}_p\boldsymbol{X}_p{}^\dagger = \mathcal{P}_p L_p/N_t \boldsymbol{I}_{N_t}$, and the optimal data structure has spatially i.i.d. inputs, *i.e.*, $\boldsymbol{Q} = \mathcal{P}_d/N_t \boldsymbol{I}_{N_t}$ [6, 54]. With the optimal pilot and data structure and i.i.d. channels, the covariance matrix of channel estimation error in (2.5) reduces to

$$\begin{aligned} \boldsymbol{R}_{\widetilde{\boldsymbol{H}}} &= \frac{1}{1+\mathcal{P}_p L_p/N_t}\boldsymbol{I}_{N_t} \\ &= \sigma^2_{\widetilde{\boldsymbol{H}}}\boldsymbol{I}_{N_t}, \end{aligned} \tag{2.11}$$

where $\sigma^2_{\widetilde{\boldsymbol{H}}} = (1+\mathcal{P}_p L_p/N_t)^{-1}$. The average capacity lower bound in (2.10) reduces to

$$\begin{aligned} \overline{C}_{\mathrm{LB}} &= \frac{L_d}{L} E_{\widehat{\boldsymbol{H}}_0}\Big\{ \log_2 \Big| \boldsymbol{I}_{N_t} + \frac{\sigma^2_{\widehat{\boldsymbol{H}}}\mathcal{P}_d}{1+\sigma^2_{\widetilde{\boldsymbol{H}}}\mathcal{P}_d} \frac{\widehat{\boldsymbol{H}}_0^\dagger \widehat{\boldsymbol{H}}_0}{N_t} \Big| \Big\} \\ &= \frac{L_d}{L} E_{\widehat{\boldsymbol{H}}_0}\Big\{ \log_2 \Big| \boldsymbol{I}_{N_t} + \rho_{\mathrm{eff}} \frac{\widehat{\boldsymbol{H}}_0^\dagger \widehat{\boldsymbol{H}}_0}{N_t} \Big| \Big\} \qquad (2.12) \\ &= n\frac{L_d}{L} E_\lambda\{\log_2(1+\rho_{\mathrm{eff}}\lambda)\}, \qquad (2.13) \end{aligned}$$

where $\sigma^2_{\widehat{\boldsymbol{H}}} = 1-\sigma^2_{\widetilde{\boldsymbol{H}}}$, $n = \min\{N_t, N_r\}$, λ is an arbitrary eigenvalue of $\widehat{\boldsymbol{H}}_0^\dagger \widehat{\boldsymbol{H}}_0/N_t$, and ρ_{eff} is named the effective SNR in [54] given by

$$\rho_{\mathrm{eff}} = \frac{(1-\sigma^2_{\widetilde{\boldsymbol{H}}})\mathcal{P}_d}{1+\sigma^2_{\widetilde{\boldsymbol{H}}}\mathcal{P}_d}. \tag{2.14}$$

The optimal transmit energy and time allocation depends on the degrees of freedom in the communication systems. In the following, we consider both variable power transmission and fixed power transmission. The former scheme achieves 5% - 10% capacity improvement over the latter scheme [54], while the latter scheme is more practical to implement.

2.4.1 Variable Power Transmission

In the variable power transmission scheme, the transmit power for pilots and data is allowed to be different. Denote the optimal training and data length as L_p^* and L_d^*, respectively. An immediate result on the optimal transmit time allocation is given below.

Lemma 2.1 The optimal transmission strategy uses all available time for transmission, *i.e.*, $L_p^* + L_d^* = L$.

Proof: For any given $\mathcal{P}_p$ and L_p, $\sigma^2_{\widetilde{\boldsymbol{H}}}$ is a constant. It can be easily shown that the derivative of $\overline{C}_{\rm LB}$ in (2.13) w.r.t. L_d is positive. Therefore, the capacity maximizing L_d takes its largest possible value, that is $L_d = L - L_p$. □

With $L_p + L_d = L$, the optimal transmit energy and time allocation in the variable power transmission scheme was derived in [54]. The optimal PSAM energy factor α^* is given by

$$\alpha^* = \begin{cases} \gamma - \sqrt{\gamma(\gamma-1)}, & \text{for } L_d > N_t \\ \frac{1}{2}, & \text{for } L_d = N_t \\ \gamma + \sqrt{\gamma(\gamma-1)}, & \text{for } L_d < N_t \end{cases} \tag{2.15}$$

where $\gamma = \frac{N_t + \mathcal{P}L}{\mathcal{P}L(1-N_t/L_d)}$. With α^*, the optimal training and data length is given by $L_p^* = N_t$ and $L_d^* = L - N_t$.

2.4.2 Fixed Power Transmission

In the fixed power transmission scheme, the transmit power for pilots and data is the same, *i.e.*, $\mathcal{P}_p = \mathcal{P}_d = \mathcal{P}_t$, and hence the PSAM energy factor is not a design parameter. On the other hand, the transmit time allocation becomes a particularly important design parameter as it directly affects the actual transmit power $\mathcal{P}_t$. To study the optimal transmit time allocation, two scenarios are considered: 1) the training length L_p and the data length L_d are to be jointly optimized; 2) the data length L_d is to be optimized for a given training length L_p.

Scenario 1: Joint Training and Data Length Optimization

In general, the optimal training and data lengths, denoted by L_p^* and L_d^* respectively, need to be found numerically by evaluating the average capacity lower bound for all possible values of L_p and L_d. Also, it is not clear whether $L_p + L_d = L$ is optimal for all SNR conditions. In the following, we provide analytical bounds on L_p^* and L_d^*, and prove the optimality of $L_p + L_d = L$.

Theorem 2.1 When both L_p and L_d are allowed to be optimized, the optimal strategy is to use all available time for transmission at any SNR values, *i.e.*, $L_p^* + L_d^* = L$. Furthermore, the optimal values of L_d and L_p satisfy the following conditions.

$$\begin{cases} L_d^* = L - N_t, \;\; L_p^* = N_t, & \text{if } L \leq 2N_t \\ \zeta \leq L_d^* \leq L - N_t, \;\; N_t \leq L_p^* \leq L - \zeta, & \text{otherwise} \end{cases} \tag{2.16}$$

where

$$\zeta = \frac{2L\mathcal{P}+N_t\mathcal{P}+2N_t-\sqrt{N_t(N_t\mathcal{P}^2+4N_t\mathcal{P}+4N_t+4L\mathcal{P}^2+4L\mathcal{P})}}{2\mathcal{P}}$$

is the value of L_d at which we have

$$\sqrt{\frac{N_t L_d(L_d + L\mathcal{P})}{N_t + L\mathcal{P}}} + L_d = L, \tag{2.17}$$

Proof: see Appendix A.1.

When $\mathcal{P} \to 0$, we see from (2.17) that $\zeta \to L/2$. Therefore, the lower (upper) bound on L_d^* (L_p^*) in (2.16) approaches $L/2$. In fact, by using the first order approximation of the average capacity lower bound at sufficiently low SNR, it can be shown that the optimal training and data lengths satisfy $L_p^* = L_d^* = L/2$ [54]. Therefore, we expect that the lower (upper) bound on L_d^* (L_p^*) given in *Theorem 2.1* is tight at sufficiently low SNR. When $\mathcal{P} \to \infty$, it can be shown that L_d should be chosen as large as possible, *i.e.*, $L_d^* = L - N_t$. (The proof follows from Appendix A.1 by letting $\mathcal{P} \to \infty$.) Therefore, we expect that the upper (lower) bound on L_d^* (L_p^*) given in *Theorem 2.1* is tight at sufficiently high SNR. We note that the optimality of $L_p = N_t$ at sufficiently high SNR was also commented in [54].

Fig. 2.3 shows the upper bound on L_p^* given in (2.16) as well as the exact values of L_p^* found numerically. We see that the upper bound significantly reduces the range of possible values of L_p^* at moderate to high SNR. For example in finding L_p^* for a 2×2 system with $L = 100$ operating at 5 dB, the analytical bounds tell that one only needs to search from 2 to 15, instead of searching every possible training length from 2 to 99. In addition, we see that the upper bound becomes tighter as SNR decreases.

Scenario 2: Optimal Data Length for a Given Training Length

As we have seen, the training and data lengths, L_p and L_d, need be found numerically according to the operating SNR which usually varies with time. When

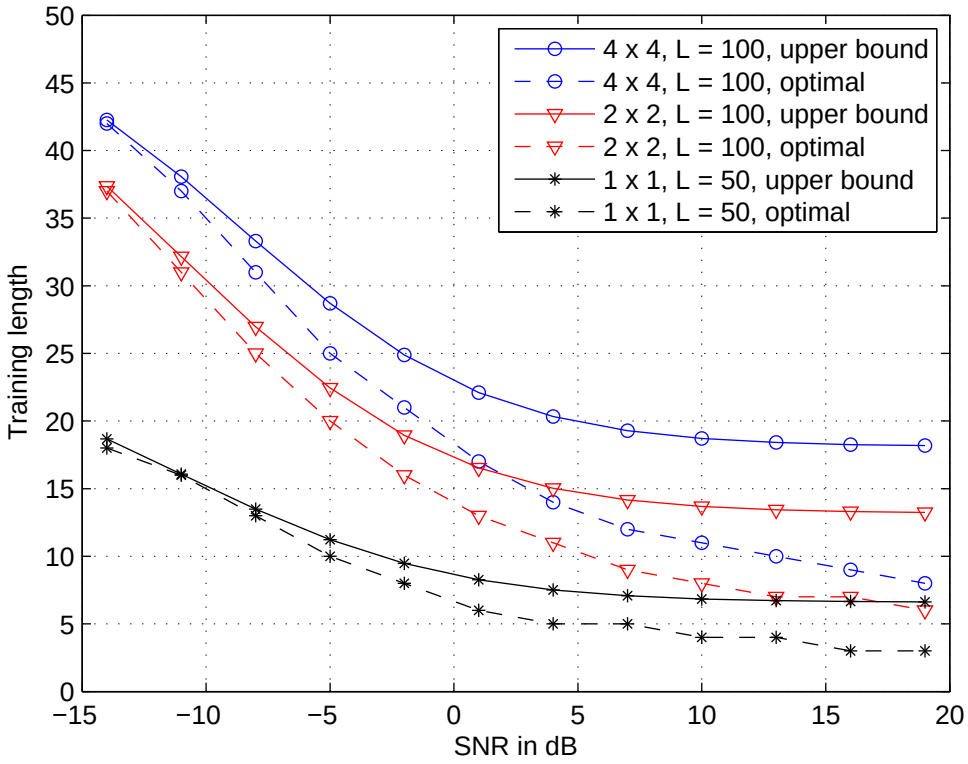

Figure 2.3: The upper bound on L_p^* given in (2.16) versus average SNR $\mathcal{P}$ for different block lengths and different numbers of transmit antennas.

the operating SNR changes rapidly with time, it is undesirable to frequently change L_p and L_d. In this scenario, we consider a transmission strategy which optimizes L_p (and L_d) for a target SNR and fix it for a certain time period before redesigning becomes crucial.

From Fig. 2.3 we see that L_p^* decreases as SNR increases. If the operating SNR is higher than the target SNR, the system is using more training resource than the optimal amount, hence it is still desirable to use all available time for data transmission, *i.e.*, $L_d^* = L - L_p$. On the other hand, if the operating SNR is lower than the target SNR, the system is using insufficient amount of training which results in a degradation in channel estimation. In this scenario, the system may need to at least reduce the data length from $L - L_p$ to boost up the SNR.[1]

In the following, we find a threshold SNR $\mathcal{P}_{\text{th}}$, such that the optimal data length is given by $L_d^* = L - L_p$ for any given L_p as long as the operating SNR is above $\mathcal{P}_{\text{th}}$. That is, for a system with a fixed L_p which has been optimized according to

[1]For a given L_p, $L_d^* < L - L_p$ can happen at sufficiently low SNR. To see this, one can apply the first order approximation of the average capacity lower bound and obtain the optimal data length by letting the derivative of the capacity approximation be zero.

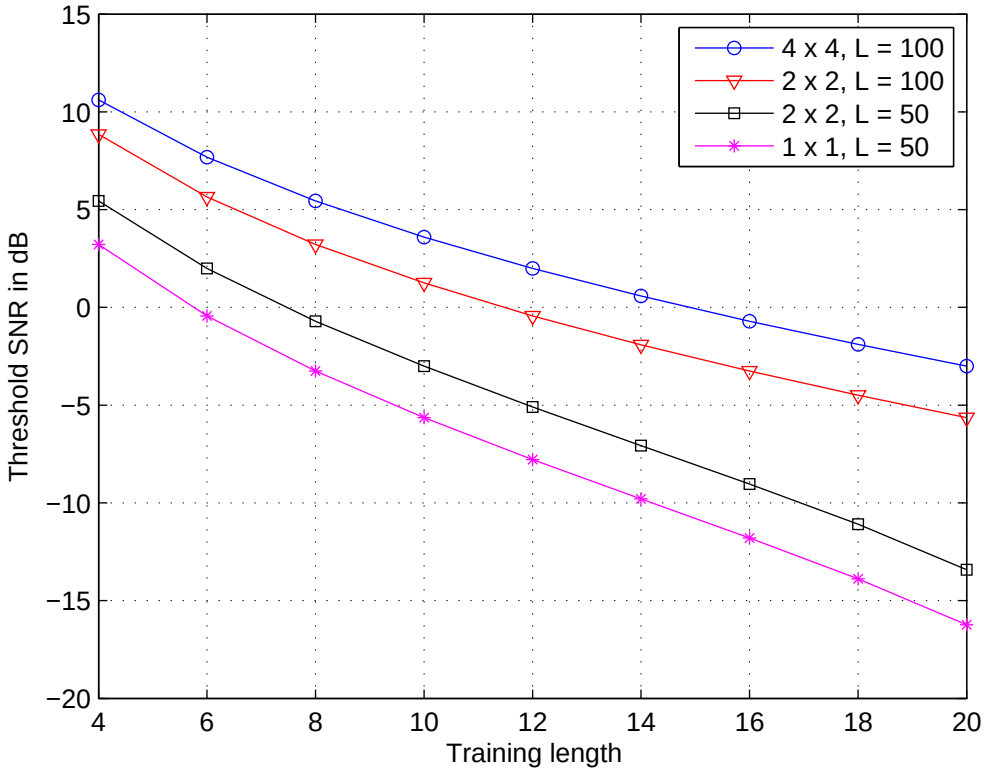

Figure 2.4: The threshold SNR $\mathcal{P}_{\text{th}}$ given in (2.18) versus training length L_p for different block lengths and different numbers of transmit antennas.

some target SNR, it can keep using the original design of $L_d = L - L_p$ as long as the operating SNR is above $\mathcal{P}_{\text{th}}$.

Lemma 2.2 For any given L_p, a threshold SNR above which the optimal L_d equals $L - L_p$ is given by

$$\mathcal{P}_{\text{th}} = \frac{N_t(L - 2L_p)}{L_p(L_p + N_t)} \tag{2.18}$$

Proof: see Appendix A.2.

Fig. 2.4 shows the threshold SNR $\mathcal{P}_{\text{th}}$ given in (2.18) versus training length L_p. If a 2×2 system with $L = 100$ has a fixed training length of $L_p = 10$ which is optimal for a target SNR of 5 dB as shown in Fig. 2.3, the data length does not need to be reduced from $L - L_p = 90$ as long as the operating SNR is above $\mathcal{P}_{\text{th}} = 1.3$ dB as shown in Fig. 2.4. Note that the optimal data length may or may not equal $L - L_p$ when the operating SNR is below $\mathcal{P}_{\text{th}}$. Therefore, it is wise to redesign the length of data transmission when the operating SNR is considerably lower than $\mathcal{P}_{\text{th}}$.

2.5 Optimal Antenna Configuration with Independent Channels

Link adaptation is an important practical problem for MIMO systems, where it is desirable to optimize the number of transmit and receive antennas for high data rate. In this section, we study the optimal antenna configuration at practical range of high SNR, *e.g.*, from 20 dB to 30 dB, in PSAM schemes from capacity maximization viewpoint. In particular, we extend the analysis in [29] and [54] to investigate the following two important problems for systems with practical antenna sizes.

- Problem 1: If an extra antenna is available to be added at either end of a MIMO system, should one add it to the transmitter or to the receiver?
- Problem 2: If it is only practical to change the number of transmit antennas, what is the optimal number of transmit antennas?

Similar to Section 2.4, we consider both variable and fixed power transmission with spatially i.i.d. channels. For variable power transmission, we assume that the optimal PSAM energy factor α^* given in (2.15) is used. In the high SNR regime, the effective SNR given in (2.14) can be approximated as $\rho_{\text{eff}} = \frac{\mathcal{P}L}{(\sqrt{L-L_p}+\sqrt{N_t})^2}$ for variable power transmission and $\rho_{\text{eff}} = \frac{\mathcal{P}}{1+N_t/L_p}$ for fixed power transmission. To ease our analysis, we first obtain a closed-form approximation of the average capacity lower bound in the high SNR regime.

For $N_t \geq N_r$, the average capacity lower bound in (2.12) is approximated as

$$\begin{aligned}\overline{C}_{\text{LB}}(N_t, N_r) &\approx E_{\widehat{\boldsymbol{H}}_0}\Big\{ \log_2 \Big|\frac{\rho_{\text{eff}}}{N_t}\widehat{\boldsymbol{H}}_0\widehat{\boldsymbol{H}}_0^\dagger\Big|\Big\} \\ &= \frac{L_d}{L}\Big[N_r \log_2 \rho_{\text{eff}} - N_r \log_2 N_t + \frac{1}{\ln 2}\Big(\sum_{j=1}^{N_r}\sum_{k=1}^{N_t-j}\frac{1}{k} - N_r\epsilon\Big)\Big], \quad (2.19)\end{aligned}$$

where $E_{\widehat{\boldsymbol{H}}_0}\{\log_2|\widehat{\boldsymbol{H}}_0\widehat{\boldsymbol{H}}_0^\dagger|\} = \frac{1}{\ln 2}\Big(\sum_{j=1}^{N_r}\sum_{k=1}^{N_t-j}\frac{1}{k} - N_r\epsilon\Big)$ [29] and $\epsilon \approx 0.577$ is the Euler's constant.

Similarly, for $N_t < N_r$, the average capacity lower bound in (2.12) can be approximated by

$$\begin{aligned}\overline{C}_{\text{LB}}(N_t, N_r) &\approx E_{\widehat{\boldsymbol{H}}_0}\Big\{ \log_2 \Big|\frac{\rho_{\text{eff}}}{N_t}\widehat{\boldsymbol{H}}_0^\dagger\widehat{\boldsymbol{H}}_0\Big|\Big\} \\ &= \frac{L_d}{L}\Big[N_t \log_2 \rho_{\text{eff}} - N_t \log_2 N_t + \frac{1}{\ln 2}\Big(\sum_{j=1}^{N_t}\sum_{k=1}^{N_r-j}\frac{1}{k} - N_t\epsilon\Big)\Big]. \quad (2.20)\end{aligned}$$

2.5.1 Solution to Problem 1

Now, we investigate the first problem: *If an extra antenna is available to be added on either end of a MIMO system, should one add it to the transmitter or to the receiver?* This question is relevant in the design of point-to-point MIMO wireless links with fixed total number of antennas [29]. It may also occur in on-the-fly link adaptation, *e.g.*, IEEE802.11n. We present the analysis for the variable power transmission scheme, while the result for the fixed power transmission scheme is similar and hence, is omitted. Note that for variable power transmission, the optimal training length is given by $L_p^* = N_t$, which is assumed in the following analysis.

We start with the case where $N_t < N_r$. Using (2.20), we compute the capacity difference between the systems having $(N_t, N_r + 1)$ and $(N_t + 1, N_r)$, i.e., $\delta\overline{C}_{\text{LB}} \triangleq \overline{C}_{\text{LB}}(N_t, N_r + 1) - \overline{C}_{\text{LB}}(N_t + 1, N_r)$ as

$$\begin{aligned}\delta\overline{C}_{\text{LB}} \approx \frac{1}{L\ln 2}\Big[&N_t(L-N_t)\ln\rho_{\text{eff}} - (N_t+1)(L-N_t-1)\ln\rho'_{\text{eff}} \\ &+N_t(L-N_t)\ln(1+\frac{1}{N_t}) - (2N_t-L+1)\ln(N_t+1) \\ &+(L-N_t)\sum_{j=1}^{N_t}\frac{1}{N_r+1-j} - (L-N_t)\sum_{k=1}^{N_r-N_t-1}\frac{1}{k} \\ &+\sum_{j=1}^{N_t+1}\sum_{k=1}^{N_r-j}\frac{1}{k} - (2N_t-L+1)\epsilon\Big],\end{aligned} \tag{2.21}$$

where $\rho_{\text{eff}} = \frac{L\mathcal{P}}{L+2\sqrt{(L-N_t)N_t}}$ is the effective SNR for a $(N_t, N_r + 1)$ system, and $\rho'_{\text{eff}} = \frac{L\mathcal{P}}{L+2\sqrt{(L-N_t-1)(N_t+1)}}$ is the effective SNR for a $(N_t + 1, N_r)$ system. We see that the sign of $(2N_t - L + 1)$ plays an important role in (2.21). In the case where $2N_t - L + 1 \geq 0$ or equivalently $L \leq 2N_t + 1$, we find that one should add the extra antenna to the receiver at moderate to high SNR. In the following, we focus on the more practical case where $L > 2N_t + 1$.

We call the SNR value at which $\delta\overline{C}_{\text{LB}} = 0$ the critical SNR, denoted by $\mathcal{P}_c$. It is the threshold SNR in determining at which end the extra antenna should be added. As the block length L approaches infinity in (2.21), we see that $\mathcal{P}_c$ approaches a limiting value, given by

$$\begin{aligned}\mathcal{P}_{c,\infty}(N_t, N_r) = \exp\Big[&N_t\ln(1+\frac{1}{N_t}) + \ln(1+N_t) \\ &+\sum_{j=1}^{N_t}\frac{1}{N_r+1-j} - \sum_{k=1}^{N_r-N_t-1}\frac{1}{k} + \epsilon\Big].\end{aligned} \tag{2.22}$$

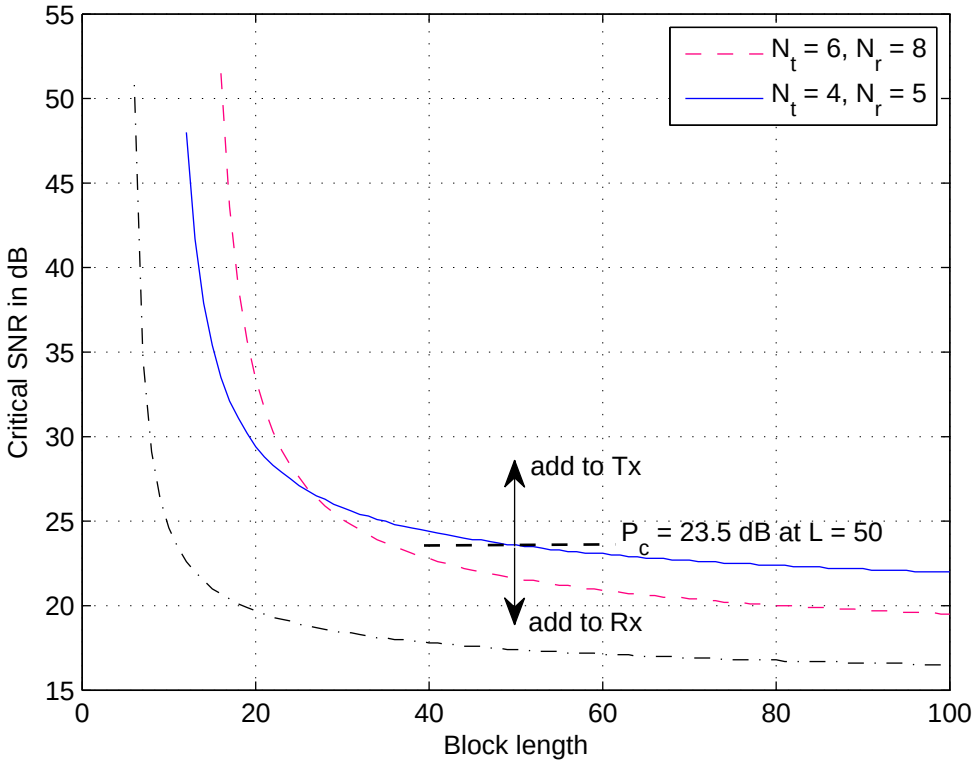

Figure 2.5: The critical SNR $\mathcal{P}_c$ found using (2.21) versus block length L in determining whether to add the extra antenna to the transmitter or to the receiver for MIMO systems with different antenna configurations.

Fig. 2.5 shows the critical SNR $\mathcal{P}_c$ for a wide range of block lengths L, where $L > 2N_t + 1$. In this case, one should add the extra antenna to the transmitter if the operating SNR is above $\mathcal{P}_c$ for any given block length, and vice versa. For example, consider the case where $(N_t = 4, N_r = 5)$ and $L = 50$. Fig. 2.5 suggests that one should add the extra antenna to the receiver when the operating SNR is below $\mathcal{P}_c = 23.5$ dB, while one should add the antenna to the transmitter when the operating SNR is above $\mathcal{P}_c = 23.5$ dB.

Furthermore, we see from Fig. 2.5 that $\mathcal{P}_c \to \infty$ as $L \to 2N_t + 1$. This suggests that one should always add the extra antenna to the receiver when L is close to $2N_t + 1$. We also see that $\mathcal{P}_c$ decreases as L increases and approaches $\mathcal{P}_{c,\infty}$ as $L \to \infty$. Therefore, $\mathcal{P}_{c,\infty}$ serves as the infimum of $\mathcal{P}_c$. This implies that one should always add the extra antenna to the receiver if the operating SNR is below $\mathcal{P}_{c,\infty}$, regardless of the block length. For example, $\mathcal{P}_{c,\infty} = 19$ dB for a $(N_t = 4, N_r = 5)$ system. These important trends of the critical SNR are not attained from the analysis in [29] under the perfect CSI assumption.[2] Therefore, the result for perfect

[2]The results in [29] showed that one should always add the extra antenna to the side with less

CSI case cannot be directly applied to the systems with imperfect CSI.

In the case where $N_t \geq N_r$, the same analysis can be carried out. We find that it is generally better to place the extra antenna at the receiver side at moderate to high SNR when $N_t \geq N_r$, regardless of the block length. This observation agrees with the result for the perfect CSI case in [29].

2.5.2 Solution to Problem 2

Now, we investigate the second problem for both variable and fixed power transmission: *If it is only practical to increase or reduce the number of transmit antennas, what is the optimal number of transmit antennas*[3] *?*

Firstly we consider the case where $N_t \geq N_r$. An example of this case would be the downlink in mobile cellular networks, where it is practical to alter the number of antennas at the base station to maximize the data rate for every mobile user, possibly without significant cost. The gain in the average ergodic capacity lower bound in (2.19) from adding an extra transmit antenna, i.e., $\triangle\overline{C}_{\mathrm{LB}} \triangleq \overline{C}_{\mathrm{LB}}(N_t+1, N_r) - \overline{C}_{\mathrm{LB}}(N_t, N_r)$, is given by

$$\begin{aligned}\triangle\overline{C}_{\mathrm{LB}} \approx \frac{1}{L\ln 2}\Big(& N_r \ln(N_t+1) + (L-N_t)\sum_{j=1}^{N_r}\frac{1}{N_t+1-j} \\ & + N_r\epsilon - N_r(L-N_t)\ln\frac{N_t+1}{N_t} - \sum_{j=1}^{N_r}\sum_{k=1}^{N_t+1-j}\frac{1}{k} \\ & + N_r(L-N_t-1)\ln\rho'_{\mathrm{eff}} - N_r(L-N_t)\ln\rho_{\mathrm{eff}}\Big), \end{aligned} \tag{2.23}$$

where ρ_{eff} and ρ'_{eff} are the effective SNR for a (N_t, N_r) system and a (N_t+1, N_r) system, respectively.

We start with the analysis on the limiting case of large block length. With $L \to \infty$, $\triangle\overline{C}_{\mathrm{LB}}$ in (2.23) reduces to

$$\begin{aligned}\triangle\overline{C}_{\mathrm{LB}} &\approx \frac{1}{\ln 2}\Big(\sum_{j=1}^{N_r}\frac{1}{N_t+1-j} - N_r\ln\frac{N_t+1}{N_t}\Big) \\ &\geq \frac{N_r}{\ln 2}\Big(\frac{1}{N_t} - \ln\frac{N_t+1}{N_t}\Big).\end{aligned}$$

Also, it is easy to show that $\frac{1}{N_t} - \ln\frac{N_t+1}{N_t} > 0$. Hence, we have $\triangle\overline{C}_{\mathrm{LB}} > 0$. This

number of antennas.

[3]From information-theoretic viewpoint, it is always beneficial to add extra antennas at the receiver, if possible, as it increases the diversity of the system without the need to sacrifice information symbols for training symbols.

Table 2.2: Optimal number of transmit antennas and its corresponding range of block length for systems with 4 receive antennas

Transmission scheme	Optimal N_t	4	5	≥ 6
Fixed power ($\mathcal{P} = 20$ dB)	Range of L	8 to 15	16 to 30	≥ 31
Variable power ($\mathcal{P} = 20$ dB)	Range of L	8 to 16	17 to 38	≥ 39
Fixed power ($\mathcal{P} = 30$ dB)	Range of L	8 to 22	23 to 46	≥ 47
Variable power ($\mathcal{P} = 30$ dB)	Range of L	8 to 26	27 to 60	≥ 61

implies adding more transmit antennas always results in higher average ergodic capacity for both power allocation schemes at sufficiently large L. However, this is not true for not so large block lengths.

Table 2.2 shows the optimal number of transmit antennas obtained from (2.23) by a linear search and its corresponding range of block length L for fixed SNR values, using L_p^* and $N_r = 4$. For variable power transmission, $L_p^* = N_t$. For fixed power transmission, L_p^* is found numerically from (2.19). We have also restricted the minimum value of L to be $2N_t$. We see that for both transmission schemes the optimum value of N_t can exceed the value of N_r and increases with L at practical high SNR values. This result is not predicted in the asymptotic high SNR analysis in [54]. Therefore, the results presented here are more accurate and provide useful insights at practical high SNRs.

Fig. 2.6 shows the average capacity lower bound computed using (2.19) versus block length for MIMO systems with $N_r = 2$ and at $\mathcal{P} = 30$ dB. We include the capacity lower bounds achieved using the (optimal) variable power transmission and/or optimal number of transmit antennas, as well as the non-optimized case (*i.e.*, the fixed power transmission and equal number of transmit and receive antennas). The optimal training length L_p^* is used in all computations. Comparing the two solid or dashed curves in Fig. 2.6, we see that the capacity percentage improvement by variable power transmission generally decreases as the block length increases. On the other hand, the capacity improvement from optimal antenna configuration increases as the block length increases. Compared with equal antenna configuration, we see in Fig. 2.6 that the capacity improvement from optimal antenna configuration reaches approximately 4.5% (or 4.3%) at $L = 100$ and around 6% (or 5.3%) at $L = 200$.

Fig. 2.7 shows the SNR saving calculated from (2.19). It indicates the amount of transmit power saved by using the (optimal) variable power transmission and/or optimal antenna configuration to achieve the same capacity as in the non-optimized case at $\mathcal{P} = 30$ dB. The optimal training length is used in all computations. From Fig. 2.7, we see that the SNR saving by variable power transmission decreases as the

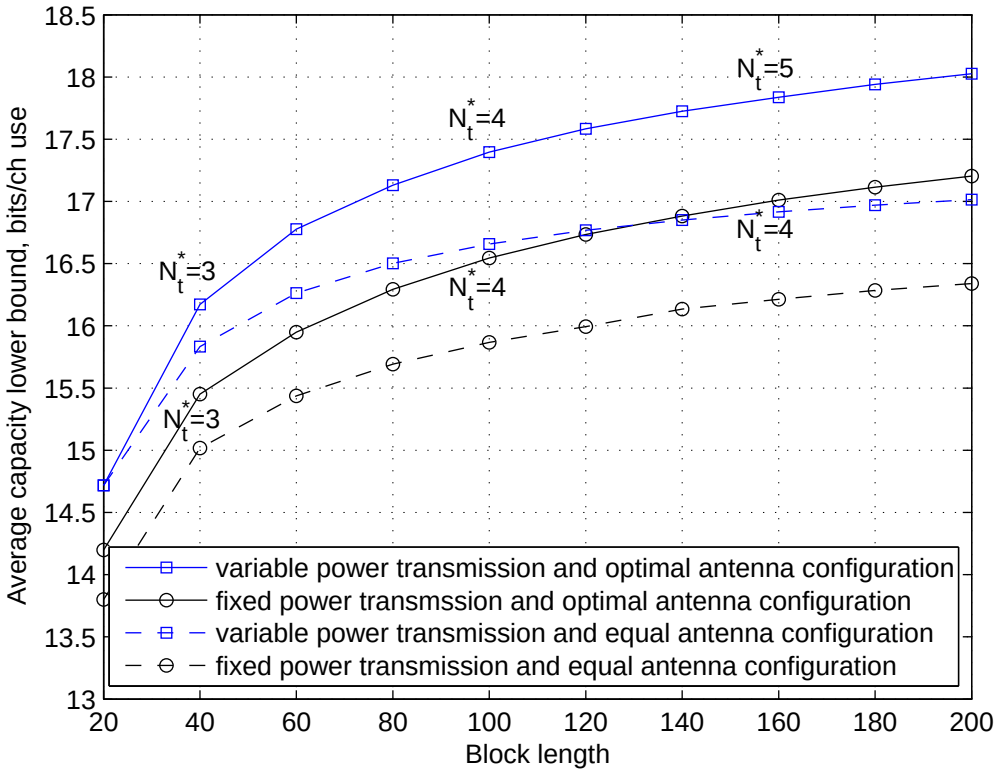

Figure 2.6: The average capacity lower bound $\overline{C}_{\mathrm{LB}}$ in (2.19) versus block length L for MIMO systems with 2 receive antennas at SNR $\mathcal{P} = 30$ dB. Equal antenna configuration represents systems with 2 transmit and receive antennas. N_t^* denotes the optimal number of transmit antennas.

block length L increases, while the SNR saving by optimal antenna configuration increases with L. For large block lengths, optimal antenna configuration generally saves more power than variable power transmission. At $L = 100$, the additional SNR saving by antenna optimization is 1.2 dB, which equals the SNR saving by power optimization alone. At $L = 200$, the additional SNR savings by antenna optimization increases to 1.6 dB, while the SNR saving by power optimization alone decreases to 1 dB. These results show that optimizing antenna configuration is more important than optimizing power allocation from an information-theoretic viewpoint.

When $N_t < N_r$, the analysis of the optimal number of transmit antennas can be carried out in the same manner as in the case where $N_t \geq N_r$. Assuming $L \gg 1$, the result in [54] suggests that adding extra transmit antennas always improves the capacity, provided N_t does not exceed N_r and $\mathcal{P} \to \infty$. Our analysis confirms that this claim is also accurate at practical range of high SNR and practical antenna sizes.

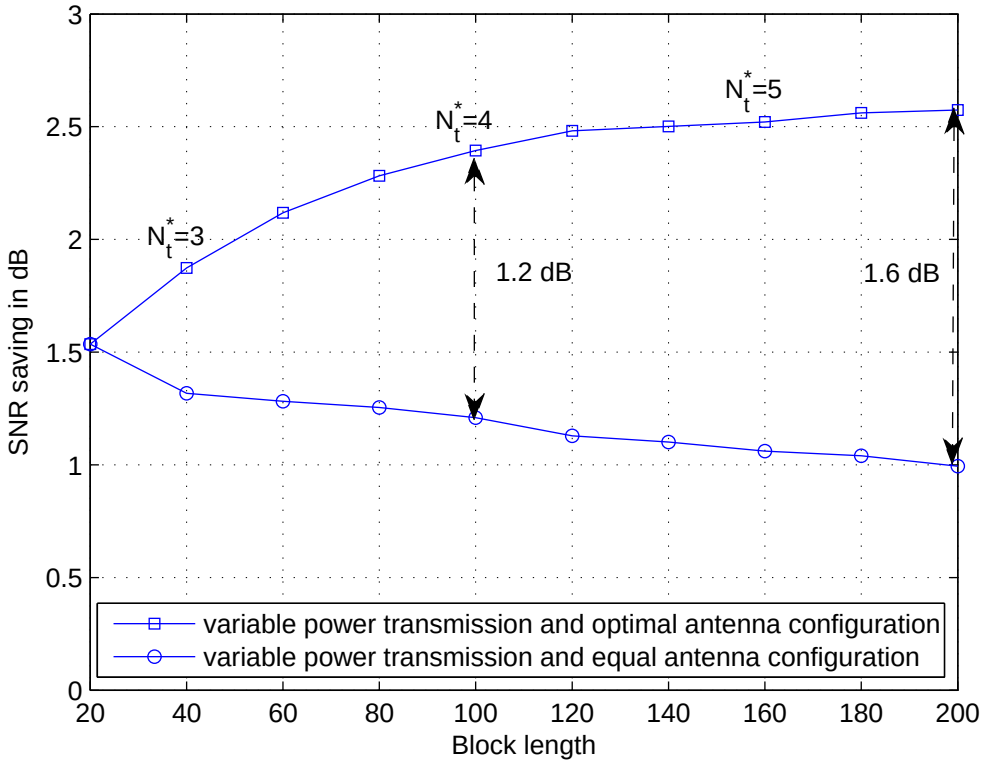

Figure 2.7: The SNR saving computed from (2.19) versus block length L for MIMO systems with 2 receive antennas at SNR $\mathcal{P} = 30$ dB. Equal antenna configuration represents systems with 2 transmit and receive antennas. N_t^* denotes the optimal number of transmit antennas.

The optimal antenna configuration at moderate SNR can be studied using the average capacity lower bound in (2.12). We find that the optimal antenna configuration obtained at high SNR can be different from that at low to moderate SNR. Particularly at sufficiently low SNR, (2.12) can be approximated as $\overline{C}_{\mathrm{LB}} = \frac{L\mathcal{P}^2}{4\ln 2}\frac{N_r}{N_t}$ for both variable and fixed power transmissions [54], which implies that the optimal number of transmit antenna is 1. However, the trends observed at high SNR on the capacity gain and SNR saving by optimizing antenna configuration are also observed at moderate SNR. That is to say, the capacity improvement from optimal antenna configuration increases as the block length increases, and optimizing antenna configuration is more beneficial than optimizing power allocation over pilot and data symbols at large block lengths.

2.6 Effects of Channel Spatial Correlation

In the previous sections, we have assumed that the channels are spatially independent. In the following, we will investigate the impact of channel spatial correlation on the transmission design and capacity performance.

2.6.1 A Robust Transmission Scheme

For non-feedback systems, the transmitter does not have any knowledge about the channel. Therefore, it is difficult to find the optimal resource allocation and transmission strategies for the unknown channel spatial correlation. Intuitively, the amount of training resource required should reduce as the channel gains become more correlated. Therefore, one good strategy may be to use the optimal PSAM energy factor α^* and the optimal training length L_p^* for i.i.d. channels so that sufficient training is ensured for all channel correlation conditions. Similarly, one may design the pilot and data transmission schemes to ensure a robust system performance for possibly correlated channels. Here, we define *robustness* to be the capability of achieving the best channel estimation and capacity performance for the least-favorable channel correlation. Note that the least or most favorable channel correlation condition depends on the choice of the training or data transmission strategy and does not necessarily imply i.i.d. or fully correlated channels.

Theorem 2.2 The transmission of orthogonal training sequences among the transmit antennas with spatially equal power allocation minimizes the channel estimation errors for the least-favorable channel correlation, *i.e.*, using $\boldsymbol{X}_p \boldsymbol{X}_p{}^\dagger = \frac{\mathcal{P}_p L_p}{N_t} \boldsymbol{I}_{N_t}$ is a robust training scheme.

Proof: see Appendix A.3.

Theorem 2.3 The transmission of i.i.d. data sequences among the transmit antennas with spatially equal power allocation, *i.e.*, $\boldsymbol{Q} = \frac{\mathcal{P}_d}{N_t} \boldsymbol{I}_{N_t}$, (a) maximizes the capacity for the least-favorable channel correlation at sufficiently low SNR, and (b) is the optimal transmission scheme at sufficiently high SNR.

Proof: see Appendix A.4.

From *Theorem 2.2* and *Theorem 2.3*, we see that the optimal transmission strategy for i.i.d. channels becomes a robust choice for correlated channels in non-feedback systems.

2.6.2 Ergodic Capacity Behavior

Having found the optimal pilot and data structure, the following result characterizes the impact of channel spatial correlation on the ergodic capacity.

Theorem 2.4 When the pilot and data structure follows $\boldsymbol{X}_p\boldsymbol{X}_p{}^{\dagger} = \frac{\mathcal{P}_p L_p}{N_t}\boldsymbol{I}_{N_t}$ and $\boldsymbol{Q} = \mathcal{P}_d/N_t\boldsymbol{I}_{N_t}$, the ergodic capacity lower bound is (a) Schur-convex on the eigenvalues of channel covariance matrix at sufficiently low SNR and (b) Schur-concave on the eigenvalues of channel covariance matrix at sufficiently high SNR. Therefore, capacity increases as the channel spatial correlation increases at sufficiently low SNR and decreases as the channel spatial correlation increases at sufficiently high SNR.

Proof: see Appendix A.5.

From *Theorem 2.4* we see that the capacity behaves very differently at low and high SNRs. At low SNR, the channel estimation error plays a crucial role and it reduces as channel spatial correlation increases, which results in an increase in the capacity. This result is in contrast with [30] which showed that the channel spatial correlation always reduces the capacity at any SNR with the assumption of perfect channel estimation at the receiver.

To validate our theoretical low and high SNR analysis, we carry out numerical studies using the ergodic capacity lower bound per channel use in (2.8). These numerical results also provide insights into the effects of channel spatial correlation on the capacity at moderate SNR values. It is noted that the ratio of energy and time allocated to pilot transmission affects the capacity results. Here we do not optimize the ratio of power and time allocated to training. We use $L_p = N_t$ and assume the fixed power transmission scheme. The trends on the capacity behavior will be similar if optimal transmit energy and time allocation is used. For numerical analysis, we choose the channel covariance matrix to be in the form of $(\boldsymbol{R}_{\boldsymbol{H}})_{ij} = \rho^{|i-j|}$, where ρ is the spatial correlation factor. For systems having two transmit antennas, ρ is referred to as the channel correlation coefficient.

Fig. 2.8 shows the ergodic capacity lower bound per channel use in (2.8) versus SNR for 2×1 MISO systems. In general, we see that fully correlated channels result in the highest capacity for a wide range of SNR values, even at 18 dB. The spatially independent channels give the lowest capacity for SNR below 6 dB. Comparing the two extreme cases, we see a significant increase in capacity from independent channels to fully correlated channels at low to moderate SNR values, which agrees with *Theorem 2.4*. For example, the increase in capacity from i.i.d.

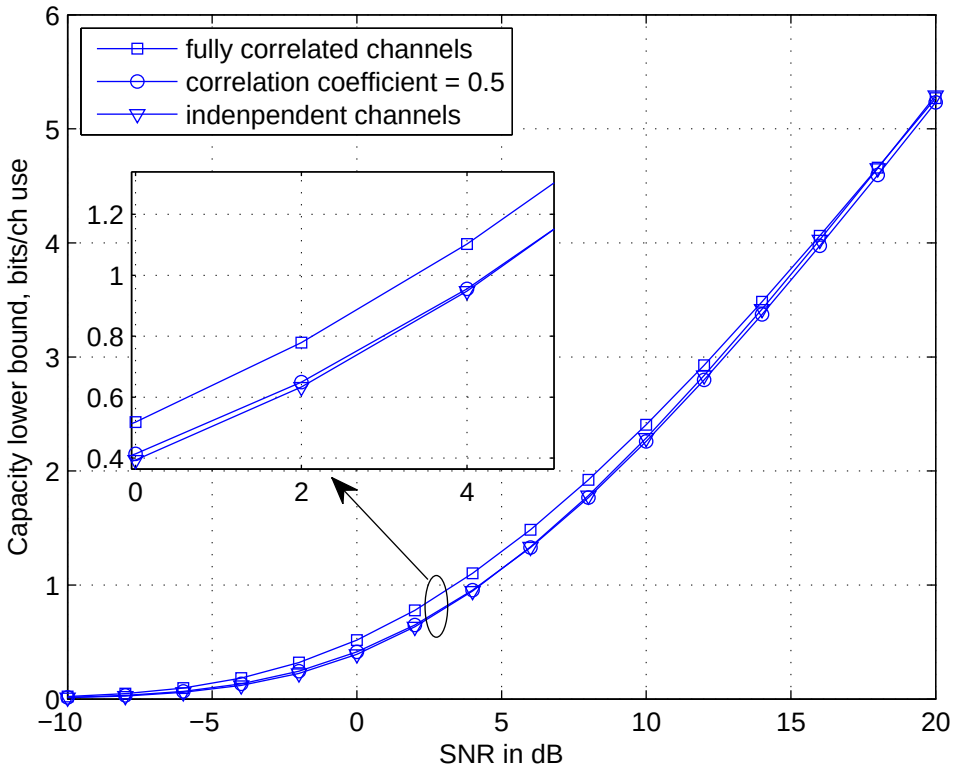

Figure 2.8: The ergodic capacity lower bound per channel use C_{LB} in (2.8) versus SNR $\mathcal{P}$, for 2×1 MISO systems with different levels of channel spatial correlation.

channels to fully correlated channels is approximately 33% at 0 dB and 13% at 5 dB.

Fig. 2.9 shows the ergodic capacity lower bound per channel use in (2.8) versus SNR for 2×2 MIMO systems. In general, we see that fully correlated channels result in the highest capacity for SNR values below 5 dB, while independent channels maximize the capacity for SNR values above 5 dB. In particular, the capacity difference between i.i.d. channels and fully correlated channels is significant at high SNRs. This is because the benefit from receiver diversity outweighs the effect of channel estimation error for i.i.d. channels at moderate to high SNRs.

2.7 Summary of Contributions

In this chapter we have investigated the optimal resource allocation in training-based MIMO systems. The important design parameters are the transmit energy and time to be distributed between training and data transmission, as well as the number of transmit and receive antennas. We have also studied the impact of

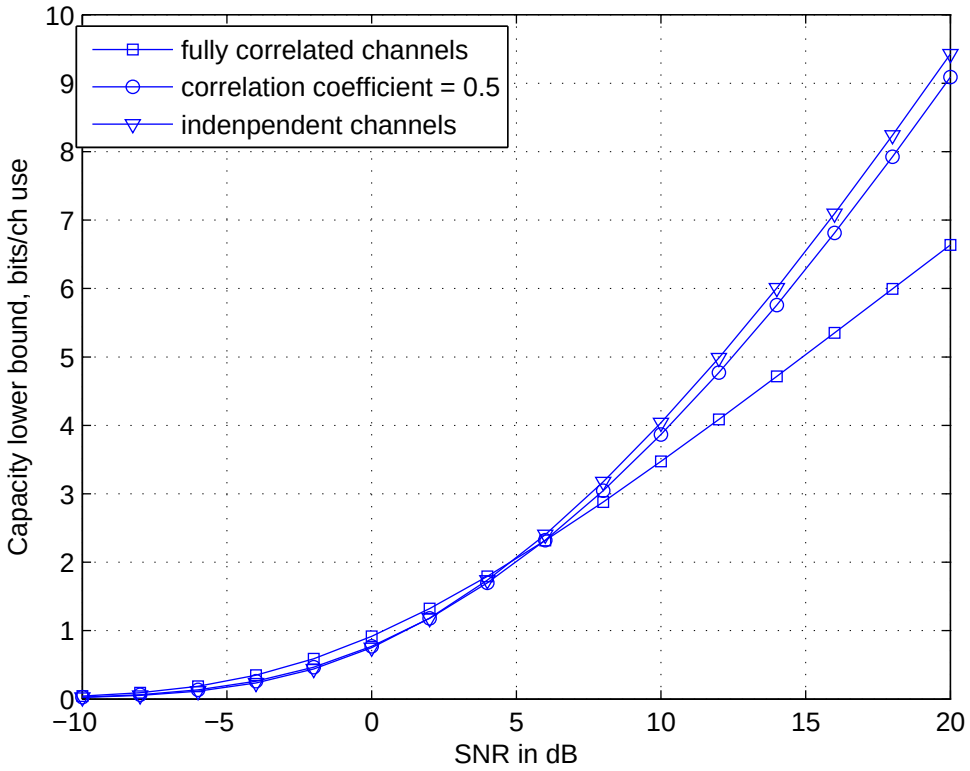

Figure 2.9: The ergodic capacity lower bound per channel use C_{LB} in (2.8) versus SNR $\mathcal{P}$, for 2×2 MIMO systems with different levels of channel spatial correlation.

channel spatial correlation on the system design and capacity performance. Some specific contributions made in this chapter are as follows.

Addressing Q1 in Section 2.1:

- For the fixed power transmission scheme where the training and data lengths can be jointly optimized, we have shown that all available time should always be used for transmission. We have also provided analytical bounds to significantly reduce the ranges of the possible values of the optimal training and data lengths.

- For the fixed power transmission scheme designed to have a fixed training length, we have derived a threshold SNR above which it is optimal to use all available time for transmission. When the actual operating SNR is considerably lower than the threshold SNR, it is wise to reduce the length of data transmission in order to avoid substantial capacity degradation.

Addressing Q2 in Section 2.1:

- When an extra antenna is available to be placed on either end of the system, one should always place it at the receiver at moderate to high SNR when the existing number of transmit antennas N_t is at least as large as that of the receive antennas N_r. When $N_t < N_r$, a critical SNR value needs to be considered, below which the extra antenna should be placed at the receiver.

- When it is only practical to change the number of transmit antennas, adding extra antennas generally improves the capacity at high SNR and large block lengths, provided $N_t < N_r$. More importantly, we have shown that the optimal number of transmit antennas can exceed the number of receive antennas, and the capacity improvement by optimal antenna configuration is significant when the block length is large. We have also shown that optimizing antenna configuration can be more beneficial than power optimization over pilot and data symbols particularly at large block lengths.

Addressing Q3 in Section 2.1:

- Extending the existing results for spatially independent channels in [53], we have derived a reasonably tight lower and upper bound on the ergodic capacity for spatially correlated MIMO systems.

- Our study on the impact of channel spatial correlation has shown that the capacity increases with channel correlation at low SNR, but decreases with correlation at high SNR. The finding at low SNR is in contrast to the existing result in [30] with perfect CSI assumption. We have also derived a robust transmission scheme which achieves the best channel estimation and capacity performance for the least-favorable channel correlation.

Chapter 3

One-Way Training-Based Systems with Feedback

3.1 Introduction

The study of MIMO communication systems can be broadly categorized based on the availability of CSI at the receiver or the transmitter side. In Chapter 2, we studied the optimal resource allocation in training-based systems with no transmitter CSI. From [6, 32, 35, 39, 95–97] we know that the MIMO information capacity can be further increased if some form of CSI is available at the transmitter. The transmitter CSI can be in the form of causal channel gain feedback (CGF) or channel covariance feedback (CCF). In this chapter, we study the optimal resource allocation in training-based systems with CGF and CCF by addressing the following questions:

Q1. What is the optimal transmission and resource allocation for systems with CGF?

Q2. What is the optimal transmission and resource allocation for systems with CCF?

Q3. What is the optimal transmission and resource allocation for systems with both CGF and CCF?

For spatially i.i.d. channels, the optimal transmission and resource allocation strategies were presented in Chapter 2 for non-feedback systems. In Section 3.3, we investigate whether the optimal solutions for the non-feedback systems also stay optimal for CGF systems. In addition, we consider the feedback delay in CGF systems and study its effect on the optimal resource allocation.

For CCF systems with spatially correlated channels, the optimal pilot structure for transmission was investigated in [61–63]. The main assumption in these works is that the number of pilot is at least as large as the number of antennas. In Section 3.4, we find that the optimal training length may be shorter than the number of transmit antennas for CCF systems. In this case, we aim to find a near optimal, yet simple transmission and resource allocation strategy.

When both CCF and CGF are available, the system can be considered as a hybrid CCF-CGF system. The optimal design for hybrid CCF-CGF systems is a very difficult research problem partially due to lack of an analytical solution for the optimal data transmission that takes both CCF and CGF into account. In addition, the complicated nature of the capacity bound expression, due to hybrid CCF-CGF, makes it nearly impossible to directly solve the problem of optimizing the training resource allocation in closed form. In Section 3.5, we utilize the optimal pilot structure and the property of LMMSE estimator to derive a closed-form solution for the optimal data transmission. We also analytically study the possible range of the optimal training length. In addition, we investigate the performance of a closed-form suboptimal solution for the energy allocation between training and data transmission.

3.2 System Model

The system model considered in this chapter is similar to that in Chapter 2. We consider a MIMO flat-fading channel model with input-output relationship given by

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

where $\boldsymbol{y}$ is the $N_r \times 1$ received symbol vector, $\boldsymbol{x}$ is the $N_t \times 1$ transmitted symbol vector, $\boldsymbol{H}$ is the $N_r \times N_t$ channel gain matrix, and $\boldsymbol{n}$ is the $N_r \times 1$ noise vector having ZMCSCG entries with unit variance. The entries of $\boldsymbol{H}$ are also ZMCSCG with unit variance.

We consider the possibility of channel spatial correlation at the transmitter side. The spatial correlation is characterized by the covariance matrix $\boldsymbol{R}_{\boldsymbol{H}} = E\{\boldsymbol{H}^{\dagger}\boldsymbol{H}\}/N_r$. Therefore, $\boldsymbol{H} = \boldsymbol{H}_0\boldsymbol{R}_{\boldsymbol{H}}^{1/2}$, where $\boldsymbol{H}_0$ has i.i.d. ZMCSCG entries with unit variance. We assume that $\boldsymbol{R}_{\boldsymbol{H}}$ is a positive definite matrix and denote the eigenvalues of $\boldsymbol{R}_{\boldsymbol{H}}$ by $\boldsymbol{g} = [g_1 \ g_2 \ \ldots \ g_{N_t}]^T$. We use the concept of majorization to characterize the degree of channel spatial correlation as explained in Section 2.2.1.

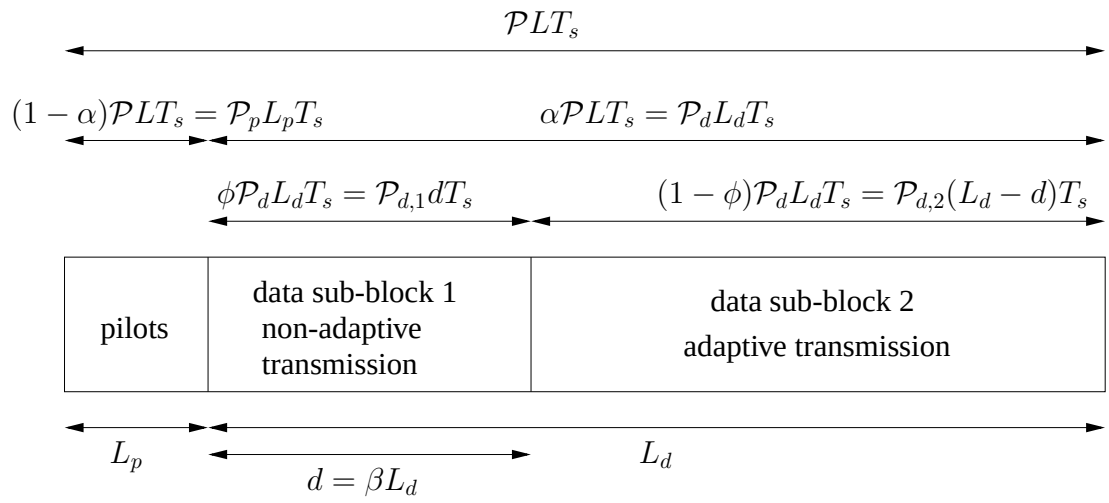

Figure 3.1: An example of a transmission block of L symbols in a system with delayed feedback. Temporal power allocations are shown at the top and the length of each sub-block is shown at the bottom.

3.2.1 Transmission Scheme

Fig. 3.1 shows an example of a transmission block of L symbol periods in a PSAM scheme. The channel gains remain constant over one block and change to independent realizations in the next block. We focus on the variable power transmission, where the pilot and data transmissions have different powers. During each transmission block, each transmit antenna sends L_p pilot symbols, followed by L_d $(= L - L_p)$ data symbols[1] as shown in Fig. 3.1. The receiver performs channel estimation during the pilot transmission. For CGF systems, the receiver feeds the channel estimates back to the transmitter once per block to allow adaptive data transmission in the form of power control. In practical scenarios, there is a time delay of d symbol periods before the transmitter receives the feedback information as shown in Fig. 3.1. Therefore, the data transmission during the first d symbol periods is not adaptive to the channel, and adaptive transmission is only available for the remaining $L_d - d$ symbol periods. We define $\beta = d/L_d$ as the feedback delay factor. For CCF systems, less frequent feedback is required as the channel correlation changes much more slowly than the channel gains. Therefore, we do not consider feedback delay in CCF systems, *i.e.*, $d = 0$. Note that for non-feedback systems, $d = L_d$.

The total transmission energy per block is given by $\mathcal{P}LT_s$ as shown in Fig. 3.1,

[1]Similar to the result for non-feedback systems in Section 2.4.1, it is easy to show that with variable power transmission, it is optimal to use all available time for transmission, *i.e.*, $L_p + L_d = L$. Therefore, we assume $L_p + L_d = L$ in this Chapter.

where $\mathcal{P}$ is the average power per symbol and T_s is the symbol duration. Similar to Chapter 2, we define the PSAM energy factor as the ratio of the total energy allocated to the data transmission, denoted by α. We also denote the power or SNR per pilot and data symbol by $\mathcal{P}_p$ and $\mathcal{P}_d$, respectively. Therefore, we have the following relationships.

$$\begin{aligned} \mathcal{P}LT_s &= \mathcal{P}_p L_p T_s + \mathcal{P}_d L_d T_s, \\ \mathcal{P}_p &= (1-\alpha)\frac{\mathcal{P}L}{L_p}, \text{ and } \mathcal{P}_d = \alpha\frac{\mathcal{P}L}{L_d}. \end{aligned} \tag{3.2}$$

For CGF systems with delay of d symbol periods as shown in Fig. 3.1, we define the data energy division factor as the ratio of the total data energy allocated to the non-adaptive sub-block, denoted by ϕ. Therefore, we have the following relationships.

$$\begin{aligned} \mathcal{P}_d L_d T_s &= \mathcal{P}_{d,1} d\, T_s + \mathcal{P}_{d,2}(L_d - d)T_s, \\ \mathcal{P}_{d,1} &= \frac{\phi}{\beta}\mathcal{P}_d, \text{ and } \mathcal{P}_{d,2} = \frac{1-\phi}{1-\beta}\mathcal{P}_d, \end{aligned} \tag{3.3}$$

where $\mathcal{P}_{d,1}$ and $\mathcal{P}_{d,2}$ are the power per symbol during the non-adaptive and adaptive sub-blocks, respectively.

3.2.2 Channel Estimation

Similar to Chapter 2, the LMMSE channel estimation method is used during the pilot transmission. Here we rewrite the key equations for ease of reference. We denote the channel estimate and estimation error as $\widehat{\boldsymbol{H}} = \widehat{\boldsymbol{H}}_0 \boldsymbol{R}_{\widehat{\boldsymbol{H}}}^{1/2}$ and $\widetilde{\boldsymbol{H}} = \widetilde{\boldsymbol{H}}_0 \boldsymbol{R}_{\widetilde{\boldsymbol{H}}}^{1/2}$, respectively, where $\widehat{\boldsymbol{H}}_0$ and $\widetilde{\boldsymbol{H}}_0$ have i.i.d. ZMCSCG entries with unit variance. $\widehat{\boldsymbol{H}}$ is given as [62]

$$\widehat{\boldsymbol{H}} = \boldsymbol{Y}(\boldsymbol{X}_p{}^{\dagger}\boldsymbol{R}_{\boldsymbol{H}}\boldsymbol{X}_p + \boldsymbol{I}_{L_p})^{-1}\boldsymbol{X}_p{}^{\dagger}\boldsymbol{R}_{\boldsymbol{H}}, \tag{3.4}$$

where $\boldsymbol{Y}$ is the $N_t \times L_p$ matrix combining the L_p received symbol vectors during pilot transmission and $\boldsymbol{X}_p$ is the $N_t \times L_p$ pilot matrix. The covariance matrix of the estimation error and the estimates are given by [62]

$$\boldsymbol{R}_{\widetilde{\boldsymbol{H}}} = \frac{E\{\widetilde{\boldsymbol{H}}^{\dagger}\widetilde{\boldsymbol{H}}\}}{N_r} = (\boldsymbol{R}_{\boldsymbol{H}}^{-1} + \boldsymbol{X}_p\boldsymbol{X}_p{}^{\dagger})^{-1}. \tag{3.5}$$

and

$$\boldsymbol{R}_{\widehat{\boldsymbol{H}}} = \frac{E\{\widehat{\boldsymbol{H}}^{\dagger}\widehat{\boldsymbol{H}}\}}{N_r} = \boldsymbol{R}_{\boldsymbol{H}} - \boldsymbol{R}_{\widetilde{\boldsymbol{H}}}. \tag{3.6}$$

3.2.3 Ergodic Capacity Bound

In Chapter 2, we derived a general capacity lower bound which is applicable to both non-feedback and feedback systems. The ergodic capacity lower bound per channel use is given by

$$C_{\mathrm{LB}} = E_{\widehat{\boldsymbol{H}}}\Big\{ \log_2 \Big| \boldsymbol{I}_{N_t} + (1 + \mathrm{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\boldsymbol{Q}\})^{-1}\widehat{\boldsymbol{H}}^{\dagger}\widehat{\boldsymbol{H}}\boldsymbol{Q} \Big| \Big\}. \tag{3.7}$$

Taking the training overhead into account, the average capacity lower bound is therefore given by

$$\begin{aligned} \overline{C}_{\mathrm{LB}} &= \frac{L_d}{L} C_{\mathrm{LB}} \\ &= \frac{L_d}{L} E_{\widehat{\boldsymbol{H}}}\Big\{ \log_2 \Big| \boldsymbol{I}_{N_t} + (1+\mathrm{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\boldsymbol{Q}\})^{-1}\widehat{\boldsymbol{H}}^{\dagger}\widehat{\boldsymbol{H}}\boldsymbol{Q} \Big| \Big\}. \end{aligned} \tag{3.8}$$

It was shown numerically in Section 2.3 that the above capacity lower bounds are reasonably tight. Therefore, we will use "capacity lower bound" and "capacity" interchangeably throughout the rest of this chapter. And the average capacity lower bound in (3.8) will be used as the figure of merit to study the optimal transmission and resource allocation strategies in the following sections.

3.3 Channel Gain Feedback (CGF) Systems

CGF systems require the receiver to feed the channel estimates back to the transmitter once per transmission block. Once the transmitter receives the estimated channel gains, it performs spatial power adaptation accordingly. When noise is present in the feedback link, the capacity that can be achieved by adaptive transmission reduces as the noise increases. The capacity reduction due to corrupted channel gain estimates was studied in [56]. It was shown that the capacity reduction can increase quickly with the noise in the estimated channel gains. Therefore, a reliable feedback scheme which minimizes the noise in the estimated channel gains is essential for CGF systems. One solution is to use low rate feedback transmission with appropriate quantization scheme for the feedback information [98]. However, the design of digital feedback is beyond the scope of this work. In this section, we

assume that the feedback link is noiseless, which is reasonable for CGF systems with reliable feedback schemes.

We consider the channels to be spatially i.i.d.. For a CGF system with correlated channels, it is fair to assume that the transmitter has the knowledge of the spatial channel correlations as well. In other words, a CGF system with correlated channels is effectively a hybrid CCF-CGF system, which will be discussed in Section 3.5. With the assumption of i.i.d. channels, the data transmission utilizes all the channels with equal probability. Hence, it is reasonable to have at least as many measurements as the number of channels for channel estimation, which implies that $L_p \geq N_t$. From [54], we know that the optimal training for i.i.d. channels consists of orthogonal pilots with equal power allocated to each antenna, *i.e.*, $\boldsymbol{X}_p \boldsymbol{X}_p{}^\dagger = \frac{\mathcal{P}_p L_p}{N_t} \boldsymbol{I}_{N_t}$.

For a given $\mathcal{P}_d$, the ergodic capacity lower bound per channel use in (3.7) can be rewritten as

$$C_{\mathrm{LB}} = E_{\widehat{\boldsymbol{H}}_0}\Big\{ \log_2 \Big| \boldsymbol{I}_{N_t} + \frac{\sigma^2_{\widehat{\boldsymbol{H}}}}{1+\sigma^2_{\widetilde{\boldsymbol{H}}}\mathcal{P}_d} \widehat{\boldsymbol{H}}_0^\dagger \widehat{\boldsymbol{H}}_0 \boldsymbol{Q} \Big| \Big\}, \tag{3.9}$$

where $\sigma^2_{\widetilde{\boldsymbol{H}}} = \left(1 + \frac{\mathcal{P}_p L_p}{N_t}\right)^{-1}$ and $\sigma^2_{\widehat{\boldsymbol{H}}} = 1 - \sigma^2_{\widetilde{\boldsymbol{H}}}$.

3.3.1 CGF Systems with No Feedback Delay

Firstly, we study an ideal (delayless) scenario in which the transmitter receives the estimated channel gains at the start of the data transmission. The delayless case can be viewed as a reasonable approximation of the delayed case when $d \ll L$. Furthermore, the results in the delayless case will be used in Section 3.3.2 for the delayed case. It was shown in [53] that the capacity is maximized when the matrix $\boldsymbol{Q}$ has the same eigenvectors as $\widehat{\boldsymbol{H}}_0^\dagger \widehat{\boldsymbol{H}}_0$. Denoting the eigenvalues of $\widehat{\boldsymbol{H}}_0^\dagger \widehat{\boldsymbol{H}}_0$ by $\boldsymbol{\lambda} = [\lambda_1 \ \lambda_2 \ \ldots \ \lambda_{N_t}]^T$ sorted in descending order, the eigenvalues of $\boldsymbol{Q}$ are found via the standard water-filling as

$$q_i = \Big[\eta - \Big(\frac{\sigma^2_{\widehat{\boldsymbol{H}}}}{1+\sigma^2_{\widetilde{\boldsymbol{H}}}\mathcal{P}_d}\lambda_i\Big)^{-1}\Big]^+ \ \text{with} \ \sum_{i=1}^{N_t} q_i = \mathcal{P}_d, \tag{3.10}$$

where η represents the water level and $[z]^+ \triangleq \max\{z, 0\}$. Therefore, (3.9) can be rewritten as

$$C_{\mathrm{LB}} \ = \ E_{\boldsymbol{\lambda}}\Big\{ \sum_{i=1}^{N_t} \log_2 \Big(1 + \frac{\sigma^2_{\widehat{\boldsymbol{H}}}}{1+\sigma^2_{\widetilde{\boldsymbol{H}}}\mathcal{P}_d}\lambda_i q_i\Big)\Big\} \tag{3.11}$$

$$= E_{\boldsymbol{\lambda}}\Big\{\sum_{i=1}^{m}\log_2\Big(\frac{\sigma^2_{\widehat{\boldsymbol{H}}}}{1+\sigma^2_{\widetilde{\boldsymbol{H}}}\mathcal{P}_d}\lambda_i\eta\Big)\Big\} \tag{3.12}$$

$$= E_{\boldsymbol{\lambda}}\Big\{\sum_{i=1}^{m}\log_2\Big(\frac{\sigma^2_{\widehat{\boldsymbol{H}}}\mathcal{P}_d}{1+\sigma^2_{\widetilde{\boldsymbol{H}}}\mathcal{P}_d}+\sum_{i=1}^{m}\lambda_i^{-1}\Big)+\sum_{i=1}^{m}\log_2\frac{\lambda_i}{m}\Big\}, \tag{3.13}$$

where m denotes the number of non-zero q_i, and (3.13) is obtained by substituting η from (3.10) into (3.12). It should be noted that $E_{\boldsymbol{\lambda}}$ in (3.12) and (3.13) is the expectation over the m largest values in $\boldsymbol{\lambda}$. Using (3.13), we now present the result for the optimal PSAM energy factor α^*.

Theorem 3.1 For delayless CGF systems with i.i.d. channels and the optimal pilot structure $\boldsymbol{X}_p\boldsymbol{X}_p{}^\dagger = \frac{\mathcal{P}_pL_p}{N_t}\boldsymbol{I}_{N_t}$, the optimal PSAM energy factor α^* coincides with α^* for non-feedback systems and is given by

$$\alpha^* = \begin{cases} \gamma-\sqrt{\gamma(\gamma-1)}, & \text{for } L_d > N_t \\ \frac{1}{2}, & \text{for } L_d = N_t \\ \gamma+\sqrt{\gamma(\gamma-1)}, & \text{for } L_d < N_t \end{cases} \tag{3.14}$$

where $\gamma = \frac{N_t+\mathcal{P}L}{\mathcal{P}L(1-N_t/L_d)}$.

Proof: see Appendix B.1.

From the study of the non-feedback systems in Section 2.4.1, we have seen that the optimal training length L_p^* is equal to N_t when the optimal PSAM energy factor α^* given in (3.14) is used. In the next theorem and corollary, we show a more general result that $L_p^* = N_t$ for any fixed value of α as well as α^* in delayless CGF systems.

Theorem 3.2 For delayless CGF systems with i.i.d. channels and the optimal pilot structure $\boldsymbol{X}_p\boldsymbol{X}_p{}^\dagger = \frac{\mathcal{P}_pL_p}{N_t}\boldsymbol{I}_{N_t}$, the optimal training length equals the number of transmit antennas for any given value of the PSAM energy factor α, *i.e.*, $L_p^* = N_t$.

Proof: see Appendix B.2.

In *Theorem 3.2*, the value of α is required to be fixed. When α is allowed to vary as L_p varies, L_p^* may or may not equal N_t. For example, α varies according to $\alpha = \frac{L-L_p}{L}$ when fixed power transmission ($\mathcal{P}_d = \mathcal{P}_p = \mathcal{P}$) is used, in which case $L_p^* = N_t$ does not hold in general.

Corollary 3.1 For delayless CGF systems with i.i.d. channels and the optimal pilot structure $\boldsymbol{X}_p\boldsymbol{X}_p{}^\dagger = \frac{\mathcal{P}_pL_p}{N_t}\boldsymbol{I}_{N_t}$ as well as the optimal PSAM energy factor α^*, the optimal training length is given by $L_p^* = N_t$.

Proof: see Appendix B.3.

Theorem 3.1 and *Corollary 3.1* show that the optimal pilot design for delayless CGF systems coincides with that for non-feedback systems discussed in Section 2.4.1. That is to say, one can use the same design to achieve optimal performance in both non-feedback and delayless CGF systems.

3.3.2 CGF Systems with Feedback Delay

For practical systems, a finite duration of d symbol periods is required before feedback comes into effect at the transmitter as shown in Fig. 3.1. Therefore, the transmitter has no knowledge about the channel during the first data sub-block of d transmissions, which is equivalent to non-feedback systems. From Section 2.4, we know that the transmitter should allocate equal power to each transmit antenna during the first data sub-block (or the non-adaptive sub-block). After receiving the estimated channel gains, the transmitter performs spatial power water-filling similar to Section 3.3.1 during the second data sub-block (or the adaptive sub-block) of length $L_d - d$. In order to optimize the PSAM energy factor α, we apply a two-stage optimization approach. Firstly, we optimize the data energy division factor ϕ for any given $\mathcal{P}_d$ and L_d. Then, we optimize the PSAM energy factor α.

Using (3.3), the capacity lower bound per channel use in (3.9) becomes

$$\begin{aligned} C_{\mathrm{LB}} &= E_{\boldsymbol{\lambda}}\Big\{\beta\sum_{i=1}^{N_t}\log_2\Big(1+\frac{\sigma^2_{\widehat{\boldsymbol{H}}}\lambda_i}{1+\sigma^2_{\widetilde{\boldsymbol{H}}}\mathcal{P}_d\phi/\beta}\frac{\mathcal{P}_d\phi}{N_t\beta}\Big) \\ &\quad +(1-\beta)\sum_{i=1}^{N_t}\log_2\Big(1+\frac{\sigma^2_{\widehat{\boldsymbol{H}}}\lambda_i q_i}{1+\sigma^2_{\widetilde{\boldsymbol{H}}}\mathcal{P}_d(1-\phi)/(1-\beta)}\Big)\Big\} \qquad (3.15) \\ &= E_{\boldsymbol{\lambda}}\Big\{\beta\sum_{i=1}^{N_t}\log_2\Big(1+\frac{\sigma^2_{\widehat{\boldsymbol{H}}}\lambda_i}{1+\sigma^2_{\widetilde{\boldsymbol{H}}}\mathcal{P}_d\phi/\beta}\frac{\mathcal{P}_d\phi}{N_t\beta}\Big) \\ &\quad +(1-\beta)\sum_{i=1}^{m}\log_2\Big(\frac{\sigma^2_{\widehat{\boldsymbol{H}}}\lambda_i\nu}{1+\sigma^2_{\widetilde{\boldsymbol{H}}}\mathcal{P}_d(1-\phi)/(1-\beta)}\Big)\Big\}, \qquad (3.16) \end{aligned}$$

where the water-filling solution for q_i with water level ν is given by

$$\begin{aligned} q_i &= \Big[\nu-\Big(\frac{\sigma^2_{\widehat{\boldsymbol{H}}}\lambda_i}{1+\sigma^2_{\widetilde{\boldsymbol{H}}}\mathcal{P}_d(1-\phi)/(1-\beta)}\Big)^{-1}\Big]^+ \qquad (3.17) \\ &\quad \text{with}\ \sum_{i=1}^{N_t} q_i = \frac{1-\phi}{1-\beta}\mathcal{P}_d. \end{aligned}$$

Lemma 3.1 The capacity lower bound per channel use in (3.16) is concave on $\phi \in [0,1]$.

Proof: From the property of water-filling solution [99], we know that q_i and ν in (3.17) are continuous on $\phi \in [0,1]$. As a result, $C_{\rm LB}$ in (3.16) is continuous on $\phi \in [0,1]$. For any fixed m, one can show that $\frac{\mathrm{d}^2 C_{\rm LB}}{\mathrm{d}\phi^2} < 0$ by directly computing the derivative using (3.16). This implies that $\frac{\mathrm{d}C_{\rm LB}}{\mathrm{d}\phi}$ given by

$$\frac{\mathrm{d}C_{\rm LB}}{\mathrm{d}\phi} = \frac{\mathcal{P}_d}{\ln 2} E_{\boldsymbol{\lambda}}\Big\{ \frac{\beta}{N_t} \sum_{i=1}^{N_t} \frac{\sigma^2_{\widehat{\boldsymbol{H}}}\lambda_i - (\sigma^2_{\widehat{\boldsymbol{H}}}\lambda_i \sigma^2_{\widetilde{\boldsymbol{H}}}\mathcal{P}_d\phi)/(\beta + \sigma^2_{\widetilde{\boldsymbol{H}}}\mathcal{P}_d\phi)}{\beta + \sigma^2_{\widetilde{\boldsymbol{H}}}\mathcal{P}_d\phi + \sigma^2_{\widehat{\boldsymbol{H}}}\lambda_i \mathcal{P}_d\phi/N_t} - \frac{1+\sigma^2_{\widehat{\boldsymbol{H}}}\sigma^2_{\widetilde{\boldsymbol{H}}}\sum_{i=1}^{m}\lambda_i^{-1}}{\nu} + \frac{m\sigma^2_{\widetilde{\boldsymbol{H}}}(1-\beta)}{1-\beta+\sigma^2_{\widetilde{\boldsymbol{H}}}\mathcal{P}_d(1-\phi)} \Big\} \tag{3.18}$$

is a decreasing function of ϕ for any given m. Furthermore, it can be shown that $C_{\rm LB}$ is differentiable on $\phi \in [0,1]$, including the points where m changes its value. To obtain differentiability, we let $\phi = \phi_0$ at which m changes between m_0 and m_0+1. At this point, the water level is given by $\nu_0 = \left(\frac{\sigma^2_{\widehat{\boldsymbol{H}}}\lambda_{(m_0+1)}}{1+\sigma^2_{\widetilde{\boldsymbol{H}}}\mathcal{P}_d(1-\phi_0)/(1-\beta)}\right)^{-1}$. With $\nu = \nu_0$, one can show that the left and right derivatives of $C_{\rm LB}$ w.r.t. ϕ equate, that is, $\frac{\mathrm{d}C_{\rm LB}}{\mathrm{d}\phi}(m = m_0) = \frac{\mathrm{d}C_{\rm LB}}{\mathrm{d}\phi}(m = m_0+1)$ at $\phi = \phi_0$. Therefore, one can conclude that $C_{\rm LB}$ is differentiable and its derivative is decreasing on $\phi \in [0,1]$, which implies concavity. □

With *Lemma 3.1*, the optimal data energy division factor ϕ^* can be found numerically. Using the Karush-Kuhn-Tucker (KKT) conditions [99], the result is given by

$$\begin{cases} \phi^* = 0, & \text{if } \frac{\mathrm{d}C_{\rm LB}}{\mathrm{d}\phi}(\phi = 0) \le 0 \\ \arg_\phi \frac{\mathrm{d}C_{\rm LB}}{\mathrm{d}\phi} = 0, & \text{otherwise} \end{cases} \tag{3.19}$$

where $\frac{\mathrm{d}C_{\rm LB}}{\mathrm{d}\phi}$ was given in (3.18).

Fig. 3.2 shows the optimal data energy division factor ϕ^* given by (3.19) versus average SNR $\mathcal{P}$ for different delay factors β. It can be seen that ϕ^* increases from 0 to β at low SNR. For moderate to high SNR, ϕ^* stays above β and converges to β as $\mathcal{P} \to \infty$.[2] As we are concerned with practical design solutions, it is desirable to have a low complexity solution for ϕ which still gives near optimal performance under practical transmission conditions. From Fig. 3.2, we see that ϕ^* is usually close to β at practical SNR range, *e.g.*, $\mathcal{P} > 0$ dB, and $\phi^* \to \beta$ as $\mathcal{P} \to \infty$. Therefore,

[2] ϕ^* for the $(N_t = 4, N_r = 2)$ system exceeds and converges back to β at a higher SNR, which is not shown in Fig. 3.2. This is because the use of spatial water-filling in data transmission gives a significant improvement in the capacity when $N_t > N_r$.

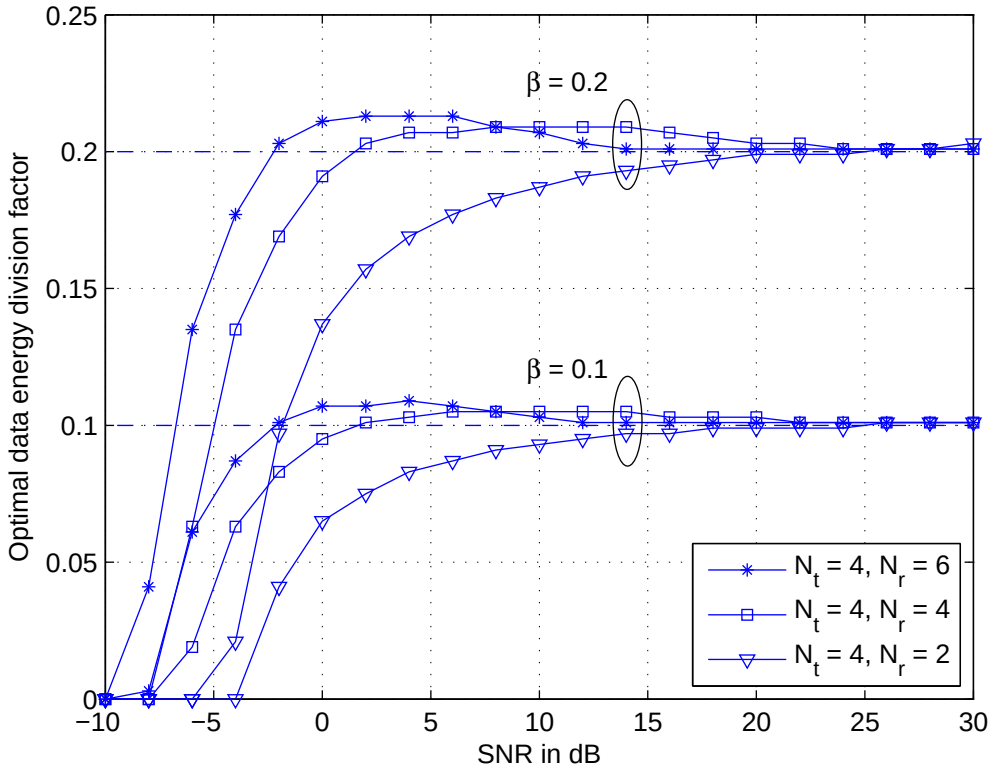

Figure 3.2: The optimal data energy division factor ϕ^* vs. SNR $\mathcal{P}$ for different values of the delay factor β and antenna sizes. In this example, a block length of $L = 100$, training length of $L_p = N_t = 4$, and PSAM energy factor given in (3.14) are used.

$\phi = \beta$ is a simple solution which provides good system performance. We will also investigate the optimality of $\phi = \beta$ using capacity results in Section 3.3.3.

From (3.3) we see that $\phi = \beta$ is actually the simplest solution which allocates the same amount of power during each data transmission in both non-adaptive and adaptive sub-blocks, *i.e.*, $\mathcal{P}_{d,1} = \mathcal{P}_{d,2} = \mathcal{P}_d$. In addition, this simple solution does not require the knowledge of the feedback delay time. Furthermore, this choice of ϕ leads to a simple closed-form solution for the optimal PSAM energy factor α^*, as well as the optimal training length L_p^* for delayed CGF system summarized in *Corollary 3.2*, which can be shown by combining the results in *Theorem 3.1*, *Corollary 3.1* and those for the non-feedback systems summarized in Section 2.4.1.

Corollary 3.2 If $\boldsymbol{X}_p \boldsymbol{X}_p{}^\dagger = \frac{\mathcal{P}_p L_p}{N_t} \boldsymbol{I}_{N_t}$ and $\phi = \beta$, the optimal PSAM energy factor α^* and the optimal training length L_p^* coincide with those in the delayless case given in *Theorem 3.1* and *Corollary 3.1*.

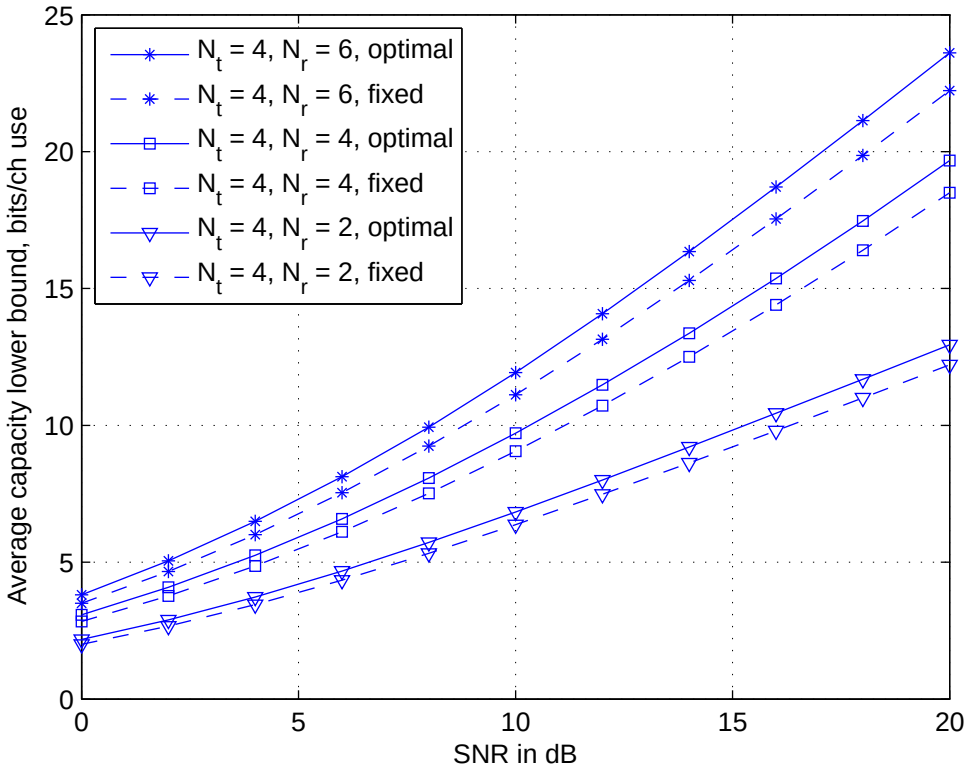

Figure 3.3: Average capacity lower bound $\overline{C}_{\mathrm{LB}}$ vs. SNR $\mathcal{P}$ for delayless CGF systems ($d = 0$ and $\beta = 0$) with i.i.d. channels and different antenna sizes. Note that $\overline{C}_{\mathrm{LB}} = (L_d/L)C_{\mathrm{LB}}$ where C_{LB} is given in (3.11). The block length is $L = 100$. Both variable power transmission with α^* given in (3.14) as well as fixed power transmission are shown for comparison. For variable power transmission, the training length is $L_p^* = 4$; while for fixed power transmission, the pilot length is optimized numerically.

3.3.3 Numerical Results

Now, we present numerical results to illustrate the capacity gain from optimizing the PSAM energy factor. Fig. 3.3 shows the average capacity lower bound $\overline{C}_{\mathrm{LB}}$ versus SNR $\mathcal{P}$ for delayless CGF systems (*i.e.*, $d = 0$ and $\beta = 0$) with i.i.d. channels and different antenna sizes. The solid lines indicate systems using α^* and L_p^* ($L_p^* = 4$ in this case). The dashed lines indicate systems using fixed power transmission and L_p^* found numerically. Comparing the solid and dashed lines, we see that the capacity gain from optimal variable power transmission is approximately 9% at 0 dB and 6% at 20 dB for all three systems. This range of capacity gain (5% to 10%) was also observed in [54] for non-feedback systems which can be viewed as an extreme case of delayed CGF system with $d = L_d$. From the results for the extreme cases, *i.e.*, $d = 0$ and $d = L_d$, we conclude that the capacity gain from optimizing

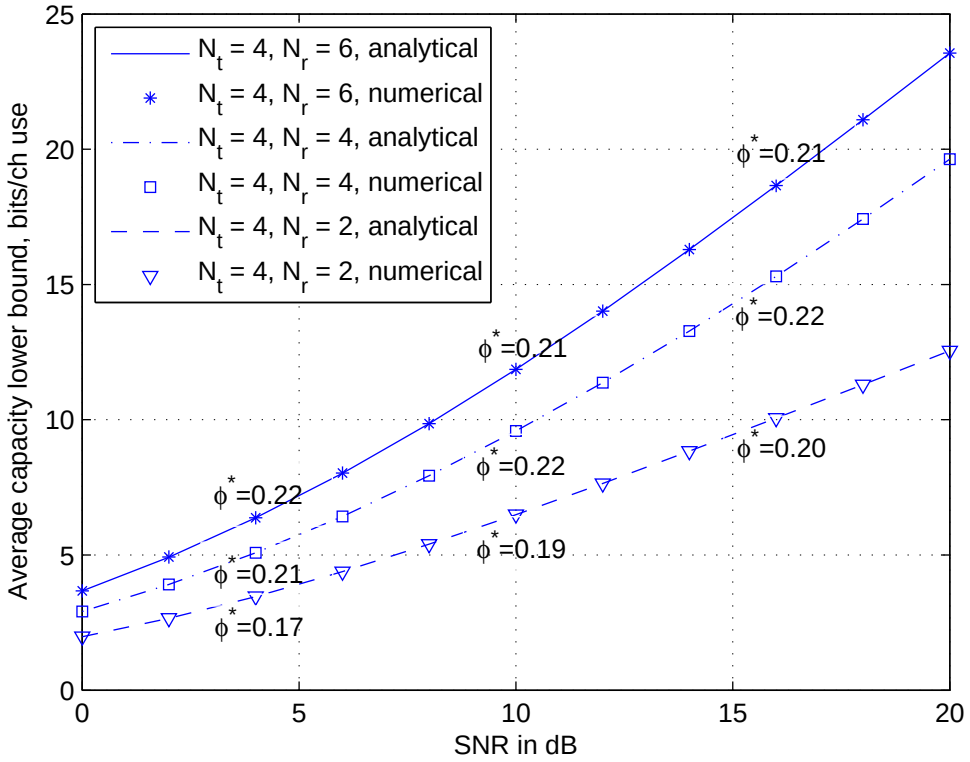

Figure 3.4: Average capacity lower bound $\overline{C}_{\mathrm{LB}}$ vs. SNR $\mathcal{P}$ for delayed CGF systems with i.i.d. channels and different antenna sizes. Note that $\overline{C}_{\mathrm{LB}} = (L_d/L)C_{\mathrm{LB}}$ where C_{LB} is given in (3.15). Within a block length of $L = 100$, the training length is $L_p = 4$, followed by a non-adaptive data transmission sub-block of length $d = 20$ and an adaptive data transmission sub-block of length 76. α in (3.14) is used. The lines indicate the use of $\phi = \beta = 0.208$, and the markers indicate optimal data energy division factor found numerically.

the PSAM energy factor is around 5% to 10% at practical SNR for delayed CGF systems with i.i.d. channels.

We now numerically verify the near optimality of $\phi = \beta$. Fig. 3.4 shows the average capacity lower bound $\overline{C}_{\mathrm{LB}}$ versus SNR $\mathcal{P}$ for delayed CGF systems with i.i.d. channels and different antenna sizes. In this example, a transmission block of length $L = 100$ consists of a training sub-block of $L_p = 4$ symbol periods, followed by a non-adaptive data sub-block of $d = 20$ symbol periods[3] and an adaptive data sub-block of $L_d - d = 76$ symbol periods. Therefore, the delay factor is $\beta = 0.208$. The lines indicate the use of $\phi = \beta$, and the markers indicate optimal data power division found through numerical optimization using $\overline{C}_{\mathrm{LB}}$. The values of ϕ^* for

[3]The delay length d takes into account the channel estimation and other processing time at the receiver and transmitter, as well as the time spent on the transmission of low-rate feedback.

SNR = 4 dB, 10 dB and 16 dB are shown in the figure as well. We see that the capacity difference between the system using $\phi = \beta$ and $\phi = \phi^*$ is negligible. That is to say the use of temporal equal power transmission over the entire data block is near optimal for systems with channel estimation errors. We have also confirmed that this trend is valid for a wide range of block lengths (results are omitted for brevity). These results validate the near optimality of $\phi = \beta$.

3.4 Channel Covariance Feedback (CCF) Systems

As discussed in Section 3.3, CGF systems require frequent use of feedback due to the rapid change in the channel gains. This requires a non-negligible amount of feedback overhead in the reverse link (from the receiver to the transmitter). On the other hand, the statistics of the channel gains change much more slowly than the channel gains themselves. As a result, it is practical for the receiver to accurately measure the channel covariance matrix and feed it back to the transmitter at a much lower frequency with negligible feedback overhead and delay. Note that for completely i.i.d. channels, there is no need for CCF. In this section, we consider CCF systems with spatially correlated channels and investigate the optimal pilot and data transmission strategy, as well as the optimal resource allocation. We assume that both the transmitter and the receiver have perfect knowledge of the channel spatial correlations.

3.4.1 Proposed Transmission Scheme

Most studies on the optimal pilot design for correlated channels assume that the training length is at least as large as the number of transmit antennas, *i.e.*, $L_p \geq N_t$ [61–63]. Intuitively, the amount of training resource can be significantly reduced as the channels become more correlated. As an extreme case where the channels are fully correlated, only one pilot transmission is needed to train all the channels. Therefore, we relax this assumption by considering $L_p \geq 1$. It was shown in [62] that the optimal training strategy is to train along the eigenvectors of the channel covariance matrix with training power being water-filled according to the eigenvalues of the channel covariance matrix, provided that $L_p \geq N_t$. In the case where $L_p < N_t$, only L_p eigen-channels can be trained. Therefore, we propose that only the L_p *strongest* eigen-channels are to be trained when $L_p < N_t$.

We perform EVD on the channel covariance matrix as $\boldsymbol{R_H} = \boldsymbol{U}\boldsymbol{G}\boldsymbol{U}^\dagger$, and let the eigenvalues of $\boldsymbol{R_H}$ be sorted in descending order in $\boldsymbol{g} = [g_1 \ g_2 \ \dots \ g_{N_t}]^T$. The optimal training sequence which minimizes the channel estimation errors (*i.e.*, $\mathrm{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\}$)

has the property that the eigenvalue decomposition of $\boldsymbol{X}_p\boldsymbol{X}_p{}^\dagger$ is given by $\boldsymbol{X}_p\boldsymbol{X}_p{}^\dagger = \boldsymbol{U}\boldsymbol{P}\boldsymbol{U}^\dagger$ [62], where $\boldsymbol{P}$ is a diagonal matrix. We denote the diagonal entries of $\boldsymbol{P}$ by $\boldsymbol{p} = [p_1\ p_2\ \ldots\ p_{N_t}]^T$. Let $n = \min\{L_p, N_t\}$, then the first n entries in $\boldsymbol{p}$ are given by

$$p_i = [\mu - g_i^{-1}]^+, \quad \text{with } \textstyle\sum_{i=1}^{n} p_i = \mathcal{P}_p L_p, \tag{3.20}$$

where μ is the water level. All the remaining entries in $\boldsymbol{p}$ (if any) are set to zero. It is not difficult to show that this choice of $\boldsymbol{X}_p\boldsymbol{X}_p{}^\dagger$ is optimal. Note that in practice, the transmitter can ensure that the number of non-zero p_i equals n by reducing L_p when needed.

For data transmission, it was shown that the optimal strategy is to transmit along the eigenvectors of $\boldsymbol{R}_{\boldsymbol{H}}$ under the perfect channel estimation [32,39,96]. With the proposed training sequence, it is easy to show from (3.5) and (3.6) that the eigenvectors of $\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}$ and $\boldsymbol{R}_{\widehat{\boldsymbol{H}}}$ are the same as those of $\boldsymbol{R}_{\boldsymbol{H}}$. Therefore, in the case of imperfect channel estimation, a reasonable strategy is to transmit data along the eigenvectors of $\boldsymbol{R}_{\widehat{\boldsymbol{H}}}$ which coincide with the eigenvectors of $\boldsymbol{R}_{\boldsymbol{H}}$. From (3.5) and (3.6), the eigenvalue decomposition of $\boldsymbol{R}_{\widehat{\boldsymbol{H}}}$ is given by $\boldsymbol{R}_{\widehat{\boldsymbol{H}}} = \boldsymbol{U}\widehat{\boldsymbol{G}}\boldsymbol{U}^\dagger$, and we set $\boldsymbol{Q} = \boldsymbol{U}\widehat{\boldsymbol{Q}}\boldsymbol{U}^\dagger$ where $\widehat{\boldsymbol{Q}}$ is a diagonal matrix with entries denoted by $\widehat{q}_i$, $\forall\, i = 1, ..., N_t$.

To the best of our knowledge, there is no closed-form solution to the optimal spatial data power allocation even with perfect channel estimation [32, 39, 96]. Instead, the optimal $\widehat{q}_i$ can be found via numerical methods [100]. As we are concerned with practical design solutions, it is desirable to have a low complexity solution for $\widehat{q}_i$ which still gives near optimal performance. Following the proposed training scheme, we propose a simple strategy which transmits data through the n trained eigen-channels with equal power. That is

$$\widehat{q}_i = \begin{cases} \mathcal{P}_d/n, & i = 1, \ldots, n \\ 0, & \text{else} \end{cases} \tag{3.21}$$

It is easy to see that only the n trained eigen-channels should be used for data transmission, since the capacity is zero for untrained eigen-channels. We will numerically investigate the optimality of (3.21) in Section 3.4.4.

For the proposed training and data transmission scheme, the capacity lower bound per channel use in (3.7) reduces to

$$C_{\text{LB}} = E_{\widehat{\boldsymbol{H}}_0}\Big\{ \log_2 \Big| \boldsymbol{I}_{N_t} + \widehat{\boldsymbol{H}}_0^\dagger \widehat{\boldsymbol{H}}_0 \widehat{\boldsymbol{G}}\widehat{\boldsymbol{Q}}(1+\mu^{-1}\mathcal{P}_d)^{-1} \Big| \Big\}, \tag{3.22}$$

where the non-zero (diagonal) entries of $\widehat{\boldsymbol{G}}$ are given by $\widehat{g}_i = g_i - \mu^{-1}$, $\forall\, i = 1, ..., n$, which is derived from (3.5), (3.6) and (3.20).

3.4.2 Optimal Training Resource Allocation

Now, we investigate the optimal training length L_p^* as well as the optimal PSAM energy factor α^* using the average capacity lower bound $\overline{C}_{\mathrm{LB}} = (L_d/L)C_{\mathrm{LB}}$ with C_{LB} given in (3.22). The results are summarized in the following theorems and corollary.

Theorem 3.3 For CCF systems in PSAM schemes with the transmission strategy proposed in Section 3.4.1, the optimal training length is at most as large as the number of transmit antennas for any given value of the PSAM energy factor α, *i.e.*, $L_p^* \leq N_t$.

Proof: see Appendix B.4.

Corollary 3.3 For CCF systems in PSAM schemes with the transmission strategy proposed in Section 3.4.1 as well as the optimal PSAM energy factor α^*, the optimal training length is given by $L_p^* \leq N_t$.

Proof: The proof is similar to that of *Corollary 3.1*.□

Although the optimal training length needs to be found numerically, the computational complexity of optimizing L_p is low due to the fact that L_p only takes integer values ranging from 1 to N_t. Hence, we will only consider $L_p \leq N_t$ in the analysis on α^*.

Theorem 3.4 For CCF systems in PSAM schemes with the transmission strategy proposed in Section 3.4.1, the optimal PSAM energy factor α^* is given by (3.14) with $\gamma = \frac{L_d}{L_d - L_p}$, provided that $\mathcal{P}L \gg \sum_{i=1}^{L_p} g_i^{-1}$.

Proof: see Appendix B.5.

The condition of $\mathcal{P}L \gg \sum_{i=1}^{L_p} g_i^{-1}$ can be easily satisfied when the block length is not too small or the SNR is moderate to high (*i.e.*, $\mathcal{P}L \gg 1$), and the spatial correlation between any trained channels is not close to 1. Therefore, the result in *Theorem 3.4* applies to many practical scenarios. It is important to note that the optimal PSAM energy factor α^* given in *Theorem 3.4* does not depend on the channel spatial correlation, provided that the condition above is met. In other words, a single design is suitable for a relatively wide range of channel spatial correlations.

3.4.3 A Special Case: Beamforming

From *Theorem 3.3* we know that $L_p^* \leq N_t$. When only the strongest eigen-channel is used, *i.e.*, $L_p = 1$, this scheme is called beamforming, which may be the optimal strategy for highly correlated channels. Furthermore, the use of beamforming significantly reduces the complexity of the system as it allows the use of well-established scalar codec technology and only requires the knowledge of the strongest eigen-channel (not the complete channel statistics) [96]. For beamforming transmission, the capacity lower bound per channel use in (3.22) reduces to

$$\begin{aligned} C_{\mathrm{LB}} &= E_{\widehat{\boldsymbol{h}}_0}\Big\{ \log_2\Big(1+\widehat{\boldsymbol{h}}_0^{\dagger}\widehat{\boldsymbol{h}}_0 \frac{(g_{\max}-\mu^{-1})\mathcal{P}_d}{1+\mu^{-1}\mathcal{P}_d}\Big)\Big\} \\ &= E_{\widehat{\boldsymbol{h}}_0}\Big\{ \log_2\Big(1+\widehat{\boldsymbol{h}}_0^{\dagger}\widehat{\boldsymbol{h}}_0 \frac{g_{\max}\mathcal{P}_p\mathcal{P}_d}{g_{\max}^{-1}+\mathcal{P}_p+\mathcal{P}_d}\Big)\Big\}, \end{aligned} \tag{3.23}$$

where $\widehat{\boldsymbol{h}}_0$ is a $N_r \times 1$ vector with i.i.d. ZMCSCG and unit variance entries, $g_{\max}$ is the largest eigenvalue in $\boldsymbol{g}$, and $\mu = \mathcal{P}_p + g_{\max}^{-1}$ which can be found by letting $L_p = 1$ in (3.20).

Lemma 3.2 For CCF systems in PSAM schemes with beamforming, the ergodic capacity lower bound per channel use in (3.23) is Schur-convex on the eigenvalues of the channel covariance matrix. Therefore, capacity increases as channel spatial correlation increases.

Proof: C_{LB} in (3.23) is an increasing function of $g_{\max}$. Also $g_{\max}$ is an increasing function of $\boldsymbol{g}$ and is Schur-convex in $\boldsymbol{g}$ [93]. Therefore, C_{LB} is also Schur-convex in $\boldsymbol{g}$, which means the capacity is an increasing function of the channel spatial correlation. □

The Schur-convexity of the capacity holds at any SNR for beamforming transmission only. However, the capacity can be neither Schur-convex nor Schur-concave when $L_p > 1$. In other words, the capacity can be neither an increasing function nor a decreasing function of the channel spatial correlation when $L_p > 1$, which will be seen in Section 3.4.4.

Theorem 3.5 For CCF systems in PSAM schemes with beamforming, the optimal PSAM energy factor α^* is given in (3.14) with $\gamma = \frac{1+g_{\max}\mathcal{P}L}{g_{\max}\mathcal{P}L(L-2)/(L-1)}$.

Proof: The proof can be obtained by letting $L_p = 1$ and $g_i = g_{\max}$ in the proof of *Theorem 3.4*. □

It can be shown for the beamforming case that $\frac{\mathrm{d}\alpha^*}{\mathrm{d}g_{\max}} > 0$. Therefore, the optimal PSAM energy factor α^* increases as the channel spatial correlation increases, that

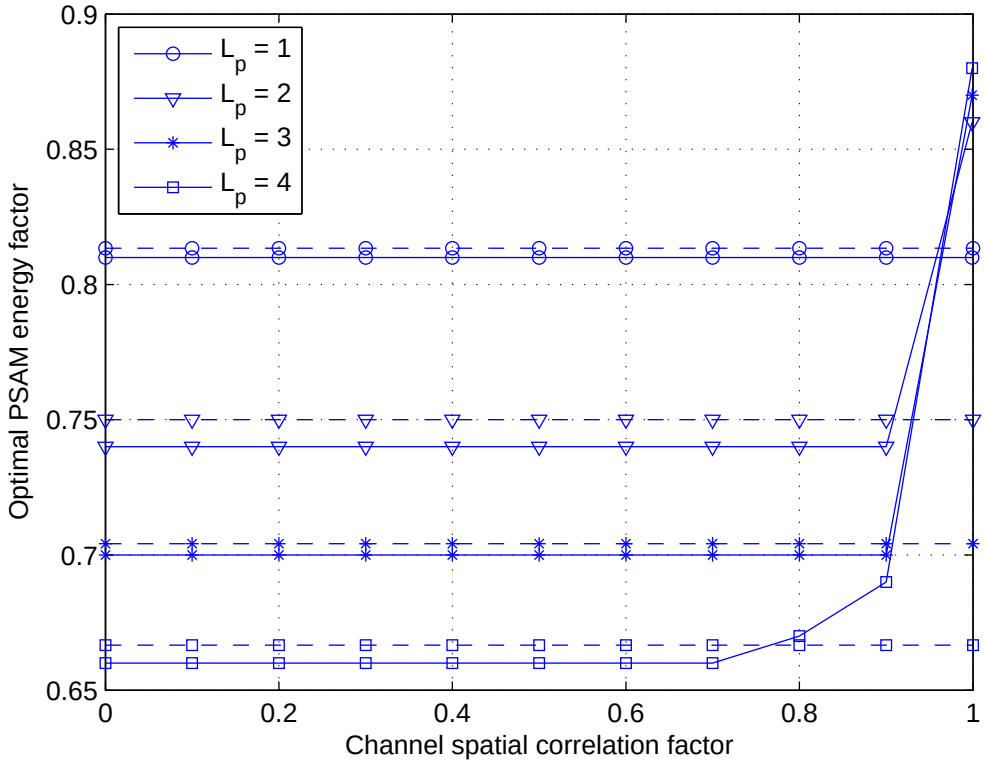

Figure 3.5: Optimal PSAM energy factor α^* vs. channel spatial correlation factor ρ for CCF 4×4 systems with a block length of $L = 20$ and SNR $= 10$ dB. The solid lines indicate the optimal α^* found numerically. The dashed lines indicate the analytical value of α given in *Theorem 3.4* and *Theorem 3.5*.

is to say, more energy should be allocated to data transmission when the channels become more correlated. When $\mathcal{P}L \gg 1$, γ reduces to $\frac{L-1}{L-2}$, hence α^* does not depend on the channel correlation.

3.4.4 Numerical Results

For numerical analysis, we choose the channel covariance matrix to be in the form of $[\boldsymbol{R}_{\boldsymbol{H}}]_{ij} = \rho^{|i-j|}$, where ρ is referred to as the spatial correlation factor [62]. Firstly we validate the solution to the optimal PSAM energy factor given in *Theorem 3.4* and *Theorem 3.5*. Fig. 3.5 shows the optimal PSAM energy factor α^* found numerically versus the channel correlation factor ρ (the solid lines) for CCF 4×4 systems with a block length of $L = 20$ and SNR of 10 dB. We see that α^* remains constant before the correlation factor gets close to 1 for $L_p > 1$, and this value of α^* is very close to the analytical value given by *Theorem 3.4* (the dashed lines). For the beamforming case where $L_p = 1$, we see that α^* does not depend on the

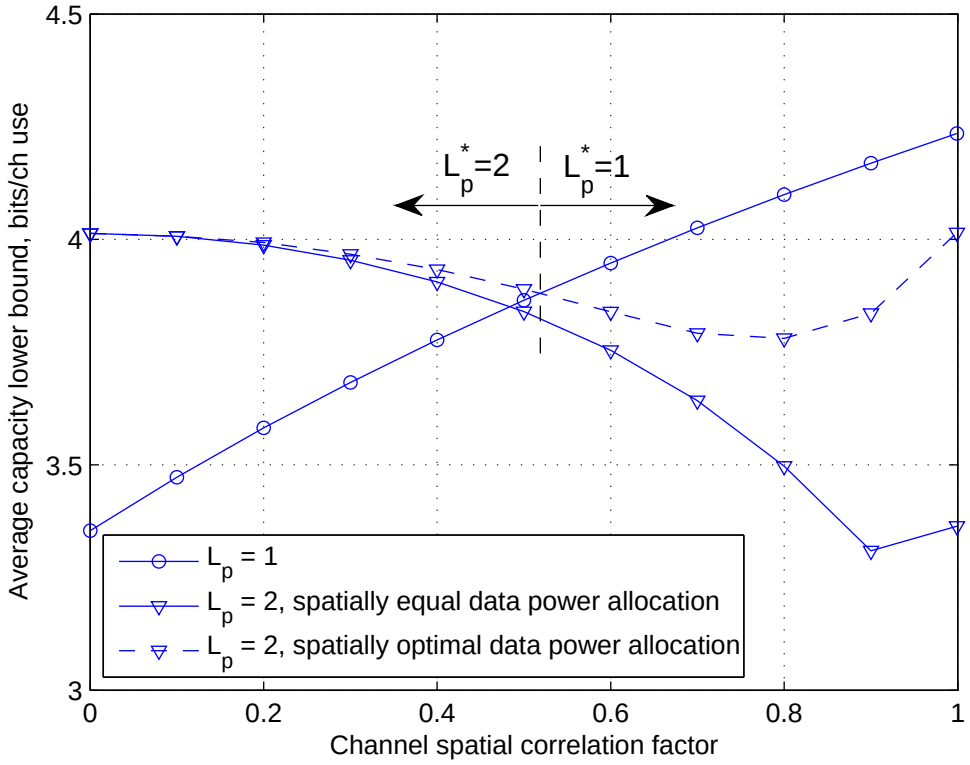

Figure 3.6: Average capacity lower bound $\overline{C}_{\mathrm{LB}}$ in (3.8) vs. channel spatial correlation factor ρ for CCF 2×2 systems with a block length of $L = 20$ and SNR of 10 dB. Training length of $L_p = 1$ and $L_p = 2$ are shown. For $L_p = 2$, both spatially equal data power allocation (dashed lines) and optimal data power allocation found numerically (solid lines) are shown.

channel correlation, which agrees with our earlier observation from *Theorem 3.5*. Similar to CGF systems, we have also compared the capacity achieved using α^* and that using fixed power transmission, and the same trend is observed (results are omitted for brevity), that is, capacity gain from optimizing PSAM energy factor is around 5% to 10% at practical SNR values.

In our proposed transmission scheme for CCF systems, spatially equal power allocation is used for data transmission. Here we illustrate the near optimality of this simple scheme in Fig. 3.6, which shows the average capacity lower bound $\overline{C}_{\mathrm{LB}}$ versus channel correlation factor ρ for CCF 2×2 systems. We compute the capacity achieved using $L_p = 1$, and $L_p = 2$ with spatially equal power allocation for data transmission (solid line) and optimal power allocation found numerically (dashed line) for a block length of $L = 20$. We also indicate the critical ρ at which L_p^* changes from 2 to 1 in Fig. 3.6. It is clear that the capacity loss from spatially optimal power allocation to spatially equal power allocation increases as ρ increases.

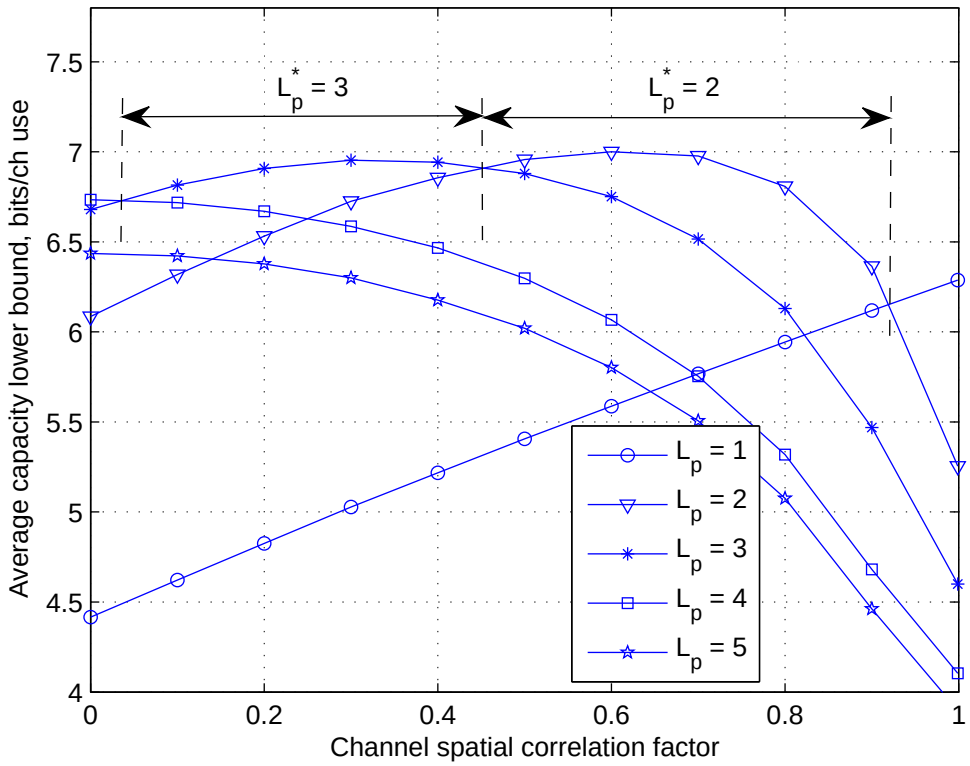

Figure 3.7: Average capacity lower bound $\overline{C}_{\mathrm{LB}}$ vs. channel spatial correlation factor ρ for CCF 4×4 systems with a block length of $L = 20$ and SNR $= 10$ dB. Note that $\overline{C}_{\mathrm{LB}} = (L_d/L)C_{\mathrm{LB}}$ where C_{LB} is given in (3.22). The optimal PSAM energy factor α^* is used in all results.

At the critical ρ, this capacity loss is less than 1.5%. We also studied the results for different values of block lengths and the same trend was found (results are omitted for brevity). These results imply that our proposed transmission scheme is very close to optimal provided that the training length is optimized.

Fig. 3.7 shows the average capacity lower bound $\overline{C}_{\mathrm{LB}}$ versus the channel correlation factor ρ for CCF 4×4 systems with a block length of $L = 20$ and SNR of 10 dB. The optimal PSAM energy factor α^* shown in Fig. 3.5 is used in the capacity computation. We also include $\overline{C}_{\mathrm{LB}}$ for $L_p = 5$ in Fig. 3.5 with α^* found numerically. It is clear that $\overline{C}_{\mathrm{LB}}$ for $L_p = 5$ is always smaller than $\overline{C}_{\mathrm{LB}}$ for $L_p^* \leq 4$, which agrees with *Theorem 3.3*. Comparing the capacity with different training lengths, we see that L_p^* decreases as the channel becomes more correlated. More importantly, the capacity gain from optimizing the training length according to the channel spatial correlation can be significant. For example, the capacity at $\rho = 0.5$ using $L_p = 4$ (which is optimal for i.i.d. channels) is approximately 6.3 bits per

channel use, while the capacity at $\rho = 0.5$ using $L_p^* = 2$ is around 7 bits per channel use, that is to say, optimizing training length results in a capacity improvement of 11% at $\rho = 0.5$. Moreover, the capacity improvement increases as channel correlation increases. The same trends are found for different values of block lengths, although the capacity improvement by optimizing the training length reduces as the block length increases (results are omitted for brevity). Therefore, it is important to numerically optimize the training length for correlated channels at small to moderate block lengths.

Furthermore, one can record the range of ρ for each value of L_p^* from Fig. 3.7, and observe the value of α^* in the corresponding range of ρ in Fig. 3.5. It can be seen that within the range of ρ where a given L_p is optimal, the value of α^* for the given L_p is a constant given by *Theorem 3.4* provided that $\mathcal{P}L \gg 1$. That is to say, the condition in *Theorem 3.4* (*i.e.*, $\mathcal{P}L \gg \sum_{i=1}^{L_p} g_i^{-1}$) can be simplified to $\mathcal{P}L \gg 1$ provided that the training length is optimized. To summarize, our numerical results show that optimizing the training length can significantly improve the capacity, and the simple transmission scheme proposed in Section 3.4.1 gives near optimal performance.

3.5 Hybrid Feedback Systems

For systems with correlated channels, one can utilize both CCF and CGF to increase the capacity. We refer to this type of system as the hybrid CCF-CGF system. In this section, we study the optimal transmission and resource allocation strategies in the hybrid system. As we will see, many results and derivations obtained in CCF system and CGF system can be used in the hybrid system. For brevity, we consider delayless feedback, while the impact of feedback delay can be analyzed in a similar way as in Section 3.3.2.

3.5.1 Optimal Transmission Scheme

First we study the optimal spatial structures of both the pilots and data, which maximize the capacity lower bound. At the beginning of each transmission block, the transmitter only knows the statistical properties of the channels. Therefore, the pilot structure design is based on the knowledge of the channel covariance matrix, and hence is identical to that outlined in Section 3.4.1.

During the data transmission, the transmitter has received the estimated channel gains fed back from the receiver. Hence, the data structure design is based

on the knowledge of both the channel covariance matrix and the estimated channel gains. The data transmission strategy characterized by $\boldsymbol{Q}$ shall be chosen to maximize the following objective function for given $\boldsymbol{R}_{\boldsymbol{H}}$ and $\widehat{\boldsymbol{H}}$ as follows.

$$\max_{\boldsymbol{Q}:\,\mathrm{tr}\{\boldsymbol{Q}\}\leq\mathcal{P}_d} \log_2 \left| \boldsymbol{I}_{N_t} + (1+\mathrm{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\boldsymbol{Q}\})^{-1}\widehat{\boldsymbol{H}}^{\dagger}\widehat{\boldsymbol{H}}\boldsymbol{Q} \right|. \tag{3.24}$$

Since $\boldsymbol{Q}$ appears in two different places in (3.24), it is generally difficult to find a closed-form solution to optimal $\boldsymbol{Q}$ and it is also not clear whether $\mathrm{tr}\{\boldsymbol{Q}\} = \mathcal{P}_d$ maximizes the objective function. The constraint set of the above optimization problem is a convex cone and hence, the solution can be found using iterative algorithms such as cone programming [99]. However, in the case where the LMMSE channel estimation and the optimal pilot structure outlined in Section 3.4.1 are used, a closed-form solution for the optimal $\boldsymbol{Q}$ can be found as follows.

Recall that the EVD of the channel covariance matrix is given by $\boldsymbol{R}_{\boldsymbol{H}} = \boldsymbol{U}\boldsymbol{G}\boldsymbol{U}^{\dagger}$. We denote the first n columns of $\widehat{\boldsymbol{H}}\boldsymbol{U}$ as $\boldsymbol{F}$ which is an $N_r \times n$ matrix, and let the EVD of $\boldsymbol{F}^{\dagger}\boldsymbol{F}$ be $\boldsymbol{F}^{\dagger}\boldsymbol{F} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\dagger}$, where the diagonal matrix $\boldsymbol{\Lambda} = \mathrm{diag}\{\chi_1,\ \chi_2,\ \ldots,\ \chi_n\}$. The following theorem gives the optimal data transmission strategy.

Theorem 3.6 The optimal data transmission structure can be written as $\boldsymbol{Q} = \boldsymbol{U}\widehat{\boldsymbol{Q}}\boldsymbol{U}^{\dagger}$ with

$$\widehat{\boldsymbol{Q}} = \begin{pmatrix} \widehat{\boldsymbol{Q}}_a & \boldsymbol{0}_{n\times(N_t-n)} \\ \boldsymbol{0}_{(N_t-n)\times n} & \boldsymbol{0}_{(N_t-n)\times(N_t-n)} \end{pmatrix}, \tag{3.25}$$

where $\widehat{\boldsymbol{Q}}_a$ is an $n \times n$ matrix and $\boldsymbol{0}_{M\times N}$ is an all-zero matrix of size $M \times N$. The optimal $\widehat{\boldsymbol{Q}}_a$ satisfies $\mathrm{tr}\{\widehat{\boldsymbol{Q}}_a\} = \mathcal{P}_d$ and is given by $\widehat{\boldsymbol{Q}}_a = \boldsymbol{V}\boldsymbol{A}\boldsymbol{V}^{\dagger}$, where $\boldsymbol{A} = \mathrm{diag}\{a_1,\ a_2,\ \ldots,\ a_n\}$ with

$$a_i = \left[\eta - \left(\frac{\chi_i}{1+\mu^{-1}\mathcal{P}_d}\right)^{-1}\right]^{+},\ \ \sum_{i=1}^{n} a_i = \mathcal{P}_d, \tag{3.26}$$

and η represents the data water level.

Proof: See Appendix B.6.

It can be seen from (3.25) that the optimal data transmission concentrates the transmit power into the n trained eigen-directions. $\mathrm{tr}\{\widehat{\boldsymbol{Q}}_a\} = \mathrm{tr}\{\boldsymbol{Q}\} = \mathcal{P}_d$ implies that all available power for data transmission should be used to achieve the capacity. As shown in (3.26), the optimal $\widehat{\boldsymbol{Q}}_a$ follows a water-filling solution

according to the estimated channel gains that are rotated by $\boldsymbol{U}$ and truncated into the n trained eigen-directions. These results agree with the intuitions obtained from CGF-only and CCF-only systems which are now rigorously proved in *Theorem 3.6*. Furthermore, *Theorem 3.6* contains an important message: only the channel gains along the trained eigen-directions have effects on the optimal data transmission design. As a result, the receiver should right rotate $\widehat{\boldsymbol{H}}$ by $\boldsymbol{U}$, and only feed back the first n columns of $\widehat{\boldsymbol{H}}\boldsymbol{U}$, *i.e.*, $\boldsymbol{F}$. Note that $n = L_p$ when $L_p < N_t$. This implies that only $N_r \times L_p$ out of $N_r \times N_t$ channel gains need to be fed back when $L_p < N_t$. Therefore, for a fixed training length, the amount of feedback overhead does not increase by adding extra transmit antennas.

With the optimal pilot and data structure, the average capacity lower bound in (3.8) can be rewritten as (derivation is similar to that in Appendix B.6)

$$\overline{C}_{\mathrm{LB}} = \frac{L_d}{L} E_{\boldsymbol{\chi}}\Big\{ \sum_{i=1}^{m} \log_2 \Big(\frac{\mathcal{P}_d}{1+\mu^{-1}\mathcal{P}_d} + \sum_{i=1}^{m} \frac{1}{\chi_i} \Big) + \sum_{i=1}^{m} \log_2 \frac{\chi_i}{m} \Big\}, \qquad (3.27)$$

where m denotes the number of non-zero a_i in (3.26).

3.5.2 Optimal Training Resource Allocation

Using the optimal transmission schemes in Section 3.5.1, we now study the optimal training length and the optimal PSAM energy factor, denoted by L_p^* and α^*, respectively, which maximize the average capacity lower bound in (3.27).

First of all, we present results for L_p^* in the following theorem and corollary.

Theorem 3.7 For hybrid CCF-CGF systems with the optimal transmission strategy described in Section 3.5.1, the optimal training length is at most as large as the number of transmit antennas for any given value of the PSAM energy factor α, i.e., $L_p^* \leq N_t$.

Proof: We first consider the range of $L_p \geq N_t$ and show that L_p should be kept to a minimum for any given α. The proof of this part follows from the proof of *Theorem 3.2*. This result implies that the optimal value of L_p cannot be greater than N_t in general. □

Corollary 3.4 For hybrid CCF-CGF systems with the optimal transmission strategy described in Section 3.5.1 as well as the optimal PSAM energy factor α^*, the optimal training length is given by $L_p^* \leq N_t$.

Proof: The proof is similar to that of *Corollary 3.1*. □

In general, the optimal training length needs to be found numerically when $L_p^* < N_t$. Nevertheless, *Theorem 3.7* and *Corollary 3.4* suggest that the computational complexity of numerically optimizing L_p is low due to the fact that L_p only takes integer values ranging from 1 to N_t. Therefore, we only consider $L_p \leq N_t$ for investigating the optimal PSAM energy factor α^*.

Generally it is difficult to find a closed-form solution for α^* due to the water-filling solution of $\widehat{\boldsymbol{Q}}_a$, although numerical methods can be applied [99]. Here, we look for a near optimal closed-form solution, that is to say, a solution with near optimal performance as well as minimal computational complexity. To this end, we consider $\widehat{\boldsymbol{Q}}_a$ to follow an equal power allocation given by $\widehat{\boldsymbol{Q}}_a = (\mathcal{P}_d/n)\boldsymbol{I}_n$ instead of the water-filling solution given in (3.26). Indeed, the equal power allocation for data transmission was also considered in Section 3.4.1 for CCF-only systems. Therefore, the optimal value of α for $\widehat{\boldsymbol{Q}}_a = (\mathcal{P}_d/n)\boldsymbol{I}_n$ is given in *Theorem 3.4*. In Section 3.5.3, we numerically investigate the optimality of α given in *Theorem 3.4* for hybrid CCF-CGF systems with data transmission following the water-filling solution.

3.5.3 Numerical Results

In the following, we numerically investigate the optimality of α given in *Theorem 3.4* for hybrid CCF-CGF systems with the optimal data transmission strategy given in *Theorem 3.6*. For numerical analysis, we choose the channel covariance matrix to be in the form of $[\boldsymbol{R}_{\boldsymbol{H}}]_{ij} = \rho^{|i-j|}$, where ρ is referred to as the spatial correlation factor.

Fig. 3.8 and Fig. 3.9 show α^* found numerically for 4×4 systems with $\mathcal{P} = 10$ dB and for $L = 20$ and $L = 50$, respectively. We see that the values of α^* shown as solid lines in both figures are very close to (*e.g.*, with 1%-3% error) the analytical solutions given in *Theorem 3.4* shown as the dashed lines for a wide range of spatial correlations, *e.g.*, $\rho < 0.9$. Therefore, we can conclude that α given in *Theorem 3.4* is a near optimal solution for hybrid CCF-CGF systems.

In practical scenarios such as the downlink of cellular systems, the cost of adding extra antennas at the base station is minimal compared with the capacity gain for every downlink transmissions. However, one may argue that adding extra transmit antennas increases the number of channels to be estimated, which results in an increase in the required training resources and feedback overhead for systems with CGF.

On the other hand, the spatial correlations between the transmit antennas usually increases as the number of antennas increases due to space constraints (at the

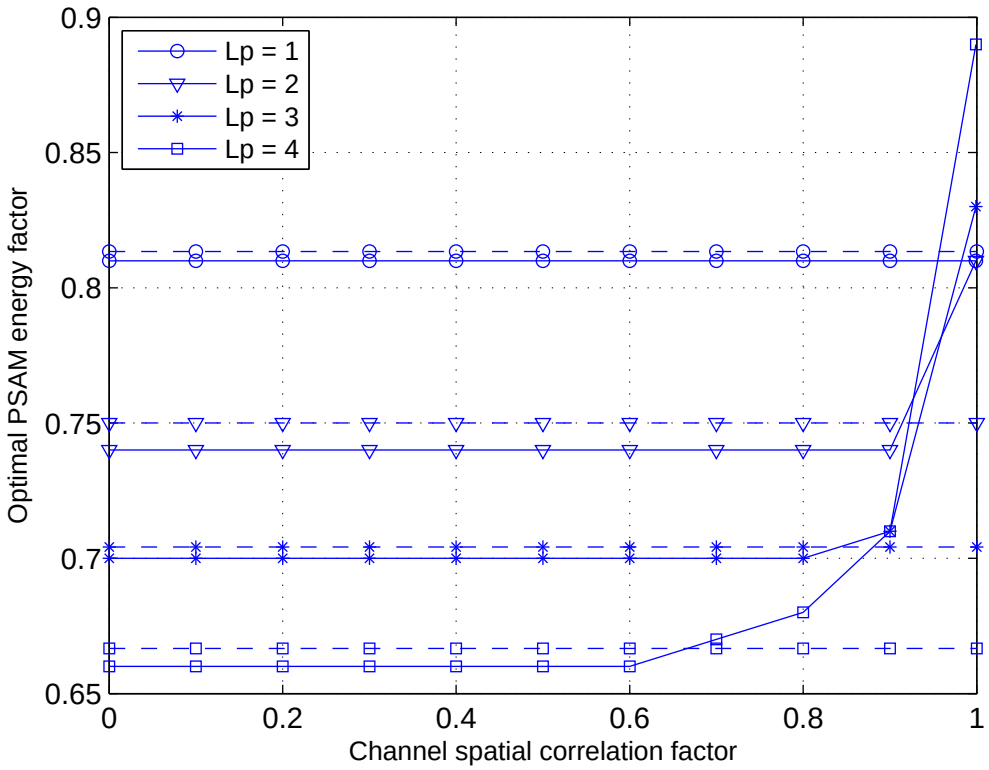

Figure 3.8: Optimal PSAM energy factor α^* vs. channel spatial correlation factor ρ for 4×4 systems with a block length of $L = 20$ and SNR at 10 dB. All values of α^* are found numerically. For comparison, the dashed lines indicate the analytical values of α given in *Theorem 3.4*.

base station). The increase in the channel spatial correlations makes the dominant eigen-directions stronger and the weak eigen-directions even weaker. Therefore, it is possible to design a system with a fixed training length, which only requires training a fixed number of strongest eigen-directions while still keeping the effect of channel estimation errors to a minimum. Note that for a fixed training length, the (near optimal) training energy is also fixed according to *Theorem 3.4*. From the discussion of *Theorem 3.6*, we also know that the amount of feedback transmission does not increase with the number of transmit antennas for a fixed training length. Hence, we numerically investigate the capacity gain from adding extra transmit antennas with fixed training resources and feedback overhead.

Fig. 3.10 shows the average capacity lower bound $\overline{C}_{\mathrm{LB}}$ versus the number of transmit antennas N_t for systems with $N_r = 4$ and $L_p = 1$, 2, 3 and 4. The transmit antennas are placed in a uniform circular array (UCA) with a fixed radius equal to half of the wavelength. The spatial correlation coefficients between the transmit antennas are calculated using the standard Jakes' model [101]. For each L_p, we see that the capacity lower bound increases as N_t increases from 4 to 20

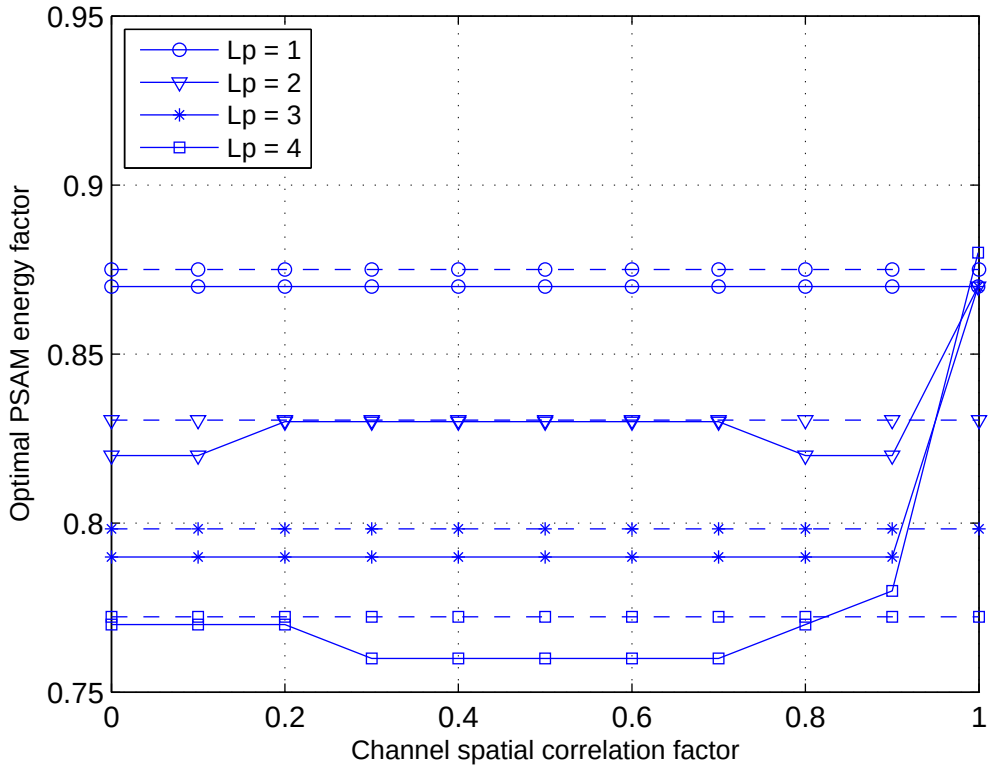

Figure 3.9: Optimal PSAM energy factor α^* vs. channel spatial correlation factor ρ for 4×4 systems with a block length of $L = 50$ and SNR at 10 dB. All values of α^* are found numerically. For comparison, the dashed lines indicate the analytical values of α given in *Theorem 3.4*.

(except for the case where $L_p = 1$ and $N_t = 6$).[4] For example, the capacity gain in using $N_t = 10$ compared to $N_t = 4$ is 24%. This gain increases to 55% when $N_t = 20$ is used. We also include the capacity lower bound for L_p^* which is found numerically, and show the value of L_p^* for $N_t = 6$, 10 and 14 in Fig. 3.10. It is clear that the optimal training length can be significantly smaller than the number of transmit antennas for correlated MIMO systems, which agrees with *Corollary 3.4*.

3.6 Summary of Contributions

In this chapter we have investigated the optimal transmission and resource allocation in training-based MIMO systems with various forms of transmitter CSI. The important design parameters are the pilot and data structures for transmission, as well as the transmit energy and time to be distributed between training and data

[4]This is caused by the decrease in the largest eigenvalue of $\boldsymbol{R_H}$ from $N_t = 4$ to $N_t = 6$ due to the UCA geometry and Jakes' model.

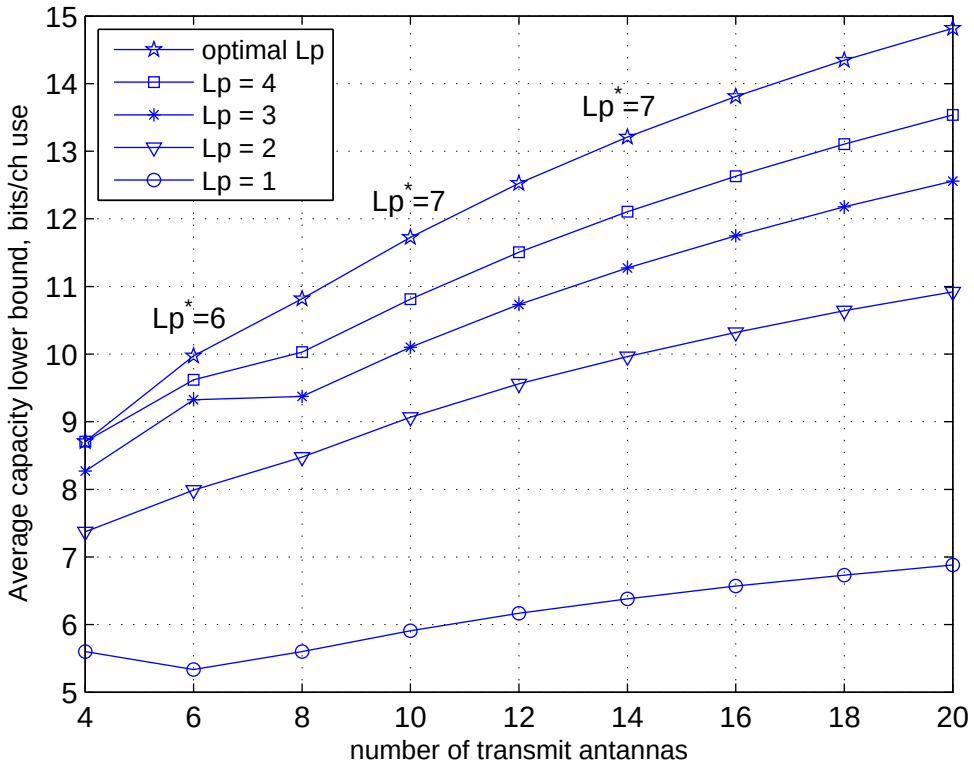

Figure 3.10: Average capacity lower bound $\overline{C}_{\mathrm{LB}}$ in (3.27) vs. number of transmit antennas N_t. The transmit antennas are placed in a UCA with radius equal to half of the wavelength. The block length is $L = 50$, SNR is 10 dB and the number of receive antennas is 4. The value of α in *Theorem 3.4* is used as a near optimal choice.

transmission. Some specific contributions made in this chapter are as follows.

Addressing Q1 in Section 3.1:

- For delayless CGF systems with i.i.d. channels, we have proved that the solutions to the optimal energy allocation to pilot and data transmission as well as the optimal training length coincide with the solutions for non-feedback systems given in Chapter 2.

- For delayed CGF systems with i.i.d. channels, our numerical results have shown that evenly distributing the power over the entire data transmission (regardless of the delay time) gives near optimal performance at practical SNR values. As a result, the solutions to the optimal energy allocation to pilot and data transmission, as well as the optimal training length for the delayless system stay nearly optimal for the delayed CGF system regardless

of the delay time.

Addressing Q2 in Section 3.1:

- For CCF systems with correlated channels, we have proposed a simple transmission scheme, taking into account the fact that the optimal training length L_p is at most as large as the number of transmit antennas. This scheme only requires numerical optimization of L_p and does not require numerical optimization of spatial or temporal power allocation over pilot and data transmission. Our numerical results have shown that this scheme is very close to optimal. In addition, optimizing L_p can result in significant capacity improvements for correlated channels.

- Using the proposed scheme for CCF systems, we have found the solution to the optimal energy allocation to pilot and data transmission, which does not depend on the channel spatial correlation under a mild condition on block length or SNR. Therefore, the proposed transmission and resource allocation schemes for CCF systems give near optimal performance while having very low computational complexity.

Addressing Q3 in Section 3.1:

- We have proved that the optimal data transmission for hybrid CCF-CGF systems is given by a water-filling solution according to the estimated channel gains, rotated and truncated into the trained eigen-directions.

- We have shown that the optimal training length for hybrid CCF-CGF systems is at most as large as the number of transmit antennas. We have considered a closed-form solution of PSAM energy factor and have numerically shown that this solution achieves near optimal performance for hybrid CCF-CGF systems.

- We have also shown that capacity can be significantly increased by adding extra transmit antennas without increasing the training or feedback overhead.

Chapter 4

Two-Way Training-Based Systems

4.1 Introduction

The availability of transmitter CSI can result in a significant increase in the data throughput in wireless communication systems. As discussed in Chapter 3, the transmitter CSI is usually obtained using various feedback transmission schemes [98]. Recently, a multi-stage training method was proposed to allow the transmitter to estimate its outgoing channel gains using pilot transmissions from both the transmitter and the receiver without using feedback [102]. The transmitter CSI can also be obtained using a bent-pipe mechanism, which requires the receiver to send back a portion of its received data to facilitate channel estimation at the transmitter [103]. These methods are designed for asymmetric channels where the outgoing and incoming channels have different characteristics.

When the channels are symmetric, such as in time-division duplex (TDD) systems, a simpler training method named two-way training was proposed in [49] for block-fading single-input multiple-output (SIMO) systems. In this scheme, the transmitter acquires the outgoing CSI using the pilots sent from the receiver (*i.e.*, reverse training) and performs block-wise power adaptation. After that, the receiver estimates the effective channel gains using the pilots sent from the transmitter (*i.e.*, forward training). The two-way training scheme was also considered in multi-user transmissions in [104, 105].

For MISO systems, the major benefit of two-way training is the reduction in the overhead of acquiring CSI at both the transmitter and the receiver. To further improve the performance in resource-constrained systems, it is crucial to optimize the transmit power for both pilot and data symbols. This issue is addressed in this chapter. We study block-fading MISO TDD communications systems with two-way training and answer the following questions:

Q1. How are the reverse training and the forward training performed for MISO TDD communication systems?

Q2. How should the system designer set the transmit power in various transmission phases in order to optimize the system performance?

Q3. Under what conditions does the use of two-way training become inefficient or redundant?

In two-way training-based systems, the important transmission phases are reverse training, forward training and data transmission. In Section 4.3, we derive the optimal *linear* channel estimator for the forward training. When investigating the optimal transmit power in different transmission phases, we consider three scenarios and obtain closed-form solutions at high SNR in Section 4.4. The SER and ergodic capacity performance of the derived power allocation strategies are shown in Section 4.6 and Section 4.7, respectively.

4.2 System Model

We consider a flat-fading wireless communication system with N_t transmit antennas and a single receive antenna. For simplicity, we refer to the transmitter as the base station (BS) and the receiver as the user terminal (UT). The received signal at the UT is given by

$$y = \boldsymbol{h}\boldsymbol{x} + n, \tag{4.1}$$

where $\boldsymbol{x}$ is the $N_t \times 1$ transmitted symbol vector from the BS, $\boldsymbol{h}$ is the $1 \times N_t$ channel gain vector and n is the noise at the UT. We assume that both n and the elements of $\boldsymbol{h}$ are i.i.d. ZMCSCG with unit variance. We also assume that the forward and reverse channels are symmetric, *i.e.*, channel reciprocity holds. The symmetric channel assumption can be justified for TDD channels in which the forward and reverse transmissions share a common frequency band, *e.g.*, in IEEE 802.11 standards [49].

4.2.1 Two-way Training-based Transmission

Similar to [104], we consider a four-stage TDD transmission scheme shown in Fig. 4.1. The total duration of the four-stage transmission of L symbols is less than the channel coherence time. Hence, we assume that the channel gains remain constant over L symbol periods.

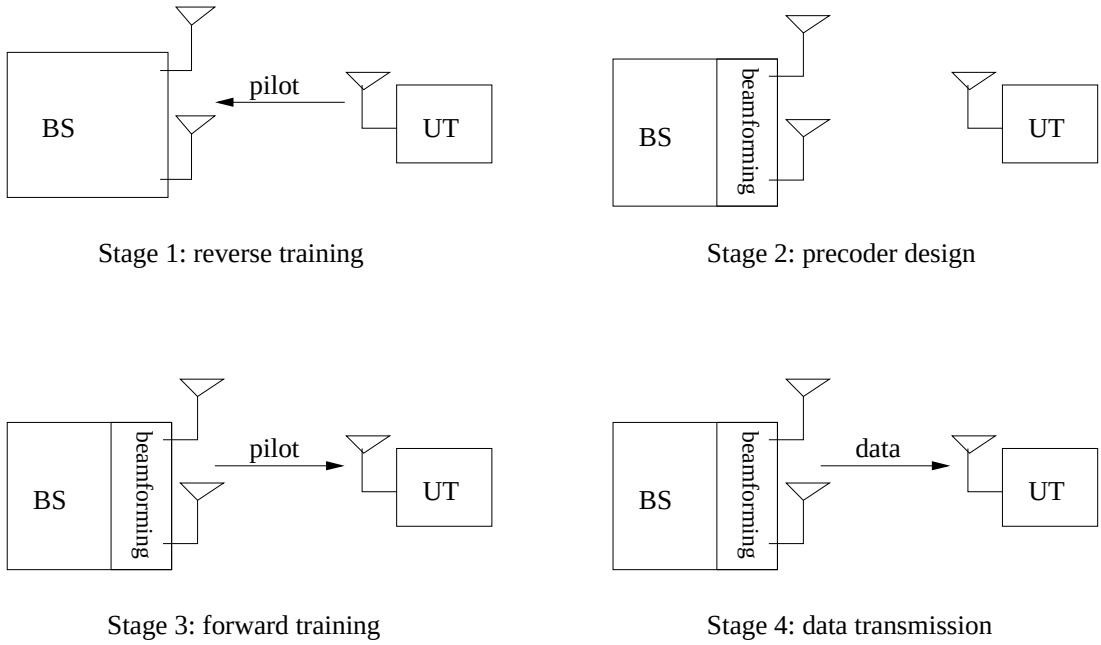

Figure 4.1: An illustration of the four-stage TDD tranmission scheme with two-way training.

- *Stage 1 (reverse training)*: The UT sends one pilot with power $\mathcal{P}_r$ and the BS estimates the channel using LMMSE estimator. Using channel reciprocity, the BS obtains the estimates of its outgoing channel gains. We denote the channel estimates and the estimation errors by $\widehat{\boldsymbol{h}}$ and $\widetilde{\boldsymbol{h}}$, respectively, and $\boldsymbol{h} = \widehat{\boldsymbol{h}} + \widetilde{\boldsymbol{h}}$. The variance of each element in $\widetilde{\boldsymbol{h}}$ is given by $\sigma^2_{\widetilde{\boldsymbol{h}}} = \frac{1}{\mathcal{P}_r+1}$. The duration of the reverse training stage is one symbol period.

- *Stage 2 (precoder design)*: The BS treats $\widehat{\boldsymbol{h}}$ as the true channel and designs the beamforming vector $\boldsymbol{w}$ as $\boldsymbol{w} = \frac{\widehat{\boldsymbol{h}}^\dagger}{\|\widehat{\boldsymbol{h}}\|} = \frac{\widehat{\boldsymbol{h}}^\dagger}{\sqrt{\widehat{\boldsymbol{h}}\widehat{\boldsymbol{h}}^\dagger}}$ to maximize the SNR at the UT. For simplicity, we assume that the duration of the second stage is negligible.

- *Stage 3 (forward training)*: The BS sends one pilot denoted by p with power $|p|^2 = pp^* = \mathcal{P}_f$ via the beamforming vector $\boldsymbol{w}$. The UT estimates the effective channel using LMMSE estimator, which will be discussed in detail in Section 4.3. The duration of the forward training stage is one symbol period.

- *Stage 4 (data transmission)*: The BS transmits data symbols s_i, $i = 1, 2, ..., L-2$ with power $E\{s_i s_i^*\} = \mathcal{P}_d$ via the beamforming vector $\boldsymbol{w}$. The duration of the data transmission stage is $L - 2$ symbol periods.

Since the noise and channel variances are normalized to unity, we also refer to $\mathcal{P}_r$, $\mathcal{P}_f$ and $\mathcal{P}_d$ as the reverse training SNR, forward training SNR and data SNR, respectively.

4.3 Forward Channel Estimation

During the forward training stage, the received signal at the UT is given by

$$\begin{aligned} y &= \boldsymbol{h}\boldsymbol{w}p + n \\ &= (\|\widehat{\boldsymbol{h}}\| + \widetilde{\boldsymbol{h}}\boldsymbol{w})p + n \\ &= fp + n, \end{aligned} \tag{4.2}$$

where $f = \|\widehat{\boldsymbol{h}}\| + \widetilde{\boldsymbol{h}}\boldsymbol{w}$ denotes the effective channel for the forward transmission. Unlike the reverse training where the Gaussian channel makes the LMMSE estimator equivalent to the MMSE estimator, the complicated non-Gaussian distribution of f makes MMSE estimator in the forward training mathematically intractable. Therefore, we consider the widely-used LMMSE channel estimation for the forward training. Denoting the first and second order statistics of f by $\mu_1 = E\{f\}$ and $\mu_2 = E\{|f|^2\}$, the LMMSE channel estimation is given by [94]

$$\widehat{f} = \mu_1 + \frac{\sigma_{fy}^2}{\sigma_{yy}^2}(y - \mu_1 p), \tag{4.3}$$

where

$$\sigma_{fy}^2 = E\{fy^*\} - \mu_1^2 p^* = (\mu_2 - \mu_1^2)p^*,$$

and

$$\sigma_{yy}^2 = E\{yy^*\} - \mu_1^2 \mathcal{P}_f = (\mu_2 - \mu_1^2)\mathcal{P}_f + 1.$$

To find the values of μ_2 and μ_1, we proceed as follows. Since the elements of $\boldsymbol{h}$ are i.i.d. Gaussian random variables, its LMMSE estimate $\widehat{\boldsymbol{h}}$ also has i.i.d. Gaussian elements, and hence $\|\widehat{\boldsymbol{h}}\|^2$ has a Gamma distribution with parameters $(N_t, 1 - \sigma_{\widetilde{\boldsymbol{h}}}^2)$. Therefore, we have

$$\begin{aligned} \mu_2 &= E\{\|\widehat{\boldsymbol{h}}\|^2\} + E\{\widetilde{\boldsymbol{h}}\boldsymbol{w}\boldsymbol{w}^\dagger\widetilde{\boldsymbol{h}}^\dagger\} \\ &= N_t(1 - \sigma_{\widetilde{\boldsymbol{h}}}^2) + \sigma_{\widetilde{\boldsymbol{h}}}^2 \end{aligned}$$

$$= N_t \frac{\mathcal{P}_r}{\mathcal{P}_r + 1} + \frac{1}{\mathcal{P}_r + 1}. \tag{4.4}$$

Furthermore, μ_1 can be calculated using the probability density function of $g = \|\widehat{\boldsymbol{h}}\|^2$ as

$$\begin{aligned} \mu_1 &= E\{g^{1/2}\} \\ &= \int_0^\infty g^{1/2} g^{N_t - 1} \frac{e^{-g/(1-\sigma_{\widetilde{\boldsymbol{h}}}^2)}}{(1-\sigma_{\widetilde{\boldsymbol{h}}}^2)^{N_t}\Gamma(N_t)} \mathrm{d}g \\ &= \frac{1}{\sqrt{(1-\sigma_{\widetilde{\boldsymbol{h}}}^2)}\Gamma(N_t)} \int_0^\infty \Big(\frac{g}{1-\sigma_{\widetilde{\boldsymbol{h}}}^2}\Big)^{N_t - 1/2} e^{-g/(1-\sigma_{\widetilde{\boldsymbol{h}}}^2)} \mathrm{d}g, \end{aligned} \tag{4.5}$$

where $\Gamma(z) = \int_0^\infty t^{-1+z} e^{-t} \mathrm{d}t$ is the Gamma function. Letting $t = g/(1-\sigma_{\widetilde{\boldsymbol{h}}}^2)$, (4.5) reduces to

$$\begin{aligned} \mu_1 &= \frac{\sqrt{(1-\sigma_{\widetilde{\boldsymbol{h}}}^2)}}{\Gamma(N_t)} \int_0^\infty t^{N_t - 1/2} e^{-t} \mathrm{d}t \\ &= \sqrt{\frac{\mathcal{P}_r}{\mathcal{P}_r + 1}} \frac{\Gamma(N_t + 1/2)}{\Gamma(N_t)}. \end{aligned} \tag{4.6}$$

Denoting the variance of f as $\sigma_f^2 = \mu_2 - \mu_1^2$, the variance of the channel estimation error $\widetilde{f} = f - \widehat{f}$ is given by

$$\sigma_{\widetilde{f}}^2 = E\{|\widetilde{f}|^2\} = \frac{\sigma_f^2}{\sigma_f^2 \mathcal{P}_f + 1}, \tag{4.7}$$

$$\text{where } \sigma_f^2 = \frac{k\mathcal{P}_r + 1}{\mathcal{P}_r + 1} \text{ and } k = N_t - \left(\frac{\Gamma(N_t + 1/2)}{\Gamma(N_t)}\right)^2.$$

We see that k characterizes the effect of N_t on $\sigma_{\widetilde{f}}^2$. By evaluating k for different N_t, we see that $0.232 < k < 0.250$ for $1 < N_t < 100$. Therefore, the value of N_t has little impact on $\sigma_{\widetilde{f}}^2$, which implies that adding or removing antennas at the BS does not considerably change the forward channel estimation error.

4.4 Power Allocation with Two-way Training

In this section, we study the optimal power allocation for two-way training-based transmission. The problem of optimizing power allocation can be formulated in different ways according to the given power constraints and the degrees of freedom in the system design. We provide a comprehensive study by solving the power optimization problem in three different scenarios.

The received signal at the UT during data transmission is given as

$$y = \widehat{f}s + \widetilde{f}s + n. \tag{4.8}$$

Hence, the received SNR for an instantaneous channel realization is given by

$$\rho_{\text{inst}} = \frac{\mathcal{P}_d|\widehat{f}|^2}{1 + \mathcal{P}_d|\widetilde{f}|^2}. \tag{4.9}$$

Due to the complicated nature of the distribution of ρ_{inst}, closed-form expressions for long-term system performance measures, such as information capacity and symbol error rate (SER), are generally very difficult to obtain, which makes the problem of power optimization mathematically intractable. Instead of directly dealing with ρ_{inst}, we define a measure of the average received SNR as

$$\rho_{\text{ave}} = \frac{\mathcal{P}_d E\{|\widehat{f}|^2\}}{1 + \mathcal{P}_d E\{|\widetilde{f}|^2\}}. \tag{4.10}$$

Using $E\{|\widehat{f}|^2\} > E\{\widehat{f}\}E\{\widehat{f}^*\}$, we obtain a lower bound on ρ_{ave} as

$$\begin{aligned} \rho_{\text{ave}}^{\text{LB}} &= \frac{\mathcal{P}_d \mu_1^2}{1 + \mathcal{P}_d \sigma_{\widetilde{f}}^2} \\ &= \nu \frac{\mathcal{P}_d \mathcal{P}_r[(k\mathcal{P}_r + 1)\mathcal{P}_f + \mathcal{P}_r + 1]}{(\mathcal{P}_r + 1)[(k\mathcal{P}_r + 1)(\mathcal{P}_f + \mathcal{P}_d) + \mathcal{P}_r + 1]}, \end{aligned} \tag{4.11}$$

which is obtained using (4.6) and (4.7), and $\nu = \left(\frac{\Gamma(N_t+1/2)}{\Gamma(N_t)}\right)^2$. We simply refer to $\rho_{\text{ave}}^{\text{LB}}$ in (4.11) as the *average SNR lower bound* and propose to use it as the objective function to obtain solutions for power optimization. Since the value of k is almost independent of the number of transmit antennas N_t, the solution for power optimization, using $\rho_{\text{ave}}^{\text{LB}}$ as the objective function, is almost independent of N_t as well. In the following, we formulate the power optimization problem in three different scenarios and derive analytical solutions using high SNR approximations. We will investigate the optimality of these solutions for SER and capacity performance in Section 4.6 and Section 4.7, respectively.

4.4.1 Optimizing Reverse and Forward Training

In the first scenario, we study the optimal power allocation between reverse and forward training for a given average training power budget. The data transmission is assumed to have a fixed SNR, *i.e.*, $\mathcal{P}_d$ is fixed. This study allows us to investigate

the relative importance of reverse training and forward training. Note that the power optimization does not depend on the block length L. We denote the average training power or SNR as $\mathcal{P}_{rf}$ and the ratio of power (or indeed the ratio of total training energy) allocated to forward training as α. Then, the power constraint can be written as

$$2\mathcal{P}_{rf} = \mathcal{P}_r + \mathcal{P}_f. \tag{4.12}$$

Therefore, we have the following relationships.

$$\mathcal{P}_f = \alpha 2\mathcal{P}_{rf} \qquad \text{and} \qquad \mathcal{P}_r = (1-\alpha)2\mathcal{P}_{rf}. \tag{4.13}$$

In the high SNR regime for training, we assume that $\mathcal{P}_r \gg 1$ and $\mathcal{P}_f \gg 1$ (which implies $k\mathcal{P}_r\mathcal{P}_f \gg \mathcal{P}_r$). We also assume that $k\mathcal{P}_r\mathcal{P}_f \gg \mathcal{P}_d$ which is valid when either $\mathcal{P}_r$ or $\mathcal{P}_f$ is much higher than $\mathcal{P}_d$. Therefore, the average SNR lower bound $\rho_{\text{ave}}^{\text{LB}}$ in (4.11) can be approximated as

$$\begin{aligned} \rho_{\text{ave}}^{\text{LB}} &\approx \nu\mathcal{P}_d \frac{\mathcal{P}_r}{\mathcal{P}_r+1}\frac{\mathcal{P}_f}{\mathcal{P}_f+\mathcal{P}_d} \\ &\approx \frac{2\nu\mathcal{P}_d\mathcal{P}_{rf}\alpha(1-\alpha)}{2\alpha(1-\alpha)\mathcal{P}_{rf}+(1-\alpha)\mathcal{P}_d+\alpha}. \end{aligned} \tag{4.14}$$

Letting the first derivative of $\rho_{\text{ave}}^{\text{LB}}$ in (4.14) w.r.t. α be zero, one can solve for the optimal α as

$$\alpha = \begin{cases} \frac{1}{2}, & \text{for } \mathcal{P}_d = 1 \\ \frac{\mathcal{P}_d - \sqrt{\mathcal{P}_d}}{\mathcal{P}_d - 1}, & \text{for } \mathcal{P}_d \neq 1 \end{cases} \tag{4.15}$$

The solution given in (4.15) is obtained using the high SNR approximation. We see that the optimal power allocation between the reverse and forward training strongly depends on the data transmit power.

4.4.2 Optimizing Forward Transmission

In the second scenario, the BS tries to optimize the power allocation between forward training and data transmission for a given average forward transmit power budget. The reverse training is assumed to have a fixed SNR, *i.e.*, $\mathcal{P}_r$ is fixed. We denote the average power or SNR for the forward link as $\mathcal{P}_{fd}$ and the ratio of power (or indeed the ratio of total forward transmit energy) allocated to data

transmission as β. Then, the power constraint can be written as

$$\mathcal{P}_{fd}(L-1) = \mathcal{P}_f + \mathcal{P}_d(L-2). \tag{4.16}$$

Therefore, we have the following relationships.

$$\mathcal{P}_d = \beta \mathcal{P}_{fd} \frac{L-1}{L-2}, \;\; \mathcal{P}_f = (1-\beta)\mathcal{P}_{fd}(L-1). \tag{4.17}$$

In the high SNR regime for forward transmission, we assume that $\mathcal{P}_d \gg 1$ and $\mathcal{P}_f \gg 1$. Therefore, the average SNR lower bound $\rho_{\text{ave}}^{\text{LB}}$ in (4.11) can be approximated as

$$\begin{aligned} \rho_{\text{ave}}^{\text{LB}} &\approx \frac{\nu \mathcal{P}_r}{\mathcal{P}_r + 1} \frac{\mathcal{P}_d \mathcal{P}_f}{\mathcal{P}_d + \mathcal{P}_f} \\ &= \frac{\nu \mathcal{P}_r \mathcal{P}_{fd}}{\mathcal{P}_r + 1} \frac{(L-1)\beta(1-\beta)}{\beta + (L-2)(1-\beta)}. \end{aligned} \tag{4.18}$$

Letting the first derivative of $\rho_{\text{ave}}^{\text{LB}}$ in (4.18) w.r.t. β be zero, one can solve for the optimal β as

$$\beta = \phi - \sqrt{\phi(\phi-1)}, \;\; \text{where} \;\; \phi = \frac{L-2}{L-3}. \tag{4.19}$$

It is clear that the optimal forward power allocation at high SNR given in (4.19) is independent of the reverse training SNR. This result implies that the power optimization in the forward link is independent of the reverse link conditions, which is an important message for system designers.

4.4.3 Optimizing Overall Transmission

In the third scenario, the system designer has the most degrees of freedom and tries to optimize the power allocation between reverse training, forward training and data transmission under an overall transmit power constraint. We denote the average power or SNR for overall (reverse and forward) transmission as $\mathcal{P}$, and the ratio of power (or indeed the ratio of total energy) allocated to forward transmission as γ. We also use β as defined in the second scenario. The power constraint can be written as

$$\mathcal{P}L = \mathcal{P}_r + \mathcal{P}_f + \mathcal{P}_d(L-2). \tag{4.20}$$

Therefore, we have the following relationships.

$$\begin{aligned} \mathcal{P}_r &= (1-\gamma)\mathcal{P}L, \\ \mathcal{P}_d &= \beta\gamma\mathcal{P}L/(L-2), \\ \mathcal{P}_f &= (1-\beta)\gamma\mathcal{P}L. \end{aligned} \tag{4.21}$$

In the high SNR regime, we apply all the assumptions stated in the previous two scenarios, and hence the optimal power allocation satisfies the relationships given in (4.15) and (4.19). Using (4.13) and (4.15), we have[1] $\mathcal{P}_r = \frac{\mathcal{P}_f}{\sqrt{\mathcal{P}_d}}$. And using (4.17), we have $\mathcal{P}_f = \frac{1-\beta}{\beta}\mathcal{P}_d(L-2)$. Therefore, we obtain

$$\begin{aligned} \mathcal{P}L &= \mathcal{P}_r + \mathcal{P}_f + \mathcal{P}_d(L-2) \\ &= \frac{1-\beta}{\beta}\sqrt{\mathcal{P}_d}(L-2) + \frac{1}{\beta}\mathcal{P}_d(L-2), \end{aligned} \tag{4.22}$$

from which one can easily solve for $\mathcal{P}_d$. Then using the relationship between $\mathcal{P}_d$ and γ in (4.21), the optimal power allocation strategies at high SNR can be obtained as

$$\gamma = \frac{L-2}{2\mathcal{P}L\beta}\left[(1-\beta)^2 + \frac{2\mathcal{P}L\beta}{L-2} - (1-\beta)\sqrt{(1-\beta)^2 + \frac{4\mathcal{P}L\beta}{L-2}}\right], \tag{4.23}$$

where β is given in (4.19). Similar to the previous two scenarios, the optimal power allocation at high SNR given in (4.23) is independent of the number of transmit antennas.

4.5 Power Allocation with Reverse Training Only

The systems with only reverse training can be regarded as a special case of two-way training based systems. In this section, we provide an analytical solution for the optimal power allocation between the reverse training and the data transmission under an overall transmit power budget. Since forward training is not used, the duration of data transmission becomes $L-1$. Similar to Section 4.4.3, we denote the average power or SNR for both links as $\mathcal{P}$ and the ratio of power allocated to the forward (data) transmission as γ. Therefore, we have the following relationships.

$$\mathcal{P}_d = \gamma\mathcal{P}L/(L-1) \qquad \text{and} \qquad \mathcal{P}_r = (1-\gamma)\mathcal{P}L. \tag{4.24}$$

[1] Here we omit the result for the special case of $\mathcal{P}_d = 1$ for brevity.

With reverse training only, the UT does not know the effective channel. However, the UT can accurately obtain the mean value of the effective channel μ_1, since it is a long-term statistic which changes much more slowly than the channel gain [106, 107]. Therefore, the average received SNR at the UT is given by

$$\begin{aligned}\rho_{\text{ave}} &= \frac{\mathcal{P}_d \mu_1^2}{1+\mathcal{P}_d \sigma_f^2} \\ &= \frac{\nu \mathcal{P}_d \mathcal{P}_r}{k\mathcal{P}_d\mathcal{P}_r + \mathcal{P}_d + \mathcal{P}_r + 1} \qquad (4.25) \\ &= \frac{\nu \mathcal{P}^2 L^2 \gamma(1-\gamma)}{k\mathcal{P}^2L^2\gamma(1-\gamma)+\mathcal{P}L\gamma+\mathcal{P}L(L-1)(1-\gamma)+L-1}, \qquad (4.26)\end{aligned}$$

where (4.25) is obtained using (4.6) and (4.7), and (4.26) is obtained using (4.24). Letting the first derivative of ρ_{ave} in (4.26) w.r.t. γ be zero, one can solve for the optimal γ as

$$\gamma = \theta - \sqrt{\theta(\theta-1)}, \;\; \text{where } \theta = \frac{(\mathcal{P}L+1)(L-1)}{\mathcal{P}L(L-2)}. \qquad (4.27)$$

It is clear that the optimal power allocation given in (4.27) is independent of the number of transmit antennas, which is an important message for system designers.

4.6 Symbol Error Rate Performance

In this section, we obtain numerical results on SER to investigate the optimality of the proposed power allocation strategies in Section 4.4. The results on ergodic capacity will be presented in the next section. We carry out Monte-Carlo simulation of a two-way training based communication according to Section 4.2.1 with a simulation length of 5×10^6 symbols. The receiver performs minimum distance detection on each received symbol. The SER is then computed as the ratio of the number of incorrectly detected symbols and the total number of transmitted symbols. The optimal ratios of power allocation, *i.e.*, α, β and γ, which minimize the SER are found from a linear search between 0 and 1 with a step size of 0.01. We will mainly use 16-QAM modulation which is a widely-considered constellation [1]. In the following, we obtain the SER performance in the three scenarios considered in Section 4.4.

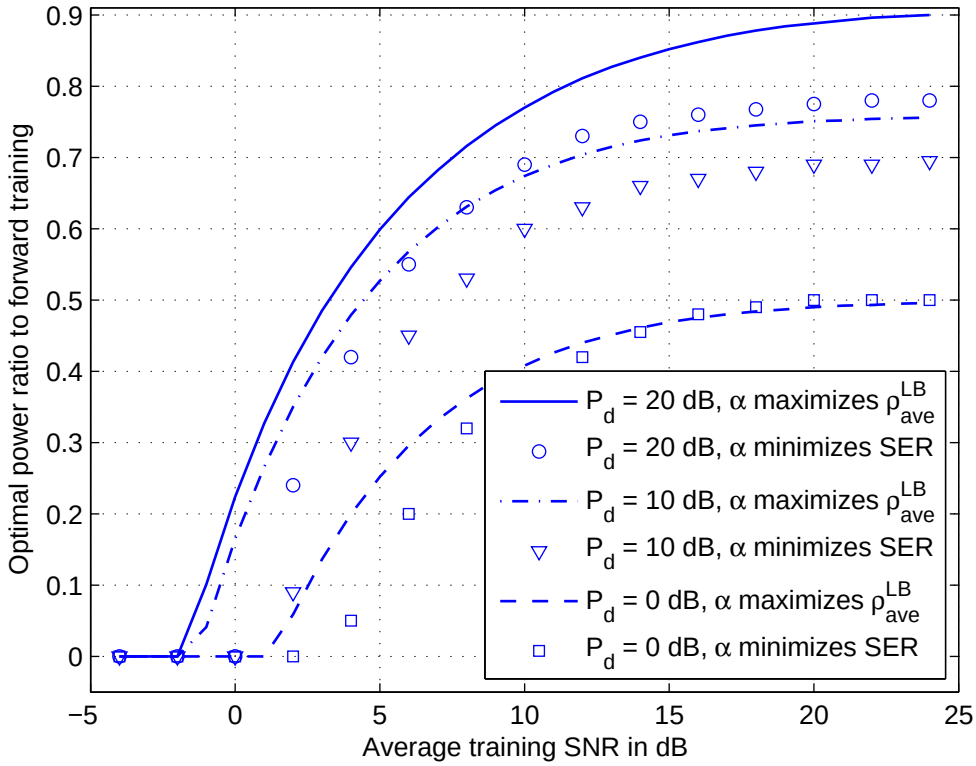

Figure 4.2: Optimal power ratio to forward training α vs. average training SNR $\mathcal{P}_{rf}$ for systems with $N_t = 4$ transmit antennas and different values of data SNR $\mathcal{P}_d$. Lines indicate the values of α that maximize the average SNR lower bound $\rho_{\text{ave}}^{\text{LB}}$ in (4.11) and the markers indicate the values of α that minimize the SER for 16-QAM modulation found via Monte-Carlo simulations.

4.6.1 Optimal Power Allocation for Reverse and Forward Training

Fig. 4.2 shows the optimal power ratio to forward training α versus average training SNR $\mathcal{P}_{rf}$ for different data SNR $\mathcal{P}_d$. We see that the values of α which minimize the SER for 16-QAM modulation follow the same trend as the values of α which maximize $\rho_{\text{ave}}^{\text{LB}}$ in (4.11), and the mismatch increases as $\mathcal{P}_d$ increases. This mismatch is mainly due to the fact that maximizing $\rho_{\text{ave}}^{\text{LB}}$ does not necessarily result in the optimal distribution of ρ_{inst} which minimizes the SER, *e.g.*, it does not minimize the probability of ρ_{inst} taking very small values. Furthermore, we see that more power should be allocated to reverse (forward) training when $\mathcal{P}_{rf}$ is low (high). In particular, we see that forward training should not be used when $\mathcal{P}_{rf}$ is sufficiently low, *e.g.*, $\mathcal{P}_{rf} < 0$ dB. As $\mathcal{P}_{rf}$ increases the optimal α is reasonably close to the value given in (4.15) for low to moderate $\mathcal{P}_d$, *e.g.*, $\alpha = 0.50$ and 0.76 for $\mathcal{P}_d = 0$ dB

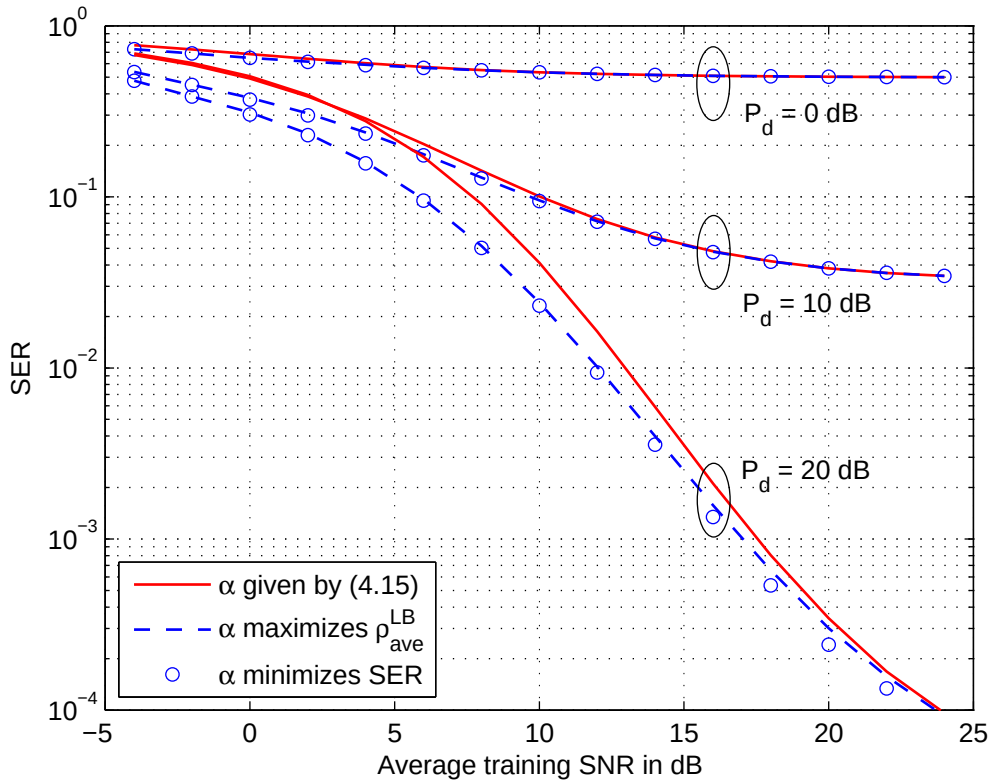

Figure 4.3: SER for 16-QAM modulation vs. average training SNR $\mathcal{P}_{rf}$ for systems with $N_t = 4$ transmit antennas and different values of the data SNR $\mathcal{P}_d$. The values of α used are found from Monte-Carlo simulations by minimizing the SER, maximizing the average SNR lower bound $\rho_{\text{ave}}^{\text{LB}}$ in (4.11), as well as the closed-form solution given in (4.15).

and 10 dB, respectively.

Fig. 4.3 shows the SER for 16-QAM modulation versus average training SNR $\mathcal{P}_{rf}$. For comparison, we use the values of α found by minimizing the SER, maximizing $\rho_{\text{ave}}^{\text{LB}}$ given in (4.11), as well as the closed-form solution given in (4.15). We see that the values of α that maximize $\rho_{\text{ave}}^{\text{LB}}$ also achieves the near optimum SER performance. Furthermore, the closed-form solution for α derived from the high SNR approximation can be used to achieve near optimum SER at moderate to high training SNR, *e.g.*, when $\mathcal{P}_{rf} > 10$ dB for $\mathcal{P}_d = 10$ dB and $\mathcal{P}_{rf} > 20$ dB for $\mathcal{P}_d = 20$ dB. We have also investigated the SER performance with other modulations. For example, the SER achieved by using the closed-form solution for α is within 0.5 dB, 1 dB and 1.4 dB from the optimum SER for 32-QAM, 16-QAM and 8-QAM, respectively, when $\mathcal{P}_{rf} > 20$ dB for $\mathcal{P}_d = 20$ dB. This suggests that the closed-form solution is more accurate for higher order modulations.

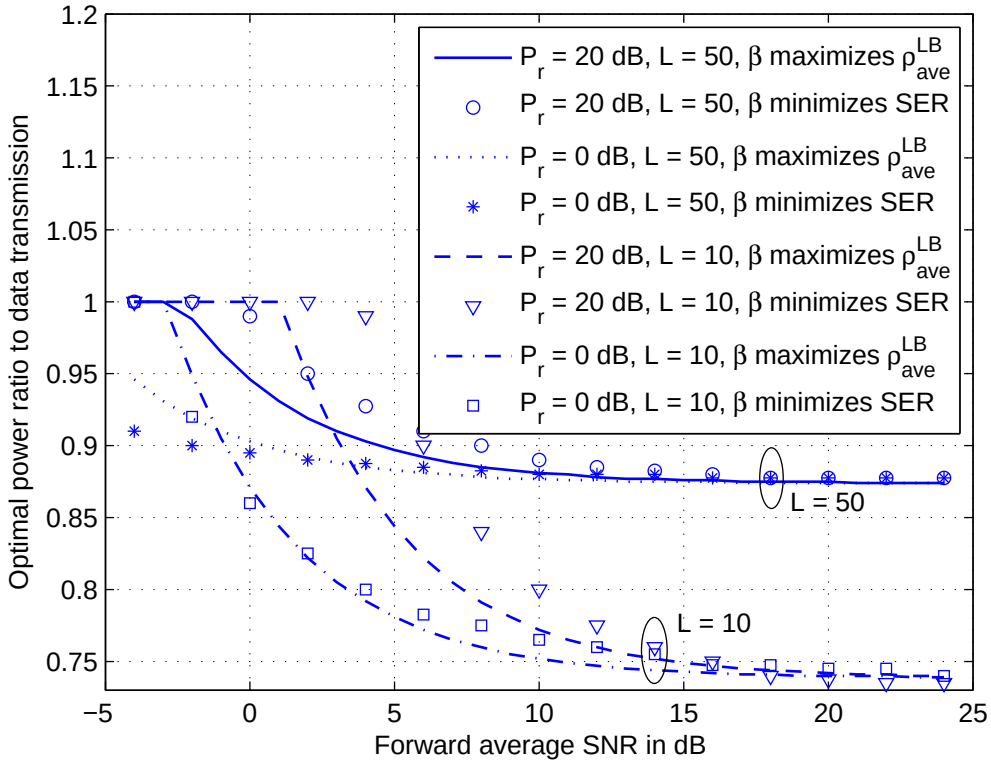

Figure 4.4: Optimal power ratio to data transmission β vs. forward average SNR $\mathcal{P}_{fd}$ for systems with $N_t = 4$ transmit antennas and different values of the reverse training SNR $\mathcal{P}_r$ and block length L. Lines indicate the values of β that maximize the average SNR lower bound $\rho_{\text{ave}}^{\text{LB}}$ in (4.11) and the markers indicate the values of β that minimize the SER for 16-QAM modulation found via Monte-Carlo simulations.

4.6.2 Optimal Power Allocation for Forward Transmission

Fig. 4.4 shows the optimal power ratio to data transmission β versus the forward average SNR $\mathcal{P}_{fd}$ for different reverse training SNR $\mathcal{P}_r$ and block lengths L. We see that the values of β which minimize the SER for 16-QAM modulation follow the same trend as the values of β which maximize $\rho_{\text{ave}}^{\text{LB}}$ in (4.11), and the mismatch occurs when $\mathcal{P}_{fd}$ is low to moderate. This mismatch is mainly due to the fact that maximizing $\rho_{\text{ave}}^{\text{LB}}$ does not necessarily result in the optimal distribution of ρ_{inst} which minimizes the SER. Similar to the first scenario, we see that forward training should not be used when $\mathcal{P}_{fd}$ is sufficiently low. As $\mathcal{P}_{fd}$ increases the optimal value of β converges to the value given in (4.19), *e.g.*, $\beta = 0.739$ and 0.874 for $L = 10$ and 50, respectively, and the value is independent of the reverse training SNR $\mathcal{P}_r$.

Fig. 4.5 shows the SER versus forward average SNR $\mathcal{P}_{fd}$. For comparison, we use the values of β found by minimizing the SER as well as the closed-form solution given in (4.19). We see that the closed-form solution for β can be used to achieve

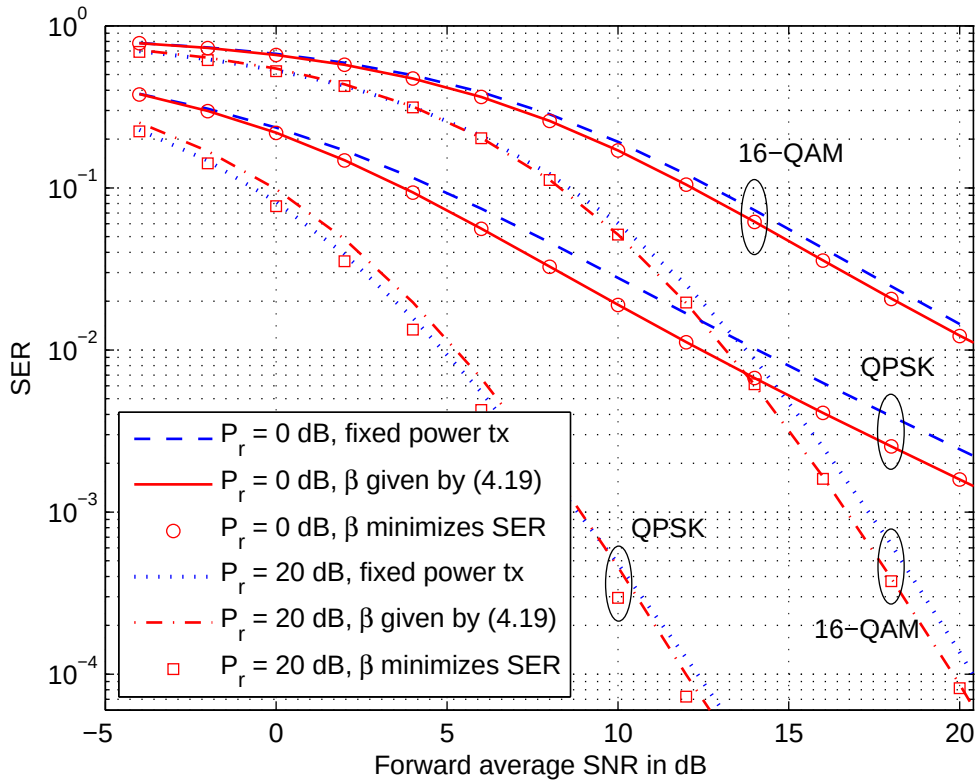

Figure 4.5: SER vs. forward average SNR $\mathcal{P}_{fd}$ for systems with $N_t = 4$ transmit antennas, reverse training SNR of $\mathcal{P}_r = 0, 20$ dB, and block length of $L = 50$. The values of β used are found from Monte-Carlo simulations by minimizing the SER, as well as the closed-form solution given in (4.19). The SER with fixed power transmission, *i.e.*, $\mathcal{P}_f = \mathcal{P}_d = \mathcal{P}_{fd}$, is also included for comparison.

near optimum SER over a wide range of SNR. We also include the SER using fixed power transmission, *i.e.*, $\mathcal{P}_f = \mathcal{P}_d = \mathcal{P}_{fd}$. We observe that power optimization only provides a maximum of 0.7 dB gain over fixed power transmission for 16-QAM modulation. Therefore, the use of fixed power transmission achieves reasonably good SER performance. On the other hand, the maximum SNR gain by using power optimization is around 1.8 dB for QPSK modulation for $\mathcal{P}_r = 0$ dB, while this gain reduces to around 0.8 dB for $\mathcal{P}_r = 20$ dB.

4.6.3 Optimal Power Allocation for Overall Transmission

Fig. 4.6 shows the optimal power ratio to forward transmission γ versus the average SNR $\mathcal{P}$ for different block lengths L.[2] The general trend is that a larger ratio of power should be allocated to the forward transmission as $\mathcal{P}$ increases. Similar to

[2]For brevity, we omit the results on the optimal values of β as they are very similar to those shown in Fig. 4.4.

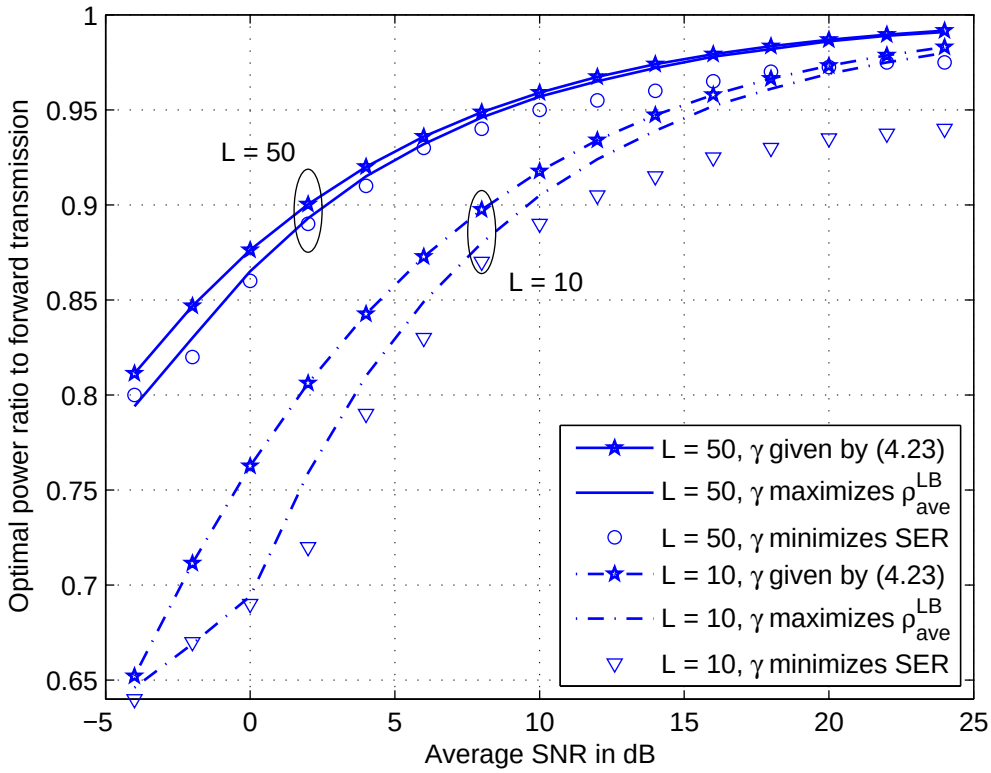

Figure 4.6: Optimal power ratio to forward transmission γ vs. average SNR $\mathcal{P}$ for systems with $N_t = 4$ transmit antennas and block length of $L = 10$ and $L = 50$. The values of γ which minimize the SER for 16-QAM modulation found via Monte-Carlo simulations as well as those maximizing the average SNR lower bound $\rho_{\text{ave}}^{\text{LB}}$ in (4.11) are shown. The closed-form solutions for the optimal γ derived from high SNR approximation in (4.23) are also included.

the previous two scenarios, we see that the values of γ which maximize $\rho_{\text{ave}}^{\text{LB}}$ in (4.11) are close to the values of γ that minimize the SER for 16-QAM modulation. We also see that the closed-form solution of γ given in (4.23) is reasonably accurate over a wide range of SNR, and the mismatch increases as $\mathcal{P}$ increases. This mismatch is mainly due to the fact that maximizing $\rho_{\text{ave}}^{\text{LB}}$ does not necessarily result in the optimal distribution of ρ_{inst} which minimizes the SER, *e.g.*, it does not minimize the probability of ρ_{inst} taking very small values.

Fig. 4.7 shows the SER versus the average SNR $\mathcal{P}$. The values of β and γ used in this plot are found both from minimizing the SER and from the closed-form solutions in (4.19) and (4.23). We see that the closed-form solutions for β and γ can be used to achieve near optimum SER over a wide range of SNR for 8-QAM and 16-QAM modulation. For QPSK modulation, the closed-form solutions result in a slight SER degradation of around 0.6 dB. In fact, we will see in the next

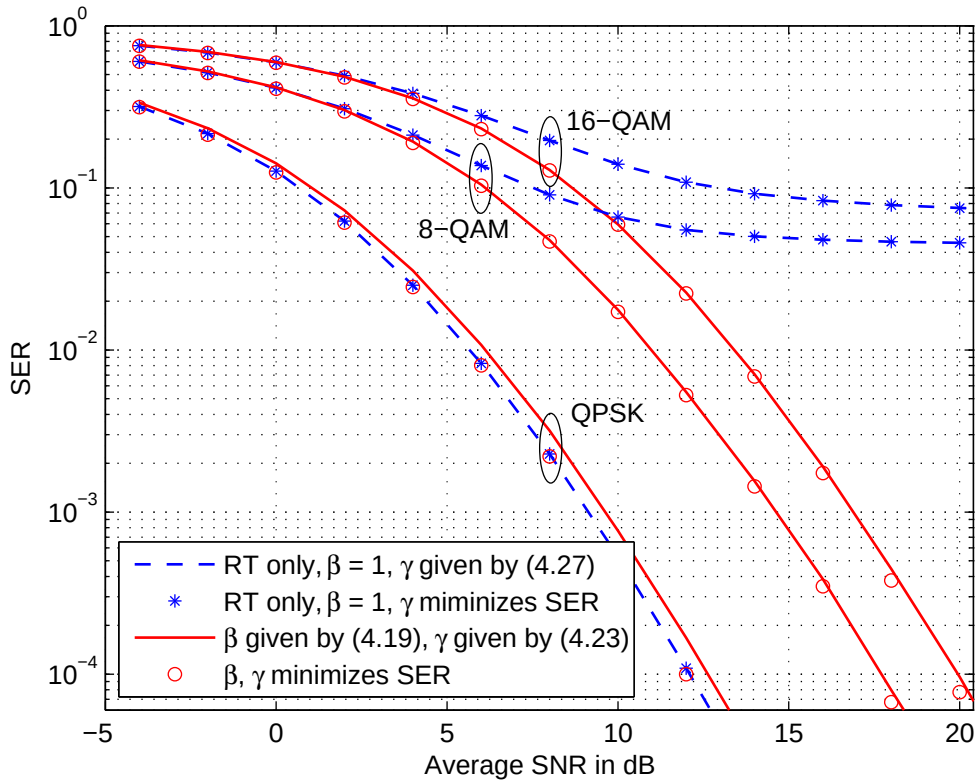

Figure 4.7: SER vs. average SNR $\mathcal{P}$ for systems with $N_t = 4$ transmit antennas and block length of $L = 50$. The values of β and γ used are found from Monte-Carlo simulations by minimizing the SER as well as the closed-form solutions given in (4.19) and (4.23). The SER with only reverse training (RT) is also included for comparison.

paragraph that the system without forward training, *i.e.*, $\mathcal{P}_f = 0$, can achieve near optimal SER performance for QPSK modulation. With $\mathcal{P}_f = 0$, the high SNR assumptions used in deriving the closed-form solutions for β and γ are no longer valid, which results in the observed SER degradation for QPSK modulation.

For comparison with two-way training based transmission, Fig. 4.7 also includes the SER performance for systems using only reverse training. We see that the closed-form solution for γ given in (4.27) achieves near optimum SER over a wide range of SNR. When moderate to high-order modulation is used, *e.g.*, 8-QAM and 16-QAM, the use of two-way training achieves a significant SER reduction over reverse training at moderate to high SNR. However, when low-order modulation is used, *e.g.*, QPSK, it is clear that the systems with two-way training achieve no or marginal performance gain over those with only reverse training. This result suggests that reverse training is sufficient at low operating SNR or when low-order modulations are used.

4.7 Ergodic Capacity Performance

In previous sections, we focused on the SER performance as a measure of quality of service (QoS) and verified the near optimality of the power allocation strategies derived in Section 4.4. In this section, we investigate the optimality of the derived power allocation strategies from an information-theoretic viewpoint. As discussed in Section 4.4, the complicated nature of the distribution of the instantaneous received SNR ρ_{inst} given in (4.9) makes it difficult to obtain any closed-form expression for information capacity. Therefore, we obtain an approximation of the ergodic capacity by treating the noise components of the received signal in (4.8) as Gaussian noise independent from the signal component. Using this approximation, the ergodic capacity per channel use with Gaussian input is given by

$$C \approx E_{\widehat{f}}\Big\{ \log_2 \Big(1 + \frac{\mathcal{P}_d |\widehat{f}|^2}{1 + \mathcal{P}_d \sigma^2_{\widehat{f}}}\Big)\Big\}. \tag{4.28}$$

Note that this capacity approximation is not guaranteed to be a lower bound because the LMMSE forward channel estimator is not the optimal MMSE estimator [108]. Taking the training overhead into account, the ergodic capacity is approximated by

$$\overline{C} \;\approx\; \frac{L-2}{L} E_{\widehat{f}}\Big\{ \log_2 \Big(1 + \frac{\mathcal{P}_d |\widehat{f}|^2}{1 + \mathcal{P}_d \sigma^2_{\widehat{f}}}\Big)\Big\}. \tag{4.29}$$

For systems with only reverse training, the UT has the knowledge of the mean value of the effective channel μ_1, and hence the ergodic capacity is approximated by

$$\overline{C} \approx \frac{L-1}{L} \log_2 \Big(1 + \frac{\mathcal{P}_d \mu_1^2}{1 + \mathcal{P}_d \sigma_f^2}\Big), \tag{4.30}$$

which is indeed a valid lower bound on the exact capacity [106, 107].

In the following, we investigate the optimality of the derived power allocation strategies in Section 4.4 using the approximations of the ergodic capacity. We will refer to (4.28), (4.29) and (4.30) as "ergodic capacity", while it should be noted that these capacity expressions are approximations only. We obtain the capacity performance in the three scenarios considered in Section 4.4.

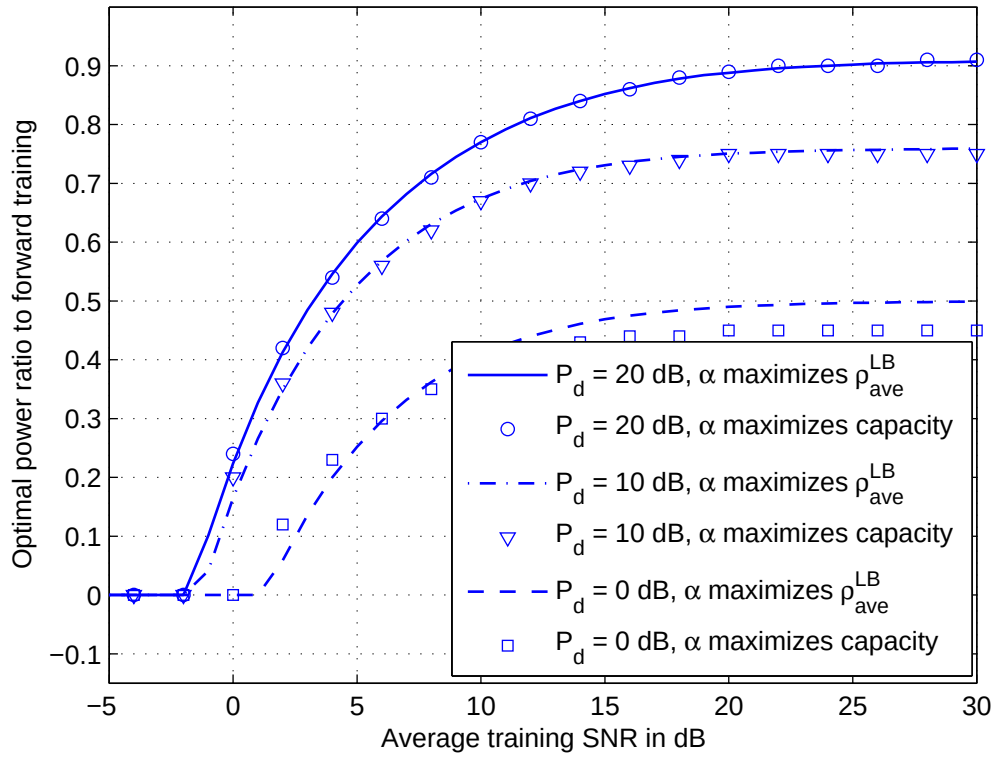

Figure 4.8: Optimal power ratio to forward training α vs. average training SNR $\mathcal{P}_{rf}$ for systems with $N_t = 4$ transmit antennas and different values of data SNR $\mathcal{P}_d$. Lines indicate the values of α that maximize the average SNR lower bound $\rho_{\text{ave}}^{\text{LB}}$ in (4.11) and the markers indicate the values of α that maximize the ergodic capacity per channel use C in (4.28) found via Monte-Carlo simulations.

4.7.1 Optimal Power Allocation for Reverse and Forward Training

Fig. 4.8 shows the optimal power ratio to forward training α versus average training SNR $\mathcal{P}_{rf}$ for different data SNR $\mathcal{P}_d$. The lines indicate the values of α which maximize the average SNR lower bound $\rho_{\text{ave}}^{\text{LB}}$ in (4.11) and the markers indicate the values of α which maximize the ergodic capacity per channel use C in (4.28) found via Monte-Carlo simulations. We see that the markers are reasonably close to the lines and the mismatch becomes insignificant at moderate to high data SNR, *e.g.*, $\mathcal{P}_d = 10$ dB and 20 dB. Furthermore, we see a general trend that more power should be allocated to reverse (forward) training when $\mathcal{P}_{rf}$ is low (high). In particular, we see that forward training should not be used when $\mathcal{P}_{rf} < -2$ dB. As $\mathcal{P}_{rf}$ increases the optimal α is close to the value given in (4.15), *e.g.*, $\alpha = 0.760$ and 0.909 for $\mathcal{P}_d = 10$ dB and 20 dB, respectively. In addition, we see that more power should be allocated to forward training as $\mathcal{P}_d$ increases.

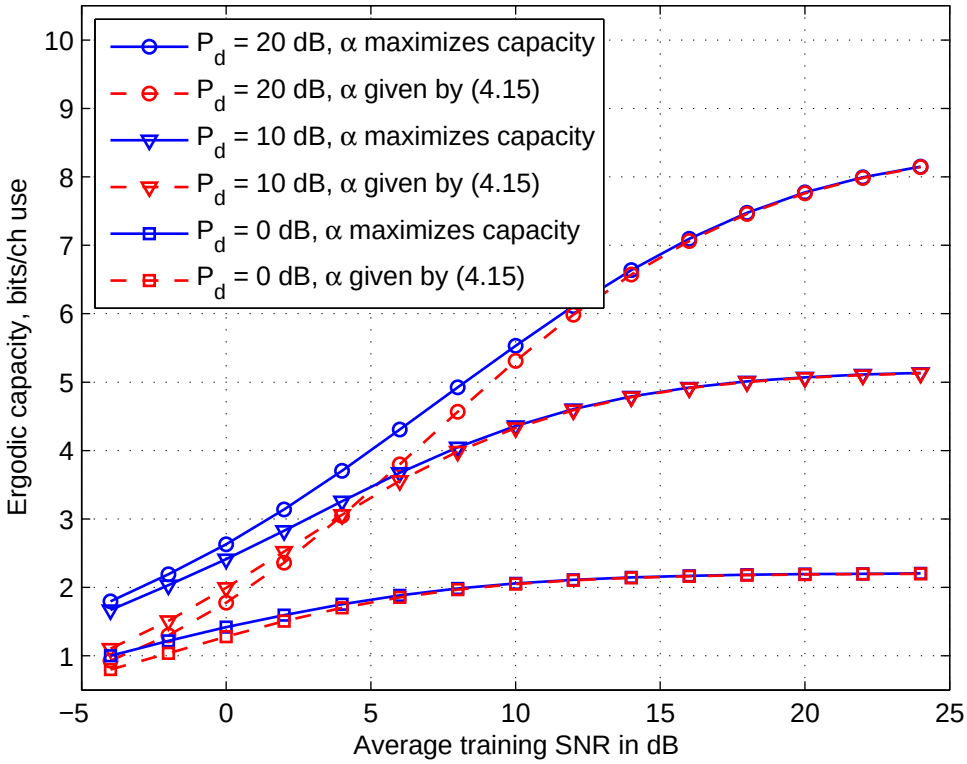

Figure 4.9: The ergodic capacity per channel use C in (4.28) vs. average training SNR $\mathcal{P}_{rf}$ for systems with $N_t = 4$ transmit antennas and different values of the data SNR $\mathcal{P}_d$. The values of α used are found from Monte-Carlo simulations by maximizing C in (4.28) as well as the high SNR approximation given in (4.15).

Fig. 4.9 shows the ergodic capacity per channel use C in (4.28) versus average training SNR $\mathcal{P}_{rf}$. For comparison, we use the values of α found by maximizing C as well as the closed-form high SNR approximation given in (4.15). We see that the closed-form solution for α can be used to achieve near maximum capacity at moderate to high training SNR, *e.g.*, when $\mathcal{P}_{rf} > 5$ dB for $\mathcal{P}_d = 0$ dB and $\mathcal{P}_{rf} > 15$ dB for $\mathcal{P}_d = 20$ dB.

4.7.2 Optimal Power Allocation for Forward Transmission

Fig. 4.10 shows the optimal power ratio to data transmission β versus the forward average SNR $\mathcal{P}_{fd}$ for different reverse training SNR $\mathcal{P}_r$ and block lengths L. The lines indicate the values of β which maximize the average SNR lower bound $\rho_{\text{ave}}^{\text{LB}}$ in (4.11) and the markers indicate the values of β which maximize the ergodic capacity $\overline{C}$ in (4.29) found via Monte-Carlo simulations. We see that the markers are close to the lines and the mismatch becomes insignificant at high reverse training SNR,

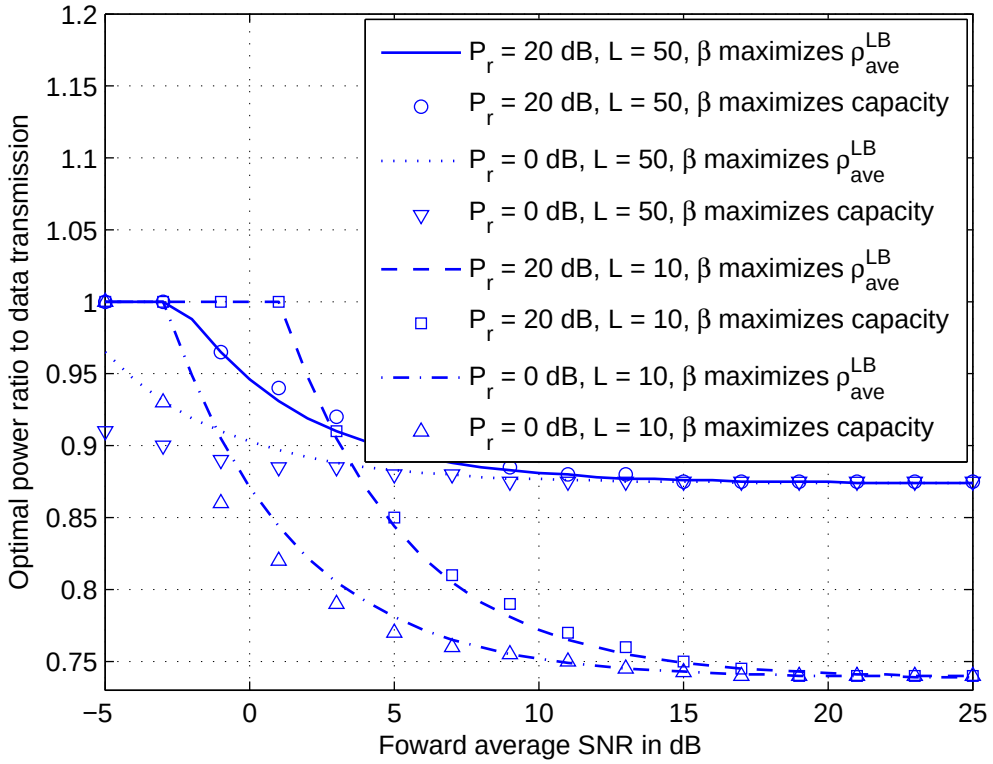

Figure 4.10: Optimal power ratio to data transmission β vs. forward average SNR $\mathcal{P}_{fd}$ for systems with $N_t = 4$ transmit antennas and different values of the reverse training SNR $\mathcal{P}_r$ and block length L. Lines indicate the values of β that maximize the average SNR lower bound $\rho_{\text{ave}}^{\text{LB}}$ in (4.11) and the markers indicate the values of β that maximize the ergodic capacity $\overline{C}$ in (4.29) found via Monte-Carlo simulations.

e.g., $\mathcal{P}_r = 20$ dB. Similar to the first scenario, we see that forward training should not be used when $\mathcal{P}_{fd}$ is low. As $\mathcal{P}_{fd}$ increases the optimal value of β converges to the value given in (4.19), *e.g.*, $\beta = 0.739$ and 0.874 for $L = 10$ and 50, respectively, and the value is independent of the reverse training SNR $\mathcal{P}_r$.

Fig. 4.11 shows the ergodic capacity $\overline{C}$ in (4.29) versus forward average SNR $\mathcal{P}_{fd}$. For comparison, we use the values of β found by maximizing $\overline{C}$ as well as the closed-form high SNR approximation given in (4.19). We see that the closed-form solution for β can be used to achieve near maximum capacity over a wide range of SNR, *e.g.*, when $\mathcal{P}_{fd} > 5$ dB for $L = 10$ and $\mathcal{P}_{fd} > 0$ dB for $L = 50$. We also include the ergodic capacity achieved using fixed power transmission, *i.e.*, $\mathcal{P}_f = \mathcal{P}_d = \mathcal{P}_{fd}$. We observe that power optimization provides 5% - 9% capacity gain over fixed power transmission at $L = 50$ and $\mathcal{P}_{fd} > 5$ dB.

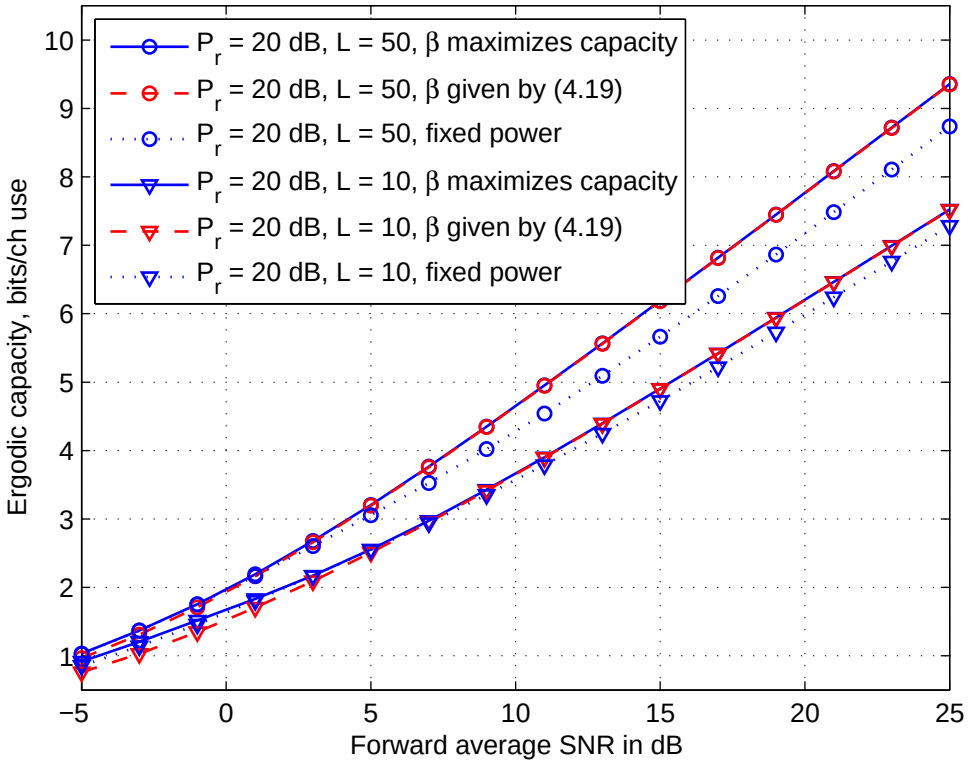

Figure 4.11: The ergodic capacity $\overline{C}$ in (4.29) vs. forward average SNR $\mathcal{P}_{fd}$ for systems with $N_t = 4$ transmit antennas, reverse training SNR of $\mathcal{P}_r = 20$ dB, and block length of $L = 10$ and $L = 50$. The values of β used are found from Monte-Carlo simulations by maximizing $\overline{C}$, as well as the high SNR approximation given in (4.19). The ergodic capacity with fixed power transmission, *i.e.*, $\mathcal{P}_f = \mathcal{P}_d = \mathcal{P}_{fd}$, is also included for comparison.

4.7.3 Optimal Power Allocation for Overall Transmission

Fig. 4.12 shows the optimal power ratio to forward transmission γ versus the average SNR $\mathcal{P}$ for different block lengths L.[3] Similar to the previous two scenarios, we see that the values of γ which maximize the average SNR lower bound $\rho_{\text{ave}}^{\text{LB}}$ in (4.11) are close to the values of γ that maximize the ergodic capacity $\overline{C}$ in (4.29). We also include the closed-form solution of γ given in (4.23) which is asymptotically optimal at high SNR. We see in Fig. 4.12 that the values of γ given in (4.23) is reasonably accurate at moderate SNR, especially at large block lengths, *e.g.*, $L = 50$.

Fig. 4.13 shows the ergodic capacity $\overline{C}$ in (4.29) versus the average SNR $\mathcal{P}$ for different block lengths L. The values of β and γ used in this plot are found both from maximizing $\overline{C}$ and from the closed-form high SNR approximations in

[3]For brevity, we omit the results on the optimal values of β as they are very similar to those shown in Fig. 4.10.

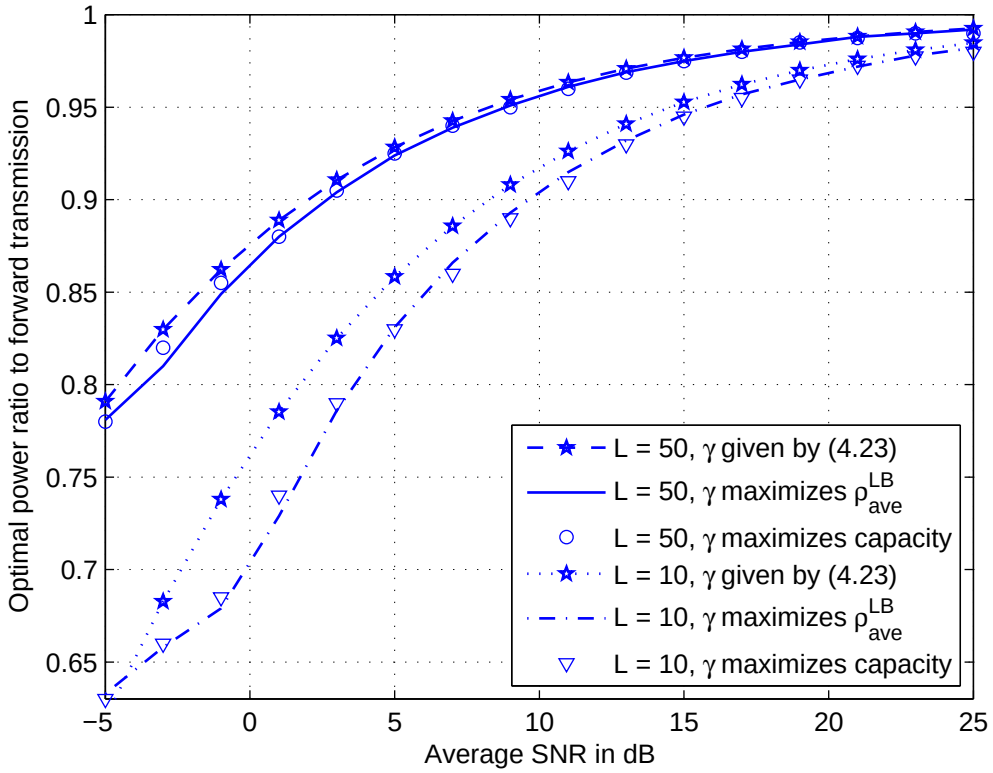

Figure 4.12: Optimal power ratio to forward transmission γ vs. average SNR $\mathcal{P}$ for systems with $N_t = 4$ transmit antennas and block length of $L = 10$ and $L = 50$. The values of γ which maximize the ergodic capacity $\overline{C}$ in (4.29) found via Monte-Carlo simulations as well as those maximizing the average SNR lower bound $\rho_{\text{ave}}^{\text{LB}}$ in (4.11) are shown. The closed-form solutions for the optimal γ derived from high SNR approximation in (4.23) are also included.

(4.19) and (4.23). We see that the closed-form solutions for β and γ can be used to achieve near maximum capacity over a wide range of SNR, *e.g.*, $\mathcal{P} > 3$ dB.

For comparison with two-way training based transmission, Fig. 4.13 also includes the ergodic capacity $\overline{C}$ in (4.30) achieved using reverse training only with γ given in (4.27). It is clear that the systems with two-way training do not achieve better performance than those with only reverse training at low SNR and small block lengths, *e.g.*, when $\mathcal{P} < 5$ dB and $L = 10$. This result shows that forward training should not be used at low operating SNR and small block lengths.

4.8 Summary of Contributions

In this chapter we have investigated the optimal transmit power allocation in MISO systems with two-way training. Depending on the degrees of freedom in the system

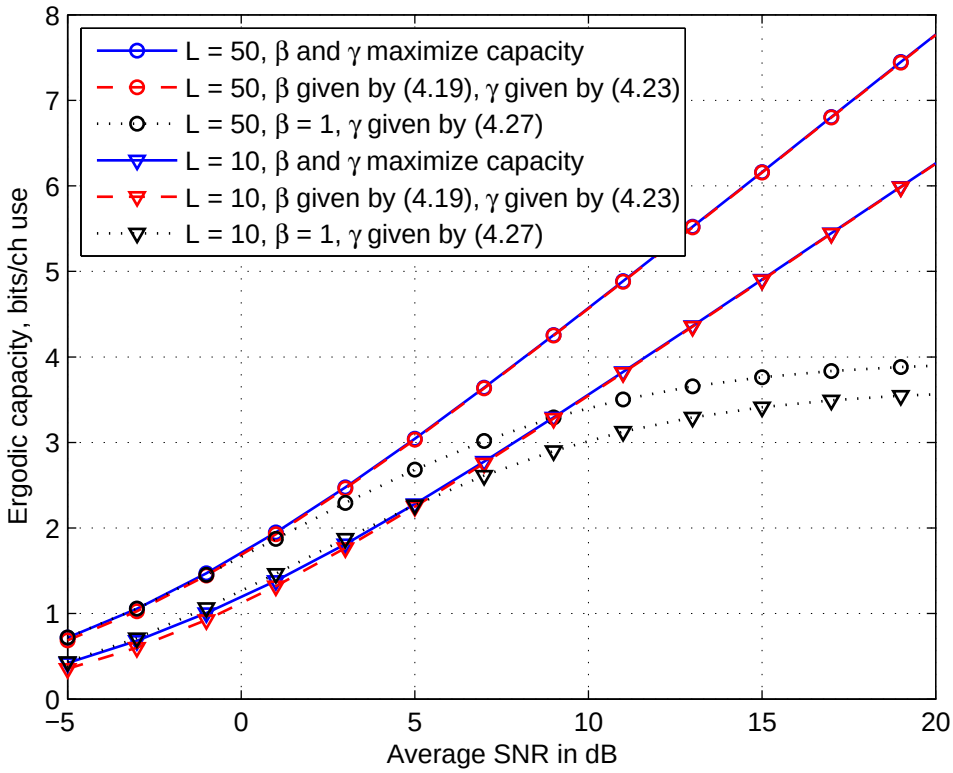

Figure 4.13: The ergodic capacity $\overline{C}$ in (4.29) vs. average SNR $\mathcal{P}$ for systems with $N_t = 4$ transmit antennas and block length of $L = 10$ and $L = 50$. The values of β and γ used are found from Monte-Carlo simulations by maximizing the ergodic capacity $\overline{C}$ in (4.29) as well as the high SNR approximation given in (4.19) and (4.23). The ergodic capacity with only reverse training, *i.e.*, $\beta = 1$ and γ given in (4.27), is also included for comparison.

design, we have solved the power optimization problem in three different scenarios. Some specific contributions made in this chapter are as follows.

Addressing Q1 in Section 4.1:

- We have considered a four-stage transmission scheme and derived the LMMSE channel estimation for the forward training. We have found that the forward channel estimation error is almost independent of the number of transmit antennas.

Addressing Q2 in Section 4.1:

- An average SNR lower bound has been used to obtain closed-form solutions to the optimal power allocation at high SNR in three different scenarios. The

derived power allocation strategies are independent of the number of transmit antennas.

- We have carried out Monte-Carlo simulations to obtain both SER and ergodic capacity performance. For both performance metrics, we have verified the near optimality of the derived power allocation strategies over a wide range of SNR values.

Addressing Q3 in Section 4.1:

- We have derived the optimal power allocation for systems with reverse training only. Our numerical results have shown that two-way training provides no or marginal performance gain over reverse training only, at low SNR or when low-order modulations are used.

Chapter 5

Physical-Layer Security-Constrained Systems

5.1 Introduction

Recently, information-theoretic security has drawn considerable attention [71–74, 79, 80]. Various physical-layer techniques were proposed to achieve secure communication in the presence of eavesdroppers [81–86]. When multiple antennas are available at the transmitter, it is possible to simultaneously transmit both the information bearing signal and artificial noise to achieve secrecy in a fading environment [24, 109]. Without knowing the eavesdroppers' channels, the artificial noise is radiated isotropically to mask the transmission of the information signal to the intended receiver. In the design of this multi-antenna technique, the transmit power allocation between the information signal and the artificial noise is an important parameter, which has not been investigated in [24, 109]. A sub-optimal power allocation strategy was considered in [110, 111], which aims to meet a target signal to interference and noise ratio (SINR) at the intended receiver to satisfy a QoS requirement.

In this chapter, we study the problem of secure communication in fading channels with a multi-antenna transmitter capable of simultaneous transmission of both the information signal and the artificial noise. We assume that the transmitter does not have any knowledge about the eavesdroppers' channels. The following important questions are addressed:

Q1. What is the achievable rate for secure communication with artificial noise transmission?

Q2. How should the system designer optimally allocate the transmit power be-

tween the information signal and the artificial noise?

Q3. What is the minimum SNR above which secure communication can be made possible?

Q4. What is the effect of channel estimation errors on the optimal power allocation strategy?

In Section 5.3, we derive a closed-form expression for an average secrecy capacity lower bound for fading channels, which is an achievable rate without knowing the noise levels at the eavesdroppers. The availability of a closed-form secrecy rate expression greatly reduces the complexity of obtaining the optimal power allocation between transmission of the information signal and the artificial noise which is studied in Section 5.4.

Without the knowledge of the eavesdropper's channel, the secrecy rate can quickly drop to zero at low to moderate SNR. This is an important problem in wideband communications in which a higher throughput is achieved by reducing the SNR per hertz while increasing the bandwidth [112]. Therefore, we study the critical SNR above which the secrecy rate is positive in Section 5.5.

Furthermore, perfect CSI at both the transmitter and the receiver is usually assumed in the existing studies on secrecy rate. With this assumption, the information signal is accurately transmitted into the intended receiver's channel and the artificial noise is accurately transmitted into the null space of the intended receiver's channel. When practical channel estimation is considered, the CSI is not perfectly known at both the receiver and the transmitter, and hence the artificial noise leaks into the receiver's channel. The effects of channel estimation errors on the secrecy rate and the aforementioned design parameters are investigated in Section 5.6.

5.2 System Model

We consider secure communication between a transmitter (Alice) and a receiver (Bob) in the presence of eavesdroppers (Eves). Alice has N_A antennas ($N_A > 1$) and Bob has a single antenna. This scenario is representative, for example, of downlink transmission in cellular systems and WLAN. In addition, each Eve is equipped with a single antenna. We consider two cases, namely non-colluding and colluding eavesdroppers. In the former case, Eves individually overhear the communication between Alice and Bob without any centralized processing. While in the latter case, there are N_E Eves capable of jointly processing their received

information. Therefore, the non-colluding case can be seen as a special colluding case where $N_E = 1$. We assume that $N_A > N_E$ for which the reason will become clear in the next section. We also assume that Eves are passive, hence they cannot transmit jamming signals.

The channel between Alice and Bob is denoted as $\boldsymbol{h}$, which is a $1 \times N_A$ vector. The channel between Alice and multiple colluding Eves is denoted as $\boldsymbol{G}$, which is an $N_E \times N_A$ matrix. The elements of $\boldsymbol{h}$ and $\boldsymbol{G}$ are independent ZMCSCG random variables. We first assume that $\boldsymbol{h}$ is accurately estimated by Bob and is also known by Alice using a noiseless feedback link from Bob[1]. The impact of channel estimation errors will be studied in Section 5.6. Similar to [24], we assume that the knowledge of both $\boldsymbol{h}$ and $\boldsymbol{G}$ is available at Eve, which makes the secrecy of communication independent of the secrecy of CSI. Note that Alice does not have any knowledge about $\boldsymbol{G}$.

The key idea of guaranteeing secure communication using artificial noise proposed in [24] is described as follows. Alice utilizes multiple antennas to transmit the information bearing signal into Bob's channel, at the same time generating a noise-like signal into the null space of Bob's channel. We let an $N_A \times N_A$ matrix $\boldsymbol{W} = [\boldsymbol{w}_1 \ \ \boldsymbol{W}_2]$ be an orthonormal basis of $\mathbb{C}^{N_A}$, where $\boldsymbol{w}_1 = \boldsymbol{h}^\dagger / \|\boldsymbol{h}\|$. The $N_A \times 1$ transmitted symbol vector at Alice is given by $\boldsymbol{x} = \boldsymbol{w}_1 u + \boldsymbol{W}_2 \boldsymbol{v}$, where the variance of u is σ_u^2 and the $N_A - 1$ elements of $\boldsymbol{v}$ are i.i.d. complex Gaussian random variables each with variance σ_v^2.

The received symbol at Bob is given by

$$\begin{aligned} y_B &= \boldsymbol{h}\boldsymbol{x} + n \\ &= \boldsymbol{h}\boldsymbol{w}_1 u + \boldsymbol{h}\boldsymbol{W}_2 \boldsymbol{v} + n \\ &= \|\boldsymbol{h}\| u + n, \end{aligned} \tag{5.1}$$

where n is the AWGN at Bob. Without loss of generality, we normalize the variance of n to unity. We see that $\boldsymbol{W}_2$ spans the null space of $\boldsymbol{h}$, hence the artificial noise $\boldsymbol{v}$ does not affect the received signal at Bob.

The received symbol vector at the multiple colluding Eves is given by

$$\begin{aligned} \boldsymbol{y}_E &= \boldsymbol{G}\boldsymbol{x} + \boldsymbol{e} \\ &= \boldsymbol{G}\boldsymbol{w}_1 u + \boldsymbol{G}\boldsymbol{W}_2 \boldsymbol{v} + \boldsymbol{e} \\ &= \boldsymbol{g}_1 u + \boldsymbol{G}_2 \boldsymbol{v} + \boldsymbol{e}, \end{aligned} \tag{5.2}$$

[1] A reliable feedback link could be achieved by using low rate transmission with appropriate quantization schemes. The design of high-quality feedback link and the effect of noisy feedback is beyond the scope of this work.

where we have defined that $\boldsymbol{g}_1 = \boldsymbol{G}\boldsymbol{w}_1$ and $\boldsymbol{G}_2 = \boldsymbol{G}\boldsymbol{W}_2$, and $\boldsymbol{e}$ is the AWGN at Eves.

We consider a total power per transmission denoted by $\mathcal{P}$, that is, $\mathcal{P} = \sigma_u^2 + (N_A - 1)\sigma_v^2$. Due to the normalization of the noise variance at Bob, we also refer to $\mathcal{P}$ as the transmit SNR. One important design parameter is the ratio of power allocated to the information bearing signal and the artificial noise. We denote the fraction of total power allocated to the information signal as ϕ. Hence, we have the following relationships:

$$\sigma_u^2 = \phi\mathcal{P}, \tag{5.3}$$

$$\sigma_v^2 = (1-\phi)\mathcal{P}/(N_A - 1). \tag{5.4}$$

Since $\boldsymbol{h}$ is known by Alice, she can adaptively change the value of ϕ according to the instantaneous realization of $\boldsymbol{h}$. We refer to this strategy as the adaptive power allocation strategy. Alternatively, Alice can choose a fixed value for ϕ regardless of the instantaneous channel realization, which we refer to as the non-adaptive power allocation strategy. Due to the uncertainty in the outgoing channel to Eves, *i.e.*, $\boldsymbol{G}$, Alice equally distributes the transmit power amongst the artificial noise signal, as given by (5.4).

5.3 Secrecy Capacity Lower Bound

The secrecy capacity is the maximum transmission rate at which the intended receiver can decode the data with arbitrarily small error while the mutual information between the transmitted message and the received signal at the eavesdropper is arbitrarily small. It is bounded from below by the difference in the capacity of the channel between Alice and Bob and that between Alice and Eve [70]. In this section, we derive a closed-form expression for an average secrecy capacity lower bound with the transmission of artificial noise.

The capacity of the channel between Alice and Bob is given by

$$\begin{aligned} C_1 &= E_{\boldsymbol{h}}\{\log_2(1 + \sigma_u^2\|\boldsymbol{h}\|^2)\} \\ &= E_{\boldsymbol{h}}\{\log_2(1 + \phi\mathcal{P}\|\boldsymbol{h}\|^2)\}. \end{aligned} \tag{5.5}$$

Without loss of generality, we normalize the variance of each element of $\boldsymbol{h}$ to unity. It is then easy to see that $\|\boldsymbol{h}\|^2$ follows a Gamma distribution with parameters $(N_A, 1)$. Therefore, for systems with non-adaptive power allocation

strategy, we can rewrite (5.5) in an integral form as

$$\begin{aligned} C_1 &= \frac{1}{\ln 2}\int_0^\infty \ln(1+\phi\mathcal{P}x)x^{N_A-1}\frac{\exp(-x)}{\Gamma(N_A)}\mathrm{d}x \\ &= \frac{1}{\ln 2}\exp\Big(\frac{z}{\mathcal{P}}\Big)\sum_{k=1}^{N_A}\mathrm{E}_k\Big(\frac{z}{\mathcal{P}}\Big), \end{aligned} \tag{5.6}$$

where $\Gamma(\cdot)$ is the Gamma function, $\mathrm{E}_n(\cdot)$ is the generalized exponential integral, (5.6) is obtained using an integral identity given in [113], and we have defined $z = \phi^{-1}$.

Next we study the capacity of the channel between Alice and multiple colluding Eves. When multiple Eves are present, the noise at each Eve may be different. In addition, the receiver noise levels at Eves may not be known by Alice and Bob. To guarantee secure communication, it is therefore reasonable to consider the worst case scenario where the noises at Eves are arbitrarily small. Note that this approach was also taken in [109]. In this case, the capacity between Alice and each Eve is determined from the signal-to-artificial-noise ratio. Considering the signal reception at a particular Eve, both the information signal and the artificial noise are generated from the same source (Alice), and hence their ratio is independent of the large scale fading from Alice to Eve. That is to say, the signal-to-artificial-noise ratios are i.i.d. random variables for all Eves, regardless of their distances from Alice. Therefore, we can normalize the distance of each Eve to make the variance of the elements of $\boldsymbol{G}$ equal to unity without loss of generality.

The noiseless eavesdropper assumption effectively gives an upper bound on the capacity of the channel between Alice and multiple colluding Eves as [109]

$$\begin{aligned} C_2 &= E_{\boldsymbol{h},\boldsymbol{g}_1,\boldsymbol{G}_2}\Big\{\log_2\Big|\boldsymbol{I}+\sigma_u^2\boldsymbol{g}_1\boldsymbol{g}_1{}^\dagger(\sigma_v^2\boldsymbol{G}_2\boldsymbol{G}_2{}^\dagger)^{-1}\Big|\Big\} \\ &= E_{\boldsymbol{h},\boldsymbol{g}_1,\boldsymbol{G}_2}\Big\{\log_2\Big(1+\frac{N_A-1}{z-1}\boldsymbol{g}_1{}^\dagger(\boldsymbol{G}_2\boldsymbol{G}_2{}^\dagger)^{-1}\boldsymbol{g}_1\Big)\Big\}, \end{aligned} \tag{5.7}$$

where we have again used $z = \phi^{-1}$. The expectation over $\boldsymbol{h}$ in (5.7) is due to the fact that z may be dependent on $\boldsymbol{h}$ (which happens when adaptive power allocation strategy is used). It is required in (5.7) that $\boldsymbol{G}_2\boldsymbol{G}_2{}^\dagger$ is invertible, which is guaranteed with the assumption of $N_A > N_E$. If the assumption is violated, the colluding eavesdroppers are able to eliminate the artificial noise, resulting $C_2 = \infty$. Hence, We assume $N_A > N_E$ for guaranteeing positive secrecy capacity.

Since $\boldsymbol{G}$ has i.i.d. complex Gaussian entries and $\boldsymbol{W}$ is a unitary matrix, $\boldsymbol{GW} = [\boldsymbol{g}_1 \;\; \boldsymbol{G}_2]$ also has i.i.d. complex Gaussian entries. Therefore, the elements of $\boldsymbol{g}_1$ and $\boldsymbol{G}_2$ are independent. As a result, the quantity $\boldsymbol{g}_1{}^\dagger(\boldsymbol{G}_2\boldsymbol{G}_2{}^\dagger)^{-1}\boldsymbol{g}_1$ is equivalent to the

signal-to-interference ratio (SIR) of a N_E-branch MMSE diversity combiner with $N_A - 1$ interferers. The complementary cumulative distribution function (CCDF) of $X = \boldsymbol{g}_1{}^{\dagger}(\boldsymbol{G}_2\boldsymbol{G}_2{}^{\dagger})^{-1}\boldsymbol{g}_1$ is given in [114] as

$$R_X(x) \;=\; \frac{\sum_{k=0}^{N_E-1}\binom{N_A-1}{k}x^k}{(1+x)^{N_A-1}}. \tag{5.8}$$

Therefore, we can rewrite (5.7) in an integral form as

$$\begin{aligned} C_2 &= E_{\boldsymbol{h}}\Big\{\int_0^{\infty}\log_2\Big(1+\frac{N_A-1}{z-1}x\Big)f_X(x)\mathrm{d}x\Big\} \\ &= E_{\boldsymbol{h}}\Big\{\frac{1}{\ln 2}\int_0^{\infty}\frac{N_A-1}{z-1}\Big(1+\frac{N_A-1}{z-1}x\Big)^{-1}R_X(x)\mathrm{d}x\Big\} \qquad (5.9)\\ &= E_{\boldsymbol{h}}\Big\{\frac{1}{\ln 2}\sum_{k=0}^{N_E-1}\binom{N_A-1}{k}\int_0^{\infty}\Big(\frac{z-1}{N_A-1}+x\Big)^{-1}(1+x)^{1-N_A}x^k\mathrm{d}x\Big\} \\ &= E_{\boldsymbol{h}}\Big\{\frac{1}{\ln 2}\sum_{k=0}^{N_E-1}\binom{N_A-1}{k}\frac{N_A-1}{z-1}\mathrm{B}(k+1,N_A-1-k)\,{}_2\mathrm{F}_1\Big(1,k+1;N_A;\frac{z-N_A}{z-1}\Big)\Big\}, \qquad (5.10)\end{aligned}$$

where $f_X(x)$ denotes the probability density function (PDF) of X, $\mathrm{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the Beta function and ${}_2\mathrm{F}_1(\cdot)$ is the Gauss hypergeometric function. Note that (5.9) is obtained using integration by parts, and (5.10) is obtained using an integration identity given in [115].

After deriving expressions for C_1 and C_2, a lower bound on the average secrecy capacity can now be obtained as $C_S = [C_1 - C_2]^+$, where $[\alpha]^+ = \max\{0,\alpha\}$. This is the maximum data rate that can be always guaranteed for the secure communication (without knowing the noise level at Eves). For systems with adaptive power allocation, the average secrecy capacity lower bound is given as

$$\begin{aligned} C_S = \frac{1}{\ln 2}\Big[E_{\boldsymbol{h}}\Big\{&\ln\Big(1+\frac{\mathcal{P}}{z}\|\boldsymbol{h}\|^2\Big) \\ &-\sum_{k=0}^{N_E-1}\binom{N_A-1}{k}\frac{N_A-1}{z-1}\mathrm{B}(k+1,N_A-1-k)\,{}_2\mathrm{F}_1\Big(1,k+1;N_A;\frac{z-N_A}{z-1}\Big)\Big\}\Big]^+, \end{aligned} \tag{5.11}$$

where z is a function of $\boldsymbol{h}$. For systems with non-adaptive power allocation, the average secrecy capacity lower bound is given as

$$C_S = \frac{1}{\ln 2}\Big[\exp\Big(\frac{z}{\mathcal{P}}\Big)\sum_{k=1}^{N_A}\mathrm{E}_k\Big(\frac{z}{\mathcal{P}}\Big)$$

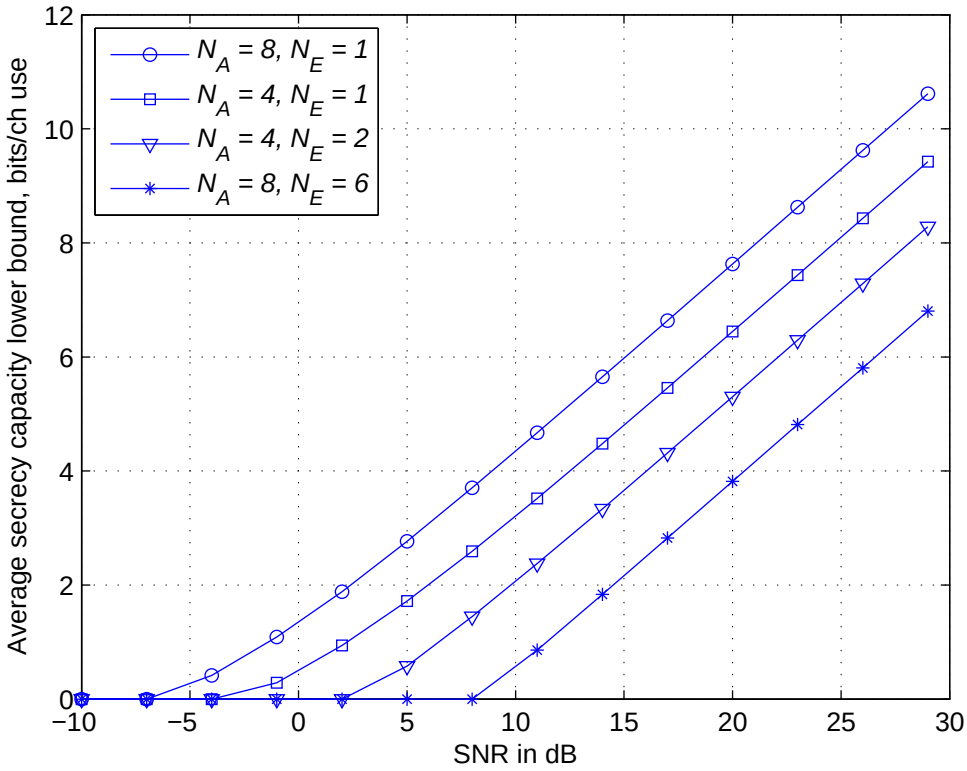

Figure 5.1: Average secrecy capacity lower bound C_S in (5.12) versus SNR $\mathcal{P}$ for systems with different numbers of antennas. The ratio of power allocation is set to $\phi = 0.5$.

$$- \sum_{k=0}^{N_E-1} \binom{N_A-1}{k} \frac{N_A-1}{z-1} \mathrm{B}(k+1, N_A-1-k)\, {}_2\mathrm{F}_1\Big(1, k+1; N_A; \frac{z-N_A}{z-1}\Big)\Big]^+, \tag{5.12}$$

where z is a constant independent of $\boldsymbol{h}$.

Fig. 5.1 shows the average secrecy capacity lower bound C_S in (5.12) for systems with different numbers of antennas. We see that the presence of multiple colluding Eves dramatically reduces the secrecy rate, compared with the case of non-colluding Eves. Furthermore, the secrecy rate quickly reduces to zero at low to moderate SNR values.

In the following subsections, we aim to give simplified or approximated expressions of the secrecy capacity lower bound in two special scenarios. These expressions will be used to obtain analytical results and useful insights on the optimal power allocation in Section 5.4. Note that the derived approximations may not be an achievable secrecy rate, although they are useful for the design of power allocation.

5.3.1 Non-colluding Eavesdroppers

In the case where Eves cannot collude, we have $N_E = 1$. Then C_2 in (5.10) reduces to

$$\begin{aligned} C_2 &= E_{\boldsymbol{h}}\Big\{\frac{1}{\ln 2}\frac{1}{z-1}\,{}_2\mathrm{F}_1\Big(1,1;N_A;\frac{z-N_A}{z-1}\Big)\Big\} \\ &= E_{\boldsymbol{h}}\left\{\frac{1}{\ln 2}\Big(\frac{N_A-1}{N_A-z}\Big)^{N_A-1}\left(\ln\Big(\frac{N_A-1}{z-1}\Big)-\sum_{l=1}^{N_A-2}\frac{1}{l}\Big(\frac{N_A-z}{N_A-1}\Big)^{l}\right)\right\}, \end{aligned} \tag{5.13}$$

where (5.13) is obtained using an identity for the Gauss hypergeometric function derived in Appendix C.1. This can then be substituted into $C_S = [C_1 - C_2]^+$ to yield simplified expressions for the average secrecy capacity lower bound.

5.3.2 Large Transmit Antenna Approximation

When the number of antennas at Alice goes large, the law of large numbers implies that $\lim_{N_A\to\infty}\|\boldsymbol{h}\|^2/N_A = 1$ and $\lim_{N_A\to\infty}\boldsymbol{G}_2\boldsymbol{G}_2{}^{\dagger}/(N_A-1) = \boldsymbol{I}$. Therefore, for sufficiently large N_A, we can approximate C_1 in (5.5) as

$$C_1 \;\approx\; \log_2\Big(1+\frac{N_A\mathcal{P}}{z}\Big). \tag{5.14}$$

It is clear from (5.14) that the optimal power allocation is independent of the channel from Alice to Bob, $\boldsymbol{h}$, when N_A becomes large. Also, C_2 in (5.7) can be approximated as

$$\begin{aligned} C_2 &\approx E_{\boldsymbol{g}_1}\Big\{\log_2\Big(1+\frac{1}{z-1}\|\boldsymbol{g}_1\|^2\Big)\Big\} \\ &= \frac{1}{\ln 2}\exp(z-1)\sum_{k=1}^{N_E}\mathrm{E}_k(z-1), \end{aligned} \tag{5.15}$$

where $\|\boldsymbol{g}_1\|^2$ has a Gamma distribution with parameters $(N_E, 1)$. It can be seen from (5.15) that C_2 is independent of N_A. Therefore, altering the number of antennas at Alice only affects the channel capacity between Alice and Bob but not the capacity between Alice and Eves, when N_A is large. The average secrecy capacity lower bound for large N_A is then approximated by

$$C_S \approx \frac{1}{\ln 2}\Big[\ln\Big(1+\frac{N_A\mathcal{P}}{z}\Big)-\exp(z-1)\sum_{k=1}^{N_E}\mathrm{E}_k(z-1)\Big]^+. \tag{5.16}$$

5.4 Optimal Power Allocation

In this section, we study the optimal power allocation between the information bearing signal and the artificial noise. As we have discussed, the power allocation strategy can be either adaptive or non-adaptive. The former depends on every realization of the channel gain while the latter is fixed for all channel realizations. The objective function for this optimization problem is the average secrecy capacity lower bound. The closed-form expressions derived in the previous section greatly reduce the computational complexity of the optimization process. In the following, we first study the case of non-colluding eavesdroppers and then look at the case of colluding eavesdroppers.

5.4.1 Non-colluding Eavesdroppers Case

The optimal value of ϕ or z can be easily found numerically using the capacity lower bound expressions derived in Section 5.3. Moveover, these expressions enable us to analytically obtain useful insights into the optimal z in the high SNR regime as follows.

In the high SNR regime, C_1 in (5.5) can be approximated as

$$\begin{aligned} C_1 &\approx E_{\boldsymbol{h}}\Big\{ \log_2 \Big(\frac{P}{z}\|\boldsymbol{h}\|^2\Big)\Big\} \\ &= E_{\boldsymbol{h}}\{\log_2(P\|\boldsymbol{h}\|^2)\} - E_{\boldsymbol{h}}\{\log_2 z\}. \end{aligned} \tag{5.17}$$

We see in (5.17) that $E_{\boldsymbol{h}}\{\log_2(P\|\boldsymbol{h}\|^2)\}$ is a constant and $E_{\boldsymbol{h}}\{\log_2 z\}$ does not directly depend on $\boldsymbol{h}$ although z may be a function of $\boldsymbol{h}$. Therefore, the high SNR approximation of the secrecy capacity lower bound does not have $\boldsymbol{h}$ in its expression (except for the expectation over $\boldsymbol{h}$). Consequently, for any value of $\boldsymbol{h}$, the optimal z that maximizes the high SNR approximation of the secrecy capacity lower bound is the same. In other words, the value of $\boldsymbol{h}$ is irrelevant in finding the optimal power allocation. Therefore, the adaptive power allocation strategy does not need to be considered at high SNR.

The optimal value of z in the high SNR regime satisfies

$$\frac{\mathrm{d}C_S}{\mathrm{d}z} = \frac{\mathrm{d}C_1}{\mathrm{d}z} - \frac{\mathrm{d}C_2}{\mathrm{d}z} = -\frac{1}{z\ln 2} - \frac{\mathrm{d}C_2}{\mathrm{d}z} = 0, \tag{5.18}$$

where the derivative of C_2 w.r.t. z can be computed in closed-form using (5.13).

In the special case of $N_A = 2$, (5.18) is reduces to

$$-\frac{1}{z} - \frac{1}{(z-2)(z-1)} + \frac{\ln(z-1)}{(z-2)^2} = 0. \tag{5.19}$$

The solution to the above equation is given by $z = 2$. It can be shown that $\lim_{z\to 2} \frac{\mathrm{d}^2 C_S}{\mathrm{d}z^2} < 0$. Hence the optimal ratio of power allocation is given by $\phi = 0.5$, that is to say, equal power allocation between the information signal and the artificial noise is the optimal strategy in the high SNR regime for $N_A = 2$.

For large N_A, using (5.15) with $N_E = 1$, we have

$$\begin{aligned} \frac{\mathrm{d}C_2}{\mathrm{d}z} &= \frac{1}{\ln 2}\Big(\exp(z-1)\mathrm{E}_1(z-1) - \exp(z-1)\mathrm{E}_0(z-1)\Big) \\ &= \frac{1}{\ln 2}\Big(\exp(z-1)\mathrm{E}_1(z-1) - (z-1)^{-1}\Big). \end{aligned} \tag{5.20}$$

Hence the optimal value of z satisfies

$$-\frac{1}{z} - e^{z-1}\mathrm{E}_1(z-1) + \frac{1}{z-1} = 0, \tag{5.21}$$

which gives $z = 1.80$. It can be shown that at $z = 1.80$, $\frac{\mathrm{d}^2 C_S}{\mathrm{d}z^2} < 0$. Hence the optimal ratio of power allocation is given by $\phi = 0.55$ in the high SNR regime for sufficiently large N_A. We see that the difference between the optimal values of ϕ for the smallest N_A (*i.e.*, $N_A = 2$) and asymptotically large N_A is very small. This observation suggests that the optimal power allocation is insensitive to the number of transmit antennas at high SNR.

Fig. 5.2 shows the optimal values of ϕ using the non-adaptive power allocation strategy for systems with different numbers of antennas at Alice N_A. The values of ϕ are shown for SNRs at which the average secrecy capacity lower bound is positive. The general trend is that more power needs to be allocated to the information signal as SNR or N_A increases. In the high SNR regime, we see that the optimal values of ϕ converge to constant values. For $N_A = 2$, the optimal value of ϕ converges to 0.5, which agrees with our analytical result. Furthermore, this constant value only increases slightly with N_A, and the maximum value is 0.55 which agrees with our large N_A analysis. These observations suggest that a near-optimal yet simple power allocation strategy at moderate to high SNR values is the equal power allocation between the information signal and the artificial noise.

Fig. 5.3 shows the average secrecy capacity lower bound C_S in (5.12) with numerically optimized ϕ using the non-adaptive power allocation strategy, indicated by the markers. For comparison, we also include the capacity lower bound with

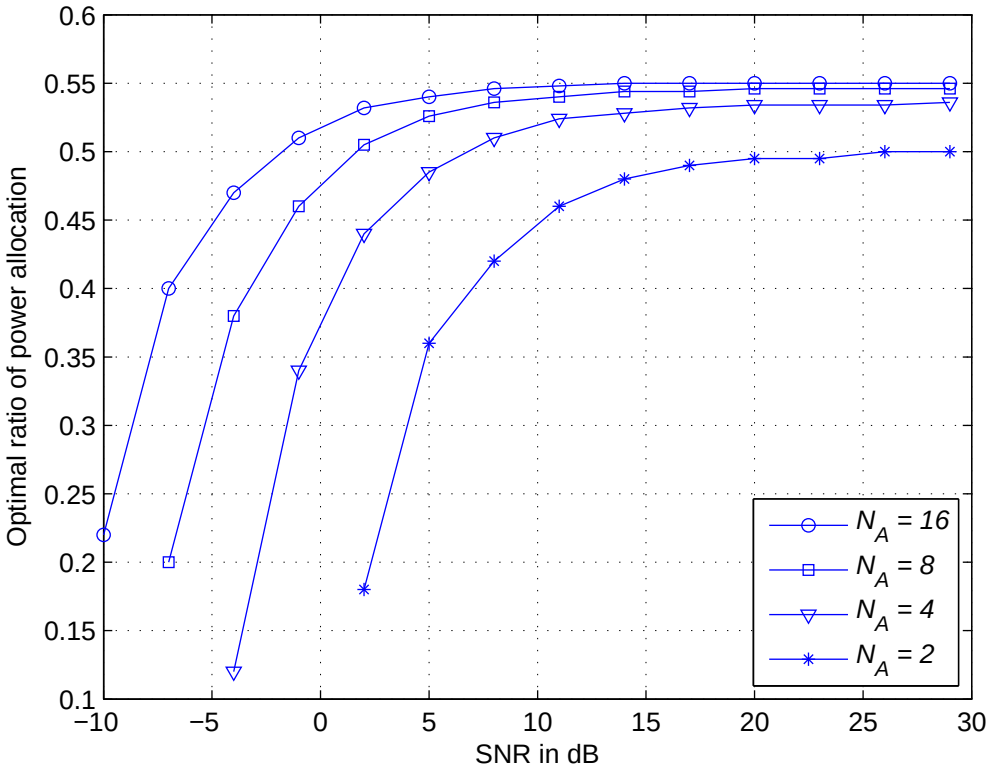

Figure 5.2: Optimal ratio of power allocation ϕ versus SNR $\mathcal{P}$ for different numbers of antennas at Alice N_A. The non-adaptive power allocation strategy is used. The values of ϕ are shown for SNRs at which the average secrecy capacity lower bound is positive.

equal power allocation, *i.e.*, $\phi = 0.5$, indicated by the solid lines. We see that the equal power allocation strategy achieves nearly the same secrecy rate as the optimal non-adaptive power allocation in all cases over a wide range of SNR values. This confirms that equal power allocation is a simple and generic strategy which yields close to optimal performance in terms of the derived achievable secrecy rate.

Fig. 5.4 shows the average secrecy capacity lower bound C_S in (5.11) and (5.12) with the optimized ϕ, using both the adaptive and non-adaptive power allocation strategies. We see that there is no difference between C_S achieved by the adaptive and non-adaptive strategies over a wide range of SNR values. The adaptive strategy only gives marginal advantage when the secrecy rate is close to zero. This result suggests that the non-adaptive power allocation strategy is sufficient to achieve almost the best possible secrecy rate performance.[2] For this reason, we will only focus on the non-adaptive scheme in the rest of this chapter.

[2]The same result is found for the colluding Eves case. The numerical results are omitted for brevity.

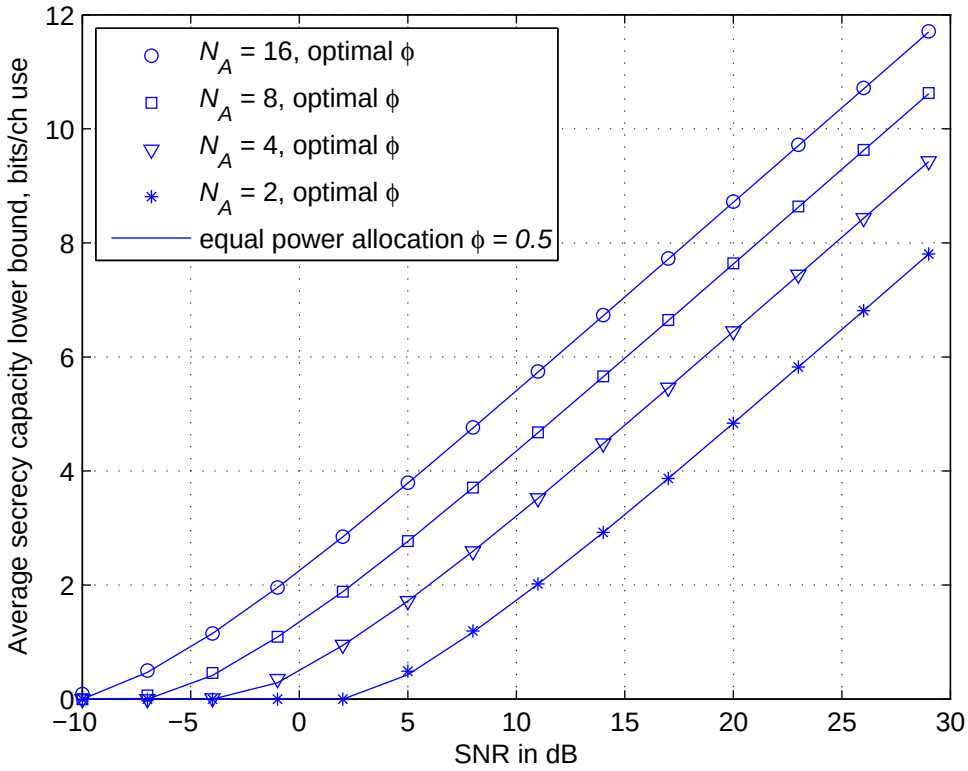

Figure 5.3: Average secrecy capacity lower bound C_S in (5.12) versus SNR $\mathcal{P}$ for different numbers of antennas at Alice N_A. The non-adaptive power allocation strategy with numerically optimized ϕ is used. The average capacity lower bound with equal power allocation for each case, indicated by the solid line, is also shown for comparison.

5.4.2 Colluding Eavesdroppers Case

As we have seen in Fig. 5.1, the presence of multiple colluding Eves severely degrades the secrecy rate. Therefore, it is essential for Alice to have a relatively large number of antennas to maintain a good secure communication link. For any value of N_E, the optimal value of ϕ or z can be easily found numerically using the closed-form capacity lower bound expression given in Section 5.3. As the number of antennas at Alice is desired to be large, we carry out large N_A analysis to obtain an asymptotic result on optimal z in the high SNR regime as follows.

In the high SNR regime with large N_A, C_S in (5.16) can be approximated as

$$C_S \approx \frac{1}{\ln 2}\Big[\ln(N_A\mathcal{P}) - \ln z - \exp(z-1)\sum_{k=1}^{N_E} \mathrm{E}_k(z-1)\Big]. \tag{5.22}$$

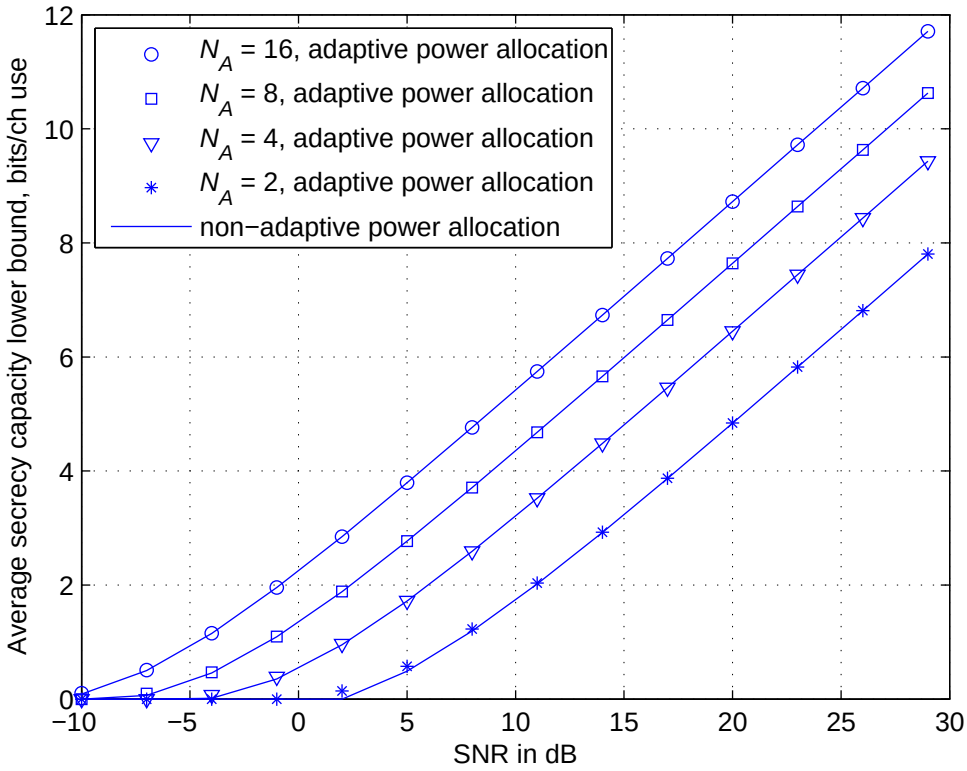

Figure 5.4: Average secrecy capacity lower bound C_S in (5.11) and (5.12) versus SNR $\mathcal{P}$ for different numbers of antennas at Alice N_A. Both the adaptive and non-adaptive power allocation strategies are used, indicated by the markers and the lines, respectively.

By taking the derivative of C_S w.r.t. z, the optimal z satisfies

$$-\frac{1}{z} - e^{z-1}\mathrm{E}_{N_E}(z-1) + \frac{1}{z-1} = 0. \tag{5.23}$$

Using $e^{z-1}\mathrm{E}_{N_E}(z-1) \approx (z-1+N_E)^{-1}$ from [116], which is accurate when either N_E or z is large, (5.23) reduces to

$$-\frac{1}{z} - \frac{1}{z-1+N_E} + \frac{1}{z-1} = 0. \tag{5.24}$$

And hence the optimal z is given by

$$z^* = 1 + \sqrt{N_E}. \tag{5.25}$$

From (5.25) we see that the optimal value of z only depends on N_E in the high SNR and large antenna regime. Moreover, (5.25) suggests that more power should be used to generate artificial noise when the number of Eves increases.

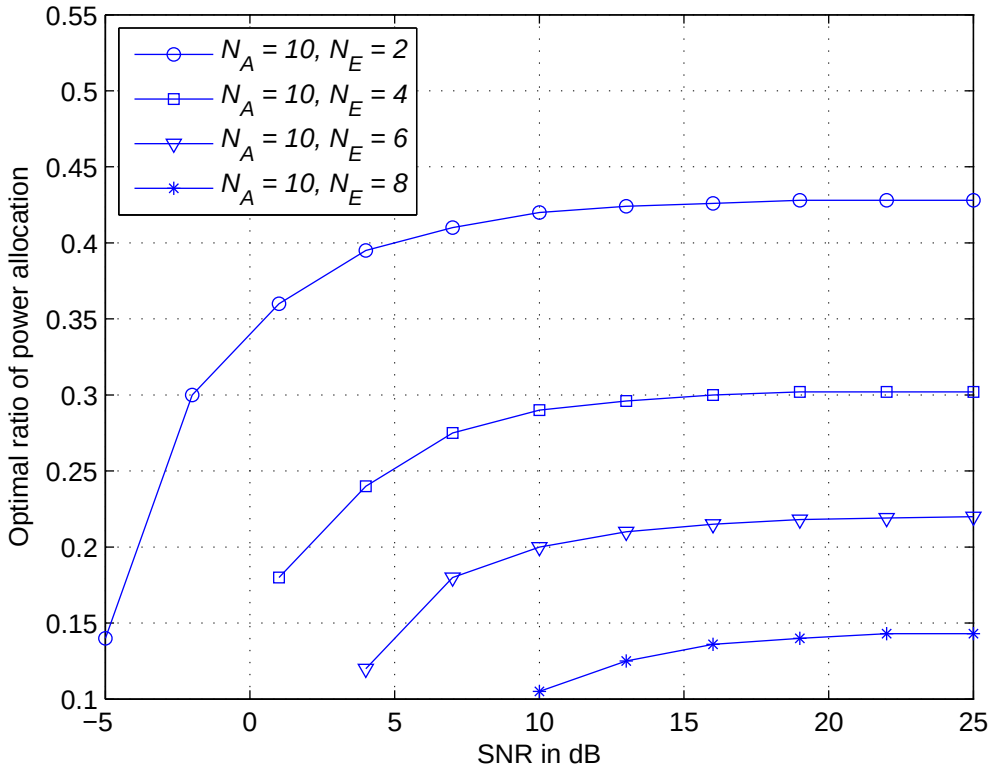

Figure 5.5: Optimal ratio of power allocation ϕ versus SNR $\mathcal{P}$ for systems with different numbers of colluding Eves N_E. The values of ϕ are shown for SNRs at which the average secrecy capacity lower bound is positive.

Fig. 5.5 shows the optimal value of ϕ for systems with different numbers of colluding Eves N_E. Similar to the non-colluding Eves case, we see that more power should be used to transmit the information signal as SNR increases. The optimal value of ϕ stays constant in the high SNR regime. Furthermore, the optimal value of ϕ for colluding Eves case is usually much smaller than 0.5, *i.e.*, equal power allocation, which is near optimal for non-colluding Eves case. In particular, the optimal ϕ reduces as N_E grows, which implies that more power should be allocated to generate the artificial noise as the number of colluding Eves increases. This observation agrees with our analytical insight and intuition.

Fig. 5.6 shows the average secrecy capacity lower bound C_S in (5.12) for systems with different N_E. Here, we investigate the sensitivity in the secrecy rate to the design of power allocation. Consider a scenario where the total number of Eves that can collude is 8, and hence Alice has optimized ϕ for $N_E = 8$. When N_E changes, the power allocation parameter ϕ does not need to be optimized again as long as N_E stays reasonably close to 8, *e.g.*, $N_E = 6$, since the value of ϕ optimized for $N_E = 8$ still works well for $N_E = 6$ (with a power loss of 0.2 dB) as shown in

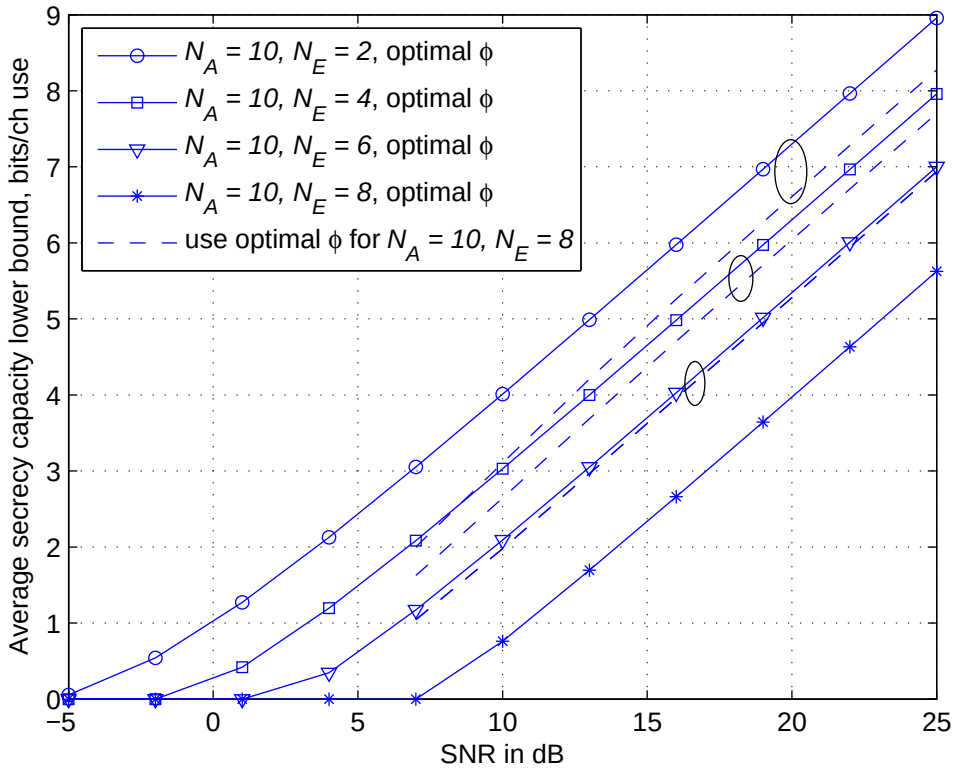

Figure 5.6: Average secrecy capacity lower bound C_S in (5.12) versus SNR $\mathcal{P}$ for systems with different numbers of colluding Eves N_E. The solid lines with markers indicate C_S achieved with optimal values of ϕ for the corresponding system. The dashed lines indicate C_S achieved with value of ϕ optimized for $N_E = 8$, which represents the case where the power allocation was initially designed for $N_E = 8$, but the current value of N_E reduces from 8 and the power allocation is not redesigned.

Fig. 5.6. However, redesigning of ϕ becomes important when N_E is considerably different from 8, *e.g.*, $N_E = 2$ to 4. For example, if N_E changes from 8 to 4, a power loss of approximately 1 dB will incur if Alice still uses the value of ϕ optimized for $N_E = 8$, as shown in Fig. 5.6.

We also provide numerical verification of the optimal power allocation obtained from the large antenna approximation in the high SNR regime. Fig. 5.7 shows the ratio of power allocation ϕ at $\mathcal{P} = 20$ dB versus the number of antennas at Alice N_A for systems with different numbers of colluding Eves N_E. For a fixed N_E, we see that the optimal value of ϕ increases with N_A and reaches a constant value when N_A is sufficiently large. This agrees with our analytical insight that the optimal power allocation does not depend on N_A when N_A is large. The asymptotic constant value of ϕ is close to the analytical value given in (5.25) obtained from

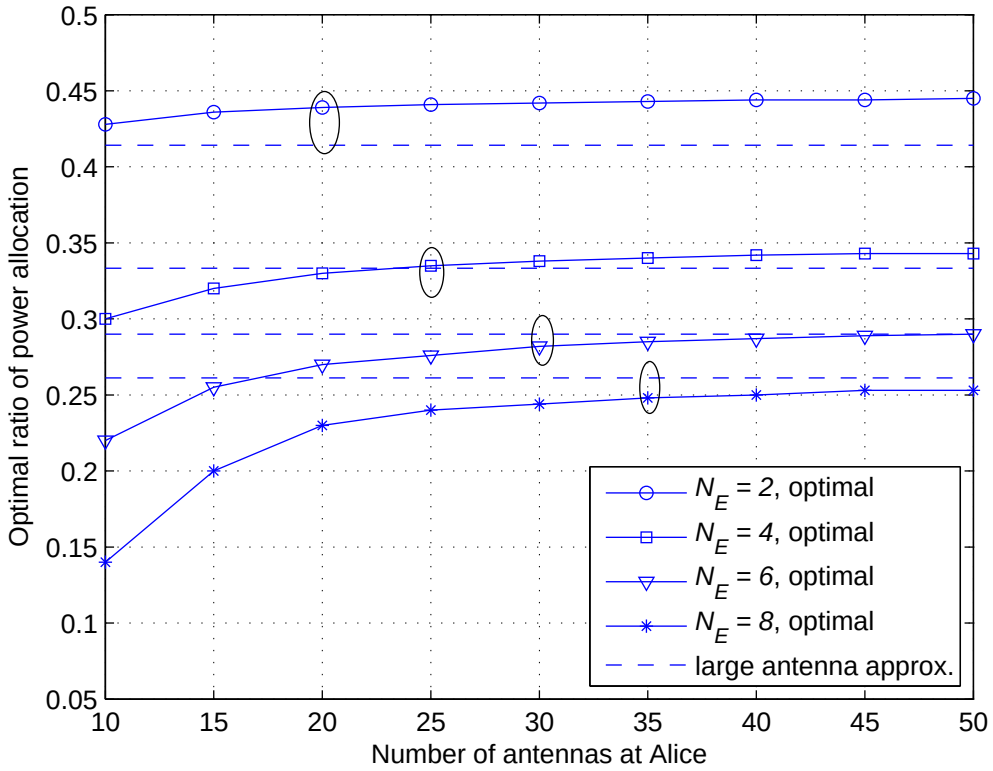

Figure 5.7: The ratio of power allocation ϕ at $\mathcal{P} = 20$ dB versus the number of antennas at Alice N_A for systems with different numbers of colluding Eves N_E. The solid lines with markers indicate the optimal values of ϕ, while the dashed lines indicate the values of ϕ from the large antenna approximation given in (5.25).

the large antenna approximation.

5.5 Critical SNR for Secure Communications

Another important aspect of secure communication is the minimum SNR required for positive secrecy rate, which is a critical parameter in wideband communications. With the closed-form expression of the secrecy capacity lower bound derived in Section 5.3, one can numerically find the critical SNR with low computational complexity. In this section, we derive a closed-form upper bound on the critical SNR which is useful in the design of wideband secure communications.

Using properties of the exponential integral function in [116], (5.6) can be bounded from below as

$$C_1 \;>\; \frac{1}{\ln 2}\sum_{k=1}^{N_A}\frac{1}{\frac{z}{\mathcal{P}}+k}, \tag{5.26}$$

which is asymptotically tight as the SNR approaches zero, *i.e.*, $\mathcal{P} \to 0$. Using the convexity of (5.26) in k, we can further bound C_1 as

$$C_1 \;>\; \frac{1}{\ln 2}\frac{N_A}{\frac{z}{\mathcal{P}}+\frac{N_A+1}{2}}, \tag{5.27}$$

which is also asymptotically tight as the SNR approaches zero. Using the lower bound on C_1 in (5.27) and C_2 in (5.10), the average secrecy capacity lower bound can be further bounded from below as

$$\begin{aligned} C_S \;>\; & \frac{1}{\ln 2}\frac{N_A}{\frac{z}{\mathcal{P}}+\frac{N_A+1}{2}} \\ & -\frac{1}{\ln 2}\sum_{k=0}^{N_E-1}\binom{N_A-1}{k}\frac{N_A-1}{z-1}\mathrm{B}(k+1,N_A-1-k)\,{}_2\mathrm{F}_1\Big(1,k+1;N_A;\frac{z-N_A}{z-1}\Big). \end{aligned} \tag{5.28}$$

The critical SNR, denoted by $\mathcal{P}_C$, is the SNR at which C_S drops to zero. With the lower bound on C_S given in (5.28), an upper bound on $\mathcal{P}_C$ can be found as

$$\mathcal{P}_C < z\left[\frac{N_A}{\sum_{k=0}^{N_E-1}\binom{N_A-1}{k}\frac{N_A-1}{z-1}\mathrm{B}(k+1,N_A-1-k)\,{}_2\mathrm{F}_1\Big(1,k+1;N_A;\frac{z-N_A}{z-1}\Big)}-\frac{N_A+1}{2}\right]^{-1}. \tag{5.29}$$

In the case of non-colluding eavesdroppers, *i.e.*, $N_E = 1$, (5.29) reduces to

$$\mathcal{P}_C < z\left[\frac{N_A}{\left(\frac{N_A-1}{N_A-z}\right)^{N_A-1}\left(\ln\left(\frac{N_A-1}{z-1}\right)-\sum_{l=1}^{N_A-2}\frac{1}{l}\left(\frac{N_A-z}{N_A-1}\right)^l\right)}-\frac{N_A+1}{2}\right]^{-1}. \tag{5.30}$$

The upper bound in (5.29) or (5.30) indicates a minimum SNR that guarantees positive secrecy rate. Since (5.29) and (5.30) are asymptotically tight at low SNR, they can be used to fine tune the power allocation parameter z to minimize $\mathcal{P}_C$.

Fig. 5.8 shows the critical SNR $\mathcal{P}_C$ versus number of antennas at Alice N_A for systems with different numbers of colluding Eves N_E. The power allocation is set to $\phi = 0.2$ in all cases. The general trend is that $\mathcal{P}_C$ decreases as N_A increases, and a higher $\mathcal{P}_C$ is required when N_E increases. These observations agree with intuition. Furthermore, we see that the analytical upper bound on $\mathcal{P}_C$ is very accurate for the case of non-colluding Eves. For the case of colluding Eves, the upper bound is reasonably accurate when $\mathcal{P}_C < 0$ dB. The difference between the exact value

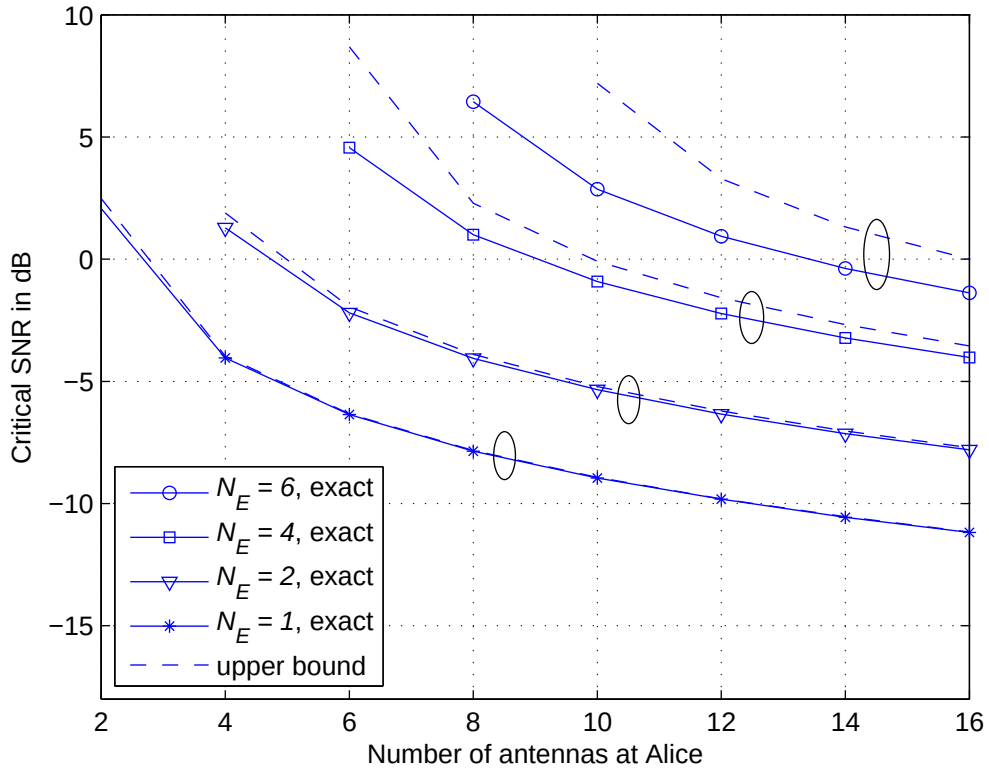

Figure 5.8: The critical SNR $\mathcal{P}_C$ versus number of antennas at Alice N_A for systems with different numbers of colluding Eves N_E. The ratio of power allocation is set to $\phi = 0.2$. The solid lines with markers indicate the exact value of $\mathcal{P}_C$, while the dashed lines indicate the analytical upper bound given in (5.29).

of $\mathcal{P}_C$ and its upper bound gradually increases as N_E increases, which is mainly due to the increase in $\mathcal{P}_C$. When N_E is relatively large, *e.g.*, $N_E = 6$, one should allocate more power to generate the artificial noise (*i.e.*, reduce ϕ), as suggested in Fig. 5.5, in order to achieve a lower $\mathcal{P}_C$, which in turn makes the bound tighter.

5.6 Effect of Imperfect Channel State Information

So far, we have assumed that the CSI can be perfectly obtained at Alice and Bob. In this section, we investigate the effect of imperfect CSI by considering channel estimation errors. With imperfect CSI, the beamforming transmission from Alice to Bob is designed based on the estimated channel gains rather than the true channel gains. Therefore, the artificial noise leaks into Bob's channel.

To incorporate channel estimation errors, we consider that Bob performs the

LMMSE channel estimation. Therefore, we have

$$\begin{aligned} \boldsymbol{h} &= \widehat{\boldsymbol{h}} + \widetilde{\boldsymbol{h}}, && (5.31)\\ \sigma^2_{\boldsymbol{h}} &= \sigma^2_{\widehat{\boldsymbol{h}}} + \sigma^2_{\widetilde{\boldsymbol{h}}}, && (5.32) \end{aligned}$$

where $\widehat{\boldsymbol{h}}$ denotes the channel estimate and $\widetilde{\boldsymbol{h}}$ denotes the estimation error. $\sigma^2_{\boldsymbol{h}}$ denotes the variance of each element in $\boldsymbol{h}$. $\sigma^2_{\widehat{\boldsymbol{h}}}$ and $\sigma^2_{\widetilde{\boldsymbol{h}}}$ denote the variance of each element in $\widehat{\boldsymbol{h}}$ and $\widetilde{\boldsymbol{h}}$, respectively. As a general property of the LMMSE estimator for Gaussian signals [117], $\widehat{\boldsymbol{h}}$ and $\widetilde{\boldsymbol{h}}$ are uncorrelated, each having i.i.d. complex Gaussian entries.

Similar to our system model in Section 5.2, we assume that the knowledge of $\widehat{\boldsymbol{h}}$ is available at Alice and Eves. Therefore, the beamforming vector becomes $\boldsymbol{w}_1 = \widehat{\boldsymbol{h}}^{\dagger}/\|\widehat{\boldsymbol{h}}\|$, and the received symbol at Bob is given by

$$\begin{aligned} y_B &= \boldsymbol{h}\boldsymbol{x} + n \\ &= \widehat{\boldsymbol{h}}\boldsymbol{x} + \widetilde{\boldsymbol{h}}\boldsymbol{x} + n \\ &= \|\widehat{\boldsymbol{h}}\|u + \widetilde{\boldsymbol{h}}\boldsymbol{W}[u \;\; \boldsymbol{v}^T]^T + n. && (5.33) \end{aligned}$$

A capacity lower bound for the channel between Alice and Bob can be obtained by considering $\widetilde{\boldsymbol{h}}\boldsymbol{W}[u \;\; \boldsymbol{v}^T]^T + n$ as the worst case Gaussian noise. Note that $\boldsymbol{W}$ is a unitary matrix, hence $\widetilde{\boldsymbol{h}}\boldsymbol{W}$ has the same distribution as $\widetilde{\boldsymbol{h}}$. Therefore, the average capacity lower bound for the channel between Alice and Bob is given by

$$\widehat{C}_1 = E_{\widehat{\boldsymbol{h}}}\Big\{ \log_2\Big(1 + \frac{\sigma_u^2\|\widehat{\boldsymbol{h}}\|^2}{\sigma^2_{\widetilde{\boldsymbol{h}}}\mathcal{P} + 1}\Big)\Big\}. \quad (5.34)$$

With $\sigma^2_{\boldsymbol{h}}$ normalized to unity, we have $\sigma^2_{\widehat{\boldsymbol{H}}} = 1 - \sigma^2_{\widetilde{\boldsymbol{H}}}$. Since the elements of $\widehat{\boldsymbol{h}}$ is i.i.d. complex Gaussian, $\|\widehat{\boldsymbol{h}}\|^2$ is a sum of i.i.d. exponential distributed random variables, which follows a Gamma distribution with parameter $(N_A, 1 - \sigma^2_{\widetilde{\boldsymbol{H}}})$. Therefore, we obtain a closed-form expression for $\widehat{C}_1$ as

$$\widehat{C}_1 = \frac{1}{\ln 2}\exp\Big(z\frac{\sigma^2_{\widetilde{\boldsymbol{h}}} + \mathcal{P}^{-1}}{1 - \sigma^2_{\widetilde{\boldsymbol{h}}}}\Big)\sum_{k=1}^{N_A}\mathrm{E}_k\Big(z\frac{\sigma^2_{\widetilde{\boldsymbol{h}}} + \mathcal{P}^{-1}}{1 - \sigma^2_{\widetilde{\boldsymbol{h}}}}\Big). \quad (5.35)$$

The presence of channel estimation errors at Bob does not affect the signal reception at Eves given in (5.2). Therefore, the average secrecy capacity lower

bound can be obtained by subtracting C_2 from $\widehat{C}_1$ as

$$C_S = \frac{1}{\ln 2}\Big[\exp\Big(z\frac{\sigma^2_{\tilde{\boldsymbol{h}}}+\mathcal{P}^{-1}}{1-\sigma^2_{\tilde{\boldsymbol{h}}}}\Big)\sum_{k=1}^{N_A}\mathrm{E}_k\Big(z\frac{\sigma^2_{\tilde{\boldsymbol{h}}}+\mathcal{P}^{-1}}{1-\sigma^2_{\tilde{\boldsymbol{h}}}}\Big) - \sum_{k=0}^{N_E-1}\binom{N_A-1}{k}\frac{N_A-1}{z-1}\mathrm{B}(k+1,N_A-1-k)\,{}_2\mathrm{F}_1\Big(1,k+1;N_A;\frac{z-N_A}{z-1}\Big)\Big]^+. \tag{5.36}$$

Following the steps in Section 5.5, we can also bound C_S from below in order to obtain an upper bound on the critical SNR for secure communication with channel estimation errors as

$$C_S > \frac{1}{\ln 2}\frac{N_A}{z\frac{\sigma^2_{\tilde{\boldsymbol{h}}}+\mathcal{P}^{-1}}{1-\sigma^2_{\tilde{\boldsymbol{h}}}}+\frac{N_A+1}{2}} - \frac{1}{\ln 2}\sum_{k=0}^{N_E-1}\binom{N_A-1}{k}\frac{N_A-1}{z-1}\mathrm{B}(k+1,N_A-1-k)\,{}_2\mathrm{F}_1\Big(1,k+1;N_A;\frac{z-N_A}{z-1}\Big). \tag{5.37}$$

And the upper bound on the critical SNR is then given by

$$\mathcal{P}_C < \Big[\frac{1-\sigma^2_{\tilde{\boldsymbol{h}}}}{z}\Big(\frac{N_A}{\sum_{k=0}^{N_E-1}\binom{N_A-1}{k}\frac{N_A-1}{z-1}\mathrm{B}(k+1,N_A-1-k)\,{}_2\mathrm{F}_1\Big(1,k+1;N_A;\frac{z-N_A}{z-1}\Big)} - \frac{N_A+1}{2}\Big) - \sigma^2_{\tilde{\boldsymbol{h}}}\Big]^{-1}, \tag{5.38}$$

which is asymptotically tight at low SNR.

We now present numerical results on the optimal power allocation as well as critical SNR in the presence of the channel estimation errors. For brevity, we focus on the case of non-colluding eavesdroppers. The trends on the effect of channel estimation errors observed in the following results also apply to the case of colluding eavesdroppers.

Fig. 5.9 shows the optimal ratio of power allocation ϕ with different channel estimation error variances $\sigma^2_{\tilde{\boldsymbol{h}}}$. We see that the channel estimation error has noticeable impact on the value of ϕ, especially for small number of antennas at Alice, *e.g.*, $N_A = 2$. The general trend is that less power should be allocated to information signal as channel estimation error increases. This is mainly due to the fact that the efficiency of improving Bob's signal reception by boosting the transmit power of the information signal reduces as the channel estimation error increases.

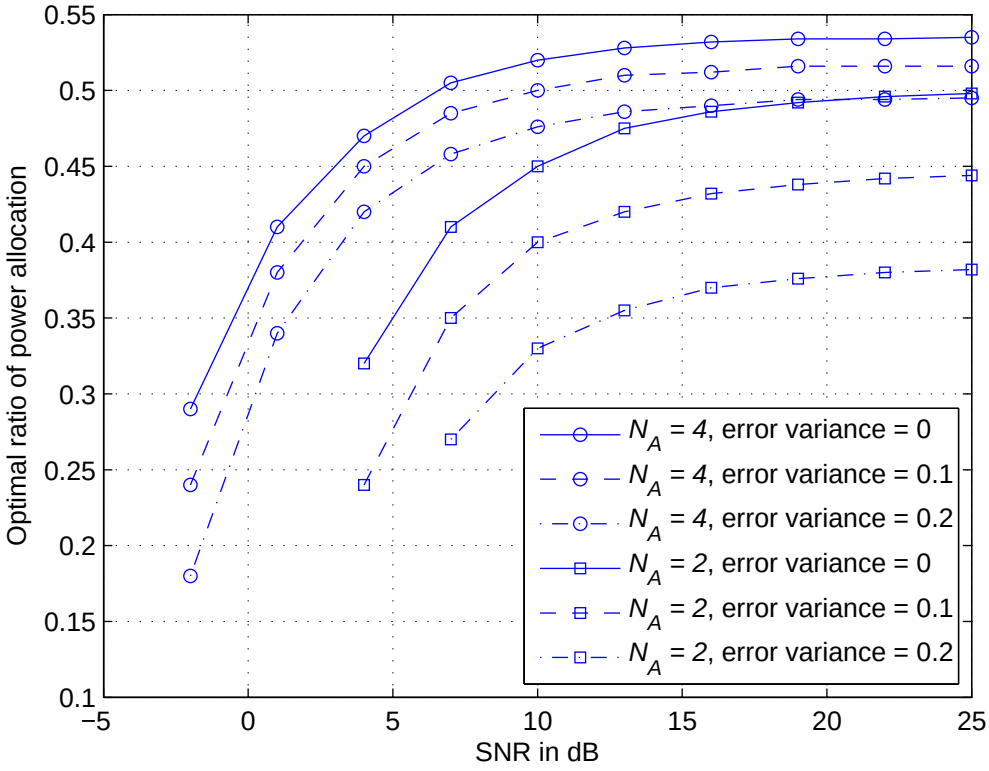

Figure 5.9: Optimal ratio of power allocation ϕ versus SNR $\mathcal{P}$ for different numbers of antennas at Alice N_A and different variances of the channel estimation errors $\sigma^2_{\tilde{\boldsymbol{h}}}$. The values of ϕ are shown for SNRs at which the average secrecy capacity lower bound is positive.

On the other hand, the efficiency of degrading Eve's signal reception by boosting the transmit power of the artificial noise stays the same regardless of the channel estimation error. Hence, it is better to create more noise for Eves than to increase the signal strength for Bob if the CSI is not accurately obtained.

In practical systems, the channel estimation error usually reduces as the SNR increases, although their exact relationship depends on the training design. From Fig. 5.9, we can expect that at low to moderate SNR where the channel estimation error is usually noticeable, the optimal power allocation between the transmission of information signal and artificial noise is very different from that in the perfect CSI case. While at high SNR where the channel estimation error is usually small, the optimal power allocation is expected to be very close to that of the perfect CSI case. Therefore, in practical systems it is important to take channel estimation error into account when designing the power allocation at relatively low SNR.

Table 5.1 lists the exact values of the critical SNR $\mathcal{P}_C$ as well as the closed-form upper bound given in (5.38) with $\phi = 0.5$. The general trend is that the

Table 5.1: Critical SNR (in dB) for Secure Communications with Equal Power Allocation

	Error variance $\sigma^2_{\tilde{h}}$	Number of antennas N_A				
		2	4	6	8	10
Exact	0	3.01	-2.62	-4.89	-6.36	-7.45
Upper bound	0	6.02	-1.97	-4.46	-6.01	-7.14
Exact	0.1	4.56	-1.88	-4.27	-5.79	-6.90
Upper bound	0.1	9.03	-1.20	-3.83	-5.43	-6.59
Exact	0.2	6.99	-1.01	-3.55	-5.13	-6.28
Upper bound	0.2	∞	-0.26	-3.08	-4.76	-5.96

critical SNR increases as the channel estimation error increases, which agrees with intuition. The upper bound gets tighter as $\mathcal{P}_C$ reduces (or N_A increases), and is accurate for $N_A \geq 4$ with an error of less than 1 dB.

5.7 Summary of Contributions

In this chapter we have considered secure communication in the wireless fading environment in the presence of non-colluding or colluding eavesdroppers. The transmitter is equipped with multiple antennas and is able to simultaneously transmit an information signal to the intended receiver and artificial noise to confuse the eavesdroppers. Some specific contributions made in this chapter are as follows.

Addressing Q1 in Section 5.1:

- We have obtained a closed-form expression for the average secrecy capacity lower bound, which is the maximum data rate that can be guaranteed for secure communication without knowing the noise level at the eavesdroppers.

Addressing Q2 in Section 5.1:

- We have studied the optimal power allocation between transmission of the information signal and the artificial noise. In particular, equal power allocation is shown to be a near optimal strategy in the case of non-colluding eavesdroppers. When the number of colluding eavesdroppers increases, more power should be used to generate artificial noise.

Addressing Q3 in Section 5.1:

- We have also derived an upper bound on the critical SNR above which the secrecy rate is positive and this bound is shown to be tight at low SNR.

Addressing Q4 in Section 5.1:

- When imperfect CSI is considered in the form of channel estimation errors, we have found that it is wise to create more artificial noise to confuse the eavesdroppers than to increase the signal strength for the intended receiver.

Chapter 6

Conclusions and Future Research Directions

In this chapter we state the general conclusions drawn from this thesis. The summary of specific contributions can be found at the end of each chapter and are not repeated here. We also outline some future research directions arising from this work.

6.1 Conclusions

This thesis has been primarily concerned with multi-antenna wireless communication systems. With limited amount of resource available for transmission, we have obtained practical design guidelines on efficient resource allocation for systems with channel uncertainty and/or security constraints.

In training-based transmission schemes, pilot symbols are periodically inserted into data blocks to facilitate channel estimation. We have extended the existing results on the optimal transmission resource allocation for systems with no transmitter-side channel knowledge to systems with various forms of transmitter-side channel knowledge. Simple closed-form solutions were obtained for transmit energy allocation between pilot and data symbols that achieve (near) optimal capacity performance for both spatially i.i.d. and correlated channels. The optimal training was shown to be at most as large as the number of transmit antennas for generally correlated channels. We have also found that the channel spatial correlation improves the information capacity at low SNR due to the reduction in the channel estimation errors.

We have extended our study on the optimal transmission resource allocation from traditional one-way training-based schemes to two-way training-based schemes,

where the transmitter and receiver channel state information is obtained from a reverse training and a forward training prior to the data transmission. Near optimal solutions on the transmit power distribution between the different training phases and the data transmission phase were obtained. Compared with systems with only reverse training, we have shown that the use of two-way training can provide noticeable performance improvement when the system is operating at moderate to high SNR and using high-order modulations. While this improvement from two-way training is insignificant at low SNR or when low-order modulations are used.

For multi-antenna transmission in the presence of eavesdroppers, we have considered the use of artificial noise for guaranteeing physical-layer wireless security. We have obtained an achievable secrecy rate and studied the optimal transmit power allocation between the information-bearing signal and the artificial noise. We have shown that equal power allocation performs well when the eavesdroppers cannot collude, while less power should be assigned to the information signal when either the eavesdroppers collude or the channel estimation error increases.

6.2 Future Research Directions

A number of future research directions arise from the work presented in this thesis.

Optimal Transmit Resource Allocation with Channel Estimation Errors and Finite-Rate Feedback: In Chapter 3 of this thesis, we assumed that the feedback link from the receiver to the transmitter is noiseless. In practical systems, however, finite-rate feedback is used and hence, feedback errors always exist. The optimal transmission resource allocation which takes into account both the channel estimation errors and feedback errors becomes an interesting and important design problem. Initial results can be found in [118,119] where beamforming transmission is considered.

Optimal Transmit Resource Allocation in Training-Based Systems in the Wideband Regime: In wideband communications, it is desirable to reduce the transmit power to improve the spectral efficiency. Hence, the system operates in the low SNR regime. Flash signalling was shown to be the optimal transmission scheme, in which only a portion of the transmit time is used [120]. In the training-based system, the portion of total available time to be used for transmission is an important design parameter. We have obtained initial results for systems with no transmitter-side channel knowledge and fixed power transmission in [121], while this problem remains challenging for systems with any forms of transmitter-side

channel knowledge.

Optimizing Two-Way Training-Based Multi-User MIMO Systems: In Chapter 4 of this thesis, we only considered the single-user case. When multiple users are considered, the precoding and scheduling come into the design problem. One starting point for future research is the case of single-user beamforming with user selection, where the analytical approach in studying the transmit power distribution in this thesis is applicable.

Physical-Layer Security in Wireless Random Networks: Very few results on information-theoretic security are concerned with large-scale wireless networks [122–125]. Unlike point-to-point communications, security in wireless networks strongly depends on the spatial distribution of both the legitimate nodes and the eavesdroppers. We have obtained initial results on secrecy outage constrained connectivity performance with different eavesdropping strategies [126]. The use of artificial noise or interference for improving wireless network secrecy is still an open problem.

Appendix A

A.1 Proof of Theorem 2.1

Finding the bounds on L_d^* and L_p^* involves two steps. First we find the optimal L_p for any given L_d. We then find the range of values in which L_d^* lies. For any given L_d, the capacity lower bound (2.13) is maximized when the effective SNR (2.14) is maximized. Letting the first derivative of ρ_{eff} w.r.t. L_p be zero, we obtain only one positive root for L_p given by

$$L_p^0 = \arg_{L_p}\left\{\frac{\mathrm{d}\rho_{\text{eff}}}{\mathrm{d}L_p} = 0\right\} = \sqrt{\frac{N_t L_d(L_d + L\mathcal{P})}{N_t + L\mathcal{P}}}. \tag{A.1}$$

Note that the training length must satisfy $N_t \leq L_p \leq L - L_d$. Therefore, for a given L_d the optimal training length is given by

$$L_p^* = \begin{cases} N_t, & \text{for } L_p^0 \leq N_t \\ L_p^0, & \text{for } N_t \leq L_p^0 \leq L - L_d \\ L - L_d, & \text{for } L - L_d \leq L_p^0, \end{cases} \tag{A.2}$$

which by referring to (A.1) is equivalent to

$$L_p^* = \begin{cases} N_t, & \text{for } 1 \leq L_d \leq N_t \\ L_p^0, & \text{for } N_t \leq L_d \leq \zeta \\ L - L_d, & \text{for } \zeta \leq L_d \leq L - N_t, \end{cases} \tag{A.3}$$

where

$$\zeta = \frac{2L\mathcal{P}+N_t\mathcal{P}+2N_t-\sqrt{N_t(N_t\mathcal{P}^2+4N_t\mathcal{P}+4N_t+4L\mathcal{P}^2+4L\mathcal{P})}}{2\mathcal{P}} \tag{A.4}$$

is the value of L_d at which $\sqrt{\frac{N_t L_d(L_d+L\mathcal{P})}{N_t+L\mathcal{P}}} + L_d = L$.

In order to find the range of values in which L_d^* lies, we consider two different

cases, *i.e.*, $L \leq 2N_t$ and $L > 2N_t$.

When $L \leq 2N_t$, using (A.4) it can be shown that $L - N_t \leq \zeta \leq N_t$. Therefore, only the first range of L_d in (A.3) is feasible and $L_p^* = N_t$. In order to show the optimal data length takes the maximum possible value, *i.e.*, $L_d^* = L - L_p^* = L - N_t$, we need to show that the derivative of $\overline{C}_{\text{LB}}$ w.r.t. L_d is positive at $L_p^* = N_t$. Using (2.13) we have

$$\begin{aligned}\frac{\mathrm{d}\overline{C}_{\text{LB}}}{\mathrm{d}L_d} &= \frac{n}{L\ln 2}E_\lambda\Big\{\ln(1+\rho_{\text{eff}}\lambda) + \frac{\lambda L_d}{1+\rho_{\text{eff}}\lambda}\frac{\mathrm{d}\rho_{\text{eff}}}{\mathrm{d}L_d}\Big\} \\ &= \frac{n}{L\ln 2}E_\lambda\Big\{\ln(1+\rho_{\text{eff}}\lambda) - \frac{\rho_{\text{eff}}\lambda}{1+\rho_{\text{eff}}\lambda}L_d\rho_{\text{eff}}\frac{\mathrm{d}\rho_{\text{eff}}^{-1}}{\mathrm{d}L_d}\Big\}.\end{aligned}$$

Since both ρ_{eff} and λ are non-negative, it can be shown that

$$\ln(1+\rho_{\text{eff}}\lambda) - \frac{\rho_{\text{eff}}\lambda}{1+\rho_{\text{eff}}\lambda} \geq 0.$$

Therefore, $L_d\rho_{\text{eff}}\frac{\mathrm{d}}{\mathrm{d}L_d}\rho_{\text{eff}}^{-1} < 1$ implies $\frac{\mathrm{d}}{\mathrm{d}L_d}\overline{C}_{\text{LB}} > 0$. Using (2.14) it is then easy to show that $L_d\rho_{\text{eff}}\frac{\mathrm{d}}{\mathrm{d}L_d}\rho_{\text{eff}}^{-1} < 1$ at $L_p^* = N_t$. Hence we obtain $L_d^* = L - N_t$.

When $L > 2N_t$, we aim to show that the optimal data length resides in the last range in (A.3), *i.e.*, $\zeta \leq L_d^* \leq L - N_t$. This is proved by showing that L_d^* cannot reside in $1 \leq L_d < N_t$ and $N_t \leq L_d < \zeta$. When $1 \leq L_d < N_t$, it is easy to show that $L_d\rho_{\text{eff}}\frac{\mathrm{d}}{\mathrm{d}L_d}\rho_{\text{eff}}^{-1} < 1$ at $L_p^* = N_t$ given by (A.3). This implies $\frac{\mathrm{d}}{\mathrm{d}L_d}\overline{C}_{\text{LB}} > 0$ and hence L_d^* is not in $1 \leq L_d < N_t$. When $N_t \leq L_d < \zeta$, it is easy to show that $L_d\rho_{\text{eff}}\frac{\mathrm{d}}{\mathrm{d}L_d}\rho_{\text{eff}}^{-1} < 1$ at $L_p^* = L_p^0$ given by (A.3). This implies $\frac{\mathrm{d}}{\mathrm{d}L_d}\overline{C}_{\text{LB}} > 0$ and hence L_d^* is not in $N_t \leq L_d < \zeta$. Therefore, we have shown that $\zeta \leq L_d^* \leq L - N_t$, which also gives $L_p^* = L - L_d^*$ from (A.3). □

A.2 Proof of Lemma 2.2

We need to show that for all SNR values above $\mathcal{P}_{\text{th}}$, we have $\frac{\mathrm{d}\overline{C}_{\text{LB}}}{\mathrm{d}L_d} > 0$ for $0 \leq L_d \leq L - L_p$. To find such a $\mathcal{P}_{\text{th}}$, we use the result in Appendix A.1 which states that $L_d\rho_{\text{eff}}\frac{\mathrm{d}\rho_{\text{eff}}^{-1}}{\mathrm{d}L_d} < 1$ implies $\frac{\mathrm{d}\overline{C}_{\text{LB}}}{\mathrm{d}L_d} > 0$. For any given L_p, we use ρ_{eff} in (2.14) to obtain

$$\begin{aligned}L_d\rho_{\text{eff}}\frac{\mathrm{d}\rho_{\text{eff}}^{-1}}{\mathrm{d}L_d} &= \frac{L_d(N_tL\mathcal{P} + 2N_tL_d + 2N_tL_p + L_pL\mathcal{P})}{(L_p+L_d)(N_tL\mathcal{P} + N_tL_p + N_tL_d + L_pL\mathcal{P})} \\ &= \frac{\tau + N_tL_d^2}{\tau + L_p^2L\mathcal{P} + N_tL_pL\mathcal{P} + N_tL_p^2},\end{aligned}$$

where $\tau = N_t L_d L\mathcal{P} + N_t L_d^2 + 2N_t L_p L_d + L_p L_d L\mathcal{P}$. We see that $L_d \rho_{\text{eff}} \frac{\mathrm{d}\rho_{\text{eff}}^{-1}}{\mathrm{d}L_d} < 1$ reduces to $L_p^2 L\mathcal{P} + N_t L_p L\mathcal{P} + N_t L_p^2 > N_t L_d^2$, which needs to hold for $0 \leq L_d \leq L - L_p$. Therefore, we need to find the values of $\mathcal{P}$ which satisfy $L_p^2 L\mathcal{P} + N_t L_p L\mathcal{P} + N_t L_p^2 > N_t (L - L_p)^2$, given by

$$\mathcal{P} > \frac{N_t L - 2N_t L_p}{L_p (L_p + N_t)}. \tag{A.5}$$

Therefore, a threshold SNR value, above which we have $\frac{\mathrm{d}\overline{C}_{\text{LB}}}{\mathrm{d}L_d} > 0$, is found as in (2.18). □

A.3 Proof of Theorem 2.2

This is a max-min problem where the mean square error (MSE) of the channel estimates is to be minimized by $\boldsymbol{X}_p \boldsymbol{X}_p{}^\dagger$ and to be maximized by $\boldsymbol{R}_{\boldsymbol{H}}$. With the constraint of $\text{tr}\{\boldsymbol{X}_p \boldsymbol{X}_p{}^\dagger\} = \mathcal{P}_p L_p$, we need to show that $\inf_{\boldsymbol{X}_p \boldsymbol{X}_p{}^\dagger} \sup_{\boldsymbol{R}_{\boldsymbol{H}}} \text{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\}$ is achieved by an orthogonal pilot sequence with equal power allocated among the transmit antennas, *i.e.*, $\boldsymbol{X}_p \boldsymbol{X}_p{}^\dagger = \frac{\mathcal{P}_p L_p}{N_t} \boldsymbol{I}_{N_t}$, assuming $L_p \geq N_t$.

From (2.5) we see that

$$\begin{aligned} \sup_{\boldsymbol{R}_{\boldsymbol{H}}} \text{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\} &\geq \text{tr}\{(\boldsymbol{I}_{N_t} + \boldsymbol{X}_p \boldsymbol{X}_p{}^\dagger)^{-1}\} \\ &= \sum_{i=1}^{N_t} (1 + p_i)^{-1}, \end{aligned} \tag{A.6}$$

where $\boldsymbol{p} = [p_1 \; p_2 \; \ldots \; p_{N_t}]^T$ are the eigenvalues of $\boldsymbol{X}_p \boldsymbol{X}_p{}^\dagger$. Since the sum of a convex function of p_i is Schur-convex in $\boldsymbol{p}$ [93], we conclude that (A.6) is Schur-convex in $\boldsymbol{p}$. Consequently, we have

$$\sup_{\boldsymbol{R}_{\boldsymbol{H}}} \text{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\} \geq \sum_{i=1}^{N_t} (1 + p_i)^{-1} \geq \sum_{i=1}^{N_t} \left(1 + \frac{\mathcal{P}_p L_p}{N_t}\right)^{-1}, \tag{A.7}$$

where $\sum_{i=1}^{N_t} (1 + p_i)^{-1} = \sum_{i=1}^{N_t} \left(1 + \frac{\mathcal{P}_p L_p}{N_t}\right)^{-1}$ when $\boldsymbol{X}_p \boldsymbol{X}_p{}^\dagger = \frac{\mathcal{P}_p L_p}{N_t} \boldsymbol{I}_{N_t}$. Note that (A.7) holds for any $\boldsymbol{X}_p \boldsymbol{X}_p{}^\dagger$. On the other hand

$$\begin{aligned} \inf_{\boldsymbol{X}_p \boldsymbol{X}_p{}^\dagger} \sup_{\boldsymbol{R}_{\boldsymbol{H}}} \text{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\} &\leq \sup_{\boldsymbol{R}_{\boldsymbol{H}}} \text{tr}\left\{\left(\boldsymbol{R}_{\boldsymbol{H}}^{-1} + \frac{\mathcal{P}_p L_p}{N_t} \boldsymbol{I}_{N_t}\right)^{-1}\right\} \\ &= \sup_{\boldsymbol{R}_{\boldsymbol{H}}} \sum_{i=1}^{N_t} \left(g_i^{-1} + \frac{\mathcal{P}_p L_p}{N_t}\right)^{-1} \end{aligned}$$

$$\leq \sum_{i=1}^{N_t} \left(1+\frac{\mathcal{P}_p L_p}{N_t}\right)^{-1}, \tag{A.8}$$

where (A.8) is obtained using the Schur-concavity of $\sum_{i=1}^{N_t} \left(g_i^{-1}+\frac{\mathcal{P}_p L_p}{N_t}\right)^{-1}$ in $\boldsymbol{g}$, and $\boldsymbol{g} = [g_1\ g_2\ \ldots\ g_{N_t}]^T$ denote the eigenvalues of $\boldsymbol{R_H}$. The equality in (A.8) holds when $\boldsymbol{R_H} = \boldsymbol{I}_{N_t}$. From (A.7) and (A.8), we conclude that

$$\inf_{\boldsymbol{X}_p \boldsymbol{X}_p{}^\dagger} \sup_{\boldsymbol{R_H}} \mathrm{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\} \;=\; \sum_{i=1}^{N_t} \left(1+\frac{\mathcal{P}_p L_p}{N_t}\right)^{-1},$$

which can be obtained by choosing $\boldsymbol{X}_p \boldsymbol{X}_p{}^\dagger = \frac{\mathcal{P}_p L_p}{N_t} \boldsymbol{I}_{N_t}$. □

A.4 Proof of Theorem 2.3

The proof of part (a) of the theorem is a max-min problem where the capacity lower bound is to be maximized by $\boldsymbol{Q}$ and to be minimized by $\boldsymbol{R_H}$ or effectively $\boldsymbol{R}_{\widehat{\boldsymbol{H}}}$. We need to show that $\sup_{\boldsymbol{Q}} \inf_{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}} C_{\mathrm{LB}}$ is achieved by $\boldsymbol{Q} = \mathcal{P}_d/N_t \boldsymbol{I}_{N_t}$. At sufficiently low SNR, the ergodic capacity lower bound per channel use in (2.8) can be approximated as

$$\begin{aligned}
C_{\mathrm{LB}} &= \frac{1}{\ln 2} E_{\widehat{\boldsymbol{H}}} \mathrm{tr}\Big\{\ln\Big(\boldsymbol{I}_{N_t}+(1+\mathrm{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\boldsymbol{Q}\})^{-1}\widehat{\boldsymbol{H}}^\dagger \widehat{\boldsymbol{H}}\boldsymbol{Q}\Big)\Big\} && \text{(A.9)}\\
&\approx \frac{1}{\ln 2} E_{\widehat{\boldsymbol{H}}} \mathrm{tr}\Big\{\ln\Big(\boldsymbol{I}_{N_t}+\widehat{\boldsymbol{H}}^\dagger \widehat{\boldsymbol{H}}\boldsymbol{Q}\Big)\Big\} && \text{(A.10)}\\
&\approx \frac{1}{\ln 2} E_{\widehat{\boldsymbol{H}}} \mathrm{tr}\Big\{\widehat{\boldsymbol{H}}^\dagger \widehat{\boldsymbol{H}}\boldsymbol{Q}\Big\} && \text{(A.11)}\\
&= \frac{N_r}{\ln 2} \mathrm{tr}\{\boldsymbol{R}_{\widehat{\boldsymbol{H}}}\boldsymbol{Q}\}, && \text{(A.12)}
\end{aligned}$$

where (A.9) is obtained using $\ln|\cdot| = \mathrm{tr}\{\ln(\cdot)\}$, (A.10) uses low SNR approximation, and (A.11) is obtained using Taylor's series expansion of $\ln(\cdot)$ together with low SNR approximation.

From (A.12) we see that

$$\inf_{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}} C_{\mathrm{LB}} \leq \frac{N_r \kappa \mathcal{P}_d}{\ln 2}, \tag{A.13}$$

where we have chosen $\boldsymbol{R}_{\widehat{\boldsymbol{H}}} = \kappa \boldsymbol{I}_{N_t}$ for some κ. Therefore,

$$\sup_{\boldsymbol{Q}} \inf_{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}} C_{\mathrm{LB}} \leq \frac{N_r \kappa \mathcal{P}_d}{\ln 2}. \tag{A.14}$$

On the other hand,

$$\sup_{\boldsymbol{Q}} \inf_{\boldsymbol{R}_{\widehat{\boldsymbol{H}}}} C_{\mathrm{LB}} \geq \inf_{\boldsymbol{R}_{\widehat{\boldsymbol{H}}}} C_{\mathrm{LB}}(\boldsymbol{Q} = \mathcal{P}_d/N_t \boldsymbol{I}_{N_t}). \tag{A.15}$$

Using (A.12), $C_{\mathrm{LB}}(\boldsymbol{Q} = \mathcal{P}_d/N_t \boldsymbol{I}_{N_t})$ reduces to

$$\begin{aligned} C_{\mathrm{LB}} &\approx \frac{N_r \mathcal{P}_d}{N_t \ln 2} \mathrm{tr}\{\boldsymbol{R}_{\widehat{\boldsymbol{H}}}\} \\ &= \frac{N_r \mathcal{P}_d}{N_t \ln 2} \Big(N_t - \mathrm{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\} \Big) \\ &= \frac{N_r \mathcal{P}_d}{N_t \ln 2} \left(N_t - \sum_{i=1}^{N_t} \left(g_i^{-1} + \frac{\mathcal{P}_p L_p}{N_t} \right)^{-1} \right). \end{aligned} \tag{A.16}$$

It can be shown that $f(g) = (g^{-1} + \frac{\mathcal{P}_p L_p}{N_t})^{-1}$ is a concave function of g. Therefore, $\mathrm{tr}\{\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}\} = \sum_{i=1}^{N_t} f(g_i)$ is Schur-concave in $\boldsymbol{g}$ [93]. And the ergodic capacity lower bound in (A.16) is Schur-convex in $\boldsymbol{g}$.

As a result, $\inf_{\boldsymbol{R}_{\widehat{\boldsymbol{H}}}} C_{\mathrm{LB}}(\boldsymbol{Q} = \mathcal{P}_d/N_t \boldsymbol{I}_{N_t})$ is achieved by $\boldsymbol{R}_{\boldsymbol{H}} = \overline{g} \boldsymbol{I}_{N_t}$, or effectively $\boldsymbol{R}_{\widehat{\boldsymbol{H}}} = \kappa \boldsymbol{I}_{N_t}$. Hence,

$$\sup_{\boldsymbol{Q}} \inf_{\boldsymbol{R}_{\widehat{\boldsymbol{H}}}} C_{\mathrm{LB}} \geq \frac{N_r \kappa \mathcal{P}_d}{\ln 2}. \tag{A.17}$$

From (A.14) and (A.17), we have $\sup_{\boldsymbol{Q}} \inf_{\boldsymbol{R}_{\widehat{\boldsymbol{H}}}} C_{\mathrm{LB}} = \frac{N_r \kappa \mathcal{P}_d}{\ln 2}$ which is achieved by equal power transmission of data among the transmit antennas.

Part (b) of the theorem deals with the high SNR regime, in which the covariance matrix of the channel estimation error in (2.5) can be approximated as $\boldsymbol{R}_{\widetilde{\boldsymbol{H}}} \approx (\frac{\mathcal{P}_p L_p}{N_t})^{-1} \boldsymbol{I}_{N_t}$. Note that $\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}$ is independent of the channel spatial correlation in the high SNR regime. Therefore, the ergodic capacity lower bound per channel use in (2.8) can be approximated as

$$\begin{aligned} C_{\mathrm{LB}} &\approx E_{\widehat{\boldsymbol{H}}} \Big\{ \log_2 \Big| \Big(1 + \frac{N_t \mathcal{P}_d}{L_p \mathcal{P}_p} \Big)^{-1} \widehat{\boldsymbol{H}}^{\dagger} \widehat{\boldsymbol{H}} \boldsymbol{Q} \Big| \Big\} \\ &= E_{\widehat{\boldsymbol{H}}} \Big\{ \log_2 \tau |\widehat{\boldsymbol{H}}^{\dagger} \widehat{\boldsymbol{H}} \boldsymbol{Q}| \Big\} \\ &= E_{\widehat{\boldsymbol{H}}_0} \Big\{ \log_2 \tau |\widehat{\boldsymbol{H}}_0^{\dagger} \widehat{\boldsymbol{H}}_0| \Big\} + \log_2 |\boldsymbol{R}_{\widehat{\boldsymbol{H}}}| + \log_2 |\boldsymbol{Q}|, \end{aligned} \tag{A.18}$$

where $\tau = \Big(1 + \frac{N_t \mathcal{P}_d}{L_p \mathcal{P}_p} \Big)^{-N_t}$. It is then clear that the optimal data transmission scheme at high SNR is equal power transmission, i.e. $\boldsymbol{Q} = \mathcal{P}_d/N_t \boldsymbol{I}_{N_t}$. $\square$

A.5 Proof of Theorem 2.4

The proof follows from the low and high SNR approximations of the ergodic capacity lower bound derived in Appendix A.4. At sufficiently low SNR, C_{LB} is given by (A.16) which is shown to be Schur-convex in $\boldsymbol{g}$.

At sufficiently high SNR, the ergodic capacity lower bound is given in (A.18). Denote the eigenvalues of $\boldsymbol{R}_{\widehat{\boldsymbol{H}}}$ as $\widehat{\boldsymbol{g}} = [\widehat{g}_1 \ \widehat{g}_2 \ \ldots \ \widehat{g}_{N_t}]^T$. One can show that C_{LB} in (A.18) is Schur-concave in $\widehat{\boldsymbol{g}}$ due to the Schur-concavity of $\Pi_{i=1}^{N_t}\widehat{g}_i$ [93]. Also with $\boldsymbol{R}_{\widetilde{\boldsymbol{H}}} \approx (\frac{\mathcal{P}_p L_p}{N_t})^{-1}\boldsymbol{I}_{N_t}$ at high SNR, it can be shown that $\widehat{\boldsymbol{g}}$ is an affine function of $\boldsymbol{g}$, hence the concavity is preserved. Therefore, C_{LB} is Schur-concave in $\boldsymbol{g}$. $\square$

Appendix B

B.1 Proof of Theorem 3.1

To prove Theorem 3.1, we begin with the following set of results.

- R1. From the property of water-filling solution [99], m is discrete and non-decreasing on $\alpha \in [0, 1]$ as the number of non-zero q_i in (3.10) cannot decrease as the data transmission power increases.
- R2. With $\sigma^2_{\widetilde{\boldsymbol{H}}} = \left(1 + \frac{\mathcal{P}_p L_p}{N_t}\right)^{-1}$, $\sigma^2_{\widehat{\boldsymbol{H}}} = 1 - \sigma^2_{\widetilde{\boldsymbol{H}}}$ and (3.2), it can be shown that $\rho_{\text{eff}} \triangleq \frac{\sigma^2_{\widehat{\boldsymbol{H}}} \mathcal{P}_d}{1+\sigma^2_{\widetilde{\boldsymbol{H}}} \mathcal{P}_d}$ is a concave function of $\alpha \in [0, 1]$.
- R3. For any fixed m, we see from (3.13) that C_{LB} is maximized when ρ_{eff} reaches its maximum.
- R4. From the property of water-filling solution [99], we know that q_i in (3.10) is continuous on $\mathcal{P}_d$ and hence, is continuous on $\alpha \in [0, 1]$.

In Fig. B.1, we show a sketch plot of ρ_{eff} versus α. This figure helps to visualize the following proof. Basically, the main objective here is to show that the optimal α is not affected by the number of active eigen-channels.

From R1, we can divide all values of α into a finite number of regions in which m is constant.

From R2, there exist a unique local optimal point of α in each of the aforementioned regions, denoted by $\alpha^*_1, \alpha^*_2, \alpha^*_3$ and so on. And these optimal points of α occur on the boundaries of these regions except for at most one of them which is the global optimal point of α, denoted by α^*. (Note that α^* may occur on the boundary of some region as well.)

From R3, we know that the local optimal point of α for ρ_{eff} is also the local optimal point for C_{LB} in each of the aforementioned regions. That is to say, α^*_1, α^*_2, ... maximize C_{LB} in the corresponding regions where m is fixed.

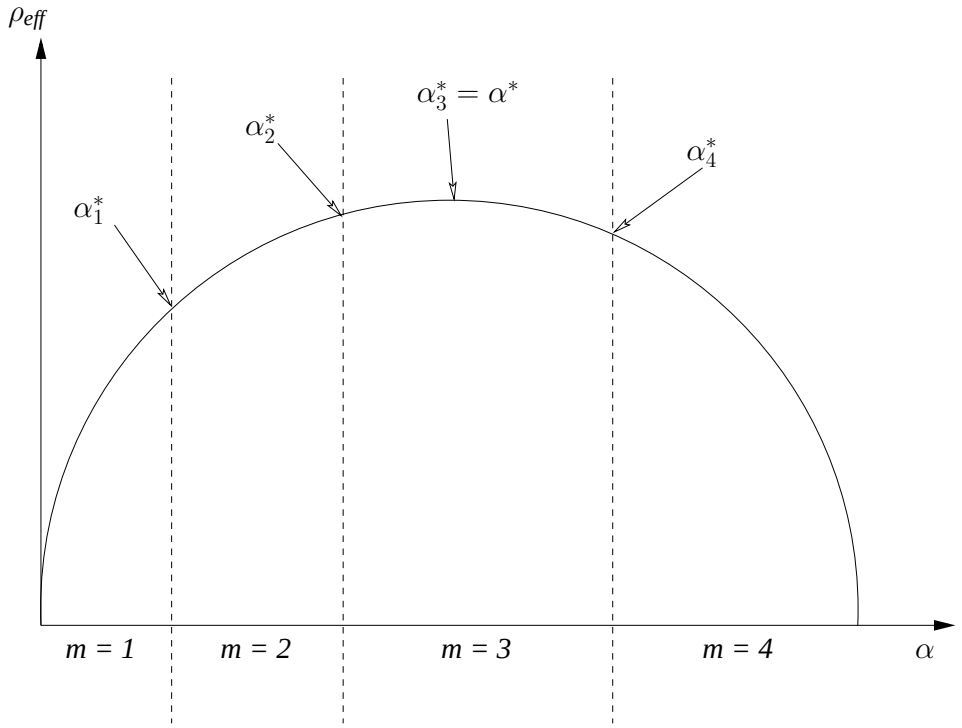

Figure B.1: A sketch example of ρ_{eff} v.s. α. The vertical dashed lines indicates the values of α at which m changes its value. α_1^*, α_2^*, α_3^* and α_4^* indicate the local optimal values of α which give local maxima of ρ_{eff}.

From R4, the continuity of q_i on $\alpha \in [0, 1]$ implies the continuity of C_{LB} in (3.11) on $\alpha \in [0, 1]$. This implies that C_{LB} is continuous across the boundaries of two different regions of α. For example in Fig. B.1, $C_{\text{LB}}(m = 1) = C_{\text{LB}}(m = 2)$ at $\alpha = \alpha_1^*$ and $C_{\text{LB}}(m = 2) = C_{\text{LB}}(m = 3)$ at $\alpha = \alpha_2^*$, *etc*.

Therefore, we can show that C_{LB} at α^* is larger than C_{LB} at α_1^*, α_2^*, ... That is to say, the global optimal point of α for ρ_{eff} is also the global optimal point for C_{LB}. Therefore, the objective function for optimizing α is reduced from C_{LB} to ρ_{eff}. It is noted that the objective function ρ_{eff} is the same as that in non-feedback systems given in [54]. Therefore, the solution of α^* coincides with the solution for non-feedback systems given in Section 2.4.1. □

B.2 Proof of Theorem 3.2

We aim to show that $\frac{\mathrm{d}\overline{C}_{\text{LB}}}{\mathrm{d}L_d} > 0$ (treating L_d as a positive real-valued variable) for any given α. First of all, it can be shown that $\overline{C}_{\text{LB}}$ is continuous on L_d regardless of the value of m, which implies that there is no discontinuity in $\overline{C}_{\text{LB}}$. Therefore, it suffices to show that $\frac{\mathrm{d}\overline{C}_{\text{LB}}}{\mathrm{d}L_d} > 0$ for any fixed m and α.

We let $\rho_{\text{eff}} = \frac{\sigma^2_{\widehat{\boldsymbol{H}}}\mathcal{P}_d}{1+\sigma^2_{\widehat{\boldsymbol{H}}}\mathcal{P}_d} = \frac{\sigma^2_{\widehat{\boldsymbol{H}}}\alpha\mathcal{P}L}{L_d+\sigma^2_{\widehat{\boldsymbol{H}}}\alpha\mathcal{P}L}$, $y = \sum_{i=1}^{m} \ln \frac{\lambda_i}{m}$, and $z = \sum_{i=1}^{m} \lambda_i^{-1}$. Then the

average capacity lower bound in (3.8) can be rewritten using (3.13) as

$$\overline{C}_{\mathrm{LB}} \;=\; \frac{L_d}{L}\frac{1}{\ln 2}E_{\boldsymbol{\lambda}}\{m\ln(\rho_{\mathrm{eff}}+z)+y\}.$$

For any given α, we know from (3.2) that $\mathcal{P}_p L_p$ is constant. Therefore, $\sigma^2_{\widetilde{\boldsymbol{H}}} = \left(1+\frac{\mathcal{P}_p L_p}{N_t}\right)^{-1}$ and $\sigma^2_{\widehat{\boldsymbol{H}}} = 1-\sigma^2_{\widetilde{\boldsymbol{H}}}$ are also constant. Differentiating $\overline{C}_{\mathrm{LB}}$ w.r.t. L_d for any fixed m gives

$$\frac{\mathrm{d}\overline{C}_{\mathrm{LB}}}{\mathrm{d}L_d}=\frac{1}{\ln 2}\frac{m}{L}\Big(E_{\boldsymbol{\lambda}}\Big\{\ln(\rho_{\mathrm{eff}}+z)+\frac{L_d}{\rho_{\mathrm{eff}}+z}\frac{\mathrm{d}\rho_{\mathrm{eff}}}{\mathrm{d}L_d}+\frac{y}{m}\Big\}\Big), \tag{B.1}$$

where

$$\frac{\mathrm{d}\rho_{\mathrm{eff}}}{\mathrm{d}L_d} \;=\; -\frac{\sigma^2_{\widehat{\boldsymbol{H}}}\alpha\mathcal{P}L}{(L_d+\sigma^2_{\widetilde{\boldsymbol{H}}}\alpha\mathcal{P}L)^2} = -\frac{\rho_{\mathrm{eff}}}{L_d+\sigma^2_{\widetilde{\boldsymbol{H}}}\alpha\mathcal{P}L}. \tag{B.2}$$

Substituting (B.2) into (B.1), we get

$$\frac{\mathrm{d}\overline{C}_{\mathrm{LB}}}{\mathrm{d}L_d}=\frac{1}{\ln 2}\frac{m}{L}E_{\boldsymbol{\lambda}}\Big\{\ln(\rho_{\mathrm{eff}}+z)-\frac{\rho_{\mathrm{eff}}}{\rho_{\mathrm{eff}}+z}\frac{L_d}{L_d+\sigma^2_{\widetilde{\boldsymbol{H}}}\alpha\mathcal{P}L}+\frac{y}{m}\Big\}.$$

Since $\frac{L_d}{L_d+\sigma^2_{\widetilde{\boldsymbol{H}}}\alpha\mathcal{P}L}<1$, it suffices to show that

$$E_{\boldsymbol{\lambda}}\Big\{\ln(\rho_{\mathrm{eff}}+z)-\frac{\rho_{\mathrm{eff}}}{\rho_{\mathrm{eff}}+z}+\frac{y}{m}\Big\}\geq 0. \tag{B.3}$$

Furthermore, one can show that

$$\frac{\mathrm{d}}{\mathrm{d}\rho_{\mathrm{eff}}}E_{\boldsymbol{\lambda}}\Big\{\ln(\rho_{\mathrm{eff}}+z)-\frac{\rho_{\mathrm{eff}}}{\rho_{\mathrm{eff}}+z}+\frac{y}{m}\Big\}=\frac{\rho_{\mathrm{eff}}}{(\rho_{\mathrm{eff}}+z)^2}\geq 0$$

for any fixed m. Since $E_{\boldsymbol{\lambda}}\Big\{\ln(\rho_{\mathrm{eff}}+z)-\frac{\rho_{\mathrm{eff}}}{\rho_{\mathrm{eff}}+z}+\frac{y}{m}\Big\}$ is an increasing function of ρ_{eff}, we only need to show (B.3) holds at $\rho_{\mathrm{eff}}=0$, that is

$$\begin{aligned}
E_{\boldsymbol{\lambda}}\Big\{\ln z+\frac{y}{m}\Big\} &= E_{\boldsymbol{\lambda}}\Big\{\ln\sum_{i=1}^{m}\lambda_i^{-1}+\frac{1}{m}\sum_{i=1}^{m}\ln\frac{\lambda_i}{m}\Big\}\\
&\geq E_{\boldsymbol{\lambda}}\Big\{\frac{1}{m}\sum_{i=1}^{m}\ln\Big(\frac{\lambda_i}{m}\Big)^{-1}+\frac{1}{m}\sum_{i=1}^{m}\ln\frac{\lambda_i}{m}\Big\} \qquad \text{(B.4)}\\
&= E_{\boldsymbol{\lambda}}\Big\{\frac{1}{m}\sum_{i=1}^{m}\ln\Big(\frac{\lambda_i}{m}\Big)^{-1}\frac{\lambda_i}{m}\Big\}\\
&= 0,
\end{aligned}$$

where (B.4) is obtained using the concavity of $\ln(\cdot)$. Therefore, we conclude that $\frac{\mathrm{d}\overline{C}_{\mathrm{LB}}}{\mathrm{d}L_d} > 0$ for any given α, which implies that L_d should be kept at its maximum. Hence, L_p should be kept at its minimum. With the assumption of $L_p \geq N_t$ the optimal training length is given by $L_p^* = N_t$. □

B.3 Proof of Corollary 3.1

Among all possible values of $\alpha \in [0, 1]$, there exists an optimal α that maximizes the capacity lower bound for any given L_p. We denote $\alpha^*(k)$ to be the optimal α for $L_p = k$ and denote $\overline{C}_{\mathrm{LB}}\Big(a, l\Big)$ to be the capacity lower bound at $\alpha = a$ and $L_p = l$. From the definition of $\alpha^*(N_t)$ we have

$$\overline{C}_{\mathrm{LB}}\Big(\alpha^*(N_t), N_t\Big) \geq \overline{C}_{\mathrm{LB}}\Big(\alpha, N_t\Big), \quad \forall \quad \alpha \in [0, 1].$$

From *Theorem 3.2* we know that when $k \geq N_t$,

$$\overline{C}_{\mathrm{LB}}\Big(\alpha, N_t\Big) \geq \overline{C}_{\mathrm{LB}}\Big(\alpha, k\Big), \quad \forall \quad \alpha \in [0, 1].$$

Combining the two inequalities and choosing $\alpha = \alpha^*(k)$, we have

$$\overline{C}_{\mathrm{LB}}\Big(\alpha^*(N_t), N_t\Big) \geq \overline{C}_{\mathrm{LB}}\Big(\alpha^*(k), k\Big),$$

that is, the capacity at $L_p = N_t$ is greater than or equal to the capacity at $L_p = k \geq N_t$ where the corresponding α^* is used in both cases. □

B.4 Proof of Theorem 3.3

Assuming $L_p \geq N_t$, we have $\widehat{\boldsymbol{Q}} = \mathcal{P}_d/N_t \boldsymbol{I}_{N_t}$. We let $\boldsymbol{\theta} = [\theta_1 \ \theta_2 \ \ldots \ \theta_r]^T$ be the non-zero eigenvalues of $\widehat{\boldsymbol{H}}_0^{\dagger}\widehat{\boldsymbol{H}}_0\widehat{\boldsymbol{G}}/N_t$, where $r = \mathrm{rank}\{\widehat{\boldsymbol{H}}_0^{\dagger}\widehat{\boldsymbol{H}}_0\widehat{\boldsymbol{G}}\} = \min\{N_t, N_r\}$. C_{LB} in (3.22) is reduced to

$$C_{\mathrm{LB}} = E_{\boldsymbol{\theta}}\{\sum_{i=1}^{r} \log_2(1 + \widehat{\rho}_{\mathrm{eff},i})\},$$

where $\widehat{\rho}_{\mathrm{eff},i} = \theta_i \mathcal{P}_d (1 + \mu^{-1}\mathcal{P}_d)^{-1}$. Note that $\theta_i > 0 \, \forall \, i$ and they are independent of L_d for any fixed α. Following the proof of *Theorem 3.2* in Appendix B.2, one can show that $\frac{\mathrm{d}\overline{C}_{\mathrm{LB}}}{\mathrm{d}L_d} > 0$ for any given α. Therefore, $\overline{C}_{\mathrm{LB}}$ reaches its maximum when L_d is maximized, which implies that $L_p^* = N_t$ under the constraint of $L_p \geq N_t$. Hence, in general we have $L_p^* \leq N_t$. □

B.5 Proof of Theorem 3.4

For any positive definite matrix $\boldsymbol{A}$, $\log_2|\boldsymbol{A}|$ is increasing in $\boldsymbol{A}$ [93]. Also, for any positive semi-definite matrix $\boldsymbol{B}$, $\boldsymbol{I}+\widehat{\boldsymbol{H}}_0^{\dagger}\widehat{\boldsymbol{H}}_0\boldsymbol{B}$ is a positive definite matrix [6]. Since $\widehat{\boldsymbol{G}}\widehat{\boldsymbol{Q}}(1+\mu^{-1}\mathcal{P}_d)^{-1}$ is a positive semi-definite matrix, the capacity lower bound in (3.22) is maximized when the diagonal entries of $\widehat{\boldsymbol{G}}\widehat{\boldsymbol{Q}}(1+\mu^{-1}\mathcal{P}_d)^{-1}$ are maximized.

The ith non-zero diagonal entry of $\widehat{\boldsymbol{G}}\widehat{\boldsymbol{Q}}(1+\mu^{-1}\mathcal{P}_d)^{-1}$ is given by

$$\begin{aligned}\rho_{\text{eff},i} &= \frac{(g_i-\mu^{-1})\mathcal{P}_d}{(1+\mu^{-1}\mathcal{P}_d)L_p} \\ &= \frac{g_i}{L_p}\frac{\mathcal{P}_p\mathcal{P}_d+\mathcal{P}_d(y-g_i^{-1})}{\mathcal{P}_p+\mathcal{P}_d+y},\end{aligned} \tag{B.5}$$

where we have used (3.20) and let $y=\mu-\mathcal{P}_p=\frac{1}{L_p}\sum_{i=1}^{L_p}g_i^{-1}$. Substituting α from (3.2) into (B.5) with some algebraic manipulations, we get

$$\rho_{\text{eff},i}=\frac{g_i\mathcal{P}L}{L_p(L_d-L_p)}\frac{\alpha(1-\alpha)+\alpha\frac{L_p}{\mathcal{P}L}(y-g_i^{-1})}{-\alpha+\frac{\mathcal{P}L+L_py}{\mathcal{P}L(1-L_p/L_d)}}. \tag{B.6}$$

Here we consider the case where $L_d>L_p$ and omit the cases $L_d=L_p$ and $L_d<L_p$ which can be handled similarly. It can be shown that $\rho_{\text{eff},i}$ in (B.6) is concave in $\alpha\in(0,1)$. Therefore, the optimal α occurs at $\frac{\mathrm{d}\rho_{\text{eff},i}}{\mathrm{d}\alpha}=0$, which is the root to

$$\alpha^2-2\alpha\gamma+\gamma+\gamma z=0,$$

where $\gamma=\frac{\mathcal{P}L+L_py}{\mathcal{P}L(1-L_p/L_d)}$ and $z=\frac{L_p}{\mathcal{P}L}(y-g_i^{-1})$. It is clear that α depends on g_i through z. Therefore, there is no unique α which maximizes all $\rho_{\text{eff},i}$. However, this dependence disappears when $\mathcal{P}L\gg L_py=\sum_{i=1}^{L_p}g_i^{-1}$. Under this condition, one can show that $\gamma\approx\frac{L_d}{L_d-L_p}$ and $z\approx0$. And there exists a unique solution of α^* which maximizes all the diagonal entries of $\widehat{\boldsymbol{G}}\widehat{\boldsymbol{Q}}(1+\mu^{-1}\mathcal{P}_d)^{-1}$, given by

$$\alpha^* = \gamma-\sqrt{\gamma(\gamma-1)},\ \text{where}\ \gamma=\frac{L_d}{L_d-L_p}.\ \square$$

B.6 Proof of Theorem 3.6

The proof consists of two steps.

Firstly, we show that the optimal data structure can be written as $\boldsymbol{Q}=\boldsymbol{U}\widehat{\boldsymbol{Q}}\boldsymbol{U}^{\dagger}$

with

$$\widehat{\boldsymbol{Q}} = \begin{pmatrix} \widehat{\boldsymbol{Q}}_a & \mathbf{0}_{n\times(N_t-n)} \\ \mathbf{0}_{(N_t-n)\times n} & \mathbf{0}_{(N_t-n)\times(N_t-n)} \end{pmatrix}.$$

The proof is trivial when $L_p \geq N_t$ in which case we have $\widehat{\boldsymbol{Q}} = \widehat{\boldsymbol{Q}}_a$. In the following we consider $L_p < N_t$. In this case, we have $n = L_p$. From (3.5) and the optimal pilot structure in Section 3.4.1, we have

$$\begin{aligned} \boldsymbol{R}_{\widetilde{\boldsymbol{H}}} &= \left((\boldsymbol{U}\boldsymbol{G}\boldsymbol{U}^\dagger)^{-1} + \boldsymbol{U}\boldsymbol{P}\boldsymbol{U}^\dagger\right)^{-1} \\ &= \boldsymbol{U}(\boldsymbol{G}^{-1} + \boldsymbol{P})^{-1}\boldsymbol{U}^\dagger \\ &= \boldsymbol{U}\widetilde{\boldsymbol{G}}\boldsymbol{U}^\dagger, \end{aligned} \tag{B.7}$$

where (B.7) is indeed the EVD of $\boldsymbol{R}_{\widetilde{\boldsymbol{H}}}$. From (3.6) and (B.7), it is easy to show that the EVD of $\boldsymbol{R}_{\widehat{\boldsymbol{H}}}$ is given by $\boldsymbol{R}_{\widehat{\boldsymbol{H}}} = \boldsymbol{U}\widehat{\boldsymbol{G}}\boldsymbol{U}^\dagger$. Substituting p_i from (3.20) into (B.7), we have

$$\widehat{\boldsymbol{G}} = \boldsymbol{G} - \widetilde{\boldsymbol{G}} = \begin{pmatrix} \widehat{\boldsymbol{G}}_a & \mathbf{0}_{L_p\times(N_t-L_p)} \\ \mathbf{0}_{(N_t-L_p)\times L_p} & \mathbf{0}_{(N_t-L_p)\times(N_t-L_p)} \end{pmatrix}, \tag{B.8}$$

where $\widehat{\boldsymbol{G}}_a = \mathrm{diag}\{g_1 - \mu^{-1},\ \ldots,\ g_{L_p} - \mu^{-1}\}$, and $\mu > 0$.

We let $\boldsymbol{Q} = \boldsymbol{U}\widehat{\boldsymbol{Q}}\boldsymbol{U}^\dagger$. It can be shown that $\widehat{\boldsymbol{Q}}$ is also a positive semi-definite matrix with $\mathrm{tr}\{\widehat{\boldsymbol{Q}}\} = \mathrm{tr}\{\boldsymbol{Q}\}$. The optimization problem in (3.24) reduces to

$$\max_{\widehat{\boldsymbol{Q}}:\,\mathrm{tr}\{\widehat{\boldsymbol{Q}}\}\leq\mathcal{P}_d} \log_2\left|\boldsymbol{I}_{N_t} + (1+\mathrm{tr}\{\widetilde{\boldsymbol{G}}\widehat{\boldsymbol{Q}}\})^{-1}(\widehat{\boldsymbol{H}}\boldsymbol{U})^\dagger\widehat{\boldsymbol{H}}\boldsymbol{U}\widehat{\boldsymbol{Q}}\right|. \tag{B.9}$$

Using $\widehat{\boldsymbol{H}} = \widehat{\boldsymbol{H}}_0\boldsymbol{R}_{\widehat{\boldsymbol{H}}}^{1/2} = \widehat{\boldsymbol{H}}_0\widehat{\boldsymbol{G}}^{1/2}\boldsymbol{U}^\dagger$, where $\widehat{\boldsymbol{H}}_0$ has i.i.d. ZMCSCG entries with unit variance, the optimization problem becomes

$$\max_{\widehat{\boldsymbol{Q}}:\mathrm{tr}\{\widehat{\boldsymbol{Q}}\}\leq\mathcal{P}_d} \log_2\left|\boldsymbol{I}_{N_t}+(1+\mathrm{tr}\{\widetilde{\boldsymbol{G}}\widehat{\boldsymbol{Q}}\})^{-1}\widehat{\boldsymbol{H}}_0^\dagger\widehat{\boldsymbol{H}}_0\widehat{\boldsymbol{G}}^{1/2}\widehat{\boldsymbol{Q}}\widehat{\boldsymbol{G}}^{1/2}\right|. \tag{B.10}$$

We express $\widehat{\boldsymbol{G}}^{1/2}\widehat{\boldsymbol{Q}}\widehat{\boldsymbol{G}}^{1/2}$ in block matrix from as

$$\begin{aligned} \widehat{\boldsymbol{G}}^{1/2}\widehat{\boldsymbol{Q}}\widehat{\boldsymbol{G}}^{1/2} &= \begin{pmatrix} \widehat{\boldsymbol{G}}_a^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\begin{pmatrix} \widehat{\boldsymbol{Q}}_a & \widehat{\boldsymbol{Q}}_b \\ \widehat{\boldsymbol{Q}}_c & \widehat{\boldsymbol{Q}}_d \end{pmatrix}\begin{pmatrix} \widehat{\boldsymbol{G}}_a^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \\ &= \begin{pmatrix} \widehat{\boldsymbol{G}}_a^{1/2}\widehat{\boldsymbol{Q}}_a\widehat{\boldsymbol{G}}_a^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \end{aligned} \tag{B.11}$$

where $\widehat{\boldsymbol{Q}}_a$ is an $L_p \times L_p$ matrix. From (B.11) we can see that $\widehat{\boldsymbol{Q}}_b$, $\widehat{\boldsymbol{Q}}_c$ and $\widehat{\boldsymbol{Q}}_d$ have no effect on $\widehat{\boldsymbol{G}}^{1/2}\widehat{\boldsymbol{Q}}\widehat{\boldsymbol{G}}^{1/2}$. Also, it can be shown from the property of trace that

$$\mathrm{tr}\{\widetilde{\boldsymbol{G}}\widehat{\boldsymbol{Q}}\} = \sum_{i=1}^{L_p} \mu^{-1}\widehat{q}_{ii} + \sum_{i=L_p+1}^{N_t} g_i \widehat{q}_{ii}, \tag{B.12}$$

where $\widehat{q}_{ii} \geq 0$ is the ith diagonal entry of $\widehat{\boldsymbol{Q}}$. We see that (B.12) does not depend on $\widehat{\boldsymbol{Q}}_b$ or $\widehat{\boldsymbol{Q}}_c$. Therefore, (B.10) is independent of $\widehat{\boldsymbol{Q}}_b$ and $\widehat{\boldsymbol{Q}}_c$. For any given $\widehat{\boldsymbol{Q}}_a$, (B.10) is achieved by minimizing (B.12), which can be done by letting $\widehat{\boldsymbol{Q}}_d = \boldsymbol{0}$. We conclude that (B.10) can be achieved by letting $\widehat{\boldsymbol{Q}}_b$, $\widehat{\boldsymbol{Q}}_c$ and $\widehat{\boldsymbol{Q}}_d$ be $\boldsymbol{0}$.

Secondly, we derive the solution for optimal $\widehat{\boldsymbol{Q}}_a$ as follows. Using the optimal structure of $\boldsymbol{Q}$, the objective function in (B.10) reduces to

$$\max_{\widehat{\boldsymbol{Q}}\,:\,\mathrm{tr}\{\widehat{\boldsymbol{Q}}\}\leq\mathcal{P}_d} \log_2 \Big| \boldsymbol{I}_{N_t} + \frac{1}{1+\mathrm{tr}\{\widehat{\boldsymbol{Q}}\}/\mu} \widehat{\boldsymbol{H}}_0^{\dagger}\widehat{\boldsymbol{H}}_0\widehat{\boldsymbol{G}}^{1/2}\widehat{\boldsymbol{Q}}\widehat{\boldsymbol{G}}^{1/2} \Big|. \tag{B.13}$$

We denote $\widehat{\boldsymbol{H}}_0$ in a block matrix form as $\widehat{\boldsymbol{H}}_0 = [\widehat{\boldsymbol{H}}_a \ \widehat{\boldsymbol{H}}_b]$, where $\widehat{\boldsymbol{H}}_a$ is a $N_r \times n$ matrix. With some algebraic manipulations, the optimization problem in (B.13) reduces to

$$\max_{\widehat{\boldsymbol{Q}}_a:\mathrm{tr}\{\widehat{\boldsymbol{Q}}_a\}\leq\mathcal{P}_d} \log_2 \Big| \boldsymbol{I}_n + \frac{1}{1+\mathrm{tr}\{\widehat{\boldsymbol{Q}}_a\}/\mu} (\widehat{\boldsymbol{H}}_a\widehat{\boldsymbol{G}}_a^{1/2})^{\dagger}\widehat{\boldsymbol{H}}_a\widehat{\boldsymbol{G}}_a^{1/2}\widehat{\boldsymbol{Q}}_a \Big|. \tag{B.14}$$

Since $\widehat{\boldsymbol{H}} = \widehat{\boldsymbol{H}}_0\widehat{\boldsymbol{G}}^{1/2}\boldsymbol{U}^{\dagger}$, it is easy to show that $\widehat{\boldsymbol{H}}_a\widehat{\boldsymbol{G}}_a^{1/2}$ contains the first n columns of $\widehat{\boldsymbol{H}}\boldsymbol{U}$. We denote the first n columns of $\widehat{\boldsymbol{H}}\boldsymbol{U}$ as $\boldsymbol{F}$, *i.e.*, $\boldsymbol{F} = \widehat{\boldsymbol{H}}_a\widehat{\boldsymbol{G}}_a^{1/2}$, then the optimization problem in (B.14) can be rewritten as

$$\max_{\widehat{\boldsymbol{Q}}_a\,:\,\mathrm{tr}\{\widehat{\boldsymbol{Q}}_a\}\leq\mathcal{P}_d} \log_2 \Big| \boldsymbol{I}_n + \frac{1}{1+\mathrm{tr}\{\widehat{\boldsymbol{Q}}_a\}/\mu} \boldsymbol{F}^{\dagger}\boldsymbol{F}\widehat{\boldsymbol{Q}}_a \Big|. \tag{B.15}$$

It is a well-known result that the optimal $\widehat{\boldsymbol{Q}}_a$ in (B.15) has the same eigenvectors as $\boldsymbol{F}^{\dagger}\boldsymbol{F}$ [6]. For a given $\mathrm{tr}\{\widehat{\boldsymbol{Q}}_a\}$, the eigenvalues of $\widehat{\boldsymbol{Q}}_a$ can be found via the standard water-filling algorithm given by

$$a_i = \Big[\widehat{\eta} - \Big(\frac{\chi_i}{1+\mu^{-1}\widehat{\mathcal{P}}_d}\Big)^{-1}\Big]^{+} \ \text{ with } \ \sum_{i=1}^{n} a_i = \widehat{\mathcal{P}}_d, \tag{B.16}$$

where $\widehat{\eta}$ represents the water level and $\widehat{\mathcal{P}}_d = \mathrm{tr}\{\widehat{\boldsymbol{Q}}_a\}$.

The remaining part is to show that $\widehat{\mathcal{P}}_d = \mathcal{P}_d$. From (B.16), we can solve the

water level as

$$\widehat{\eta} = \frac{1}{m}\widehat{\mathcal{P}}_d + (1+\mu^{-1}\widehat{\mathcal{P}}_d)\frac{1}{m}\sum_{i=1}^{m}\chi_i^{-1}, \tag{B.17}$$

where m denotes the number of non-zero a_i in (B.16). Therefore, (B.15) reduces to

$$\max_{\widehat{\mathcal{P}}_d:\widehat{\mathcal{P}}_d\leq\mathcal{P}_d}\sum_{i=1}^{n}\log_2\left(1+\frac{\chi_i a_i}{1+\mu^{-1}\widehat{\mathcal{P}}_d}\right) \tag{B.18}$$

$$=\max_{\widehat{\mathcal{P}}_d:\widehat{\mathcal{P}}_d\leq\mathcal{P}_d}\sum_{i=1}^{m}\log_2\left(\frac{\chi_i\widehat{\eta}}{1+\mu^{-1}\widehat{\mathcal{P}}_d}\right) \tag{B.19}$$

$$=\max_{\widehat{\mathcal{P}}_d:\widehat{\mathcal{P}}_d\leq\mathcal{P}_d}\sum_{i=1}^{m}\log_2\left(\frac{\widehat{\mathcal{P}}_d}{1+\mu^{-1}\widehat{\mathcal{P}}_d}+\sum_{i=1}^{m}\chi_i^{-1}\right)+\sum_{i=1}^{m}\log_2\frac{\chi_i}{m}, \tag{B.20}$$

where (B.20) is obtained by substituting $\widehat{\eta}$ from (B.17) into (B.19). It is easy to show that the first derivative of $\frac{\widehat{\mathcal{P}}_d}{1+\mu^{-1}\widehat{\mathcal{P}}_d}$ w.r.t. $\widehat{\mathcal{P}}_d$ is positive. Therefore, the objective function in (B.20) is an increasing function of $\widehat{\mathcal{P}}_d$ for any fixed m. From the property of water-filling solution [99], we know that a_i, $\forall\, i$ in (B.16) are continuous on $\widehat{\mathcal{P}}_d$. Therefore, the objective function in (B.18) is also continuous on $\widehat{\mathcal{P}}_d$. We can conclude that the objective function is an increasing function of $\widehat{\mathcal{P}}_d$ regardless of the value of m. Hence, the objective function is maximized by letting $\widehat{\mathcal{P}}_d = \mathcal{P}_d$. □

Appendix C

C.1 Identity for a Special Class of Gauss Hypergeometric Function

Here we obtain a simplified expression for the Gauss hypergeometric function in the form of ${}_2\mathrm{F}_1(1,1;N+1;x)$ or ${}_2\mathrm{F}_1(N,N;N+1;x)$ for integer $N \geq 1$. From [116], we know that these two forms of the Gauss hypergeometric function are related to each other by

$$ {}_2\mathrm{F}_1(1,1;N+1;x) = (1-x)^{N-1}\, {}_2\mathrm{F}_1(N,N;N+1;x). \tag{C.1} $$

Also, we know from [116] that

$$ \frac{\mathrm{d}^{N-1}}{\mathrm{d}x^{N-1}}\, {}_2\mathrm{F}_1(1,1;2;x) = \frac{(1)_{N-1}(1)_{N-1}}{(2)_{N-1}}\, {}_2\mathrm{F}_1(N,N;N+1;x), \tag{C.2} $$

where $(a)_b$ is the rising factorial. Therefore, we have

$$ \begin{aligned} {}_2\mathrm{F}_1(N,N;N+1;x) &= \frac{(2)_{N-1}}{(1)_{N-1}(1)_{N-1}} \frac{\mathrm{d}^{N-1}}{\mathrm{d}x^{N-1}}\, {}_2\mathrm{F}_1(1,1;2;x) \\ &= -\frac{N}{(N-1)!} \sum_{l=0}^{N-1} \binom{N-1}{l} \frac{\mathrm{d}^l}{\mathrm{d}x^l} \ln(1-x) \frac{\mathrm{d}^{N-1-l}}{\mathrm{d}x^{N-1-l}} x^{-1}, \end{aligned} \tag{C.3} $$

where we have used the identity ${}_2\mathrm{F}_1(1,1;2;x) = -\ln(1-x)/x$ from [116]. It is easy to show that

$$ \frac{\mathrm{d}^k}{\mathrm{d}x^k} \ln(1-x) = -\frac{\mathrm{d}^{k-1}}{\mathrm{d}x^{k-1}}(1-x)^{-1} = -\frac{(k-1)!}{(1-x)^k}, \quad k = 1,2,3,... \tag{C.4} $$

$$ \frac{\mathrm{d}^k}{\mathrm{d}x^k} x^{-1} = \frac{(-1)^k k!}{x^{k+1}}, \quad k = 0,1,2,3,... \tag{C.5} $$

Substituting the above expressions for the derivatives into (C.3), we obtain an identity expression as

$$\begin{aligned} {}_2\mathrm{F}_1(N,N;N+1;x) &= -\frac{N}{(N-1)!}\Big(\ln(1-x)\frac{(-1)^{N-1}(N-1)!}{x^N} \\ &\quad -\sum_{l=1}^{N-1}\frac{(N-1)!}{l!(N-1-l)!}\frac{(l-1)!}{(1-x)^l}\frac{(-1)^{N-1-l}(N-1-l)!}{x^{N-l}}\Big) \\ &= \frac{(-1)^N N}{x^N}\Big(\ln(1-x)-\sum_{l=1}^{N-1}\frac{1}{l}\frac{x^l}{(x-1)^l}\Big). \end{aligned} \tag{C.6}$$

Using (C.1), we also have

$${}_2\mathrm{F}_1(1,1;N+1;x) = \frac{(-1)^N N(1-x)^{N-1}}{x^N}\Big(\ln(1-x)-\sum_{l=1}^{N-1}\frac{1}{l}\frac{x^l}{(x-1)^l}\Big). \tag{C.7}$$

Bibliography

[1] T. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall, 2002.

[2] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, 1995.

[3] J. H. Winters, "On the capacity of radio communication systems with diversity in a Rayleigh fading environment," *IEEE J. Select. Areas Commun.*, vol. 5, no. 5, pp. 871–878, 1987.

[4] B. Lindmark and M. Nilsson, "On the available diversity gain from difference dualpolarized antennas," *IEEE J. Select. Areas Commun.*, vol. 19, no. 2, pp. 287–294, Feb. 2001.

[5] G. J. Foschini and M. J. Gans, "On the limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[6] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[7] D. Gesbert, M. Shafi, D. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: an overview of MIMO space-time coded wireless systems," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 281–302, Apr. 2003.

[8] J. Mietzner, R. Schober, L. Lampe, W. H. Gerstacker, and P. A. Hoeher, "Multiple-antenna techniques for wireless communications - a comprehensive literature survey," *IEEE Commun. Surveys & Tutorials*, vol. 11, no. 2, pp. 87–105, 2009.

[9] WWiSE, "WWiSE proposal: high throughput extension to the 802.11 standard," *IEEE 802.11-04,886r0*, 2004, available at http://www.802wirelessworld.com:8802/.

[10] TGn sync, "TGn sync proposal technical specification," *IEEE 802.11-04,889r0*, 2004, http://www.802wirelessworld.com:8802/.

[11] S. Nanda, R. Walton, J. Ketchum, M. Wallace, and S. Howard, "A high-performance MIMO-OFDM wireless LAN," *IEEE Commun. Mag.*, vol. 43, no. 2, pp. 101–109, 2005.

[12] M. Ho, "Closed-loop MIMO framework for IEEE 802.16e (v0.1)," *IEEE S802.16x04/Intr0*, 2004, http://www.802wirelessworld.com:8802/.

[13] A. Ghosh, D. Wolter, J. Andrews, and R. Chen, "Broadband wireless access with WiMax/802.16: current performance benchmarks and future potential," *IEEE Commun. Mag.*, vol. 43, no. 2, pp. 129–136, 2005.

[14] 3GPP Technical Specifications Group Radio Access Network, "Physical layer aspects of UTRA high speed downlink packet access," *Technical Report TR25.848*.

[15] Nokia, "Closed loop MIMO with 4 Tx and 2 Rx antennas," *3GPP TSG RAN WG1 #36*, http://www.3gpp.org.

[16] Digital Video Broadcasting (DVB), "Frame structure channel coding and modulation for a second generation digital terrestrial television broadcasting system (DVB-T2)," *European Norm EN 302 755*.

[17] L. Tong and S. Perreau, "Blind channel estimation: from subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, pp. 1951–1968, Oct. 1998.

[18] J. Liang and Z. Ding, "Blind MIMO system identification based on cumulant subspace decomposition," *IEEE Trans. Signal Processing*, vol. 51, no. 6, pp. 1457–1468, June 2003.

[19] G. Li and Z. Ding, "Semi-blind channel identification for individual data bursts in GSM wireless systems," *Signal Processing*, vol. 80, no. 10, pp. 2017–2031, Oct. 2000.

[20] A. Grant, "Joint decoding and channel estimation for linear MIMO channels," in *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC)*, Chicago, IL, Sept. 2000, pp. 1009–1012.

[21] T. L. Marzetta, "BLAST training: estimating channel characteristics for high-capacity space-time wireless," in *Proc. 37th Annu. Allerton Conf. Commun., Control, and Comput.*, Monticello, IL, Sept. 1999, pp. 22–24.

[22] C. Budianu and L. Tong, "Channel estimation for space-time block coding systems," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2515–2528, Oct. 2002.

[23] A. Scaglione and A. Vosoughi, "Turbo estimation of channel and symbols in precoded MIMO systems," in *Proc. IEEE Int. Conf. on Acoustic, Speech and Signal Processing (ICASSP)*, Montreal, Canada, May 2004, pp. 413–416.

[24] R. Negi and S. Goel, "Secret communications using artificial noise," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Dallas, TX, Sept. 2005, pp. 1906–1910.

[25] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 684–702, June 2003.

[26] D. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multi-element antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[27] D. Chizhik, F. Rashid-Farrokhi, J. Ling, and A. Lozano, "Effect of antenna separation on the capacity of BLAST in correlated channels," *IEEE Commun. Lett.*, vol. 4, no. 11, pp. 337–339, Nov. 2000.

[28] A. Abdi and M. Kaveh, "A space-time correlation model for multielement antenna systems in mobile fading channels," *IEEE J. Select. Areas Commun.*, vol. 20, no. 3, pp. 550–561, Apr. 2002.

[29] O. Oyman, R. U. Nabar, H. Bolsckei, and A. J. Paulraj, "Characterizing the statistical properties of mutual information in MIMO channels," *IEEE Trans. Signal Processing*, vol. 51, no. 11, pp. 2784–2795, Nov. 2003.

[30] E. A. Jorswieck and H. Boche, "Optimal transmission strategies and impact of correlation in multiantenna systems with different types of channel state information," *IEEE Trans. Signal Processing*, vol. 52, no. 12, pp. 3440–3453, Dec. 2004.

[31] H. Boche and E. A. Jorswieck, "On the ergodic capacity as a function of the correlation properties in systems with multiple transmit antennas without CSI at the transmitter," *IEEE Trans. Commun.*, vol. 52, no. 10, pp. 1654–1657, Oct. 2004.

[32] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2632–2639, Sept. 2001.

[33] 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 Trans. Inform. Theory*, vol. 49, no. 10, pp. 2770–2780, Oct. 2003.

[34] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, July 2004.

[35] J. Li and Q. T. Zhang, "Transmitter optimization for correlated MISO fading channels with generic mean and covariance feedback," *IEEE Trans. Wireless Commun.*, vol. 7, no. 9, pp. 3312–3317, Sept. 2008.

[36] S. A. Jafar, S. Vishwanath, and A. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," in *Proc. IEEE Int. Conf. Commun. (ICC)*, St. Petersburg, Russia, 2001, vol. 7, pp. 2266–2270.

[37] S. H. Simon and A. L. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE J. on Sel. Areas Commun.*, vol. 21, no. 3, pp. 406–417, Apr. 2003.

[38] E. A. Jorswieck and H. Boche, "Optimal transmission with imperfect channel state information at the transmit antenna array," *Wireless Personal Commun.*, vol. 27, pp. 33–56, 2003.

[39] E. A. Jorswieck and H. Boche, "Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1543–1553, Sept. 2004.

[40] S. K. Jayaweera and H. V. Poor, "Capacity of multiple-antenna systems with both receiver and transmitter channel state information," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2697–2709, Oct. 2003.

[41] A. Grant, "Capacity of ergodic MIMO channels with complete transmiter channel knowledge," in *Proc. Australian Commun. Theory Workshop (AusCTW)*, Brisbane, Australia, Feb. 2005, pp. 125–129.

[42] J. Liu, J. Chen, A. Høst-Madsen, and M. P. C. Fossorier, "Correlated MIMO Rayleigh fading systems with transmit channel state informatino," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Los Angeles, CA, Sept. 2004, pp. 1513–1517.

[43] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.

[44] S. M. Redl, M. K. Weber, and M. W. Oliphan, *An introduction to GSM*, Boston: Artech House, 1st edition, 1995.

[45] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: general model, design criteria, and signal processing," *IEEE Signal Processing Mag.*, pp. 12–25, Nov. 2004.

[46] P. Hoeher and F. Tufvesson, "Channel estimation with superimposed pilot sequence," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Rio de Janeiro, Brazil, Oct. 1999, pp. 2162–2166.

[47] J. Manton, I. Mareels, and Y. Hua, "Affine precoders for reliable communications," in *Proc. IEEE Int. Conf. on Acoustic, Speech and Signal Processing (ICASSP)*, Istanbul, Turkey, 2000, pp. 2749–2752.

[48] G. T. Zhou, M. Viberg, and T. McKelvey, "A first-order statistical method for channel estimation," *IEEE Signal Processing Lett.*, vol. 10, no. 3, pp. 57–60, Mar. 2003.

[49] C. Steger and A. Sabharwal, "Single-input two-way SIMO channel: Diversity-multiplexing tradeoff with two-way training," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4877–4885, Dec. 2008.

[50] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[51] A. Lapidoth and S. Shamai (Shitz), "Fading channels: how perfect need "perfect side information" be?," *IEEE Trans. Inform. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

[52] Q. Sun, D. C. Cox, H. C. Huang, and A. Lozano, "Estimation of continuous flat fading MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 549–553, Oct. 2002.

[53] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[54] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[55] E. Baccarelli and M. Biagi, "Power-allocation policy and optimized design of multiple-antenna systems with imperfect channel estimation," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 136–145, Jan. 2004.

[56] E. Baccarelli, M. Biagi, and C. Pelizzoni, "On the information throughput and optimized power allocation for MIMO wireless systems with imperfect channel estimation," *IEEE Trans. Signal Processing*, vol. 53, no. 7, pp. 2335–2347, July 2005.

[57] T. Yoo, E. Yoon, and A. Goldsmith, "MIMO capacity with channel uncertainty: Does feedback help?," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Dallas, TX, May 2004, vol. 1, pp. 96–100.

[58] J. Pang, J. Li, and L. Zhao, "Optimal training length for MIMO systems with transmit antenna correlation," in *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC)*, Las Vegas, NV, Mar. 2008, pp. 425–429.

[59] J.-C. Guey, M. P. Fitz, M. R. Bell, and W.-Y. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 4, pp. 527–537, Apr. 1999.

[60] M. Kiessling, J. Speidel, and Y. Chen, "MIMO channel estimation in correlated fading environments," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Orlando, FL, Oct. 2003, pp. 1187–1191.

[61] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Processing*, vol. 52, no. 2, pp. 546–557, Feb. 2004.

[62] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Processing*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[63] J. Pang, J. Li, L. Zhao, and Z. Lu, "Optimal training sequences for MIMO channel estimation with spatial correlation," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Baltimore, MD, Sept. 2007, pp. 651–655.

[64] G. Fock J. Baltersee and H. Meyr, "An information theoretic foundation of synchronized detection," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2115–2123, Dec. 2001.

[65] G. Fock J. Baltersee and H. Meyr, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *IEEE J. Select. Areas Commun.*, vol. 19, no. 12, pp. 2358–2368, Dec. 2001.

[66] S. Ohno and G. B. Giannakis, "Average-rate optimal PSAM transmissions over time-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 712–720, Oct. 2002.

[67] P. Sadeghi, Y. Liu, R. A. Kennedy, and P. B. Rapajic, "Pilot symbol transmission for time-varying fading channels: An information-theoretic optimization," in *Proc. Int. Conf. on Signal Processing and Commun. Syst. (ICSPCS)*, Gold Coast, Australia, Dec. 2007.

[68] H. Imai, G. Hanaoka, U. Maurer, and Y. Zheng, "Introduction to the special issue on information theoretic security," *IEEE Trans. Inform. Theory*, vol. 54, no. 6, pp. 2405–2407, June 2008.

[69] A.D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.

[70] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[71] Z. Li, W. Trappe, and R. Yates, "Secret communication via multi-antenna transmission," in *Proc. 41st Annual Conf. on Inform. Sciences and Syst. (CISS)*, Baltimore, MD, Mar. 2007, pp. 905–910.

[72] S. Shafiee and S. Ulukus, "Achievable rates in Gaussian MISO channels with secrecy constraints," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Nice, France, June 2007, pp. 2466–2470.

[73] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas: the MIMOME channel," *to appear in IEEE Trans. Inform. Theory*, Available at http://allegro.mit.edu/pubs/posted/journal/2008-khisti-wornell-it.pdf.

[74] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Toronto, Canada, July 2008, pp. 524–528.

[75] Y. Liang and H. V. Poor, "Multiple-access channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. 54, no. 3, pp. 976–1002, Mar. 2008.

[76] E. Tekin and A. Yener, "The general Gaussian multiple-access and two-way channels: achievable rates and cooperative jamming," *IEEE Trans. Inform. Theory*, vol. 54, no. 6, pp. 2735–2751, June 2008.

[77] Y. Liang, A. Somekh-Baruch, H. V. Poor, S. Shamai (Shitz), and S. Verdú, "Capacity of cognitive interference channels with and without secrecy," *IEEE Trans. Inform. Theory*, vol. 55, no. 2, pp. 604–619, Feb. 2009.

[78] E. Ekrem and S. Ulukus, "Secrecy capacity of a class of broadcast channels with an eavesdropper," *EURASIP J. Wireless Commun. and Net.*, 2009.

[79] P. K. Gopala, L. Lai, and H. E. Gamal, "On the secrecy capacity of fading channels," *IEEE Trans. Inform. Theory*, vol. 54, no. 10, pp. 4687–4698, Oct. 2008.

[80] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas: the MISOME channel," *to appear in IEEE Trans. Inform. Theory*, Available at http://arxiv.org/abs/0708.4219.

[81] M. L. Jorgensen, B. R. Yanakiev, F. E. Kirkelund, P. Popovski, H. Yomo, and T. Larsen, "Shout to secure: physical-layer wireless security with known interference," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Washington, DC, Nov. 2007, pp. 33–38.

[82] O. Simeone and P. Popovski, "Secure communications via cooperating base stations," *IEEE Commun. Lett.*, vol. 12, no. 3, pp. 188–190, Mar. 2008.

[83] X. Tang, R. Liu, P. Spasojevic, and H. V. Poor, "Interference-assisted secrect communication," in *Proc. IEEE Inform. Theory Workshop (ITW)*, Porto, Portugal, May 2008, pp. 164–168.

[84] L. Lai and H. E. Gamal, "The relay-eavesdropper channel: cooperation for secrecy," *IEEE Trans. Inform. Theory*, vol. 54, no. 9, pp. 4005–4019, Sept. 2008.

[85] L. Dong, Z. Han, A. P. Petropulu, and H. V. Poor, "Secure wireless communications via cooperation," in *Proc. 46th Annual Allerton Conf. Commun., Control, and Computing*, Monticello, IL, Sept. 2008, pp. 1132–1138.

[86] L. Dong, Z. Han, A. P. Petropulu, and H. V. Poor, "Amplify-and-forward based cooperation for secure wireless communications," in *Proc. IEEE Int. Conf. on Acoustic, Speech and Signal Processing (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 2613–2616.

[87] V. Pohl, P. H. Nguyen, V. Jungnickel, and C. Helmolt, "Continuous flat-fading MIMO channels: Achievable rate and optimal length of the training and data phases," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1889–1900, July 2005.

[88] S. Savazzi and U. Spagnolini, "Optimizing training lengths and training intervals in time-varying fading channels," *IEEE Trans. Signal Processing*, vol. 57, no. 3, pp. 1098–1112, Mar. 2009.

[89] A. Lozano and A. Tulino, "Capacity of multiple-transmit multiple-receive antenna architectures," *IEEE Trans. Inform. Theory*, vol. 48, no. 12, pp. 3117–3128, Dec. 2002.

[90] J. Du and Y. Li, "Optimization of antenna configuration for MIMO systems," *IEEE Trans. Commun.*, vol. 53, no. 9, pp. 1451–1454, Sep 2005.

[91] E. A. Jorswieck, A. Sezgin, and H. Boche, "Outage probability of OSTBC: optimal transmit strategy and suboptimality of odd number of transmit antennas," in *Proc. IEEE Int. Conf. on Acoustic, Speech and Signal Processing (ICASSP)*, Toulouse, France, May 2006, pp. 177–180.

[92] C. Chuah, D. Tse, J. Kahn, and R. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

[93] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, 1979.

[94] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice hall, 1993.

[95] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New Jersey, 2nd edition, 2006.

[96] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, July 2004.

[97] J. C. Roh and B. D. Rao, "Multiple antenna channels with partial channel state information at the transmitter," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 677–688, Mar. 2004.

[98] D. J. Love, R. W. Heath Jr., W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?," *IEEE Commun. Mag.*, pp. 54–59, Oct. 2004.

[99] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 1st edition, 2004.

[100] A. Soysal and S. Ulukus, "Joint channel estimation and resource allocation for MIMO systems - part I: single-user analysis," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 624–631, Feb. 2010.

[101] T. A. Lamahewa, R. A. Kennedy, T. D. Abhayapala, and T. Betlehem, "MIMO channel correlation in general scattering environments," in *Proc. 7th Australian Communication Theory Workshop (AusCTW'06)*, Perth, Australia, Feb. 2006, pp. 93–98.

[102] L. P. Withers Jr., R. M. Taylor Jr., and D. M. Warme, "Echo-MIMO: A two-way channel training method for matched cooperative beamforming," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4419–4432, Sept. 2008.

[103] H. Xu, S. Dasgupta, and Z. Ding, "A novel channel-identification method for wireless communication systems," *IEEE Trans. Commun.*, vol. 52, no. 10, pp. 1767–1776, Oct. 2004.

[104] K. S. Gomadam, H. C. Papadopoulos, and C. W. Sundberg, "Techniques for multi-user MIMO with two-way training," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, Beijing, China, May 2008, pp. 3360–3366.

[105] R. Osawa, H. Murata, K. Yamamoto, and S. Yoshida, "Performance of two-way channel estimation technique for multi-user distributed antenna systems with spatial precoding," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Anchorage, AK, Sept. 2009, pp. 1–5.

[106] T. L. Marzetta, "How much training is required for multiuser MIMO?," in *Proc. Asilomar Conference on Signals, Systems and Computers (ACSSC)*, Pacific Grove, CA, Oct. 2006, pp. 359–363.

[107] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Scheduling and preconditioning in multi-user MIMO TDD systems," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, Beijing, China, May 2008, pp. 4100–4105.

[108] T. A. Lamahewa, P. Sadeghi, and X. Zhou, "On lower bounding the information capacity of amplify and forward wireless relay channels with channel estimation errors," submitted to IEEE Trans. Wireless Commun.

[109] S. Goel and R. Negi, "Guaranteeing secrecy using artificial noise," *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2180–2189, June 2008.

[110] A. L. Swindlehurst, "Fixed SINR solutions for the MIMO wiretap channel," in *Proc. IEEE Int. Conf. on Acoustic, Speech and Signal Processing (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 2437–2440.

[111] A. Mukherjee and A. L. Swindlehurst, "Fixed-rate power allocation strategies for enhanced secrecy in MIMO wiretap channels," in *Proc. IEEE Int. Workshop Signal Processing Advances for Wireless Commun. (SPAWC)*, Perugia, Italy, June 2009, pp. 344–348.

[112] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[113] G. Alfano, A. Lozano, A. M. Tulino, and S. Verdú, "Mutual information and eigenvalue distribution of MIMO Ricean channels," in *Proc. Int. Symp. on Inform. Theory and its Appl. (ISITA)*, Parma, Italy, Oct. 2004.

[114] H. Gao, P. J. Smith, and M. V. Clark, "Theoretical reliability of MMSE linear diversity combining in Rayleigh-fading additive interference channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 666–672, May 1998.

[115] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products*, Academic Press, 7th edition, 2007.

[116] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulaes, Graphs, and Mathematical Tables*, Dover Publications Inc., New York, 1974.

[117] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Dover Publications, Inc., 1979.

[118] C. K .Au-Yeung and D. J. Love, "Optimization and tradeoff analysis of two-way limited feedback beamforming systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2570–2579, May 2009.

[119] W. Santipach and M. L. Honig, "Optimization of training and feedback overhead for beamforming over block fading channels," submitted to IEEE Trans. Inform. Theory. Available at http://arxiv.org/abs/0908.2277.

[120] L. Zheng, D. N. C. Tse, and M. Médard, "Channel coherence in the low-SNR regime," *IEEE Trans. Inform. Theory*, vol. 53, no. 3, pp. 976–997, Mar. 2007.

[121] X. Zhou, P. Sadeghi, and T. Lamahewa, "Optimizing training-based MIMO systems: how much time is needed for actual transmission?," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Taipei, Taiwan, May 2010, pp. 1–5.

[122] O. O. Koyluoglu, C. E. Koksal, and H. E. Gamal, "On secrecy capacity scaling in wireless networks," submitted. available at http://arxiv.org/abs/0908.0898.

[123] Y. Liang, H. V. Poor, and L. Ying, "Secrecy throughput of MANETs with malicious nodes," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Seoul, Korea, Jun 2009, pp. 1189–1193.

[124] M. Haenggi, "The secrecy graph and some of its properties," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Toronto, Canada, Jul 2008, pp. 539–543.

[125] P. C. Pinto, J. Barros, and M. Z. Win, "Secure communication in stochastic wireless networks," submitted. available at http://arxiv.org/abs/1001.3697.

[126] X. Zhou, R. K. Ganti, and J. G. Andrews, "Secure wireless network connectivity with multi-antenna transmission," submitted to IEEE Trans. Wireless Commun.