AN ABSTRACT OF THE THESIS OF

Tao Xu for the degree of Doctor of Philosophy in
Electrical and Computer Engineering presented on May 10, 2010.

Title: Feedback in Multiple Antenna Wireless Communication Systems

Abstract approved: _____

Huaping Liu

Multiple-input multiple-output wireless systems promise significant capacity gain and/or diversity gain over single antenna systems. If channel state information (CSI) is available at both the transmitter and the receiver, the performance can be further improved. In this thesis, first, we study binary index feedback problem in beamforming systems when the feedback channel is not error free. Feedback errors lead to incorrect beamforming vectors to be applied at the transmitter and thus degrade beamforming performance. Index-assignment algorithms that minimize the impact of feedback errors are proposed. Second, in the limited feedback beamforming scheme the receiver has to determine the best codeword from the beamforming codebook. Exhaustive codeword search for large-size codebooks becomes a burden when the receiver is a mobile device with limited computational power. We propose an algorithm to drastically reduce codeword selection complexity with negligible performance loss. Third, we compare angle feedback scheme and transmit antenna shuffling feedback scheme for double space-time transmit diversity systems. We show that the 1-bit angle feedback scheme does not provide a better performance than the 1-bit antenna shuffling feedback scheme. Fourth, we consider training power allocation for a closed-loop MIMO system in i.i.d. Rayleigh

flat-fading channels with power constraint. We derive the optimal solution and asymptotic optimal solution of training power allocation for spatial power control and spatial and fading power control. Lastly, we analyze the optimal diversity-multiplexing tradeoff of multiple beamforming systems and compare it with the well known result for MIMO channels with channel state information at the receiver (CSIR) only and with the optimal diversity-multiplexing tradeoff of spatial multiplexing system with channel state information at the transmitter (CSIT), but without coding over space and time.

Feedback in Multiple Antenna Wireless Communication Systems

by

Tao Xu

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented May 10, 2010
Commencement June 2010

Doctor of Philosophy thesis of Tao Xu presented on May 10, 2010.

APPROVED:

Major Professor, representing Electrical and Computer Engineering

---

Director of the School of Electrical Engineering and Computer Science

---

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Tao Xu, Author

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my advisor, Professor Huaping Liu, for his support, inspiration, patience, and encouragement throughout my graduate study. This Ph.D thesis would not have been possible without his guidance. His hardwork, energy, and enthusiasm to academic research set a great role model for me. I have learned a lot from him; I am sure I will take the benefits of this experience through my whole life.

My thanks also goes to Professor Bechir Hamdaoui, Professor Mario E. Magaña, Professor Raviv Raich, and Professor Thomas A. Schmidt (Department of Mathematics) for serving as the committee members. It is my great honor to have them on my PhD committee.

Finally, I could not have done anything without my family. It was not an easy decision for me to go back to school after working for more than 10 years in industry. Without my wife Shuang's endless love and support, I could not have gone through this chapter of my life with all the challenges. My two lovely boys, Brian and Andrew, made my graduate study joyful. I hope they will learn from my experience and understand it is never too late to achieve our goals and never give up on our dreams.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

NOTATION

$\doteq$	exponential equality, i.e., denote $f(x) \doteq x^b$ if $\lim_{x\to\infty} \frac{\log f(x)}{\log x} = b$
$\dot{\leq}$	exponentially less than, i.e., denote $f(x) \dot{\leq} x^b$ if $\lim_{x\to\infty} \frac{\log f(x)}{\log x} \leq b$
$\dot{\geq}$	exponentially greater than, i.e., denote $f(x) \dot{\geq} x^b$ if $\lim_{x\to\infty} \frac{\log f(x)}{\log x} \geq b$
$\approx$	approximately equal to
$\stackrel{\Delta}{=}$	defined as equal to
$\ll$	much small than
$\mathcal{CN}(\mu, \sigma^2)$	circularly symmetric complex Gaussian distribution with mean μ and variance σ^2
$\boldsymbol{I}_m$	$m \times m$ identity matrix
$\Re\{\cdot\}$	real part of a complex number
$\mathrm{E}\{\cdot\}$	the expectation
$\mathrm{Tr}(\cdot)$	the trace of a matrix
$(\cdot)^T$	the transpose
$(\cdot)^*$	the conjugate
$(\cdot)^H$	the conjugate transpose
$\mathcal{C}^m$	the $m \times 1$ or $1 \times m$ dimensional complex space
$\mathcal{C}^{m\times n}$	the complex space with dimension $m \times n$
$(x)^+$	$\max\{x, 0\}$
$\lceil x \rceil$	the smallest integer not less than real number x
$\lfloor x \rfloor$	the largest integer not greater than real number x
$[\boldsymbol{H}]_{ij}$	the element at the ith row and the jth column of matrix $\boldsymbol{H}$
$\|\cdot\|_F^2$	the Frobenius norm of a matrix

ABBREVIATIONS

BER	bit error rate
BLAST	Bell Laboratories Layered Space Time
CSI	channel state information
CSIT	channel state information at the transmitter
CSIR	channel state information at the receiver
DMT	diversity-multiplexing tradeoff
edf	empirical distribution function
EF	equiangular frame
FDD	frequency division duplexing
GIA	group-based index assignment
GSM	global system for mobile communications
GLP	Grassmannian line packaging
i.i.d.	independent and identically distributed
MB	multiple beamforming
MIMO	multiple-input multiple-output
MISO	multiple-input single-output
ML	maximum-likelihood
MMSE	minimum mean square error
MS	mobile station
OFDM	orthogonal frequency division multiplexing
OSTBC	orthogonal space-time block code
pdf	probability density function
QOSTBC	quasi-orthogonal space-time block code

ABBREVIATIONS (Continued)

SIC	successive interference cancellation
SIMO	single input multiple output
SISO	single input single output
SNR	signal-to-noise ratio
STBC	space-time block code
STTC	space-time trellis code
TDD	time division duplexing
ZF	zero forcing
WLAN	wireless local area networks

Feedback in Multiple Antenna Wireless Communication Systems

Chapter 1 – Introduction

1.1 Background

Wireless communication systems with multiple transmit antennas and multiple receive antennas have been shown, by Telatar, Foschini and Gan [1,2], to provide high spectrum efficiency under independent and identically distributed (i.i.d.) Rayleigh flat-fading between all transmit and receive antenna pairs with channel state information (CSI) at the receiver. This information-theoretic foundation of multiple-input multiple-output (MIMO) wireless systems indicates how much information can be reliably delivered over a MIMO wireless link. The earlier work to explore high throughput/high data rate of MIMO systems is the Bell Laboratories Layered Space Time (BLAST) architecture [4,5]. BLAST is a spatial multiplexing scheme. In its layered structure, the input data stream is demultiplexed into substreams, coded independently using one-dimensional coding, and sent over different transmit antennas simultaneously. The received signal from each substream can be separated by nulling according to a zero-forcing (ZF) or minimum mean-square-error (MMSE) criterion and successive interference cancellation (SIC).

Multiple antennas can be used not only to increase the system capacity but also to lower the probability of error via diversity combining. In [7] the multiplexing gain r is defined as $r = \lim_{\mathrm{SNR}\to\infty} R(\mathrm{SNR})/\log \mathrm{SNR}$, where $R(\mathrm{SNR})$ is the information rate delivered

over the link. The diversity gain d is defined as $d = -\lim_{\mathrm{SNR}\to\infty} \log P_e(\mathrm{SNR})/\log \mathrm{SNR}$, where $P_e(\mathrm{SNR})$ is the probability of codeword error. Both $R(\mathrm{SNR})$ and $P_e(\mathrm{SNR})$ are a function of the signal-to-noise ratio (SNR). With a block fading model [3,6], the channel coefficients hold constant for a block of T symbol intervals and change independently in the next block and hold another T symbol intervals. When the channel is perfectly known at the receiver but not known at the transmitter, the optimal diversity gain achievable by any coding scheme of block length T with $T \geq M + N - 1$ in a MIMO system with M transmit antennas and N receive antennas is given by the piecewise-linear function connecting the points $(k, d^*(k))$ with $k = 0, ..., \min(M, N)$, where $d^*(k) = (M-k)(N-k)$ [7], which implies: (1) The maximum diversity gain is MN and maximum multiplexing gain is $\min(M, N)$; and (2) A system may increase multiplexing gain at the expense of diversity gain or vice versa.

For any fixed-rate transmission, in a system with one transmit antenna and N receive antennas a maximum diversity gain N is achieved when the receiver uses maximum ratio combining (MRC) technique [8], where the signals from the received antenna elements are weighted such that SNR of their sum is maximized. The MRC technique so far has been used exclusively for applications to achieve full receive diversity. As more emerging wireless services continue to emerge, more and more applications may require diversity at the transmitter or at both the transmitter and the receiver to combat fading. One of such applications is the cellular wireless system, where base stations in the cellular cells are normally equipped with multiple antennas and mobile stations (e.g., cell phones) may have fewer antennas due to its small size. The downlink, the transmission from base stations to mobile stations, needs to explore transmit diversity to improve the downlink quality. For a system with M transmit antennas, when CSI is available only at the receiver, space-time coding has been designed to explore full transmit diversity advan-

tage [10–13]. Space-time trellis code (STTC) is one type of space-time codes. In STTC scheme, the data stream is trellis encoded into M substreams, which are transmitted over M transmit antennas. The received signals are decoded using maximum-likelihood (ML) sequence estimation via the Viterbi algorithm. STTCs could provide full diversity gain, but a major disadvantage for STTC is that ML decoding complexity grows exponentially with the diversity level and transmission rate [14], which could be a challenge for real time decoding. An attractive alternative is space-time block codes (STBC) with good diversity gain but with linear receiver complexity. One important class of STBC codes is the orthogonal STBC. The well known Alamouti code [10], a special case of orthogonal space time block code (OSTBC) designed for two transmit antenna systems, has the advantage of full rate (full rate means that one symbol is transmitted per symbol time) and full diversity, and also enjoys simple ML decoding on a symbol-by-symbol basis. This scheme is generalized to an arbitrary number of transmit antennas in [12]. The optimal ML decoder for this class of codes is a simple linear receiver followed by decoupled ML detection, a single-dimensional ML detection (i.e., each of symbols in the code matrix can be detected independently [15]). This linear receiver can be viewed as a matched filter (MF) that uses the knowledge of the channel matrix to maximize the SNR for each data symbol. Unfortunately, OSTBCs with full rate of complex symbols do not exist when the number of transmit antennas $M > 2$, although full diversity can be achieved [12]. In order to increase the transmission rate in STBCs, Quasi-orthogonal STBC (QOSTBC) is proposed in [16] for systems with four transmit antennas. This code loses full diversity but remains to have the transmission rate of one. In order to achieve full diversity in QOSTBC, constellation rotation is proposed [17] in QOSTBC code structure but transmission rate is still not greater than one. Further rate improvement in STBC is to use group-wise STBC structure [18, 19, 82] where the transmit

antennas are divided into several groups with each group of transmit antennas transmitting its own STBC codeword independently; thus the total transmission rate is simply the sum of the individual transmission rates from all groups. Since the group-wise STBC code structure is no longer orthogonal, the interference between groups is inevitable.

Space-time coding essentially explores transmit diversity in an open-loop system with multiple transmit antennas, where CSI is not available at the transmitter. When full CSI is available at both the transmit and receive sides to form a closed-loop system, the optimal linear precoders and decoders can be used to explore diversity gain and/or multiplexing gain in MIMO systems and are shown to have significant performance improvement over the open-loop systems [22–26]. In linear transmit precoding of beamforming or spatial multiplexing, the transmitted symbol or symbol vector is premultiplied by a beamforming vector or precoding matrix that adapts to CSI to combat channel ill-conditioning. In beamforming, one data stream is transmitted along the strongest channel eigenmode and is shown to have a diversity order MN in a MIMO system with M transmit antennas and N receive antennas assuming all MN channel coefficients are i.i.d. Rayleigh faded [20, 21]. Spatial multiplexing mode supports m data streams on the first m largest channel eigenmodes ($m \leq \min(M, N)$). The design criterion for spatial multiplexing could be multiple [22–24], for example, minimizing the mean-square errors, maximizing capacity, etc.

Linear precoding for beamforming and spatial multiplexing requires CSI at the transmitter. For time division duplexing (TDD) wireless systems, the transmitter can acquire the forward-link CSI using training signals embedded in the reverse-link data stream by reciprocal property of the channel. For frequency division duplexing (FDD) wireless systems, however, the forward and reverse links are separated in frequency. Thus the channels for the forward and reverse links could be highly uncorrelated. How does the

transmitter acquire forward-link CSI in FDD wireless systems in order to do adaptive transmission? CSI feedback schemes provide one solution. However, in practical situations, the amount of feedback from the receiver to the transmitter should be kept as small as possible to minimize transmission overhead. In this sense, the assumption of full channel knowledge at the transmitter is not realistic because feedback of full CSI needs infinite bandwidth within finite time. To address this issue, limited feedback must be explored to obtain satisfactory performance [27, 28]. CSI must be quantized before feedback. Grassmannian line packing and subspace packing have been shown to provide excellent performance in quantized beamforming [28–30, 80, 81] and quantized unitary precoding [31, 32], respectively. A finite set of precoding vectors for beamforming (also called beamformers) or unitary precoding matrices for multiple beamforming (also called unitary precoders) known as codebook, is generated in Grassmann manifold according to the predefined distance metrics. Each beamformer or unitary precoder in the codebook is called a codeword. The codebook is known to both the transmitter and the receiver. The receiver selects the optimal codeword from the codebook with a selection criterion based on the current CSI at the receiver, and reports the index of the optimal codeword to the transmitter over a limited feedback channel. In addition to beamforming and unitary precoding with limited feedback, there exist many other limited feedback schemes for MIMO systems with different design objectives, such as transmit antenna subset selection [33], quantized covariance feedback [34, 35], non-unitary precoding [37], transmit antenna grouping/shuffling [38] and angle feedback [39–41]. The limited-feedback transmit antenna shuffling scheme [82] and the angle feedback scheme have a low complexity and are attractive in terms of hardware implementation.

Coherent reception requires CSI at the receiver. For non-coherent communications, the receiver does not need CSI but ML decoding must be performed [6]; the complexity

of ML could be very high, which limits its application. In coherent reception, the receiver needs to learn CSI before data detection. In a fading environment, the radio waves from the transmitter are reflected by numerous surrounding objects, and the reflected replicas with different magnitudes and phases are added either constructively or destructively at the receiver, resulting in fading. Imperfect channel estimate could result in degradation of system performance [42–48, 109] compared with perfect CSI at the receiver. There are two common types of channel estimation techniques: blind channel estimation [49–57] and training-based (i.e., pilot-aided) channel estimation[1] [61–78]. Blind channel estimation has the advantage of no training overhead; however it relies on the stochastic property of the transmitted data, and requires observing a large amount of data to obtain reliable channel estimate. This is clearly a disadvantage in a time-varying fading environment. Training-based approaches use pilot signals to perform channel estimation. They are widely used in the commercial wireless systems, for example, global system for mobile communications (GSM), wireless local area networks (WLAN) designed according to IEEE 802.11 standard, local and metropolitan area networks designed according to IEEE 802.16 standard. Although training-based channel estimation sacrifices some power and bandwidth for pilot signals, it can be optimal in terms of approaching the channel capacity when SNR is sufficiently high, provided that the number of transmit antennas is optimized [84]. Optimal training design in non-selective (flat in both time and frequency), block Rayleigh-fading channels for MIMO systems is considered in [84] and optimal training design for OFDM systems over frequency-selective, block Rayleigh-fading channels is considered in [105], both aiming at maximizing the open-loop achievable rate (the channel capacity lower bound).

[1] We classify semi-blind channel estimation (see [58–60]) as pilot aided channel estimation since it is not purely blind.

1.2 Objective and Thesis Outline

In this thesis, we focus on multiple antenna wireless systems with CSI feedback. Through CSI feedback, a closed-loop system is constructed. many problems related to the closed-loop systems are studied and presented in this thesis, which is outlined as follows.

In Chapter 2, we study the problem of binary index assignment for beamforming codewords in a noisy feedback channel. Beamforming in FDD wireless systems requires CSI at the transmitter through a limited-rate feedback channel. With limited-rate feedback, a common codebook containing a finite set of beamforming vectors is shared at both the transmitter and the receiver. Each codeword is assigned a binary index to represent a beamforming vector. The receiver selects the best beamforming vector and sends its binary index to the transmitter. The transmitter applies the beamforming vector indicated by the feedback index for beamforming. All previous work assumes an error-free feedback channel for performance evaluation. But in a noisy feedback channel, feedback errors are inevitable in the feedback indices. They lead to incorrect beamforming vectors to be applied at the transmitter and thus degrade beamforming performance causing a loss of full diversity. In this chapter, we present two index-assignment algorithms that minimize the impact of feedback errors. The first algorithm requires exhaustive search to find the best binary index mapping. When the codebook size is large, the complexity of the algorithm becomes prohibitive. We thus propose a group-based index assignment (GIA) that has a low complexity while still performs better than random index assignments.

Chapter 3 still focuses on limited-feedback beamforming systems. The receiver determines the best codeword from the beamforming codebook and sends its index to the

transmitter. For mobile stations with limited computational power, exhaustive codeword search for large-size codebooks becomes impractical, especially for OFDM systems where multiple subcarriers or subcarrier groups must select their own beamforming vectors. We propose an algorithm to reduce codeword-selection complexity with negligible performance loss, reducing codeword-search time. The proposed algorithm provides a codeword ordering mechanism based on group-wise chordal distance. It achieves better performance than existing complexity-reduction algorithms. Simulations validate the effectiveness of the proposed algorithm.

In Chapter 4, we compare angle feedback scheme and transmit antenna shuffling feedback scheme for double space-time transmit diversity (DSTTD) systems with four transmit antennas and at least two receive antennas. Transmit antenna shuffling feedback scheme is used for spatial multiplexing with low complexity. Angle feedback scheme has been shown to suppress the interference in QOSTBC [40] and DSTTD [41] with four transmit antennas. It is of interest to compare these two schemes with limited feedback. We show that (a) when an MMSE equalizer or a ZF equalizer is applied, minimizing the mean-square-error of the detected symbols is equivalent to minimizing the sum of the squared amplitude of the off-diagonal (interference) terms, which is proposed by existing work; (b) it is sufficient to perform rotation on only one transmit antenna, which can be arbitrarily chosen in advance; and (c) a DSTTD system with 1-bit of angle feedback does not provide better performance than the same system with 1-bit of transmit antenna shuffling. We also present a simplified general result about the selection of antenna shuffling matrix from six permutation matrices to facilitate our arguments. In simulation, we observe that antenna shuffling outperforms angle feedback.

In Chapter 5, we consider training design for a closed-loop MIMO system in i.i.d. Rayleigh flat-fading channels. The channel follows block-fading model with a block of

T symbol intervals. Channel coefficients remain constant within a block and take an independent value in the next block. Each fading block is partitioned into two phases, a training phase and a data transmission phase. After training, the transmitter controls the data transmit power based on feedback of channel information to maximize a lower bound on the closed-loop system capacity. We first consider power allocation between the training phase and the data transmission phase when all fading blocks are assigned the same power so that only spatial power control is performed (i.e., no fading power control is applied), and show that the optimal percentage of the power used for data transmission is the same as that in open-loop systems. Second, we consider power allocation between the training phase and the data transmission phase when the transmitter varies data power to adapt to fading using both spatial and fading power control. We show that the optimal percentage of the average power per block used for data transmission asymptotically converges to its open-loop counterpart.

In Chapter 6, we derive the optimal diversity-multiplexing tradeoff (DMT) of multiple beamforming systems with M transmit antennas and N receive antennas. We assume that only the right singular vectors of channel matrix corresponding to the first K ($K \leq \min\{M, N\}$) largest singular values are available at the transmitter. The resulting optimal DMT provides a performance limit for limited feedback unitary precoding [32]. We show that multiple beamforming with singular vector information at the transmitter does not increase the diversity gain when compared with the fundamental DMT of MIMO channel with CSI at the receiver (CSIR) [7]. Compared with full CSIT without coding over channel eigenmodes and time [118], the optimal DMT of multiple beamforming is still better owing to coding over both eigenmodes and time.

Chapter 7 gives the conclusions of this thesis and the direction of the future work.

Chapter 2 – Index Assignment for Beamforming with Limited-Rate Imperfect Feedback

2.1 Introduction

Wireless systems with multiple transmit antennas could use beamforming for reliable communications by exploring spatial diversity, which is effective to combat fading. Beamforming requires CSI at the transmitter, which is typically obtained through limited feedback from the receiver in an FDD system where the property of channel reciprocity usually does not hold. Due to the finite rate of the feedback, CSI must be quantized into a finite beamforming vecotr set, called a codebook, which is shared at both the transmitter and the receiver. Beamforming with limited-rate feedback has been studied extensively [30,79–81]. The Grassmannian line packing technique in beamforming vector quantization has been shown to have excellent performance [79, 80]. In this limited-rate feedback scheme, the receiver selects the best beamformer vector based on the instantaneous CSI and sends its index to the transmitter.

Existing work on Grassmannian beamforming assumes an error-free feedback channel. In this case, the indices can be arbitrarily assigned to the sets of beamforming vectors without incurring performance loss. In practice, the indices could be corrupted by feedback errors, causing the transmitter to apply the undesired beamforming vectors. In this chapter we propose two algorithms to optimize the index assignment so that when feedback is not error-free, the performance degradation of beamforming is minimized. We will compare the performance of the proposed scheme with that of the scheme used in [82].

2.2 System Model

Consider a wireless system with M_t transmit antennas and M_r receive antennas. We assume that one data stream is transmitted using beamforming and the receiver uses maximum ratio combining (MRC). The received signal is expressed as

$$y = \boldsymbol{c}^H \boldsymbol{H} \boldsymbol{w} x + \boldsymbol{c}^H \boldsymbol{n} \tag{2.1}$$

where $x \in \mathcal{C}$ is the transmitted symbol with average symbol energy $E_s = E[|x|^2]$, y is the observation after MRC, $\boldsymbol{c}$ is the $M_r \times 1$ MRC weight vector, $\boldsymbol{n}$ is the $M_r \times 1$ noise vector whose elements are i.i.d., circularly symmetric, complex Gaussian noise samples with mean zero and variance N_0, $\boldsymbol{w}$ is the $M_t \times 1$ beamforming vector, and $\boldsymbol{H}$ is the $M_r \times M_t$ channel matrix whose entries, h_{ij}, $1 \leq i \leq M_r$, $1 \leq j \leq M_t$, are circularly symmetric, complex Gaussian random variables with mean zero and unit variance. The beamforming vector $\boldsymbol{w}$ satisfies $\|\boldsymbol{w}\| = 1$ ($\|\cdot\|$ denotes L2 norm) to ensure that the total signal power allocated among M_t transmit antennas is normalized. The MRC vector $\boldsymbol{c}$ has the form

$$\boldsymbol{c} = a\boldsymbol{H}\boldsymbol{w} \tag{2.2}$$

to maximize the received SNR, where a is a scalar that is typically chosen as $a = 1/\|\boldsymbol{H}\boldsymbol{w}\|$, so that $\|\boldsymbol{c}\| = 1$. After MRC, the instantaneous received SNR is

$$\gamma = (E_s/N_0)\|\boldsymbol{H}\boldsymbol{w}\|^2.$$

The optimal beamforming vector $\boldsymbol{w}$ that maximizes the instantaneous SNR is ex-

pressed as

$$\boldsymbol{w}^* = \arg\max_{\|\boldsymbol{w}\|=1} \|\boldsymbol{H}\boldsymbol{w}\| \tag{2.3}$$

which turns out to be the right singular vector corresponding to the largest singular value of the channel matrix $\boldsymbol{H}$ [83]. When $\boldsymbol{H}$ has i.i.d. entries and the channel obeys the block-fading law [84], the optimal beamforming vector is uniformly distributed on the unit hypersphere Ω^{M_t} [79]. For systems with limited-rate feedback, it is impossible to send the optimal beamforming vector in (2.3) for each channel realization. A feasible solution is to partition the unit hypersphere Ω^{M_t} into N non-overlaying and exhaustive Voronoi regions, each of which is represented by a vector $\boldsymbol{w}_i$, $0 \leq i \leq N-1$. Thus there are N such vectors, which form a beamforming codebook $\boldsymbol{W} = \{\boldsymbol{w}_0, \boldsymbol{w}_1, \cdots, \boldsymbol{w}_{N-1}\}$ that is shared by both the transmitter and the receiver. Let B be the number of feedback bits required for each channel realization. Obviously $B = \lceil \log_2(N) \rceil$, where $\lceil x \rceil$ denotes the smallest integer that is not less than x. Each codeword is a B-bit index, which represents the beamforming vector in the codebook $\boldsymbol{W}$ for the corresponding Voronoi region. For each channel realization, the receiver chooses one vector from the codebook that maximizes the metric given in (2.3), and sends the corresponding index to the transmitter. The transmitter applies the beamforming vector indicated by the received index for beamforming.

2.3 Index Assignment

2.3.1 Near-optimum Index Assignment

When feedback is error-free, the index-vector mapping can be chosen arbitrarily; otherwise, different index-vector mappings could result in different performance. We use the term "good codebook" as the codebook that not only has the best beamforming vector set, but also the optimized index-vector mapping to minimize performance degradation due to feedback errors. While the construction of the best vector set of a codebook has been well studied [79, 80, 85], optimization of index assignment has not been addressed. We focus on case where $N \geq M_t$ and assume that $N = 2^B$ for simplicity. $N < M_t$ is undesirable because the codebook with size of $N < M_t$ can not have full transmit diversity order of M_t [79, 90].

The bit errors in the feedback index might cause the transmitter to apply the incorrect beamforming vector, thus reducing the instantaneous SNR at the receiver. For example, let $\boldsymbol{b}_i = [b_{i1}, \cdots, b_{iB}]$ be the index for vector $\boldsymbol{w}_i$, and $\boldsymbol{b}_j = [b_{j1}, \cdots, b_{jB}]$ for $\boldsymbol{w}_j$. If the receiver sends $\boldsymbol{b}_i$, but the transmitter receives $\boldsymbol{b}_j$, the performance will degrade as a result of incorrectly applying the weight vector $\boldsymbol{w}_j$. However, assuming that the index bits have an equal probability to be in error, the probabilities of receiving index $\boldsymbol{b}_j$ given the desired index $\boldsymbol{b}_i$, $P(\boldsymbol{b}_j|\boldsymbol{b}_i)$, will be different for different values of j when the error rate is not unrealistically high (e.g., not greater than 10^{-2}). For example, $P([1, \cdots, 1]|[0, \cdots, 0]) \ll P([0, \cdots, 0, 1]|[0, \cdots, 0])$. Thus SNR degradation will be minimized if indices with higher transition probabilities are assigned to beamforming vectors that have a larger square magnitude of mutual inner products. This ensures that the degradation in the instantaneous received SNR due to occasional index-bit errors is

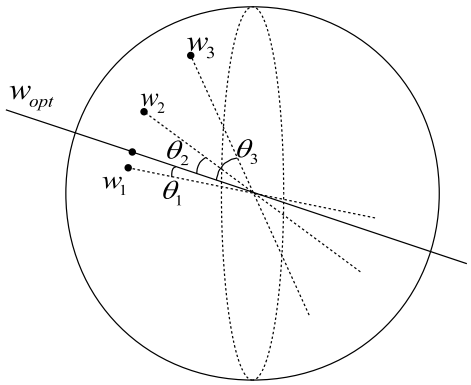

Figure 2.1: Illustration of index assignment in a real three dimensional unit hypersphere (ball). $\theta_2 < \theta_3$.

minimized. This is illustrated in Fig. 2.1, where $\boldsymbol{w}_{opt}$ is the ideal beamforming vector and $\boldsymbol{w}_1$ is best quantized beamforming vector. If the index of $\boldsymbol{w}_1$ changes its value due to feedback errors when it is received by the transmitter, the transmitter receives an erroneous index which represents the beamforming vector other than $\boldsymbol{w}_1$. We wish the beamforming vector indicated by the corrupted index be $\boldsymbol{w}_2$, not $\boldsymbol{w}_3$, since $\boldsymbol{w}_2$ is closer to $\boldsymbol{w}_{opt}$ than $\boldsymbol{w}_3$, and $\boldsymbol{w}_2$ can provide a better received SNR than $\boldsymbol{w}_3$. Our proposed index assignment algorithm is described below.

We line up the beamforming vectors in a codebook in an arbitrary but fixed sequence in the form $[\boldsymbol{w}_0, \boldsymbol{w}_1, ...\boldsymbol{w}_{N-1}]$ without assigning indices. We then pick an index sequence from the $N!$ possible permutations in the form $[I_0, I_1, ..., I_{N-1}]$ to match $[\boldsymbol{w}_0, \boldsymbol{w}_1, ...\boldsymbol{w}_{N-1}]$, that is, I_k represents $\boldsymbol{w}_k$, where each I_k is a B-bit index. The best

index-vector mapping is the one which maximizes the following target function:

$$C = \sum_{i=0}^{N-1} P(I_i) \sum_{\substack{j=0 \\ j \neq i}}^{N-1} P_{I_i I_j} |\boldsymbol{w}_i^H \boldsymbol{w}_j|^2 \tag{2.4}$$

where $P(I_i)$ is the probability that the receiver sends I_i for feedback and $P_{I_i I_j}$ is the transition probability from I_i to I_j.

There exists a class of codebooks, named equiangular frame (EF) codebooks [86–88], which have the property that the absolute values of the inner product of any pair of distinct vectors in the codebook are identical. For EF codebooks, (2.4) becomes constant; thus there is no need to optimize index-vector mapping. However, it is shown in [87] that equiangular vector sets do not exist for $N > M_t^2$. Additionally, EF codebooks work well only when the channel is modeled as quasi-static [84] and $\boldsymbol{H}$ has i.i.d. entries. When the channel is correlated in spatial and/or temporal domains, EF codebooks are far from being optimal [81, 89, 91].

We consider the most commonly accepted scenario that $\boldsymbol{H}$ has i.i.d. entries and the channel fading can be modeled as quasi-static. Under these conditions, we have $P(I_0) = P(I_1) = \cdots = P(I_{N-1}) = 1/N$. Additionally, under normal operation conditions the bit error rate (BER) of the feedback channel p_e would not be too high (e.g., not higher than 10^{-2}). Hence the probability that two or more index bits are in error will be much smaller than the probability that one index bit is in error. We can thus ignore the $O(p_e^2)$ terms in (2.4). Applying the symmetry property between $P_{I_i I_j} |\boldsymbol{w}_i^H \boldsymbol{w}_j|^2$ and $P_{I_j I_i} |\boldsymbol{w}_j^H \boldsymbol{w}_i|^2$ and after some mathematical manipulations, we rewrite (2.4) as

$$C = \sum_{i=0}^{N-1} \sum_{j>i}^{N-1} I(d_{I_i I_j}) |\boldsymbol{w}_i^H \boldsymbol{w}_j|^2 \tag{2.5}$$

where $d_{I_iI_j}$ is the Hamming distance between index I_i and index I_j, $I(d_{I_iI_j}) = 1$ if $d_{I_iI_j} = 1$ and $I(d_{I_iI_j}) = 0$ otherwise. With (2.5), we can search for the optimized index numerically.

2.3.2 Group Index Assignment Scheme

When N is large, the computational load of exhaustive search becomes prohibitive. We propose a group-based index assignment (GIA) method that has a reduced search load and still outperforms random index assignment. The GIA method starts with a smaller-size good codebook of size N_p ($N_p < N$ but is made as large as possible) with optimized index-vector mapping, named the parent codebook expressed as $\boldsymbol{C}_p = [\boldsymbol{c}_0, \boldsymbol{c}_1, \cdots, \boldsymbol{c}_{N_p-1}]$. Any larger codebook with size N, named a child codebook, is partitioned into N_p non-overlapping groups. The partitioning procedure is given as

1. Initialize $i = 0$.
2. With $j = [i]_{N_P}$, where $[\cdot]_{N_P}$ denotes modulo-N_P operation, select a beamforming vector from the child codebook that has the largest magnitude of inner product with vector $\boldsymbol{c}_j$ and add it in the j-th group. A new child codebook with a reduced-size is then formed by removing this vector from the current child codebook.
3. Let $i = i + 1$. If $i < N$, repeat step 2); otherwise the partition process ends.

We assume that $N = 2^B$ and $N_p = 2^{B'}$, where B' and B are positive integers and $B' < B$. When B and/or B' are not integers, the extension is straightforward. After the partition, there are $2^{B'}$ groups, each of which has $2^{B-B'}$ beamforming vectors. For the j-th group, $0 \leq j \leq N_P - 1$, we copy the index of $\boldsymbol{c}_j$ as the B' most significant bits of the index of each beamforming vector within this group. For the rest $B - B'$ unassigned index

bits, we perform random index-vector mapping for simplicity; additional performance improvement due to further optimization within the group is negligible. The reason is that if we consider that there is one bit of error and ignore cases when there are two and more bit errors in each feedback index, then the possibility to take wrong indices in adjacent groups is B'/B and the possibility to take the other indices within same group $1 - B'/B$. The performance degradation by taking indices in adjacent groups is much larger than the performance improvement by the index optimization within the same group so that the effect by performing index optimization within groups is negligible. The resulting index for each beamforming vector in the child codebook has B' most significant bits to indicate which group it belongs to, and $B - B'$ least significant bits for index mapping within the group.

2.4 Diversity Order

For a quantized MIMO beamforming system with M_t transmit antennas and M_r receive antennas, if the elements of $\boldsymbol{H}$ are i.i.d., circularly symmetric, complex Gaussian random variables with zero mean and unit variance, then the maximum diversity order is $M_t M_r$ [90], if $N \geq M_t$. Next, we show that when there exist feedback errors, the same system can only provide a diversity order M_r with a fixed date rate.

Proof. The diversity order is defined by $d = -\lim_{\mathrm{SNR}\to\infty} \log P_e(\mathrm{SNR}) / \log \mathrm{SNR}$, where SNR is the forward channel SNR and $P_e(\mathrm{SNR})$ is the forward channel symbol error rate (SER) averaged over fading. The SER can be expressed as

$$P_e(\mathrm{SNR}) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \left[P_{e,\boldsymbol{w}_j}(\mathrm{SNR}) | I_i \right] \cdot P_{I_i I_j} \cdot P(I_i), \tag{2.6}$$

where $P(I_i)$ and $P_{I_iI_j}$ are defined in (2.4); $\left[P_{e,\boldsymbol{w}_j}(\text{SNR})|I_i\right]$ is SER when the transmitter uses the deterministic beamforming vector $\boldsymbol{w}_j$ averaged only on channel realizations that generate feedback index I_i. Define $P_{\min} = \min\limits_{0\le i,j\le N-1} P_{I_iI_j}$ and $P_{\max} = \max\limits_{0\le i,j\le N-1} P_{I_iI_j}$. When the feedback channel is not error free, $0 < P_{\min} \le P_{\max} < 1$. Substituting $P_{\min}$ and $P_{\max}$ into (2.6), we have the following inequality

$$P_{\max}\sum_{j=0}^{N-1} P_{e,\boldsymbol{w}_j}(\text{SNR}) \ge P_e(\text{SNR}) \ge P_{\min}\sum_{j=0}^{N-1} P_{e,\boldsymbol{w}_j}(\text{SNR}) \tag{2.7}$$

where $P_{e,\boldsymbol{w}_j}(\text{SNR})$ is SER averaged over all possible channel realizations when the transmitter uses the deterministic beamforming vector $\boldsymbol{w}_j$. Note that $\forall j$, $0 \le j \le N-1$, when the transmitter uses $\boldsymbol{w}_j$ for deterministic transmission, the $M_r \times 1$ vector $\boldsymbol{H}\boldsymbol{w}_j$ (the equivalent channel) has the same distribution as column vector $\boldsymbol{h}$, where $\boldsymbol{h}$ is $M_r \times 1$ vector with i.i.d., circularly symmetric, complex Gaussian random variables with zero mean and unit variance. Since the distributions of $\boldsymbol{H}\boldsymbol{w}_i$ and $\boldsymbol{H}\boldsymbol{w}_j$ are the same ($0 \le i \ne j \le N-1$), no matter which deterministic beamforming vector is applied, the system model (2.1) reduces to

$$y = \boldsymbol{c}^H\boldsymbol{h}x + \boldsymbol{c}^H\boldsymbol{n} \tag{2.8}$$

which represents a single-input multiple-output (SIMO) system with one transmit antenna and M_r receive antennas using MRC receiver. Define $P_o(\text{SNR})$ as the probability of error of (2.8), we have

$$P_{e,\boldsymbol{w}_i}(\text{SNR}) = P_{e,\boldsymbol{w}_j}(\text{SNR}) = P_o(\text{SNR}) \tag{2.9}$$

holds for $1 \le i, j \le N-1$. It is well known that with MRC this SIMO system has a

diversity order of M_r [8], that is

$$M_r = -\lim_{\mathrm{SNR}\to\infty} \frac{\log P_o(\mathrm{SNR})}{\log \mathrm{SNR}}, \qquad \forall\ j. \tag{2.10}$$

Combining (2.7), (2.9) and (2.10), we have

$$d = -\lim_{\mathrm{SNR}\to\infty} \frac{\log P_e(\mathrm{SNR})}{\log \mathrm{SNR}} \leq -\lim_{\mathrm{SNR}\to\infty} \frac{\log\big[N \cdot P_{\min} \cdot P_o(\mathrm{SNR})\big]}{\log \mathrm{SNR}} = M_r. \tag{2.11}$$

$$d = -\lim_{\mathrm{SNR}\to\infty} \frac{\log P_e(\mathrm{SNR})}{\log \mathrm{SNR}} \geq -\lim_{\mathrm{SNR}\to\infty} \frac{\log\big[N \cdot P_{\max} \cdot P_o(\mathrm{SNR})\big]}{\log \mathrm{SNR}} = M_r. \tag{2.12}$$

Combining (2.11) and (2.12), we have $d = M_r$. ■

2.5 Simulation Results

We simulate the performance of two codebooks from [82] with different index assignment strategies. The simulations assume uncoded binary phase-shift keying modulation and perfect knowledge of $\boldsymbol{H}$ to the receiver. We use BER to denote the average bit error rate of information bits and FBER to denote the feedback channel BER. BER is averaged over 2×10^4 channel realizations. For each channel realization BER is calculated by considering all possible received indices and the corresponding probabilities of receiving them.

Simulation 1: The system has two transmit antennas and one receive antenna and applies codebook $V(2, 1, 3)$ from Table $298m$ in [82] with 3-bit feedback. Fig. 2.2 shows the BER performance with different index assignments. The curve labeled 'perfect feedback' is obtained assuming that the transmitter has perfect knowledge of the channel coefficients. The case of '3-bit feedback without feedback error' means that the indices

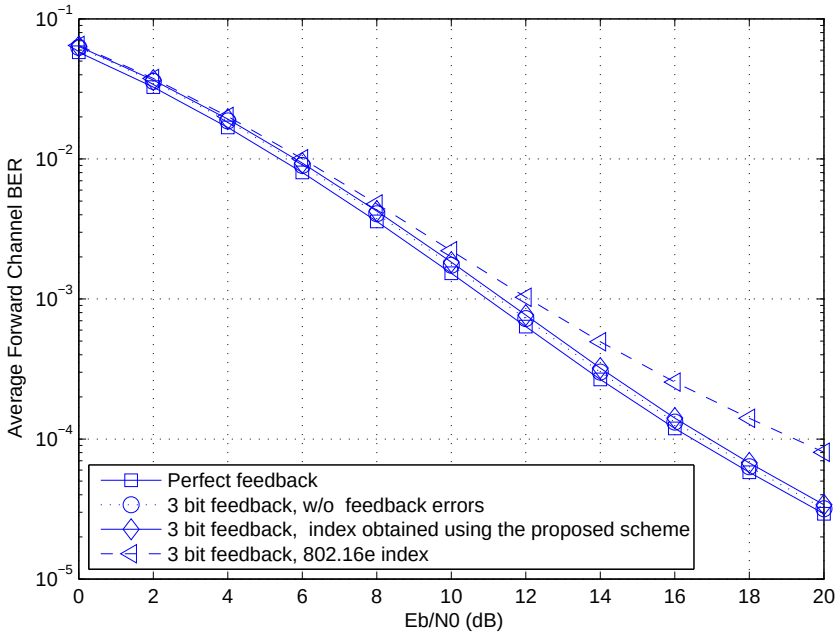

Figure 2.2: BER performance of various schemes (3-bit feedback, FBER=10^{-2}, and $(M_t, M_r)=(2,1)$).

are always received correctly by the transmitter. The optimized index assignment is obtained by applying the proposed method to maximize the target function in (2.5). The proposed scheme is found to achieve nearly the same performance as the case of no feedback errors when FBER is at 10^{-2}. At a BER of 10^{-4}, the proposed scheme achieves a gain of about 2 dB over the scheme in [82], and more at BER values below 10^{-4}. The original index assignment and proposed index assignment are shown in Table 2.1.

Simulation 2: The system has three transmit antennas and one receive antenna, and applies codebook $V(3,1,6)$ from Table 298u in [82] with 6-bit feedback. Since the computational load to determine the optimal index sequences for 6-bit codebook is prohibitively high, we use the proposed GIA method. We adopt the 3-bit codebook from

Table 2.1: Index assignment for codebook V(2,1,3) from table 298m in [82].

Codeword	Index assignments	
	802.16e	Proposed
[1,0]	000	000
[0.7940, -0.5801+j0.1818]	001	110
[0.7940, 0.0576+j0.6051]	010	100
[0.7941, -0.2978-j0.5298]	011	010
[0.7941, 0.6038+j0.0689]	100	001
[0.3289, 0.6614+j0.6740]	101	101
[0.5112, 0.4754-j0.7160]	110	011
[0.3289, -0.8779-j0.3481]	111	111

[79] as the parent codebook. It is easy to verify that $|\boldsymbol{w}_i^H \boldsymbol{w}_j| = 0.5$ for $0 \leq i \neq j \leq 7$. It is an EF codebook, and thus we can randomly select an index sequence to maximize (2.5) and then follow the procedure given in Section 2.3.2 to complete the index assignment. Fig. 2.3 shows the performance of various index assignment schemes at different FBERs for this scenario. Comparison with the scheme adopted in [82]: at FBER of 10^{-2}, the proposed GIA scheme achieves a gain of about 3 dB at BER of 10^{-3}; at FBER 10^{-3}, the gain is about 1.8 dB at BER of 10^{-4}; at FBER=10^{-4}, the gain is about 0.75 dB at BER of 10^{-5}.

Note that although the improvement in performance decreases as the FBER reduces, there is no penalty implementing the codebook optimized using the proposed exhaustive-search-based and GIA methods for any scenarios.

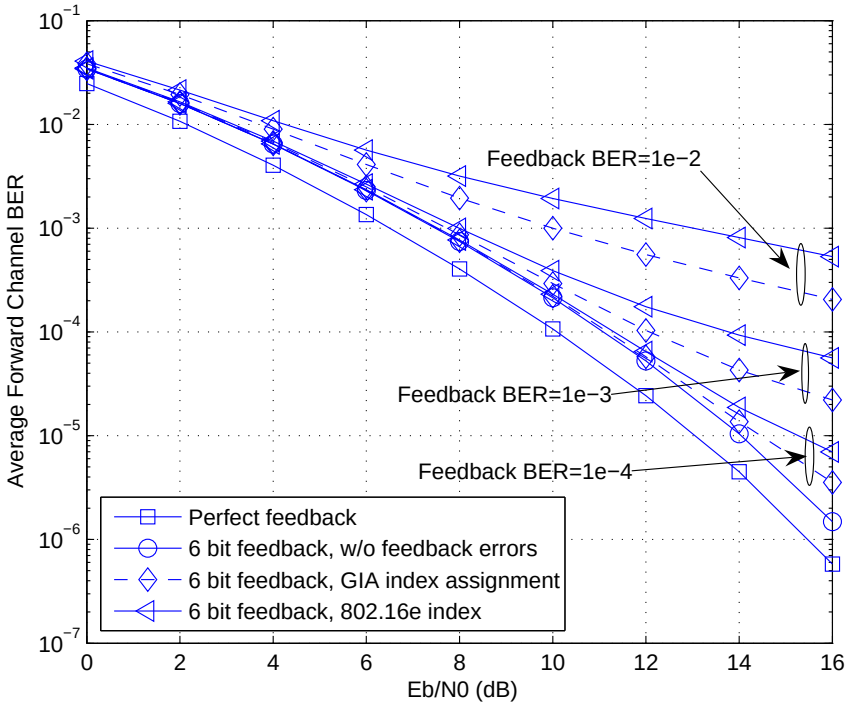

Figure 2.3: BER of various schemes (6-bit feedback, $(M_t, M_r) = (3, 1)$).

2.6 Conclusion

We have presented two index-assignment methods, an exhaustive-search based algorithm and the GIA scheme, to improve beamforming performance in the presence of feedback errors. When the complexity of exhaustive search becomes prohibitively high, the GIA scheme can be applied. The improved codebooks outperform the codebooks adopted in [82]. The gain over the schemes in [82] decreases when the feedback channel BER decreases; once the optimization is complete, however, there is no penalty implementing the codebook obtained using the proposed methods.

Chapter 3 – Reduced-Complexity Codeword Selection for MISO Wireless Systems with Limited Feedback

3.1 Introduction

Beamforming in wireless systems with multiple transmit antennas can significantly improve link quality through diversity gains. Optimal beamforming requires the transmitter to have perfect CSI. In FDD wireless systems, a commonly used method to acquire CSI at the transmitter is to use a low-rate feedback channel to send CSI from the receiver. Because of the limited bandwidth of the feedback channel, obtaining full CSI at the transmitter is impossible; thus limited-rate feedback in beamforming has been extensively studied [79–81]. Among the various schemes, Grassmannian line packaging (GLP) in beamforming vector quantization has been shown to have excellent performance.

The beamforming codebook design criterion [79] requires that the minimal chordal distance between any pair of codewords be as large as possible. Existing codebook design methods can be broadly classified into two categories: non-structural codebook construction and structural codebook construction. Non-structural codebook construction uses a vector quantization (VQ) approach (Lloyd algorithm, for example) to generate the codebook from randomly generated training sets. The number of training sets must be large enough to closely match the channel distribution. Since the codebook is refined by using the different training sets, the codewords generated by Lloyd algorithm have no special structrues except for the unit norm to keep the total transmit power across all transmit antennas constant. Structural codebook construction does not require training sets; it systematically generates the codebooks by certain procedures [79, 92]. The codebooks

generated from structural codebook construction may have elements of all codewords with equal modulus [92] to impose the constraint of equal average transmit power on each transmit antenna. Also, their performance in term of chordal distance measure is in general inferior to that generated by Lloyd algorithm [93]. Compared with codebooks generated by Lloyd algorithm, however, they need less memory for codebook storage.

In limited-rate feedback beamforming systems, one common codebook containing quantized beamformers (i.e., codewords) is shared by the transmitter and the receiver. After channel estimation, the receiver selects the best codeword to maximize the instantaneous received SNR and then feeds back its index to the transmitter. When the codebook size is large, exhaustive search of codewords at the receiver becomes impractical for mobile stations (MS) that have limited computational power. This is especially true for OFDM systems since codeword selection and feedback might be required for each subcarrier or subcarrier group.

A complexity-reduction technique for codeword selection called order and bound algorithm is proposed in [92] for MIMO wireless systems. This algorithm is effective in reducing complexity when N_r is not much smaller than N_t, where N_t and N_r are the number of transmit antennas and the number of receive antennas, respectively. Simulations in [92] show that for a system with $N_t = 4$ and $N_r = 5$, the complexity is reduced by 68% when the number of feedback bits is $N_b = 7$; when the number of feedback bits is $N_b = 12$, complexity is reduced by 80%. But this algorithm has limitations. First, in the downlink of a typical wireless system where the receiver has fewer number of antennas than the transmitter, especially for multiple-input single-output (MISO) systems, this algorithm is not effective in reducing complexity; it actually increases search complexity due to Cholesky decomposition. Second, its ordering mechanism could be further improved.

This chapter focuses on reducing the codeword-search complexity in MISO wireless systems, for which the order and bound algorithm in [92] is not effective. Motivated by the codeword partitioning method used in group index assignment in Chapter 2, we propose a new algorithm to expedite the codeword-selection process. The new algorithm is a codeword ordering algorithm. It allows one to search only a portion of the codebook rather than the whole codebook, thus reducing the codeword-selection complexity. Because the proposed codeword-search algorithm only searches a certain portion of codebook, there is a small probability to miss the best codeword, which causes a performance loss. But this performance loss (compared with exhaustive search) can be controlled to a negligible level by adjusting the size of search portion of the codebook. We also find that by replacing the ordering algorithm in [92] with the proposed ordering algorithm, its codeword-selection complexity can be further reduced without performance loss, and if a small performance loss is allowed, complexity reduction with our proposed ordering algorithm becomes even more significant.

We will introduce the system model in Section 3.2. The details of the proposed codeword-selection algorithm are described in Section 3.3. Simulation results and discussion are given in Section 3.4, which is followed by conclusions in Section 3.5.

3.2 System Model

Consider a MISO wireless system with N_t transmit antennas and one receive antenna. The channel coefficients between each transmit antenna and the receive antenna are modeled as i.i.d., circularly symmetric, complex Gaussian random variable with zero mean and unit variance. The N_t channel coefficients are written in a column vector as $\boldsymbol{h} = [h_1, h_2, \cdots, h_{N_t}]^T$. For beamforming, an $N_t \times 1$ beamforming vector $\boldsymbol{w}$ is applied at

the transmitter. With beamforming at the transmitter, the received signal y is expressed as

$$y = \boldsymbol{h}^T \boldsymbol{w} x + n \tag{3.1}$$

where $x \in \mathcal{C}$ is the transmitted symbol with average symbol energy $E_s = E[|x|^2]$ and n, a circularly symmetric, complex Gaussian random variable, is the noise sample with mean zero and variance N_0. The beamforming vector $\boldsymbol{w}$ satisfies $\|\boldsymbol{w}\| = 1$ ($\|\cdot\|$ denotes L2 norm) to ensure that the total signal power allocated among the N_t transmit antennas is a constant. The instantaneous received SNR is $\gamma = (E_s/N_0)|\boldsymbol{h}^T \boldsymbol{w}|^2$.

The optimal beamforming vector $\boldsymbol{w}$ that maximizes the instantaneous SNR is expressed as

$$\boldsymbol{w}_{\text{opt}} = \arg\max_{\|\boldsymbol{w}\|=1} |\boldsymbol{h}^T \boldsymbol{w}| \tag{3.2}$$

and $\boldsymbol{w}_{\text{opt}}$ is given by $\boldsymbol{h}^*/\|\boldsymbol{h}\|$. We have

$$\boldsymbol{h}^T = \|\boldsymbol{h}\| \boldsymbol{w}_{\text{opt}}^H \tag{3.3}$$

When $\boldsymbol{h}$ has i.i.d. entries and the channel obeys the block-fading model [84], the optimal beamforming vector is uniformly distributed on the unit hypersphere Ω^{M_t} [79]. For systems with limited-rate feedback, it is impossible to send the optimal beamforming vector in (3.2) for each channel realization. A feasible solution is to partition the unit hypersphere Ω^{N_t} into N non-overlapping Voronoi regions, each of which is represented by a vector $\boldsymbol{w}_i$, $0 \le i \le N-1$. Thus there are N such vectors, which form a beamforming codebook $\boldsymbol{W} = \{\boldsymbol{w}_0, \boldsymbol{w}_1, \cdots, \boldsymbol{w}_{N-1}\}$ that is known to both the transmitter and the receiver. Let B be the number of feedback bits required for each channel realization. Obviously, $B = \lceil \log_2(N) \rceil$, where $\lceil a \rceil$ denotes the smallest integer that is not less than

a. In practice, $N = 2^B$. Each codeword is a B-bit index, which represents a quantized beamforming vector in the codebook $\boldsymbol{W}$.

The optimal beamforming codebook design criterion is to minimize the distortion metric

$$\mathrm{E}_{\boldsymbol{h}}\left[\min_{\boldsymbol{w}\in\boldsymbol{W}}\left(|\boldsymbol{h}^T\boldsymbol{w}_{\mathrm{opt}}|^2-|\boldsymbol{h}^T\boldsymbol{w}|^2\right)\right]. \tag{3.4}$$

This is equivalent to maximizing

$$\mathrm{E}_{\boldsymbol{h}}\left(\max_{\boldsymbol{w}\in\boldsymbol{W}}|\boldsymbol{h}^T\boldsymbol{w}|^2\right)$$

by observing in (3.4) that $|\boldsymbol{h}^T\boldsymbol{w}_{\mathrm{opt}}|^2 = \|\boldsymbol{h}\|^2$ is independent of codebook $\boldsymbol{W}$ and its expectation over channel realization $\boldsymbol{h}$ is a constant.

It is shown in [79] that maximizing the minimum of chordal distance among all pairs of two distinct codewords effectively reduces the distortion in (3.4). Chordal distance is defined by

$$d_c(\boldsymbol{w}_k,\boldsymbol{w}_l)=\sqrt{1-|\boldsymbol{w}_k^H\boldsymbol{w}_l|^2},\quad 0\le k<l\le N-1. \tag{3.5}$$

This leads to the Grassmannian beamforming criterion, which requires that the codewords $\{\boldsymbol{w}_i\}_{i=0}^{N-1}$ in the codebook $\boldsymbol{W}$ maximize the following quantity:

$$\min_{0\le k<l\le N-1}\sqrt{1-|\boldsymbol{w}_k^H\boldsymbol{w}_l|^2}. \tag{3.6}$$

Once the codebook $\boldsymbol{W}$ is generated, it is shared by both the transmitter and the receiver. The receiver selects the best codeword from the codebook $\boldsymbol{W}$ to maximize the

instantaneous received SNR

$$\gamma = \max_{0 \le i \le N-1} (E_s/N_0) |\boldsymbol{h}^T \boldsymbol{w}_i|^2 \tag{3.7}$$

and sends its index to the transmitter, usually through a dedicated, low-rate channel. The transmitter applies the quantized beamforming vector indicated by the received index for beamforming.

3.3 Codeword Selection

For each channel realization of $\boldsymbol{h}$, the index of the best codeword to reach (3.7) is

$$\beta = \arg \max_{0 \le i \le N-1} |\boldsymbol{h}^T \boldsymbol{w}_i|^2 = \arg \max_{0 \le i \le N-1} |\boldsymbol{w}_{\text{opt}}^H \boldsymbol{w}_i|. \tag{3.8}$$

The last step is derived using (3.3). When the codebook size is small, an exhaustive search for the index β is feasible. When the codebook size is large, however, exhaustive search could become impractical for mobile devices that have limited computational power. The order and bound algorithm in [92] effectively reduces the codeword-selection complexity when $N_r \ge N_t$. When N_r is much smaller than N_t, however, especially in a MISO system this algorithm becomes ineffective. In the rest of this chapter, we develop a codeword-selection algorithm for MISO systems.

Our proposed codeword-selection algorithm is a new ordering algorithm. We first select a well-designed codebook with size of N_P, where $N_P \ll N$, as the pilot codebook. For simplicity of discussion, we call the codewords in the pilot codebook pilot codewords. We partition the codewords in the beamforming codebook $\boldsymbol{W}$ into N_P non-overlapping groups with each group related to one pilot codeword according to chordal distance. The

partition process makes sure that in each group the codewords from the beamforming codebook $\boldsymbol{W}$ have the smallest chordal distance to its own pilot codeword than to any other pilot codewords[1]. After this partition, the codewords within each group are arranged in ascending order according to the chordal distance to its own pilot codeword to form a row vector. There are a total of N_P row vectors corresponding to the N_P groups of beamforming codewords. This partition process is completed off-line.

For each channel realization, after the receiver acquires the channel coefficients, the ideal beamforming vector $\boldsymbol{w}_{\text{opt}}$ can be obtained, and then the receiver does real time codeword searching. Before codeword searching, our algorithm first computes the chordal distance between $\boldsymbol{w}_{\text{opt}}$ and each pilot codeword, and sorts the N_P pilot codewords in ascending order based on the computed chordal distance. The N_P row vectors are concatenated one by one in the same order as the sorted pilot codewords to form a long row vector. This process is called the codeword ordering process which needs to be done for each channel realization. In the proposed codeword-selection algorithm, after codeword ordering process searching of the codewords starts from the left of the concatenated vector to its right and terminates when K, the predetermined number of codewords being searched, is reached. In Section 3.4, we will provide simulation results to show that even though only a small number of codewords are searched, the performance can quickly approach that with exhaustive search.

The proposed algorithm is built upon the fact that if $\boldsymbol{w}_{\text{opt}}$ has a smaller chordal distance to some pilot codewords than to the rest, the best codeword is very likely to have the same distance property because the best codeword is the closest one to $\boldsymbol{w}_{\text{opt}}$ in chordal distance among all codewords, see (3.8). The groups corresponding to the pilot

[1]Note that the number of codewords in each group might not be the same due to non-uniform allocation of codewords (from both pilot codebook and target codebook) generated by practical quantization methods on the surface of the unit hypersphere.

codewords that have a small chordal distance to $\boldsymbol{w}_{\text{opt}}$ could contain the global optimal beamformer with a high probability.

Next, we discuss two important aspects related to the proposed ordering algorithm.

1. What is the optimal size of the pilot codebook?

   First, we consider complexity reduction for two extreme cases: (a) when $N_P = 1$ and (b) when $N_P = N$. For case (a), there is only one large group−the beamformer codebook. The exhaustive search must be conducted. Although the ordering algorithm in [92] could be used, in simulation we find that K must be very close to N in order to achieve an acceptable performance. Thus, the complexity is virtually not reduced. For case (b), let the pilot codebook be the beamforming codebook itself for simplicity. Ordering of pilot codewords needs to check every pilot codeword, i.e., every beamformer in the beamforming codebook. This is an exhaustive search approach; therefore, complexity is not reduced for this case either. From these two extreme cases, we conclude that N_P should take a value between 1 and N. We suggest that N_P takes the value $2^{\lfloor B/2 \rfloor}$, where $B = \log_2 N$ and $\lfloor a \rfloor$ denotes the maximum integer not greater than a.

2. How is the pilot codebook generated?

   The pilot codebook could be any appropriately selected codebooks. In order to minimize the memory required for storage, the pilot codebooks should be generated by using a structured construction method that will minimize the memory requirement [92].

3.3.1 Fixed Search Length

When the proposed algorithm searches only the first K beamformers in the ordered beamforming codebook for any channel realizations, the complexity reduction compared with the exhaustive search is given by

$$R = 1 - (K + N_P)/N \tag{3.9}$$

where N_P represents the computation needed to calculate the chordal distance between $\boldsymbol{w}_{\text{opt}}$ and all pilot codewords.

3.3.2 Adjustable Search Length

In order to further reduce the codeword search complexity, the search length could be adjusted, instead of taking the fixed value K. For any given pilot codebook, the smallest distance between $\boldsymbol{w}_{\text{opt}}$ and the pilot codewords, denoted by d_c^*, is in the range of $[0, d_{\max}]$, where $d_{\max}$ is the upper bound of d_c^* and is determined by the specific pilot codebook. Let us denote $\mu_0 = 0$ and $\mu_M = d_{\max}$. We partition the interval $[0, d_{\max}]$ into M smaller intervals such that $[0, d_{\max}] = [\mu_0, \mu_1) \cup [\mu_1, \mu_2) \cup \cdots \cup [\mu_{M-1}, \mu_M]$, where μ_i, $1 \leq i \leq M-1$, are predefined thresholds. We limit the search length as follows: when d_c^* is in the range $[\mu_j, \mu_{j+1})$, the codeword search length is set to K_j, $0 \leq j \leq M-2$; when d_c^* is in the range $[\mu_{M-1}, \mu_M]$, the codeword search length is set to K_{M-1}. Let $K_0 < K_1 < \cdots < K_{M-1} = K$. The values of M, μ_i, $1 \leq i \leq M-1$, and K_j, $0 \leq j \leq M-1$, mainly depend on the size of the beamforming codebook and the size of the pilot codebook. Although there are no close-form solutions, suboptimal solutions can be obtained by using simulations.

3.4 Simulation and Discussion

In this section, we provide simulations to evaluate the performance of the proposed algorithm. We adopt BPSK modulation. The uncoded BER performance is averaged over 10^5 independent channel realizations for each simulated SNR value. We use the number of floating point operations (FLOPS) to measure the algorithm complexity as in [92], but we only count real multiplications[2] and omit addition/subtraction operations since multiplications in general need much more computations than additions/subtractions. The system in the simulations has four transmit antennas, $N_t = 4$. The exhaustive search is conducted according to (3.8) for MISO systems.

Simulation 1: In this simulation, we compare the exhaustive search and order and bound algorithm in [92]. A MISO system with $N_t = 4$ and the number of feedback bits ranging from 3 to 7 is evaluated. The results are shown in Fig. 3.1. The number of feedback bits is capped at 7 because increasing it to over 7 drastically increases the size of the codebook but does not significantly gain further in performance. Practical systems rarely use more than 7 feedback bits[3]. Notice that the complexity of the order and bound algorithm is much higher than that of the exhaustive search in MISO systems. This is because Cholesky decomposition in [92] increases the number of variables from $N_t \times 1$ in $\boldsymbol{h}$ to $N_t(N_t + 1)/2$ in an $N_t \times N_t$ upper triangular matrix. Although the bounding technique is applied, search complexity is not effectively reduced. Compared with exhaustive search, the complexity of the order and bound algorithm is 51% and 34% higher at 6 and 7 feedback bits, respectively.

Simulation 2: Fig. 3.2 shows the effectiveness of the proposed ordering algorithm

[2]One complex multiplication includes four real multiplications.

[3]In [82, 94, 95], the beamforming codebooks for 4 transmit antennas are provided up to 6 feedback bits only.

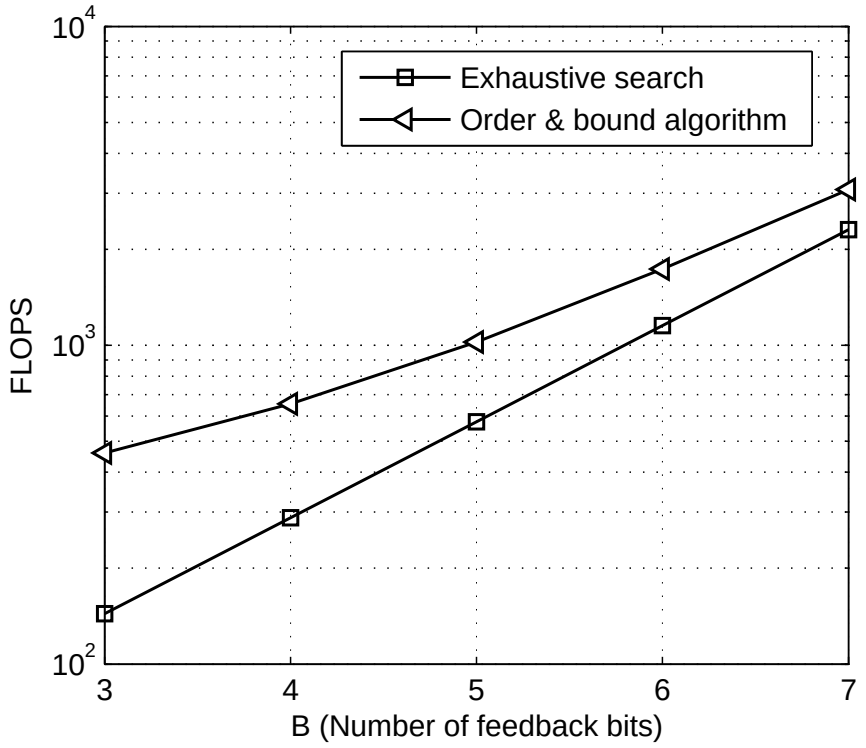

Figure 3.1: Complexity of the order and bound algorithm in [92] and exhaustive search in a MISO system with $(N_t, N_r) = (4, 1)$.

in reducing the search complexity with a performance comparable to that of exhaustive search. In this simulation, a 7-bit beamforming codebook is used and a 3-bit codebook is selected as the pilot codebook, i.e., $N_P = 8$. Both the beamforming codebook and the pilot codebook are generated by the genetic algorithm in [92] to reduce the memory storage. The search length is fixed at $K = 24$. The performance loss of the proposed ordering algorithm is within 0.1 dB of exhaustive search in the simulated SNR range, and the codeword-search complexity reduces by 75% compared with exhaustive search. In order to maintain the same low performance loss of 0.1 dB for the order and bound

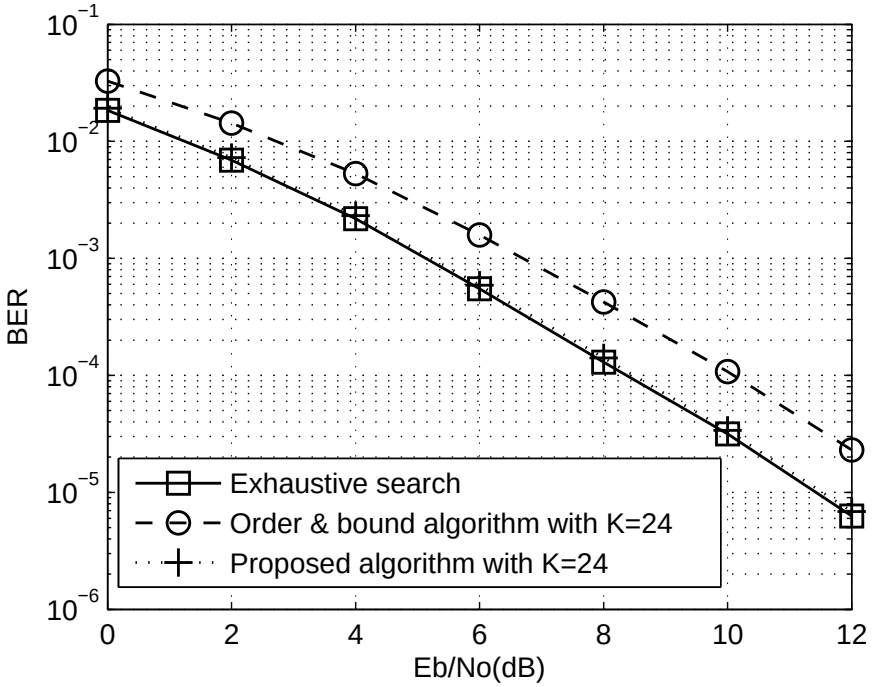

Figure 3.2: Performance comparison: the proposed algorithm, the order and bound algorithm and the exhaustive search for a 7-bit beamforming codebook, $(N_t, N_r) = (4, 1)$.

algorithm proposed in [92], we have to search at least 102 codewords and the complexity increases by 13% compared with exhaustive search. This shows that the order and bound algorithm does not work in MISO systems in term of complexity reduction. If only the first 24 codewords are searched by the order and bound algorithm, as shown in the figure, about 1.7 dB of SNR loss is observed.

In Fig. 3.3, we simulate a beamforming codebook with 6 bits of feedback. The same pilot codebook as used in Fig. 3.2 is used. With $K = 24$, the performance loss of our proposed ordering algorithm is within 0.05 dB of exhaustive search and its complexity is reduced by 50%, while the order and bound algorithm in [92] has a performance loss of

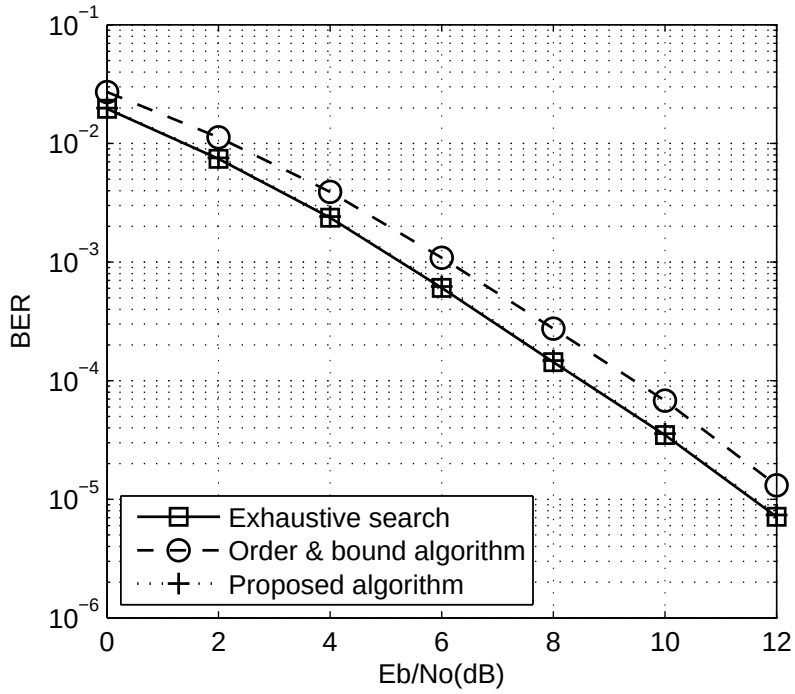

Figure 3.3: Performance comparison: proposed algorithm, order and bound algorithm in [92] and exhaustive search for a 6-bit beamforming codebook, $(N_t, N_r) = (4, 1)$.

Table 3.1: Performance comparison. Data collected from Fig. 3.2 and Fig. 3.3.

$(N_t, N_r) = (4, 1)$	N_p	N	K	complexity	degradation
proposed ordering	8 (3bits)	128 (7 bits)	24	-75%	0.1dB
order and bound	N/A	128 (7 bits)	102	13%	0.1dB
order and bound	N/A	128 (7 bits)	24	-59%	1.7dB
proposed ordering	8 (3bits)	64 (6 bits)	24	-50%	0.05dB
order and bound	N/A	64 (6 bits)	24	N/A	0.8dB
proposed ordering	8 (3bits)	64 (6 bits)	18	-59%	0.1dB

about 0.8 dB with $K = 24$. If a performance loss of 0.1 dB is allowed, the codeword search length K can be further reduced to 18 for our proposed algorithm and the complexity is

reduced by 59%. Performence comparison results are summerized in Table 3.1.

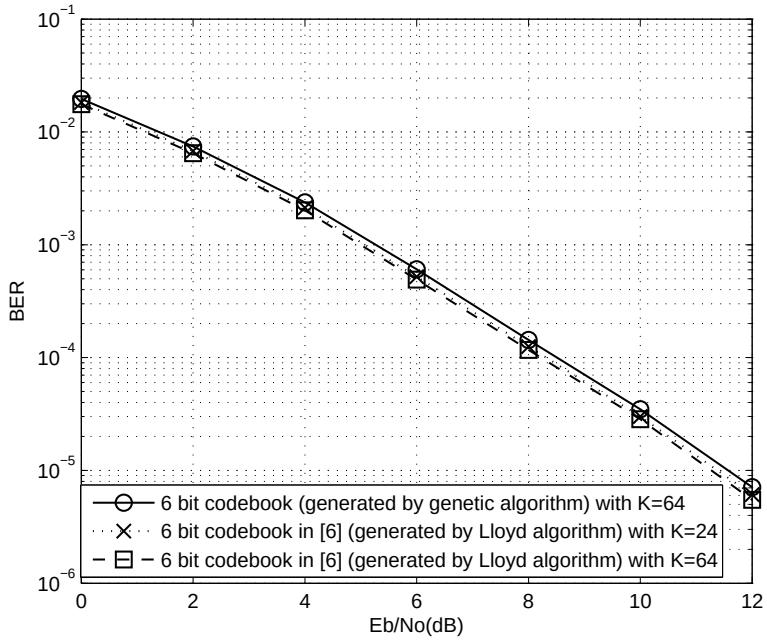

Figure 3.4: Performance comparison: the proposed ordering algorithm and exhaustive search for two 6-bit beamforming codebooks. one in [94] generated by Lloyd algorithm and the other generated by genetic algorithm in [92], $(N_t, N_r) = (4, 1)$.

Simulation 3: Our proposed algorithm is not sensitive to the codebook construction methods since it only depends on the chordal distance. In Fig. 3.4, we simulate the proposed algorithm using the 6-bit codebook in [94], which is generated by Lloyd algorithm. The 3-bit pilot codebook from Simulation 2 is used as the pilot codebook in this simulation. In general, codebooks generated by Lloyd algorithm have better performance than those generated by the genetic algorithm because genetic algorithm proposed in [92] has modulus constraint on entries of codewords which assures the average transmit powers on all transmit antennas are the same. On the contrary, Lloyd algorithm does not have

modulus constraint on the codewords and it allows the average transmit power on each transmit antenna to be different. Thus, the codebooks generated by the genetic algorithm provide quantized equal gain transmission while the codebooks generated by the Lloyd algorithm provide quantized maximum ratio transmission. When the transmitter has full channel knowledge, i.e., $B = \infty$ in the feedback channel, the performance of maximum ratio transmission is better than that of equal gain transmission. When the feedback channel is a low-rate channel, since both types of codebooks are generated according to the Grassmanian beamforming criterion in (3.6), the performance difference is typically small or even diminishes [79]. We observe that the performance of the codebook in [94] is around 0.3 dB better than that of the codebook generated by the genetic algorithm, both with exhaustive search, i.e., $K = 64$. In order to assess the effectiveness of the proposed algorithm, we let $K = 24$. The performance of the former codebook with the reduced-complexity search using our proposed ordering algorithm (with 50% complexity reduction) still has a gain of about 0.2 dB over the performance of the latter with exhaustive search.

Simulation 4: Fig. 3.5 shows the performance of the proposed algorithm with adjusted search length. The beamforming codebook and the pilot codebook are the same with the ones used for Fig. 3.3. We partition the interval $[0, d_{\max}]$ into four subintervals with thresholds $\mu_1 = 0.3162$, $\mu_2 = 0.4472$, $\mu_3 = 0.6325$. The codeword search lengths are $K_0 = 4$, $K_1 = 8$, $K_2 = 16$, and $K_3 = K = 24$. It is observed that the performance of the adjusted search length algorithm is virtually the same with the performance of fixed search length algorithm (with $K = 24$), but the former one has 57% in complexity reduction while the latter one has 50%. In this simulation, the number of subintervals and thresholds are not optimized. If they are further optimized, more reduction in complexity will be obtained.

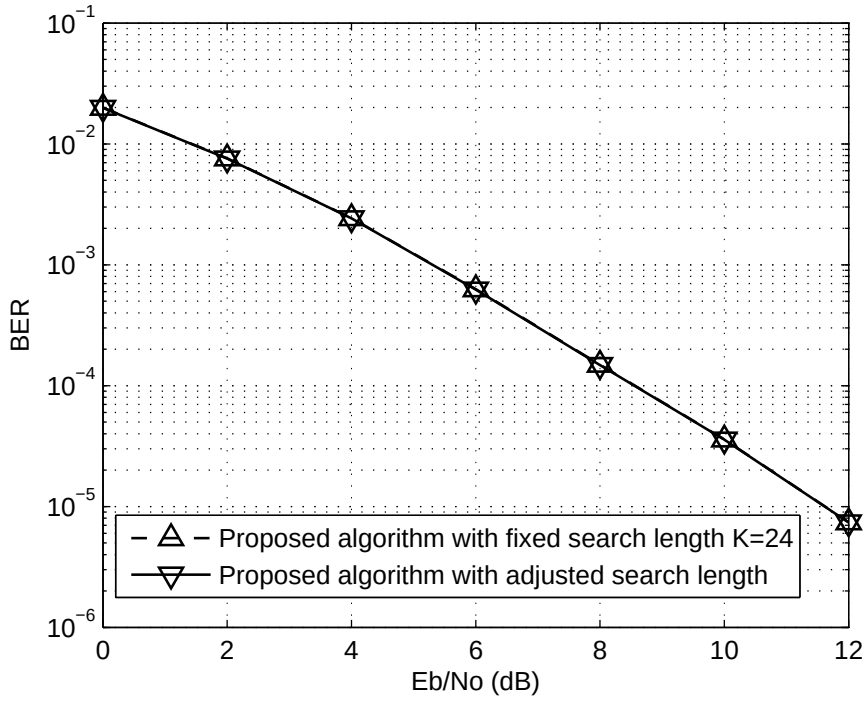

Figure 3.5: Performance of the proposed algorithm with adjusted search length, $(N_t, N_r) = (4, 1)$.

Simulation 5: Although our algorithm is proposed for MISO systems, through simulations, we find that the proposed ordering algorithm still works well in MIMO systems. Combining the proposed ordering algorithm with the bounding algorithm in [92] to form a modified order and bound algorithm, we can achieve better performance in complexity reduction than the original order and bound algorithm. In Fig. 3.6, a MIMO system with $N_t = 4$, $N_r = 5$ is considered. A 7-bit beamforming codebook is adopted and a 3-bit pilot codebook is used for the modified order and bound algorithm. When $K = 128$, that is, an exhaustive search is performed[4], the original order and bound algorithm and

[4]In MIMO case, the exhaustive search needs to compute $\|\boldsymbol{H}\boldsymbol{w}_i\|$ for $0 \leq i \leq N-1$ rather than (3.8),

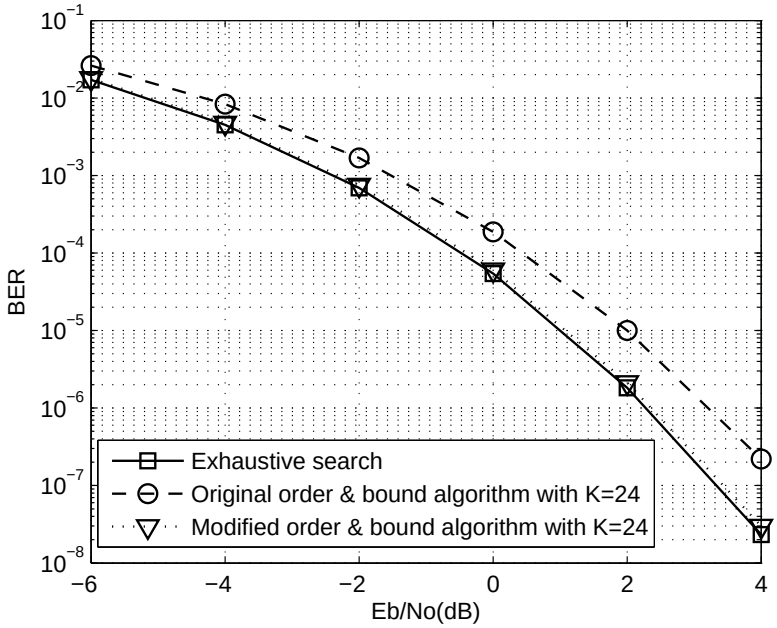

Figure 3.6: Performance comparison between the modified order and bound algorithm and original order and bound algorithm in [92] in a MIMO system with $(N_t, N_r) = (4, 5)$. A 7-bit beamforming codebook is adopted.

our modified order and bound algorithm reduce the search complexity by 71% and 74%, respectively. When $K = 24$, the search complexity reduces by 90% for both algorithms. But at BER of 10^{-6}, the modified order and bound algorithm has less than 0.15 dB of SNR degradation while the original order and bound algorithm has about 1 dB of SNR degradation. In order to limit the performance loss of the original order and bound algorithm within 0.15 dB, the corresponding search length K must be increased to 96 and the search complexity is only reduced by 74% as shown in Fig. 3.7. Performence comparison results are summerized in Table 3.2.

where $\boldsymbol{H}$ is an $N_r \times N_t$ channel matrix.

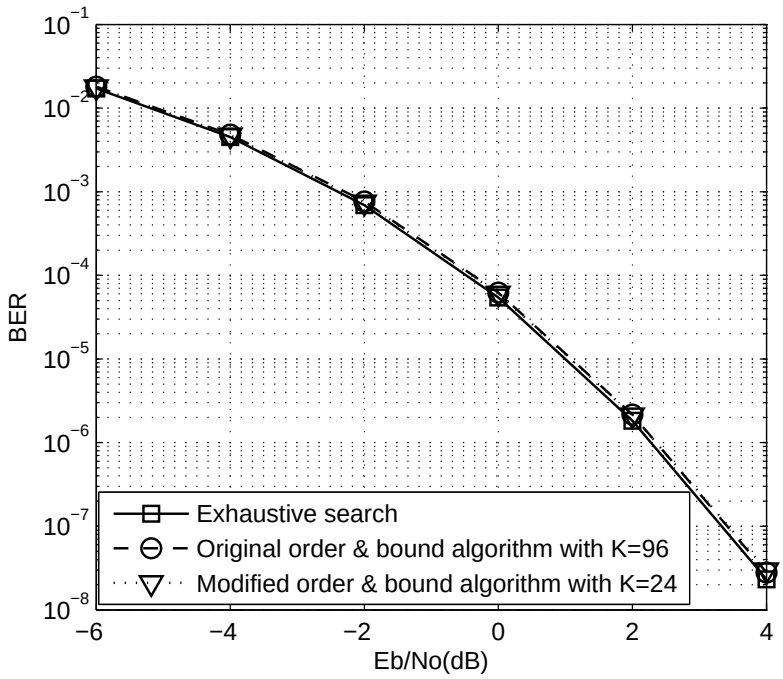

Figure 3.7: Re-plot of Fig. 3.6 with $K = 96$ in original order and bound algorithm.

Table 3.2: Performance comparison. Data collected from Fig. 3.6 and Fig. 3.7. O&B is the abbreviation of order and bound algorithm.

$(N_t, N_r) = (4, 5)$	N_p	N	K	complexity	degradation
modified O&B	8 (3 bits)	128 (7 bits)	128	-74%	0 dB
original O&B	N/A	128 (7 bits)	128	-71%	0 dB
modified O&B	8 (3 bits)	128 (7 bits)	24	-90%	0.15dB @ 10^{-6}
original O&B	N/A	128 (7 bits)	24	-90%	1.0dB @ 10^{-6}
original O&B	N/A	128 (7 bits)	96	-74%	0.15dB @ 10^{-6}

3.5 Conclusion

A complexity-reduction algorithm at the receivers for codeword search in MISO beamforming systems is proposed and analyzed in this chapter. The proposed algo-

rithm effectively reduces the computational complexity and still maintains comparable performance while the order and bound algorithm proposed in [92] does not work due to Choleskey decomposition in addition to the fact that the codeword sequence is not properly ordered. We introduce a pilot codebook to partition the beamforming codebook into N_P smaller groups and line up the groups to form a search sequence. The proposed ordering algorithm allows one to search only a small portion of the beamforming codebook and still provides a performance that is very close to the performance of the exhaustive search algorithm. Although our ordering algorithm is proposed for MISO beamforming systems, when combined with the bounding algorithm in [92], the modified order and bound algorithm still works well for MIMO beamforming systems. When the search length is reduced, a comparable performance and significant complexity reduction are still observed in MIMO beamforming systems.

Chapter 4 – On Angle Feedback and Antenna Shuffling in Double Space-Time Transmit Diversity Systems

4.1 Introduction

Systems employing OSTBC coding scheme can explore full transmit and receive diversity when there is no CSI at the transmitter and the symbols in code matrix can be decoded symbol-by-symbol using ML decoding with linear complexity. Full rate (rate one) OSTBCs with complex symbols, however, do not exist when the number of transmit antennas is greater than two [12]. In order to increase the transmission rate, DSTTD systems have been studied for high data rate transmission [41, 82, 97–99]. Due to feedback, a closed-loop DSTTD system has much better performance than the one without feedback.

In WiMAX 802.16e [82], transmit antenna shuffling scheme is adopted to improve the system performance in a DSTTD system with four transmit antennas. Recently, the angle feedback scheme is proposed [39–41]. It is shown that for QOSTBC with four transmit antennas the feedback of the exact angle information can be used to achieve full diversity [39, 40]. For DSTTD systems with four transmit antennas, the angle feedback scheme is proposed in [41] to suppress interference caused by the non-orthogonal structure of quadratic product of the effective channel, where a rotation factor $c = e^{j\theta}$, which is controlled by the receiver through feedback, is applied to the transmit symbols transmitted by the second transmit antenna.

Because rotating the symbols transmitted by the transmit antennas is equivalent to rotating the channel coefficients, we can consider the effective channel coefficients

obtained by multiplying the channel coefficients with the rotation factor c. Note that angle rotation can be applied to all transmit antennas. Consider a DSTTD system with four transmit antennas. The natural questions are: 1) If a scheme with four rotation factors applied to four transmit antennas is used, will it perform better than the scheme with the rotation applied to only one transmit antenna which is proposed in [41] with linear receivers? 2) To what extent will linear receivers suppress interference as proposed in [41]? In this chapter, we first answer these two questions that will help us further understand angle feedback scheme, and then we compare angle feedback scheme with transmit antenna shuffling scheme since both schemes are feedback scheme for DSTTD systems with four transmit antennas and no comparison has been made before. We show that with linear equalizers the transmit antenna shuffling scheme outperforms the angle feedback scheme.

4.2 Angle Feedback versus Transmit Antenna Shuffling

4.2.1 Angle Feedback

Consider a four-transmit-antenna system that employs angle feedback. An example system with two receive antennas is shown in Fig. 4.1.

In angle feedback scheme, four transmit antennas are divided into two groups with each group containing two transmit antennas. Each group of antennas are driven by its own Alamouti OSTBC encoder. We provide four angle rotation factors $c_1 = e^{j\theta_1}$, $c_2 = e^{j\theta_2}$, $c_3 = e^{j\theta_3}$ and $c_4 = e^{j\theta_4}$ and apply rotation factor c_i to the symbols transmitted by the ith transmit antenna, $1 \leq i \leq 4$. Since $|c_i| = 1$, the transmit power of each transmit antenna is unchanged. Combine two Alamouti code matrices to form a 4×2

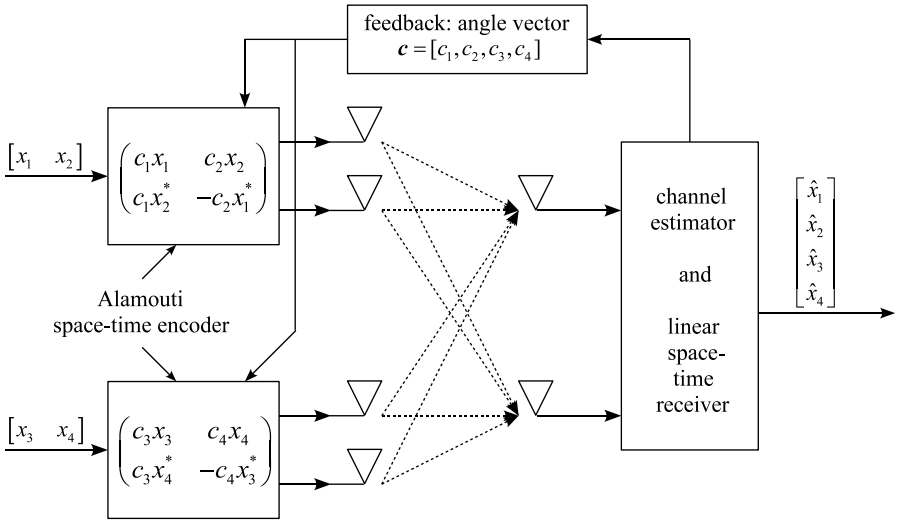

Figure 4.1: Angle feedback scheme [41] of a DSTTD system with four transmit antennas and two receive antennas. Four angle rotation factors, $c_i = e^{j\theta_i} (1 \leq i \leq 4)$, are used. c_i is applied to the symbols transmitted by the ith transmit antenna.

code matrix

$$\boldsymbol{X} = \begin{bmatrix} c_1 x_1 & c_2 x_2 & c_3 x_3 & c_4 x_4 \\ c_1 x_2^* & -c_2 x_1^* & c_3 x_4^* & -c_4 x_3 * \end{bmatrix}^T . \tag{4.1}$$

where column dimension represents transmit symbol intervals and row dimension represents transmit antennas. Since there are four symbols in a code matrix which is transmitted over two successive time intervals. In order to decode transmitted symbols in a code matrix, at least two receive antennas are needed, i.e., $N_r \geq 2$, where N_r is the number of receive antennas. Let's assume that channel coefficients are perfectly known at the receiver and remain constant during one code matrix. The received signal

$\boldsymbol{R} \in \mathcal{C}^{N_r \times 2}$ can be expressed as

$$\boldsymbol{R} = \bar{\boldsymbol{H}}\boldsymbol{X} + \boldsymbol{N} \tag{4.2}$$

where

$$\begin{aligned}
\boldsymbol{R} &= [\boldsymbol{r}_1 \quad \boldsymbol{r}_2 \quad \cdots \quad \boldsymbol{r}_{N_r}]^T && (4.3)\\
\boldsymbol{r}_j &= [r_{j,1} \quad r_{j,2}]^T && (4.4)\\
\boldsymbol{N} &= [\boldsymbol{n}_1 \quad \boldsymbol{n}_2 \quad \cdots \quad \boldsymbol{n}_{N_r}]^T && (4.5)\\
\boldsymbol{n}_j &= [n_{j,1} \quad n_{j,2}]^T && (4.6)\\
\bar{\boldsymbol{H}} &= [\boldsymbol{h}_1 \quad \boldsymbol{h}_2 \quad \cdots \quad \boldsymbol{h}_{N_r}]^T && (4.7)\\
\boldsymbol{h}_j &= [h_{1j} \quad h_{2j} \quad g_{1j} \quad g_{2j}]^T && (4.8)
\end{aligned}$$

and $r_{j,k}$ and $n_{j,k}$ are, respectively, the received signal and circularly symmetric, complex Gaussian noise at the jth receive antenna at the kth time interval (k=1,2). $E(|n_{j,k}|^2) = \sigma^2$ and $E[n_{j,k}n_{i,l}^*] = 0\ \forall i \neq j$ or $l \neq k$. h_{1j} and h_{2j} are the channel coefficients from the first and the second transmit antennas (the first group) to the jth receive antenna. g_{1j} and g_{2j} are the channel coefficients from the third and the fourth transmit antennas (the second group) to the jth receive antenna. In order to decode the transmitted symbols, we rewrite (4.2) into a more convenient format given by

$$\boldsymbol{r} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \tag{4.9}$$

where

$$\boldsymbol{x} = [x_1 \;\; x_2 \;\; x_3 \;\; x_4]^T \tag{4.10}$$

$$\boldsymbol{H} = [\boldsymbol{H}_1 \;\; \boldsymbol{H}_2 \;\; \cdots \;\; \boldsymbol{H}_{N_r}]^T \tag{4.11}$$

$$\boldsymbol{r} = [\bar{\boldsymbol{r}}_1 \;\; \bar{\boldsymbol{r}}_2 \;\; \cdots \;\; \bar{\boldsymbol{r}}_{N_r}]^T \tag{4.12}$$

$$\boldsymbol{n} = [\bar{\boldsymbol{n}}_1 \;\; \bar{\boldsymbol{n}}_2 \;\; \cdots \;\; \bar{\boldsymbol{n}}_{N_r}]^T \tag{4.13}$$

$$\bar{\boldsymbol{r}}_j = [r_{j,1} \;\; r_{j,2}^*] \tag{4.14}$$

$$\bar{\boldsymbol{n}}_j = [n_{j,1} \;\; n_{j,2}^*] \tag{4.15}$$

and the effective channel matrix, $\boldsymbol{H}_j$, from the transmitter to the jth receive antenna is

$$\boldsymbol{H}_j = \begin{bmatrix} c_1 h_{1j} & c_2 h_{2j} & c_3 g_{1j} & c_4 g_{2j} \\ -c_2^* h_{2j}^* & c_1^* h_{1j}^* & -c_4^* g_{2j}^* & c_3^* g_{1j}^* \end{bmatrix}^T, \tag{4.16}$$

$E[\boldsymbol{n}\boldsymbol{n}^H] = \sigma^2 \boldsymbol{I}$ since $\bar{\boldsymbol{n}}_j$ and $\boldsymbol{n}_j$ have the same distribution. With the notation $\boldsymbol{c} = [c_1, c_2, c_3, c_4]$, the quadratic channel product in [41, eq. (5)] becomes

$$\boldsymbol{H}^H \boldsymbol{H} = \begin{bmatrix} \rho \boldsymbol{I}_2 & \boldsymbol{U} \\ \boldsymbol{U}^H & \mu \boldsymbol{I}_2 \end{bmatrix}$$

where $\boldsymbol{I}_2$ is the 2×2 identify matrix and

$$\rho = \sum_{i=1}^{2} \sum_{j=1}^{N_r} |h_{ij}|^2 , \tag{4.17}$$

$$\mu = \sum_{i=1}^{2} \sum_{j=1}^{N_r} |g_{ij}|^2 , \tag{4.18}$$

$$\boldsymbol{U} = \begin{bmatrix} \delta_1(\boldsymbol{c}) & \delta_2(\boldsymbol{c}) \\ -\delta_2^*(\boldsymbol{c}) & \delta_1^*(\boldsymbol{c}) \end{bmatrix}, \tag{4.19}$$

$$\delta_1(\boldsymbol{c}) = c_1^* c_3 \sum_{j=1}^{N_r} h_{1j}^* g_{1j} + c_2 c_4^* \sum_{j=1}^{N_r} h_{2j} g_{2j}^*, \tag{4.20}$$

$$\delta_2(\boldsymbol{c}) = c_1^* c_4 \sum_{j=1}^{N_r} h_{1j}^* g_{2j} - c_2 c_3^* \sum_{j=1}^{N_r} h_{2j} g_{1j}^* \tag{4.21}$$

and ρ and μ are independent of c_i, $1 \leq i \leq 4$. Let $\eta = |\delta_1(\boldsymbol{c})|^2 + |\delta_2(\boldsymbol{c})|^2$. Matrix $\boldsymbol{U}$ satisfies the following relationship:

$$\boldsymbol{U}^H \boldsymbol{U} = (|\delta_1(\boldsymbol{c})|^2 + |\delta_2(\boldsymbol{c})|^2)\boldsymbol{I}_2 = \eta \boldsymbol{I}_2, \tag{4.22}$$

In [41], the interference suppression criterion is to minimize η. The following lemma relates minimization of MSE of a ZF or MMSE equalizer to minimization of η. This connection is not explored in [41].

Lemma 1 *When a ZF or MMSE equalizer is applied at the receiver, minimizing the MSE of the detected symbols is equivalent to minimizing η.*

Proof. Applying an MMSE equalizer to the received signal $\boldsymbol{r}$ in (4.9), the estimate of

the transmitted symbol vector is expressed as

$$\hat{\boldsymbol{x}}_{\mathrm{MMSE}} = \left(\boldsymbol{H}^H\boldsymbol{H} + \frac{1}{\zeta}\boldsymbol{I}_4\right)^{-1}\boldsymbol{H}^H\boldsymbol{r} \tag{4.23}$$

where $\zeta \triangleq E_s/\sigma^2$ is defined as SNR, E_s is the average transmit symbol energy. Using matrix inversion lemma [83, 0.7.3] for partitioned matrices [1], we obtain the MSE of the MMSE equalizer as

$$\begin{aligned}\mathrm{E}\left[\|\hat{\boldsymbol{x}}_{\mathrm{MMSE}} - \boldsymbol{x}\|^2\right] &= \sigma^2\mathrm{tr}\left(\left[\boldsymbol{H}^H\boldsymbol{H} + \frac{1}{\zeta}\boldsymbol{I}_4\right]^{-1}\right) \\ &= \frac{\sigma^2\left[2(\rho+\mu)+4/\zeta\right]}{1/\zeta^2 + (\rho+\mu)/\zeta + \rho\mu - \eta}.\end{aligned} \tag{4.24}$$

For a ZF equalizer, we have

$$\hat{\boldsymbol{x}}_{\mathrm{ZF}} = (\boldsymbol{H}^H\boldsymbol{H})^{-1}\boldsymbol{H}^H\boldsymbol{r}. \tag{4.25}$$

The MSE of the ZF equalizer is expressed as

$$\begin{aligned}\mathrm{E}\left[\|\hat{\boldsymbol{x}}_{\mathrm{ZF}} - \boldsymbol{x}\|^2\right] &= \sigma^2\mathrm{tr}\left(\left[\boldsymbol{H}^H\boldsymbol{H}\right]^{-1}\right) \\ &= \frac{2(\rho+\mu)\sigma^2}{\rho\mu - \eta}.\end{aligned} \tag{4.26}$$

Our goal is to minimize the MSE with an appropriate rotation vector $\boldsymbol{c}$. Since ζ, σ^2, ρ,

[1] Let a $(m+n)\times(m+n)$ matrix $\boldsymbol{M}$ be partitioned into a block form: $\boldsymbol{M} = \left[\begin{smallmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{D}\end{smallmatrix}\right]$, where $\boldsymbol{A} \in \mathcal{C}^{m\times m}$, $\boldsymbol{B} \in \mathcal{C}^{m\times n}$, $\boldsymbol{C} \in \mathcal{C}^{n\times m}$ and $\boldsymbol{D} \in \mathcal{C}^{n\times n}$. Assume all relevant matrix inversions exist, then $\boldsymbol{M}^{-1} = \left[\begin{smallmatrix}(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} & \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}-\boldsymbol{D})^{-1}\\ (\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}-\boldsymbol{D})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & (\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\end{smallmatrix}\right]$.

and μ are independent of $\boldsymbol{c}$, we have

$$\arg\min_{\boldsymbol{c}} \mathrm{E}\left[\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|^2\right] = \arg\min_{\boldsymbol{c}} \eta$$

for both MMSE and ZF equalizers. ■

Lemma 1 shows that we cannot further minimize the MSE of a ZF or MMSE equalizer if η is minimized.

Note that η is a function of $\boldsymbol{c}$ and can be written as

$$\begin{aligned}
\eta &= |\delta_1(\boldsymbol{c})|^2 + |\delta_2(\boldsymbol{c})|^2 \\
&= \delta_1(\boldsymbol{c})\delta_1(\boldsymbol{c})^* + \delta_2(\boldsymbol{c})\delta_2(\boldsymbol{c})^* \\
&= \alpha \cdot c + \alpha^* \cdot c^* + \xi \\
&= 2\Re\{\alpha \cdot c\} + \xi
\end{aligned}$$

where α and ξ are

$$\begin{aligned}
\alpha &= (\sum_{l=1}^{N_r} h_{1l} g_{1l}^*) \sum_{j=1}^{N_r} (h_{2j} g_{2j}^*) - (\sum_{l=1}^{N_r} h_{1l} g_{2l}^*)(\sum_{j=1}^{N_r} h_{2j} g_{1j}^*) \\
\xi &= (\sum_{l=1}^{N_r} h_{2l}^* g_{2l}) \sum_{j=1}^{N_r} (h_{2j} g_{2j}^*) + (\sum_{l=1}^{N_r} h_{1l} g_{1l}^*)(\sum_{j=1}^{N_r} h_{1j}^* g_{1j}) \\
&+ (\sum_{l=1}^{N_r} h_{1l} g_{2l}^*) \sum_{j=1}^{N_r} (h_{1j}^* g_{2j}) + (\sum_{l=1}^{N_r} h_{2l}^* g_{1l})(\sum_{j=1}^{N_r} h_{2j} g_{1j}^*)
\end{aligned}$$

and

$$c = c_1 c_2 c_3^* c_4^* = e^{j(\theta_1 + \theta_2 - \theta_3 - \theta_4)}. \tag{4.27}$$

Note that ξ and η are nonnegative and real, and ξ does not change with c. The solution for c that minimizes η, equivalent to minimizing $\Re\{\alpha \cdot c\}$, is expressed as [41]

$$c = e^{j(\pi - \psi)} \tag{4.28}$$

where ψ is the phase of α. The resulting minimum of η is given as $\eta = \xi - 2|\alpha|$. From (4.27), we observe that the optimum rotation angle of c can be realized by any θ_i, $i \in \{1, 2, 3, 4\}$, with $\theta_j = 0, \quad 1 \leq j \neq i \leq 4$. We have proved next lemma.

Lemma 2 *With a ZF or MMSE equalizer, it is sufficient to perform rotation on only one transmit antenna rather on four transmit antennas to suppress interference. When the ith transmit antenna ($1 \leq i \leq 4$) is selected for rotation, the optimum rotation angle is $\theta_i = (-1)^{\lceil i/2 \rceil}(\psi - \pi)$, where $\lceil \cdot \rceil$ represents the ceiling function, and $\forall\, j$, $1 \leq j \neq i \leq 4$, $\theta_j = 0$.*

Lemma 2 shows that in terms of interference suppression, angle rotation performed on only one transmit antenna can have the same performance as rotation on all transmit antennas; thus we will rotate the data transmitted by the second antenna as in [41] for performance comparison. When the feedback channel is rate limited, the quantized angle needs to be fed back from the receiver to the transmitter. Since the optimum feedback angle θ is distributed in $[0, 2\pi]$, for B bits of feedback, the codebook is $C_B = [c_0, \cdots, c_m, \cdots, c_{2^B-1}] = [1, e^{j\pi/2^{B-1}}, \cdots, e^{jm\pi/2^{B-1}}, \cdots, e^{j(2^B-1)\pi/2^{B-1}}]$ [41]. The best index to be fed back is

$$I = \arg \min_{m \in \{0, \cdots, 2^B - 1\}} \Re\{\alpha \cdot c_m\}. \tag{4.29}$$

4.2.2 Transmit Antenna Shuffling

Transmit antenna shuffling (permutation) in the DSTTD system with four transmit antennas has been adopted in WiMAX [82] for high data rate transmission. The DSTTD system with four transmit antennas and two receive antennas is illustrated in Fig. 4.2. The four outputs of two Alamouti encoders go through shuffling matrix and then drive the physical antennas. For simplicity of notation, we treat h_{ij} and g_{ij} in (4.2), $i = 1, 2$ and $1 \leq j \leq N_r$, as logical channels, and denote $\hbar_{kj}$ as the physical channel between the kth ($1 \leq k \leq 4$) transmit antenna and the jth receive antenna. The mapping between physical channels and logical channels is expressed as [82]

$$[h_{1j} \quad h_{2j} \quad g_{1j} \quad g_{2j}] = [\hbar_{1j} \quad \hbar_{2j} \quad \hbar_{3j} \quad \hbar_{4j}]\boldsymbol{W} \tag{4.30}$$

where $\boldsymbol{W}$ is the permutation matrix and $\boldsymbol{W} \in S_{\mathrm{W}} = \{\boldsymbol{W}_1, \boldsymbol{W}_2, \boldsymbol{W}_3, \boldsymbol{W}_4, \boldsymbol{W}_5, \boldsymbol{W}_6\}$ with $\boldsymbol{W}_1 = [\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3, \mathbf{i}_4]$, $\boldsymbol{W}_2 = [\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_4, \mathbf{i}_3]$, $\boldsymbol{W}_3 = [\mathbf{i}_1, \mathbf{i}_3, \mathbf{i}_2, \mathbf{i}_4]$, $\boldsymbol{W}_4 = [\mathbf{i}_1, \mathbf{i}_3, \mathbf{i}_4, \mathbf{i}_2]$, $\boldsymbol{W}_5 = [\mathbf{i}_1, \mathbf{i}_4, \mathbf{i}_2, \mathbf{i}_3]$ and $\boldsymbol{W}_6 = [\mathbf{i}_1, \mathbf{i}_4, \mathbf{i}_3, \mathbf{i}_2]$. $\mathbf{i}_k$ is a 4×1 vector having only one non-zero element in the kth position with value of one.

In this section, we first generalize the permutation matrix selection criterion from $N_r = 2$ [99, *Property 1*] to $N_r \geq 2$ using a ZF or MMSE equalizer. This helps facilitate permutation matrix selection. Using (4.17) we can verify that permutations change the values of ρ and μ but do not change the sum of ρ and μ. Note there are no angle rotations in transmit antenna shuffling scheme, so c_i takes value of one. In order to minimize the MSE of MMSE and ZF equalizers in (4.24) and (4.26), we must maximize $\rho\mu - \eta$ using one of the six permutation matrices in S_W. A simple permutation matrix selection criterion for $N_r = 2$ is given in [99]. We generalize the matrix selection criterion

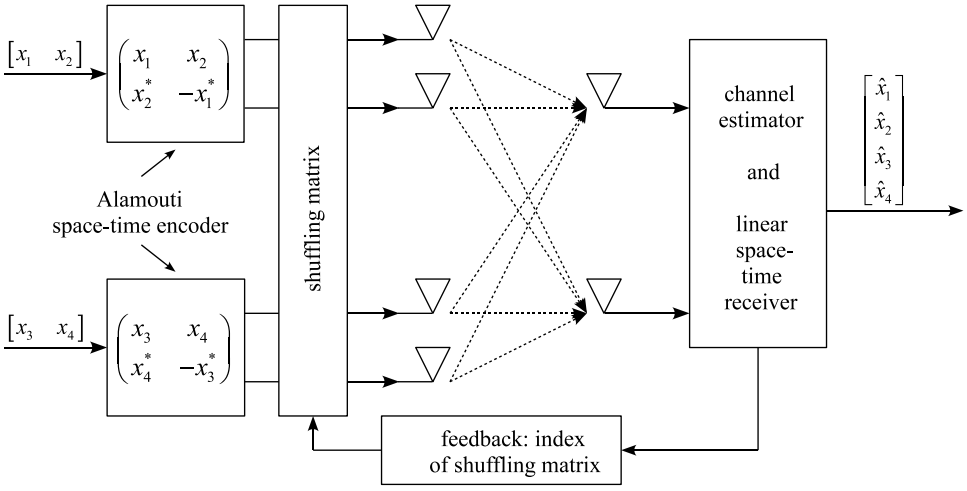

Figure 4.2: Transmit antenna shuffling scheme in WiMAX [82] of a DSTTD system with four transmit antennas and two receive antennas.

to cases of $N_r \geq 2$ through the next lemma.

Lemma 3 *In order to minimize MSE of received symbols, the permutation matrix selection criterion for $N_r \geq 2$ for the DSTTD systems with a ZF or MMSE equalizer is expressed as*

$$\arg\max_{\boldsymbol{W} \in S_{\boldsymbol{W}}} (\rho\mu - \eta) = \arg\min_{\boldsymbol{W} \in S_{\boldsymbol{W}}} \sum_{m=1}^{M} |\Gamma_m|^2 \tag{4.31}$$

where

$$\Gamma_m = \det\left(\begin{bmatrix} h_{1\boldsymbol{\ell}_m(1)} & h_{2\boldsymbol{\ell}_m(1)} \\ h_{1\boldsymbol{\ell}_m(2)} & h_{2\boldsymbol{\ell}_m(2)} \end{bmatrix}\right) + \det\left(\begin{bmatrix} g_{1\boldsymbol{\ell}_m(1)} & g_{2\boldsymbol{\ell}_m(1)} \\ g_{1\boldsymbol{\ell}_m(2)} & g_{2\boldsymbol{\ell}_m(2)} \end{bmatrix}\right)$$

and $M = \binom{N_r}{2} = \frac{N_r!}{2(N_r-2)!}$ *is the number of sets of combination with each set containing two receive antenna indices.* 1×2 *vector* $\boldsymbol{\ell}_m = [\boldsymbol{\ell}_m(1)\ \boldsymbol{\ell}_m(2)]$ *denotes the* m*th set with indices* $1 \le \boldsymbol{\ell}_m(1) < \boldsymbol{\ell}_m(2) \le N_r$, $1 \le m \le M$.

Proof. With some mathematical manipulations, we have

$$\rho\mu - \eta = \sum_{m=1}^{M} (\rho_m \mu_m - \eta_m), \tag{4.32}$$

where

$$\begin{aligned} \rho_m &= \sum_{i=1}^{2}\sum_{k=1}^{2} \left|h_{i\boldsymbol{\ell}_m(k)}\right|^2, & (4.33)\\ \mu_m &= \sum_{i=1}^{2}\sum_{k=1}^{2} \left|g_{i\boldsymbol{\ell}_m(k)}\right|^2, & (4.34)\\ \eta_m &= \left|\delta_{1,m}\right|^2 + \left|\delta_{2,m}\right|^2, & (4.35) \end{aligned}$$

with

$$\begin{aligned} \delta_{1,m} &= \sum_{k=1}^{2} (h^*_{1\boldsymbol{\ell}_m(k)} g_{1\boldsymbol{\ell}_m(k)} + h_{2\boldsymbol{\ell}_m(k)} g^*_{2\boldsymbol{\ell}_m(k)}), & (4.36)\\ \delta_{2,m} &= \sum_{k=1}^{2} (h^*_{1\boldsymbol{\ell}_m(k)} g_{2\boldsymbol{\ell}_m(k)} - h_{2\boldsymbol{\ell}_m(k)} g^*_{1\boldsymbol{\ell}_m(k)}). & (4.37) \end{aligned}$$

We can show that

$$\rho_m \mu_m - \eta_m = \Lambda_m - |\Gamma_m|^2 \tag{4.38}$$

with

$$\Lambda_m = \sum_{i=1}^{4} \sum_{j=1, j\neq i}^{4} |\hbar_{i\boldsymbol{\ell}_m(1)} \hbar_{j\boldsymbol{\ell}_m(2)}|^2 - 2\Re\left\{ \sum_{i=1}^{4} \left(\hbar^*_{i\boldsymbol{\ell}_m(1)} \hbar_{i\boldsymbol{\ell}_m(2)} \sum_{j=i+1}^{4} \hbar_{j\boldsymbol{\ell}_m(1)} \hbar^*_{j\boldsymbol{\ell}_m(2)} \right) \right\}. \quad (4.39)$$

Since Λ_m is independent of $\boldsymbol{W}$, we have (4.31). A more detailed proof is provided in Appendix A ■

When $N_r = 2$, the right-hand side of (4.31) has only one term, which can be written as $\arg\min_{\boldsymbol{W}\in S_W} |\Gamma_1|$. This has the same form as in [99, *Property 1*], a special case of lemma 3.

Next, we show that the 1-bit angle feedback scheme in [41] is a special case of transmit antenna shuffling. For angle feedback, we assume that a 1-bit angle feedback is applied to the second antenna with $c = \pm 1$. We have $[h_{1j}\ h_{2j}\ g_{1j}\ g_{2j}] = [\hbar_{1j}\ c\hbar_{2j}\ \hbar_{3j}\ \hbar_{4j}]$ and h_{2j} could be either $\hbar_{2j}$ or $-\hbar_{2j}$. $|\Gamma_m(c)|$ is expressed as

$$|\Gamma_m(c=1)| = \left| \det\left(\begin{bmatrix} \hbar_{1\boldsymbol{\ell}_m(1)} & \hbar_{2\boldsymbol{\ell}_m(1)} \\ \hbar_{1\boldsymbol{\ell}_m(2)} & \hbar_{2\boldsymbol{\ell}_m(2)} \end{bmatrix} \right) + \det\left(\begin{bmatrix} \hbar_{3\boldsymbol{\ell}_m(1)} & \hbar_{4\boldsymbol{\ell}_m(1)} \\ \hbar_{3\boldsymbol{\ell}_m(2)} & \hbar_{4\boldsymbol{\ell}_m(2)} \end{bmatrix} \right) \right|. \quad (4.40a)$$

$$|\Gamma_m(c=-1)| = \left| \det\left(\begin{bmatrix} \hbar_{1\boldsymbol{\ell}_m(1)} & \hbar_{2\boldsymbol{\ell}_m(1)} \\ \hbar_{1\boldsymbol{\ell}_m(2)} & \hbar_{2\boldsymbol{\ell}_m(2)} \end{bmatrix} \right) - \det\left(\begin{bmatrix} \hbar_{3\boldsymbol{\ell}_m(1)} & \hbar_{4\boldsymbol{\ell}_m(1)} \\ \hbar_{3\boldsymbol{\ell}_m(2)} & \hbar_{4\boldsymbol{\ell}_m(2)} \end{bmatrix} \right) \right|. \quad (4.40b)$$

For transmit antenna shuffling, we select shuffling matrices $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ for 1-bit feedback. We have

$$|\Gamma_m(\boldsymbol{W} = \boldsymbol{W}_1)| = |\Gamma_m(c=1)|, \quad (4.41a)$$

$$|\Gamma_m(\boldsymbol{W} = \boldsymbol{W}_2)| = |\Gamma_m(c=-1)|. \quad (4.41b)$$

(4.41a) is easy to verify. (4.41b) can be verified as follows: When $c = -1$ (applied to second transmit antenna), we have

$$\begin{aligned}
|\Gamma_m(c=-1)| &= \left| \det\left(\begin{bmatrix} \hbar_{1\boldsymbol{\ell}_m(1)} & -\hbar_{2\boldsymbol{\ell}_m(1)} \\ \hbar_{1\boldsymbol{\ell}_m(2)} & -\hbar_{2\boldsymbol{\ell}_m(2)} \end{bmatrix} \right) + \det\left(\begin{bmatrix} \hbar_{3\boldsymbol{\ell}_m(1)} & \hbar_{4\boldsymbol{\ell}_m(1)} \\ \hbar_{3\boldsymbol{\ell}_m(2)} & \hbar_{4\boldsymbol{\ell}_m(2)} \end{bmatrix} \right) \right| \\
&= \left| \det\left(\begin{bmatrix} \hbar_{1\boldsymbol{\ell}_m(1)} & \hbar_{2\boldsymbol{\ell}_m(1)} \\ \hbar_{1\boldsymbol{\ell}_m(2)} & \hbar_{2\boldsymbol{\ell}_m(2)} \end{bmatrix} \right) - \det\left(\begin{bmatrix} \hbar_{3\boldsymbol{\ell}_m(1)} & \hbar_{4\boldsymbol{\ell}_m(1)} \\ \hbar_{3\boldsymbol{\ell}_m(2)} & \hbar_{4\boldsymbol{\ell}_m(2)} \end{bmatrix} \right) \right| \\
&\overset{a}{=} \left| \det\left(\begin{bmatrix} \hbar_{1\boldsymbol{\ell}_m(1)} & \hbar_{2\boldsymbol{\ell}_m(1)} \\ \hbar_{1\boldsymbol{\ell}_m(2)} & \hbar_{2\boldsymbol{\ell}_m(2)} \end{bmatrix} \right) + \det\left(\begin{bmatrix} \hbar_{4\boldsymbol{\ell}_m(1)} & \hbar_{3\boldsymbol{\ell}_m(1)} \\ \hbar_{4\boldsymbol{\ell}_m(2)} & \hbar_{3\boldsymbol{\ell}_m(2)} \end{bmatrix} \right) \right| \\
&= |\Gamma_m(\boldsymbol{W} = \boldsymbol{W}_2)| \,.
\end{aligned}$$

Equality (a) follows the fact that the det of matrix changes sign when two columns are swapped.

From the proof of lemma 3, we have $\rho\mu - \eta = \sum_{m=1}^{M}(\Lambda_m - |\Gamma_m|^2)$ and Λ_m does not change value with $c = \pm 1$, $\boldsymbol{W}_1$, and $\boldsymbol{W}_2$. Considering (4.41a) and (4.41b), we have

$$(\rho\mu - \eta)|_{c=1} = (\rho\mu - \eta)|_{\boldsymbol{W}=\boldsymbol{W}_1}, \tag{4.42a}$$

$$(\rho\mu - \eta)|_{c=-1} = (\rho\mu - \eta)|_{\boldsymbol{W}=\boldsymbol{W}_2}. \tag{4.42b}$$

It can be verified that no matter on which transmit antenna the 1-bit rotation factor $c = \pm 1$ is applied, (4.42a) and (4.42b) are always valid. Since ρ and μ are invariant with $c = \pm 1$ or $\boldsymbol{W} = \boldsymbol{W}_1/\boldsymbol{W}_2$, $\eta|_{c=1} = \eta|_{\boldsymbol{W}=\boldsymbol{W}_1}$ and $\eta|_{c=-1} = \eta|_{\boldsymbol{W}=\boldsymbol{W}_2}$. From (4.24) and (4.26), we have angle rotation with $c = -1$ (1) and transmit antenna shuffling with $\boldsymbol{W} = \boldsymbol{W}_2$ ($\boldsymbol{W}_1$) have the same MSE. This means selection of $c = -1$ (1) and selection of $\boldsymbol{W} = \boldsymbol{W}_2$ ($\boldsymbol{W}_1$) have to be in pair under the same channel realization.

With one-to-one mapping selections for any channel realization $\boldsymbol{H}$, let us further consider the BER performance of both schemes with 1-bit feedback. We only need to consider angle rotation $c = -1$ and shuffling matrix $\boldsymbol{W} = \boldsymbol{W}_2$ since angle feedback system with $c = 1$ and transmit antenna shuffling system with $\boldsymbol{W} = \boldsymbol{W}_1$ are exactly the same, thus have the same BER with a ZF or MMSE equalizer. Define $\eta|_{c=-1} = \eta|_{\boldsymbol{W}=\boldsymbol{W}_2} = \bar{\eta}$ and $\boldsymbol{\Pi} = \boldsymbol{H}^H\boldsymbol{H}$. The post-processing SNR of kth ($1 \le k \le 4$) data stream $\hat{x}_k$ of ZF and MMSE equalizers can be expressed as [33]

$$\begin{aligned} \gamma_k^{ZF} &= \frac{\zeta}{[\boldsymbol{\Pi}]_{k,k}^{-1}}, \\ \gamma_k^{MMSE} &= \frac{\zeta}{[(\boldsymbol{\Pi} + \frac{1}{\zeta}\boldsymbol{I}_4)^{-1}]_{k,k}} - 1. \end{aligned}$$

Using matrix inversion lemma for partitioned matrices we can verify that for any channel realization $\boldsymbol{H}$ resulting in $c = -1$ and $\boldsymbol{W} = \boldsymbol{W}_2$, the following equalities always hold:

$$[\boldsymbol{\Pi}^{-1}]_{k,k}|_{c=-1} = [\boldsymbol{\Pi}^{-1}]_{k,k}|_{\boldsymbol{W}=\boldsymbol{W}_2} = (\rho - \bar{\eta}/\mu)^{-1}, \qquad k = 1, 2 \tag{4.43}$$

$$[\boldsymbol{\Pi}^{-1}]_{k,k}|_{c=-1} = [\boldsymbol{\Pi}^{-1}]_{k,k}|_{\boldsymbol{W}=\boldsymbol{W}_2} = (\mu - \bar{\eta}/\rho)^{-1}, \qquad k = 3, 4 \tag{4.44}$$

$$\begin{aligned} [(\boldsymbol{\Pi} + \frac{1}{\zeta}\boldsymbol{I}_4)^{-1}]_{k,k}|_{c=-1} &= [(\boldsymbol{\Pi} + \frac{1}{\zeta}\boldsymbol{I}_4)^{-1}]_{k,k}|_{\boldsymbol{W}=\boldsymbol{W}_2} \\ &= \begin{cases} [\rho + 1/\zeta - \bar{\eta}/(\mu + 1/\zeta)]^{-1}, & k = 1, 2 \\ [\mu + 1/\zeta - \bar{\eta}/(\rho + 1/\zeta)]^{-1}, & k = 3, 4. \end{cases} \end{aligned} \tag{4.45}$$

Therefore, we have

$$\gamma_k^{ZF}|_{c=-1} = \gamma_k^{ZF}|_{\boldsymbol{W}=\boldsymbol{W}_2}, \tag{4.46}$$

$$\gamma_k^{MMSE}|_{c=-1} = \gamma_k^{MMSE}|_{\boldsymbol{W}=\boldsymbol{W}_2}. \tag{4.47}$$

Combining the scenarios $c = 1$ and $\boldsymbol{W} = \boldsymbol{W}_1$, we claim that both schemes (angle rotation with $c = \pm 1$ and transmit antenna shuffling with $\boldsymbol{W} = \boldsymbol{W}_1/\boldsymbol{W}_2$) have the same BER performance with a ZF or MMSE receiver. Following the same arguments we can also verify that this result holds no matter on which transmit antenna the angle rotation $c = \pm 1$ is applied.

Because any pair of shuffling matrices in S_W could be selected for 1-bit transmit antenna shuffling, we have the following lemma:

Lemma 4 *With a ZF or MMSE equalizer, the 1-bit angle rotation scheme proposed in [41] has the same performance with 1-bit transmit antenna shuffling with shuffling matrices* $\boldsymbol{W}_1 = [\boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{i}_3, \boldsymbol{i}_4]$ *and* $\boldsymbol{W}_2 = [\boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{i}_4, \boldsymbol{i}_3]$. *Furthermore, it does not perform better than the 1-bit transmit antenna shuffling scheme with a pair of properly selected shuffling matrices.*

Lemma 4 gives the relative performance between angle feedback scheme and transmit antenna shuffling scheme with 1-bit feedback. With two or more feedback bits, it is difficult to compare the performance of the two schemes analytically. We thus resort to simulations, through which we observe that transmit antenna shuffling with only two bits of feedback performs better than angle feedback with infinite number of feedback bits.

4.3 Simulation Results

We simulate the bit-error rate (BER) performance of an uncoded DSTTD system with 4 transmit antennas, and 2 and 3 receive antennas. The system employs quadrature phase-shift keying modulation and an MMSE receiver. Channel coefficients are i.i.d.,

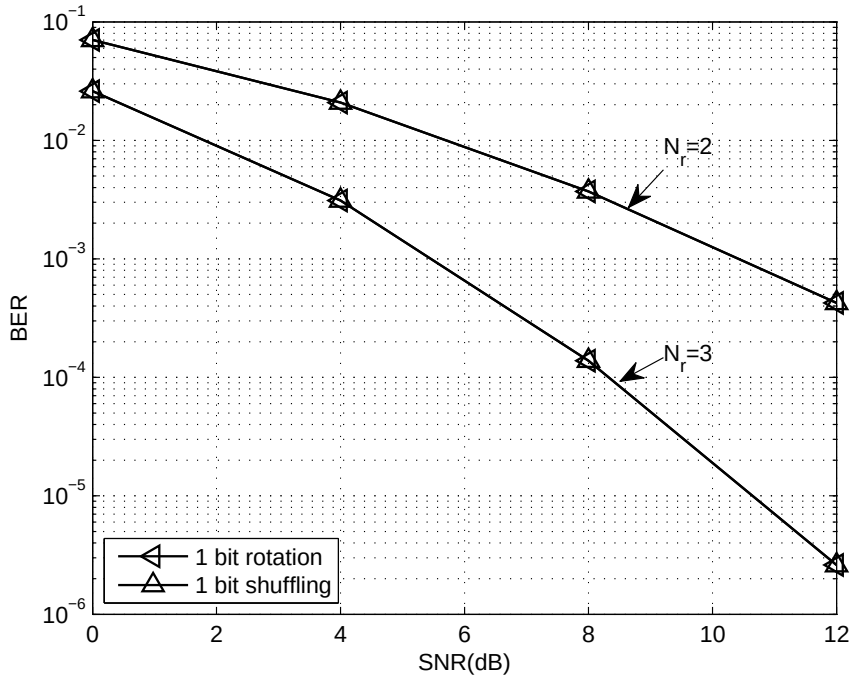

Figure 4.3: Performance of 1-bit angle feedback scheme with $c = \pm 1$ on the second transmit antenna and 1-bit transmit antenna shuffling scheme with shuffling set $\{\boldsymbol{W}_1, \boldsymbol{W}_2\}$.

circularly symmetric, complex Gaussian random variables with zero mean and unit variance. BER is averaged over 10^6 independent channel realizations. 4×10^3 QPSK symbols are transmitted over each channel realization.

First we verify that 1-bit angle feedback scheme with $c = \pm 1$ has the same performance with 1-bit transmit antenna shuffling scheme with shuffling matrix set $\{\boldsymbol{W}_1, \boldsymbol{W}_2\}$ which we claimed at Lemma 4. Angle rotation is applied on the second transmit antenna. Fig. 4.3 shows the BER curves of both schemes with above 1-bit feedback setting match each other exactly and this confirms our analytical results.

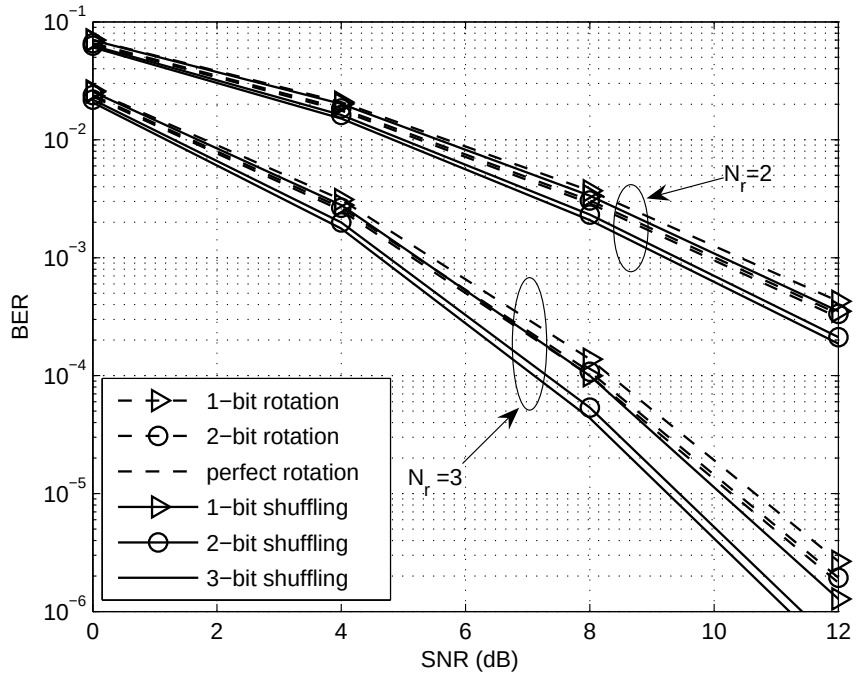

Figure 4.4: Performance comparison between angle feedback scheme (1-bit: $c = \pm 1$, 2-bit: $c = \pm 1, \pm j$) and transmit antenna shuffling scheme (1-bit: $\{\boldsymbol{W}_2, \boldsymbol{W}_3\}$, 2-bit: $\{\boldsymbol{W}_2, \boldsymbol{W}_3, \boldsymbol{W}_4, \boldsymbol{W}_5\}$, 3-bit: S_W) .

We further conduct simulations with different settings. For transmit antenna shuffling scheme, 1-bit transmit antenna shuffling set is $\{\boldsymbol{W}_2, \boldsymbol{W}_3\}$; two-bit shuffling matrix set is $\{\boldsymbol{W}_2, \boldsymbol{W}_3, \boldsymbol{W}_4, \boldsymbol{W}_5\}$; three-bit transmit antenna shuffling matrix set is S_W. The angle rotation scheme uses $c = \pm 1$ for 1-bit angle rotation and $c = \pm 1, \pm j$ for 2-bit angle rotation. Angle rotation is applied on the second transmit antenna. The results are shown in Fig. 4.4. We observe that the 1-bit transmit antenna shuffling scheme with new shuffling set has a better performance than the 1-bit angle feedback scheme; 2-bit transmit antenna shuffling performs better than angle feedback scheme with infinite

number of feedback bits. This demonstrates that transmit antenna shuffling scheme is much more effective than the angle rotation scheme in term of feedback gain.

4.4 Conclusion

A comprehensive comparison of the performance of angle feedback scheme and transmit antenna shuffling scheme for DSTTD systems is provided in this chapter. We have shown that transmit antenna shuffling outperforms angle feedback under the same feedback bits. The reason is that transmit antenna shuffling effectively boosts the minimum post-processing SNR among all data streams by maximizing the quantity $\rho\mu - \eta$ defined in this chapter, whereas angle feedback only minimizes η; thus the potential achievable by balancing ρ and μ for maximizing the minimum post-processing SNR is not explored.

Chapter 5 – Pilot Power in Training-Based MMSE Channel Estimation for Closed-Loop MIMO Systems

5.1 Introduction

Coherent communications need channel information at the receiver. One way for the receiver to acquire channel information is through training. In the training process, the training symbols (also called pilot symbols) that are known to the receiver are transmitted for estimation of the unknown channel coefficients at the receiver.

Training design for various wireless systems has been studied extensively. In [78] training symbols constructed using space-time codes for the MIMO systems over the quasi-static, frequency-selective channels is proposed. In [100], the optimal pilot sequences and the optimal placement of pilot tones are derived in the sense of minimizing the MSE of the channel estimate in MIMO-OFDM systems. General classes of optimal training signals for channel estimation in MIMO-OFDM systems are developed in [101] using the MSE criterion. Optimal placement of training signals for frequency-selective channels is derived by maximizing a capacity lower bound and minimizing the outage probability, in [102] and [103], respectively. The training design regarding the number of training intervals, the training sequence, and the training power is optimized to maximize a capacity lower bound in MIMO frequency-flat, block-fading channels in [84], and in the single-input single-output (SISO) frequency-selective block-fading channels in [104]. A similar problem regarding the optimal number of training symbols, the training sequence, and the training symbol placement and the power allocation between data symbols and training symbols is solved for OFDM systems in frequency-selective, block-

fading channels in [105]. For the doubly-selective fading channels, the optimal training design is addressed in [106–108].

These works address training design for the open-loop wireless systems. It is well known that if channel knowledge is available at the transmitter, the transmitter could adapt to the channel condition to improve the system performance, e.g., capacity. Feedback schemes have been adopted in some wireless standards (e.g., [82]) to form the closed-loop systems to improve performance. But in the training-based systems with limited training power, channel estimation error is inevitable and it introduces channel uncertainty, which could significantly affect the achievable rate - the capacity lower bound [109, 110].

In this chapter, we focus on power allocation between training and data for the closed-loop MIMO systems over flat Rayleigh-fading channels. We follow the block fading model in [84] and analyze two power control schemes: spatial power control, and spatial and fading power control. In the spatial power control scheme, all fading blocks are assigned equal power; data power is optimized within each block, but no adaptation is made to compensate fading across block boundaries. In the spatial and fading power control scheme, power adapts to fading variations in addition to spatial power control; thus power assigned to fading blocks varies depending on the fading condition. Both schemes optimize the power allocation between training and data in a approach of maximizing the ergodic capacity lower bound. We first show that power allocation in the open-loop MIMO systems is still optimal in closed-loop MIMO systems with only spatial power control. Then we show that the power allocation scheme for the open-loop MIMO systems is asymptotically optimal for the closed-loop MIMO systems with both spatial and fading power control.

This chapter is structured as follows. Section 5.2 describes the system model and the

training design. In Section 5.3, the capacity lower bound for closed-loop training-based MIMO systems is derived. In Section 5.4, we consider power allocation between training and data for closed-loop systems. First, we focus on power allocation for the case of spatial power control. Then we extend it to the case of spatial and fading power control. Simulation results are provided in Section 5.5, followed by conclusion in Section 5.6.

5.2 System Model

Consider a MIMO system with M transmit antennas and N receive antennas over a frequency-flat block-fading channel, where channel coefficients keep constant over a block of T symbol intervals and change independently in the next block. We denote the channel coefficients by matrix $\mathbf{H}$, whose (i,j)th element represents the channel gain between the ith transmit antenna and the jth receive antenna. The elements of $\mathbf{H}$ are i.i.d., circularly symmetric, complex Gaussian random variables following $\mathcal{CN}(0,1)$, $\mathbf{H} \in \mathcal{C}^{M\times N}$. In the training-based channel estimation schemes, transmission in each fading block is partitioned into two phases [84]: training phase and data transmission phase, whose durations are denoted by T_p and T_d symbol intervals, respectively. The data transmission phase follows the training phase. Clearly, $T_p + T_d = T$.

We assume the same training signal is used in every fading block. In the training phase within a fading block, the received signals can be expressed as

$$\mathbf{Y}_p = \mathbf{X}_p\mathbf{H} + \mathbf{N}_p \tag{5.1}$$

where matrix $\mathbf{Y}_p \in \mathcal{C}^{T_p\times N}$ is the received signals over T_p training intervals, $\mathbf{X}_p \in \mathcal{C}^{T_p\times M}$ is the matrix of training signals known to the receiver, P_p is the power allocated for

training symbols which satisfies $P_p = \mathrm{Tr}(\mathbf{X}_p\mathbf{X}_p^H)$, and $\mathbf{N}_p \in \mathcal{C}^{T_p \times N}$ is the matrix of noise samples at the receiver, independent of $\boldsymbol{X}_p$. The elements of $\mathbf{N}_p$ are i.i.d., circularly symmetric, complex Gaussian random variables following $\mathcal{CN}(0,1)$.

In the data transmission phase, the received signals can be expressed as

$$\mathbf{Y}_d = \mathbf{X}_d\mathbf{H} + \mathbf{N}_d \tag{5.2}$$

where $\mathbf{Y}_d \in \mathcal{C}^{T_d \times N}$ is the matrix of the received signals over T_d data intervals, $\mathbf{X}_d \in \mathcal{C}^{T_d \times M}$ is the matrix of the transmit signals, $\mathrm{E}\left\{\mathrm{Tr}(\mathbf{X}_d\mathbf{X}_d^H)\right\}$ is the data power in the block, and $\mathbf{N}_d \in \mathcal{C}^{T_d \times N}$ is the matrix of noise samples at the receiver. The elements of $\mathbf{N}_d$ are i.i.d., circularly symmetric, complex Gaussian random variables following $\mathcal{CN}(0,1)$.

In the training phase the receiver performs linear MMSE channel estimation using received signal $\mathbf{Y}_p$ and the known pilot data $\mathbf{X}_p$ to estimate the channel. Denote $\hat{\mathbf{H}}$ as channel estimate. Because channel coefficients independently change from block to block in the considered block-fading model, the receiver must estimate channel at the beginning of each fading block. In closed-loop system model, we assume that there is an error-free feedback channel from the receiver to the transmitter with unlimited bandwidth (thus, there is no delay) so that the receiver can send channel estimate $\hat{\mathbf{H}}$ to the transmitter right after the training phase. We adopt the training signal designed in [84] because before training the transmitter does not have channel knowledge of current block in closed-loop systems[1].

[1]Although the channel information fed back in the previous blocks can not be used for data transmission in the current block, it can be exploited by the transmitter for channel statistics, e.g., covariance of channel coefficients, to facilitate the training signal design. It is shown in [73] that the optimal training for spatially correlated MIMO channels is to pour the training signal power along the eigen directions of the transmit covariance matrix using a waterfilling solution. In our i.i.d. channel model, the transmit covariance of channel coefficients is an identity matrix. Following the training signal design rule in [73], it can be shown that the training signal designed in [84] is optimal in terms of waterfilling on eigenmodes of the transmit covariance matrix.

It is shown in [84] that the optimal training signal $\mathbf{X}_p$ is in the form

$$\mathbf{X}_p^H \mathbf{X}_p = \frac{P_p}{M} \boldsymbol{I}_M. \tag{5.3}$$

Because $\boldsymbol{I}_M$ is of rank M, $T_p \geq M$. In other words, a meaningful estimation of MN channel coefficients needs at least MN observations. Since the N receivers provide N independent observations per training interval, at least M training intervals are needed. Let $\tilde{\mathbf{H}}$ represents the MMSE channel estimation error. $\mathbf{H}$ can be written as

$$\mathbf{H} = \hat{\mathbf{H}} + \tilde{\mathbf{H}}. \tag{5.4}$$

$\hat{\mathbf{H}}$ and $\tilde{\mathbf{H}}$ are independent under MMSE estimation. The elements of $\hat{\mathbf{H}}$ and $\tilde{\mathbf{H}}$ are i.i.d., following $\mathcal{CN}(0, \sigma_{\hat{H}}^2)$ and $\mathcal{CN}(0, \sigma_{\tilde{H}}^2)$, respectively, and their variances satisfy

$$\sigma_H^2 = \sigma_{\hat{H}}^2 + \sigma_{\tilde{H}}^2 = 1. \tag{5.5}$$

With the training signal given in (5.3), the variance of the MMSE channel estimation error is given by [84]

$$\sigma_{\tilde{H}}^2 = \frac{1}{1 + P_p/M}. \tag{5.6}$$

Eq. (5.6) shows that given a total training power P_p, MMSE channel estimation error is independent of training interval T_P.

After the transmitter receives $\hat{\mathbf{H}}$ through the feedback, it starts data transmission conditioned on $\hat{\mathbf{H}}$. We explicitly express the data power assigned to the fading block with channel estimate $\hat{\mathbf{H}}$ as $P_d(\hat{\mathbf{H}})$, since the transmitter can adjust its power with the knowledge of $\hat{\mathbf{H}}$ using fading control. The average data power per fading block satisfies

$\mathrm{E}[P_d(\hat{\mathbf{H}})] = \bar{P}_d$. With training power P_p, the average power for each fading block is $\bar{P} = P_p + \bar{P}_d$. In a closed-loop system with limited $\bar{P}$, the length of training intervals T_p and power allocation between P_p and $\bar{P}_d$ affect the overall system capacity. If too much time is dedicated to training, there will not be enough time for data transmission; if too much power is dedicated to training, there will not be enough power for data transmission. Thus, we optimize T_p and $\bar{P}_d$ (or P_p) by maximizing a capacity lower bound given T and $\bar{P}$.

5.3 Capacity Lower Bound for the Closed-loop Training-based MIMO Systems

Each row of $\mathbf{X}_d$ represents the data transmitted at one data transmission interval. The row vectors of $\mathbf{X}_d$ are assumed to be mutually independent. In one data transmission interval, the received signal $\boldsymbol{y}$ can be written as

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{x}\hat{\mathbf{H}} + \boldsymbol{x}\tilde{\mathbf{H}} + \boldsymbol{n} \\ &= \boldsymbol{x}\hat{\mathbf{H}} + \boldsymbol{v}. \end{aligned} \tag{5.7}$$

Note that only $\hat{\mathbf{H}}$ is known at both the receiver and the transmitter in the closed-loop system model. The product term $\boldsymbol{x}\tilde{\mathbf{H}}$ is an additional noise term at the receiver since $\tilde{\mathbf{H}}$ is unknown to the receiver. $\boldsymbol{v} = \boldsymbol{x}\tilde{\mathbf{H}} + \boldsymbol{n}$ denotes the combined noise term, which is no longer Gaussian distributed and can be shown to be uncorrelated with the term $\boldsymbol{x}\hat{\mathbf{H}}$ [84].

The exact channel capacity in (5.7) under channel uncertainty $\boldsymbol{v}$ is unknown in the literature. Since $\boldsymbol{v}$ contains the input signal $\boldsymbol{x}$, the capacity-achieving distribution of input signal $\boldsymbol{x}$ is unknown. The widely used technique is to derive the capacity lower bound, a guaranteed achievable rate by treating noise $\boldsymbol{v}$ as a Gaussian noise with same

power [84, 102, 104]. With the knowledge of channel estimate $\hat{\mathbf{H}}$ at the transmitter and the assumption that the distribution of $\boldsymbol{x}$ conditioned on $\hat{\mathbf{H}}$ is Gaussian, given channel estimation error $\sigma^2_{\tilde{H}}$ and transmit power p for the considered transmission interval, a lower bound of the instantaneous mutual information between $\boldsymbol{x}$ and $\boldsymbol{y}$ is given by [111]

$$I_{\text{lower}}(\boldsymbol{x};\boldsymbol{y}|\hat{\mathbf{H}},\sigma^2_{\tilde{H}},p) = \log_2 \det\left(\boldsymbol{I}_N + \frac{1}{1+\sigma^2_{\tilde{H}}p}\hat{\mathbf{H}}^H \boldsymbol{R}_{x|\hat{\mathbf{H}}}\hat{\mathbf{H}}\right) \tag{5.8}$$

where the transmit signal covariance matrix is expressed as $\boldsymbol{R}_{x|\hat{\mathbf{H}}} = \mathrm{E}[(\boldsymbol{x}|\hat{\mathbf{H}})(\boldsymbol{x}|\hat{\mathbf{H}})^H]$ and the transmit power is $p = \mathrm{Tr}(\boldsymbol{R}_{x|\hat{\mathbf{H}}})$.

Since all elements of $\hat{\mathbf{H}}$ follow $\mathcal{CN}(0,\sigma^2_{\hat{H}})$, we normalize $\hat{\mathbf{H}}$ and let $\bar{\mathbf{H}} = \hat{\mathbf{H}}/\sigma_{\hat{H}}$. The elements of $\bar{\mathbf{H}}$ are i.i.d. and follow $\mathcal{CN}(0,1)$. In the rest of this chapter, we use $\bar{\mathbf{H}}$, instead of $\hat{\mathbf{H}}$, to simplify the analysis. For example, the data power is expressed as $P_d(\hat{\mathbf{H}}) = P_d(\bar{\mathbf{H}})$. We rewrite (5.8) in a more compact form as

$$I_{\text{lower}}(\boldsymbol{x};\boldsymbol{y}|\bar{\mathbf{H}},\sigma^2_{\tilde{H}},p) = \log_2 \det\big(\boldsymbol{I}_N + \bar{\mathbf{H}}^H \boldsymbol{R}_{x|\bar{\mathbf{H}}}\bar{\mathbf{H}}\big) \tag{5.9}$$

where $\boldsymbol{R}_{x|\bar{\mathbf{H}}} = \frac{\sigma^2_{\hat{H}}}{1+\sigma^2_{\tilde{H}}p}\boldsymbol{R}_{x|\hat{\mathbf{H}}}$.

5.3.1 Instantaneous Capacity Lower Bound Given $\bar{\mathbf{H}}$, $\sigma^2_{\tilde{H}}$ and p

We define the effective transmit power as

$$p_{\text{eff}}(\sigma^2_{\tilde{H}},p) \triangleq \mathrm{Tr}(\boldsymbol{R}_{x|\bar{\mathbf{H}}}) = \frac{\sigma^2_{\hat{H}}p}{1+\sigma^2_{\tilde{H}}p}. \tag{5.10}$$

Maximizing (5.9) by finding the input signal covariance $\boldsymbol{R}_{x|\bar{\mathbf{H}}}$ to obtain an instantaneous

capacity lower bound, which is expressed by

$$C_{\text{inst}}(\bar{\mathbf{H}}, \sigma^2_{\tilde{H}}, p) = \max_{\boldsymbol{R}_{x|\bar{\mathbf{H}}}} \log_2 \det\left(\boldsymbol{I}_N + \bar{\mathbf{H}}^H \boldsymbol{R}_{x|\bar{\mathbf{H}}} \bar{\mathbf{H}}\right), \tag{5.11}$$

subject to $\text{Tr}\left(\boldsymbol{R}_{x|\bar{\mathbf{H}}}\right) \leq p_{\text{eff}}(\sigma^2_{\tilde{H}}, p)$, is straightforward (for example, see standard water-filling technique in [2] and [112, Th. 7]). Singular value decomposition of the channel matrix $\bar{\mathbf{H}}$ in (5.9) is expressed as

$$\bar{\mathbf{H}} = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{V}^H$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary matrices with dimension $M \times M$ and $N \times N$, respectively, and $\boldsymbol{\Lambda}$ is an $M \times N$ diagonal matrix with K_1 ordered singular values of $\bar{\mathbf{H}}$, denoted as $\lambda_1^{\frac{1}{2}}(\bar{\mathbf{H}}) \geq \lambda_2^{\frac{1}{2}}(\bar{\mathbf{H}}) \geq \cdots \geq \lambda_{K_1}^{\frac{1}{2}}(\bar{\mathbf{H}})$, where $K_1 = \min(M, N)$. If $M > K_1$, we add $\lambda_{K_1+1}(\bar{\mathbf{H}}) = \cdots = \lambda_M(\bar{\mathbf{H}}) = 0$ whenever needed. The optimal covariance is in the form of

$$\boldsymbol{R}_{x|\bar{\mathbf{H}}} = \boldsymbol{U} \text{diag}\big(p_{\text{eff},1}(\sigma^2_{\tilde{H}}, p), \cdots, p_{\text{eff},M}(\sigma^2_{\tilde{H}}, p)\big) \boldsymbol{U}^H$$

with

$$\sum_{k=1}^{M} p_{\text{eff},k}(\sigma^2_{\tilde{H}}, p) = p_{\text{eff}}(\sigma^2_{\tilde{H}}, p) \;\; , \;\; p_{\text{eff},k}(\sigma^2_{\tilde{H}}, p) \geq 0 \tag{5.12}$$

where $p_{\text{eff},k}(\sigma^2_{\tilde{H}}, p)$ denotes the effective power on the eigen-direction of the kth ordered eigenvalue $\lambda_k(\bar{\mathbf{H}})$ (of $\bar{\mathbf{H}}\bar{\mathbf{H}}^H$) and is determined by

$$p_{\text{eff},k}(\sigma^2_{\tilde{H}}, p) = \begin{cases} 0 & \text{if } \lambda_k(\bar{\mathbf{H}}) = 0 \\ \left(\nu(\bar{\mathbf{H}}, \sigma^2_{\tilde{H}}, p) - \frac{1}{\lambda_k(\bar{\mathbf{H}})}\right)^+ & \text{otherwise} \end{cases} \tag{5.13}$$

where the operator $(x)^+$ returns $\max\{x, 0\}$, $\nu(\bar{\mathbf{H}}, \sigma^2_{\tilde{H}}, p)$ is the water-filling level and is determined by (5.12) as

$$\nu(\bar{\mathbf{H}}, \sigma^2_{\tilde{H}}, p) = \frac{1}{m}\left(\sum_{k=1}^{m} \frac{1}{\lambda_k(\bar{\mathbf{H}})} + p_{\text{eff}}(\sigma^2_{\tilde{H}}, p) \right) \tag{5.14}$$

with m being the number of non-zero $p_{\text{eff,k}}(\sigma^2_{\tilde{H}}, p)'s$. The solution of $C_{\text{inst}}(\bar{\mathbf{H}}, \sigma^2_{\tilde{H}}, p)$ in terms of eigenvalue $\lambda_i(\bar{\mathbf{H}})$'s is

$$C_{\text{inst}}(\bar{\mathbf{H}}, \sigma^2_{\tilde{H}}, p) = \sum_{k=1}^{m} \left\{ \log_2\left[\lambda_k(\bar{\mathbf{H}}) \nu(\bar{\mathbf{H}}, \sigma^2_{\tilde{H}}, p) \right] \right\}^+. \tag{5.15}$$

We assume the transmitter can vary its transmit power over any intervals in both the training phase and the data transmission phase, as required by fading power control. On the other hand, if the transmitter can not vary the transmit power in any transmission intervals, then the optimal length of the training intervals must be found numerically [84, 104].

Denote $P_{d,i}(\bar{\mathbf{H}})$ as the data power assigned to the ith data transmission interval with power constraint $\sum_{i=1}^{T_d} P_{d,i}(\bar{\mathbf{H}}) = P_d(\bar{\mathbf{H}})$. The following lemma gives the instantaneous capacity lower bound of fading block $\bar{\mathbf{H}}$.

Lemma 5 *When transmit power at any transmission intervals can be adjusted, given training power P_p and data power $P_d(\bar{\mathbf{H}})$ for the block with channel estimate $\bar{\mathbf{H}}$, the maximum sum rate of the instantaneous capacity lower bound over the block is achieved by $T_p = M$, $T_d = T - M$, and $P_{d,i}(\bar{\mathbf{H}}) = P_d(\bar{\mathbf{H}})/(T - M)$.*

Proof. The variance of channel estimation error, $\sigma^2_{\tilde{H}}$, is determined by P_p through (5.6). When the power at any transmission intervals can be adjusted, given training power P_p, M is the smallest meaningful training length. Letting $T_p = M$ will not degrade the

estimation accuracy. Thus data transmission can take any left transmission intervals up to $T-M$. Let us assume $T_d = T-M$. The sum rate of the instantaneous capacity lower bound in the block $\bar{\mathbf{H}}$ can be expressed as

$$C_{\text{inst-block}}(\bar{\mathbf{H}}, P_p, P_d(\bar{\mathbf{H}})) = \max_{P_{d,i}(\bar{\mathbf{H}})'s} \sum_{i=1}^{T-M} C_{\text{inst}}(\bar{\mathbf{H}}, \sigma^2_{\tilde{H}}, P_{d,i}(\bar{\mathbf{H}}))$$
$$\text{subject to:} \quad \sum_{i=1}^{T-M} P_{d,i}(\bar{\mathbf{H}}) = P_d(\bar{\mathbf{H}})$$

This is a standard optimization problem and can be solved using Lagrange multipliers [113]. The solution is expressed as

$$P_{d,i}(\bar{\mathbf{H}}) = \frac{P_d(\bar{\mathbf{H}})}{T-M} \triangleq p_d(\bar{\mathbf{H}}), \quad i = 1, ..., T-M \tag{5.16}$$

which implies that all available data transmission intervals of $T-M$ must be used with equal power $p_d(\bar{\mathbf{H}})$ and $T_d = T-M$, furthermore

$$C_{\text{inst-block}}(\bar{\mathbf{H}}, P_p, P_d(\bar{\mathbf{H}})) = (T-M) C_{\text{inst}}\big(\bar{\mathbf{H}}, \sigma^2_{\tilde{H}}, p_d(\bar{\mathbf{H}})\big) \tag{5.17}$$

with $\sigma^2_{\tilde{H}} = \frac{1}{1+P_p/M}$. ■

5.3.2 Ergodic Capacity Lower Bound Given P_p and $\bar{P}_d$

For systems operating in fading environments, we aim at maximizing ergodic capacity lower bound. Given training power P_p and training interval $T_p = M$, under an average data transmit power $\mathrm{E}[P_d(\bar{\mathbf{H}})] = \bar{P}_d$, the ergodic capacity lower bound per transmission

interval (channel use) with consideration of training effect is

$$\begin{aligned} C_{lower}(P_p, \bar{P}_d) &= \max_{\{P_d(\bar{\mathbf{H}})\}} \mathrm{E}\left[\frac{1}{T} C_{\text{inst-block}}(\bar{\mathbf{H}}, P_p, P_d(\bar{\mathbf{H}}))\right] \\ &= \max_{\{p_d(\bar{\mathbf{H}})\}} \mathrm{E}\left[\frac{T-M}{T} C_{\text{inst}}\big(\bar{\mathbf{H}}, \sigma^2_{\tilde{H}}, p_d(\bar{\mathbf{H}})\big)\right], \end{aligned} \tag{5.18}$$

subject to $\mathrm{E}[p_d(\bar{\mathbf{H}})] = \bar{P}_d/(T-M)$, which is equivalent to $\mathrm{E}[P_d(\bar{\mathbf{H}})] = \bar{P}_d$. $\big\{P_d(\bar{\mathbf{H}})\big\}$, or equivalently $\big\{p_d(\bar{\mathbf{H}})\big\}$, represents all possible power loading policies. Once again $\sigma^2_{\tilde{H}} = \frac{1}{1+P_p/M}$. The second equality follows from (5.16) and (5.17). The solution of $p_d(\bar{\mathbf{H}})$ can be shown, by using Lagrange multipliers, to meet

$$\frac{\partial C_{\text{inst}}\big(\bar{\mathbf{H}}, \sigma^2_{\tilde{H}}, p_d(\bar{\mathbf{H}})\big)}{\partial p_d(\bar{\mathbf{H}})} = \gamma \tag{5.19}$$

where γ is a constant, representing the global marginal capacity gain and is determined by power constraint $\mathrm{E}[P_d(\bar{\mathbf{H}})] = \bar{P}_d$. It can be derived using (5.15) combined with (5.10) and (5.14) that the instantaneous marginal capacity gain over the assigned power is

$$\frac{\partial C_{\text{inst}}(\bar{\mathbf{H}}, \sigma^2_{\tilde{H}}, p)}{\partial p} = \frac{1}{\ln 2} \frac{1}{\nu(\bar{\mathbf{H}}, \sigma^2_{\tilde{H}}, p)} \frac{\sigma^2_{\hat{H}}}{\Big(1 + \sigma^2_{\tilde{H}} p\Big)^2} > 0. \tag{5.20}$$

Since $\nu(\bar{\mathbf{H}}, \sigma^2_{\tilde{H}}, p)$ is an increasing function of p, which can be derived by using (5.10) and (5.14), (5.20) shows that the instantaneous marginal capacity gain decreases when the assigned power increases. Condition (5.19) shows that the ergodic capacity lower bound is maximized if the instantaneous marginal capacity gain in any data transmission intervals (if assigned power) reduces to γ. It also implies that there could be no power for some data transmission intervals because at these intervals their instantaneous marginal capacity gains are always less than γ. Note that in the block-fading channel, if no power

is allocated for one data transmission interval within a block, by Lemma 5 the whole block does not have data power either[2]. Clearly, there is no close-form expression for ergodic capacity lower bound although bound achieving condition (5.19) is simple.

5.4 Power Allocation Between Training Phase and Data Transmission Phase

Given average power per transmission block $\bar{P} = P_p + \bar{P}_d$, where P_p is the training power reserved for each fading block and $\mathrm{E}[P_d(\bar{\mathbf{H}})] = \bar{P}_d$, the percentage of $\bar{P}$ for data transmission that maximizes the ergodic capacity lower bound is unknown for closed-loop MIMO systems. In this section, we solve this problem in two scenarios. In the first scenario, all fading blocks are assigned the same power $\bar{P}$, there is no power adaptation to the fading, i.e., $P_d(\bar{\mathbf{H}}) = \bar{P}_d$. In the second scenario, data transmit power $P_d(\bar{\mathbf{H}})$ is adjusted according to channel $\bar{\mathbf{H}}$, subject to $\mathrm{E}[P_d(\bar{\mathbf{H}})] = \bar{P}_d$.

5.4.1 Optimal Power Allocation When $P_d(\bar{\mathbf{H}}) = \bar{P}_d$: Spatial Power Control

When $P_d(\bar{\mathbf{H}}) = \bar{P}_d$, that is, there is no fading power control and only spatial power control is performed, (5.18) becomes

$$\begin{aligned} C_{lower}(P_p, \bar{P}_d) &= \mathrm{E}\left[\frac{1}{T} C_{\text{inst-block}}(\bar{\mathbf{H}}, P_p, \bar{P}_d)\right] \\ &= \mathrm{E}\left[\frac{T-M}{T} C_{\text{inst}}\big(\bar{\mathbf{H}}, \sigma^2_{\tilde{H}}, \frac{\bar{P}_d}{T-M}\big)\right]. \end{aligned} \tag{5.21}$$

Let $\bar{P} = P_p + \bar{P}_d$ denote the total power for one fading block, and let

$$\bar{P}_d = \alpha \bar{P}, \qquad P_p = \bar{P} - \bar{P}_d = (1-\alpha)\bar{P}$$

[2]Training is still performed at this block.

where $0 < \alpha < 1$. parameter α indicates how much power is used for data transmission. We need to optimize α to maximize $C_{lower}(P_p, \bar{P}_d)$ in (5.21), or equivalently, to maximize $\mathrm{E}\big[C_{\mathrm{inst}}(\bar{\mathbf{H}}, \sigma^2_{\tilde{H}}, \frac{\bar{P}_d}{T-M})\big]$. From (5.14) and (5.15), we observe that α affects C_{inst} only through p_{eff} and we can show that

$$\frac{\partial C_{\mathrm{inst}}}{\partial p_{\mathrm{eff}}} = \frac{1}{\ln 2}\frac{1}{\nu} > 0 \tag{5.22}$$

since $\nu > 0$ in (5.14). C_{inst} is maximized if p_{eff} is maximized.

Substituting $\sigma^2_{\tilde{H}} = \frac{1}{1+(1-\alpha)P/M}$ and $p = \frac{\alpha\bar{P}}{T-M}$ into p_{eff} in (5.10), and letting

$$\frac{\partial p_{\mathrm{eff}}}{\partial \alpha} = 0$$

We obtain the solution of α which is given by

$$\alpha_s = \begin{cases} \frac{1}{2}, & \text{if } T = 2M; \\ \dfrac{b - \sqrt{b^2 - ab}}{a}, & \text{if } T \neq 2M \end{cases} \tag{5.23}$$

where $a = \bar{P}(T - 2M)$ and $b = (T - M)(\bar{P} + M)$. We verify that the solution α_{s} here is identical to the one given in [84] for open-loop systems with $T_p = M$.

Note that α_{s} is derived by maximizing $C_{\mathrm{inst}}(\bar{\mathbf{H}}, \sigma^2_{\tilde{H}}, \frac{\bar{P}_d}{T-M})$ in a specific fading block with channel estimate $\bar{\mathbf{H}}$, but the solutions do not depend on $\bar{\mathbf{H}}$. This implies that the solution is not only local optimum for specific fading block $\bar{\mathbf{H}}$, but also global optimum for any fading blocks. Therefore, the ergodic capacity lower bound in (5.21) is maximized under (5.23).

5.4.2 Power Allocation When $P_d(\bar{\mathbf{H}})$ Adapts to Fading: Spatial and Fading Power Control

We have considered the first scenario where a fixed power $\bar{P}_d$ is assigned to data transmission in each fading block and only spatial power control is performed within each fading block. We derive the optimal power allocation and training intervals. Since no restrictions are imposed on the power of each transmission interval, naturally the transmitter should also be able to adjust its power in different fading blocks. This allows the transmitter to perform fading power control to further improve the channel capacity lower bound.

Note that in our model, $\bar{P}$ is the average power per fading block used for both training and data transmissions. At the training phase of each fading block, we allocate a fixed percentage of $\bar{P}$ to P_p, $P_p < \bar{P}$, for training and adjust the data power $P_d(\bar{\mathbf{H}})$ according to the channel condition $\bar{\mathbf{H}}$. Intuitively, when the channel is in deep fading, less power should be allocated; when the channel condition is good, more power should be assigned. $P_d(\bar{\mathbf{H}})$ must meet the following power constraint

$$\mathrm{E}[P_d(\bar{\mathbf{H}})] = \bar{P}_d = \bar{P} - P_p. \tag{5.24}$$

Once a fading block with channel estimate $\bar{\mathbf{H}}$ is assigned data power $P_d(\bar{\mathbf{H}})$, by Lemma 5 $P_d(\bar{\mathbf{H}})$ should be evenly distributed among all $T - M$ data intervals. Within each interval, waterfilling is performed along the eigen directions. With spatial and fading power control, $P_d(\bar{\mathbf{H}})$ needs to meet (5.19). The optimal power allocation parameter α, denoted by $\alpha_{\mathrm{sf}} = \bar{P}_d/\bar{P}$, needs to be solved in the sense of maximizing the capacity lower bound in (5.18). The subscript sf of α_{sf} denotes that α is optimized under spatial and fading power control.

Unfortunately, there are no close-form expressions for (5.18). Although (5.19) provides a necessary and sufficient condition to reach the capacity lower bound given P_p and $\bar{P}_d$, it is rather difficult to find the solution $P_d(\bar{\mathbf{H}})$ without resorting to a numerical approach. If the capacity lower bound is further maximized by optimally allocating power between the training phase and the data transmission phase, it would be virtually impossible to find an analytical solution. For this reason, it is of interest to find the asymptotic solution to gain some insights about the behavior of power allocation.

5.4.2.1 Asymptotic Solution

When $\min(M, N)$ increases while the ratio of M and N is fixed, the distribution of the normalized eigenvalues of a large random central Wishart matrix tends to converge to a non-random distribution. For any square matrix $\boldsymbol{A}$ with only real eigenvalues, let $F^{\mathbf{A}}$ denote the empirical distribution function (edf) of the eigenvalues of $\boldsymbol{A}$, that is, $F^{\mathbf{A}}(x)$ is the proportion of eigenvalues of $\boldsymbol{A}$ that is less than or equal to x. An important result from [114]:

Lemma 6 *Assume on a common probability space: Let $\mathbf{X}_M$ be an $M \times N$ matrix with i.i.d. complex entries and unit variance. $N = N(M)$ with $M/N \to d$, d is a positive constant, as $M \to \infty$. $\boldsymbol{T}_M$, $M \times M$ random Hermitian nonnegative definite, with $F^{\mathbf{T}_M}$ converging almost surely in distribution to a probability distribution function (pdf) $\mathcal{T}$ on $[0, \infty)$ as $M \to \infty$. $\mathbf{X}_M$ and $\boldsymbol{T}_M$ are independent. Let $\boldsymbol{T}_M^{1/2}$ be the square root of a nonnegative Hermitian matrix $\boldsymbol{T}_M$, and let*

$$\boldsymbol{B}_M = (1/N)\boldsymbol{T}_M^{1/2}\mathbf{X}_M\mathbf{X}_M^H\boldsymbol{T}_M^{1/2}$$

then, $F^{\mathbf{B}_M}$ converges in distribution almost surely to a non-random pdf f as $M \to \infty$.

Based on Lemma 6, we have the following theorem.

Theorem 7 *When* $\min(M,N) \to \infty$*, and the ratio of* M *and* N *is a fixed constant,* $T > M$ *and* $\bar{P}/T \nrightarrow 0$*. The optimal power allocation parameter* α_{sf} *under spatial and fading power control converges to* α_s *in (5.23) almost surely and* $P_d(\bar{\mathbf{H}}) = \bar{P}_d = \alpha_{sf}\bar{P} = \alpha_s\bar{P}$ *almost surely holds.*

Proof. See Appendix B. ■

Theorem 7 provides an asymptotic solution which shows that fading power control is unnecessary when $\min(M,N) \to \infty$ and M/N is a constant. All fading blocks should be assigned the same data power $\bar{P}_d$ and the problem of the optimal power allocation under spatial and fading power control policy reduces to the problem of spatial power control only, which is addressed in Section 5.4.1. Although this is an asymptotic solution, we find in simulations that α_{sf} for systems with a finite number of antennas at both sides is very close to its counterpart α_{s}. Furthermore, extensive simulations show that for finite MIMO antenna array, α_{sf} is upper bounded by α_{s}. We conjecture that this bound is true, although we have not yet been able to prove it.

Conjecture 8 *For any size of MIMO antenna array,* $\alpha_{sf} \leq \alpha_s$*, where equality holds when* $\min(M,N) \to \infty$ *and* M/N *is a constant.*

5.5 Simulation and Discussion

In this section, we simulate the capacity lower bound of several MIMO systems with different number of antennas. SNR is defined as $\rho = \bar{P}/T$.

We clarify the quantities to be used in the simulations: α_{sf} denotes the optimal α when spatial and fading power control is applied. It is obtained by numerical search

since there is no analytical solution when the number of antennas is finite; α_{s} denotes the optimal α when spatial power control is applied; C_{open} denotes the open-loop capacity lower bound [84] using α_{s} from (5.23); $C_{\mathrm{perfect,sf}}$ represents the closed-loop capacity with perfect channel knowledge at both the transmitter and the receiver. The perfect channel knowledge is provided by a genie so that no training is needed. $C_{\mathrm{perfect,sf}}$ is obtained by using both spatial and fading power control; C_{s} is the closed-loop capacity lower bound with spatial power control only in Section 5.4.1 using α_{s}; C_{sf} is the closed-loop capacity lower bound with spatial and fading power control, and is obtained by using α_{sf}, unless explicitly specified otherwise.

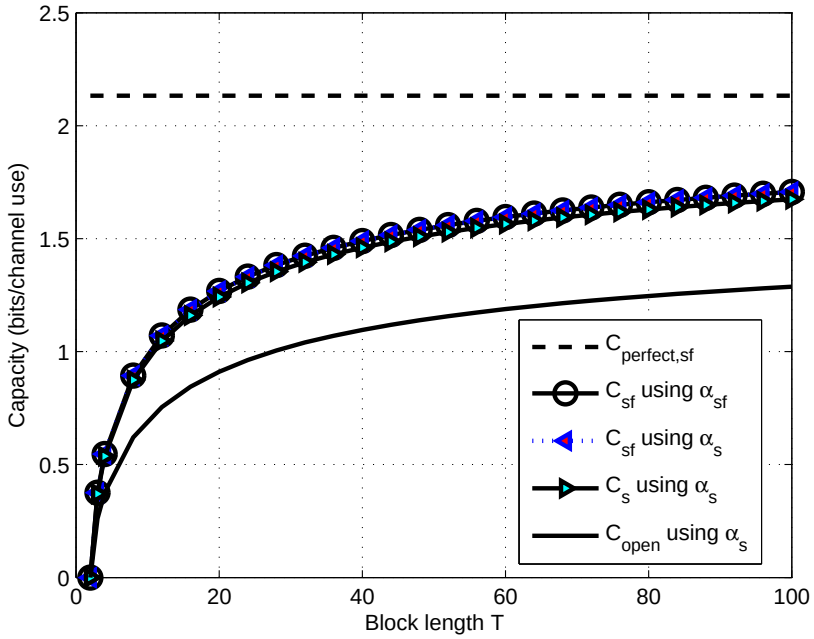

Figure 5.1: Capacity comparison of various schemes ($M = N = 2$, $\rho = 0$ dB).

We first simulate the performance of a 2×2 MIMO system. Fig. 5.1 shows that when only spatial power control is performed, the capacity gain of C_{s} over C_{open} at

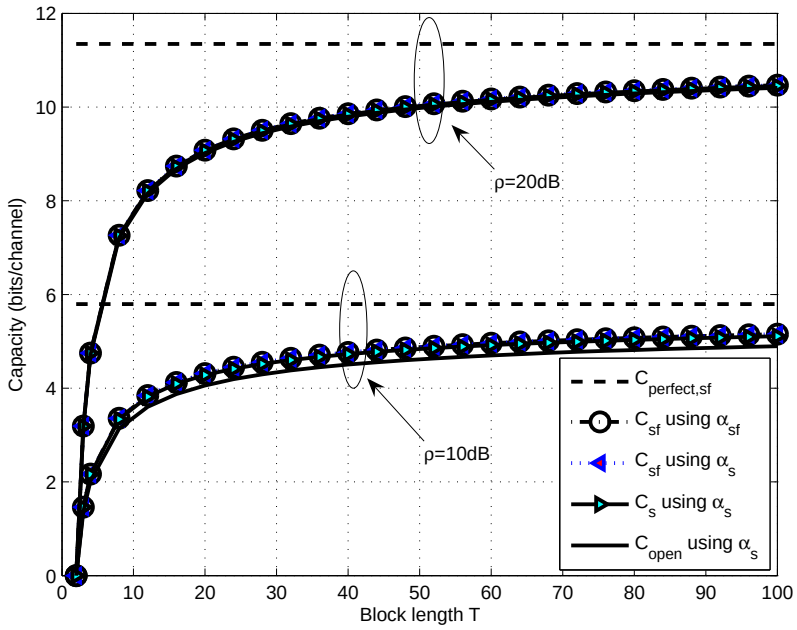

Figure 5.2: Capacity with $M = N = 2$, $\rho = 10$ dB and $\rho = 20$ dB.

$\rho = 0$ dB is greater than the gains at $\rho = 10$ dB and $\rho = 20$ dB as shown in Fig. 5.2. The gain almost vanishes when at $\rho = 20$ dB; this shows that channel knowledge at the transmitter becomes unnecessary in the high-SNR regime when only spatial power control is performed.

We also plot the capacity of C_{sf} using α_{sf} obtained by numerical search. The gain of C_{sf} using α_{sf} over C_{sf} using α_{s} is hardly noticeable. This shows that even if the number of antennas is not large, α_{s} could provide a suboptimal solution in spatial and fading power control policy, instead of α_{sf}, which must be determined via numerical search. On the other hand, both Fig. 5.1 and Fig. 5.2 show that the gain of C_{sf} using α_{s} or α_{sf} over C_{s} is very limited. This shows that spatial and fading power control does not provide significant gain over spatial power control only and suggests that capacity

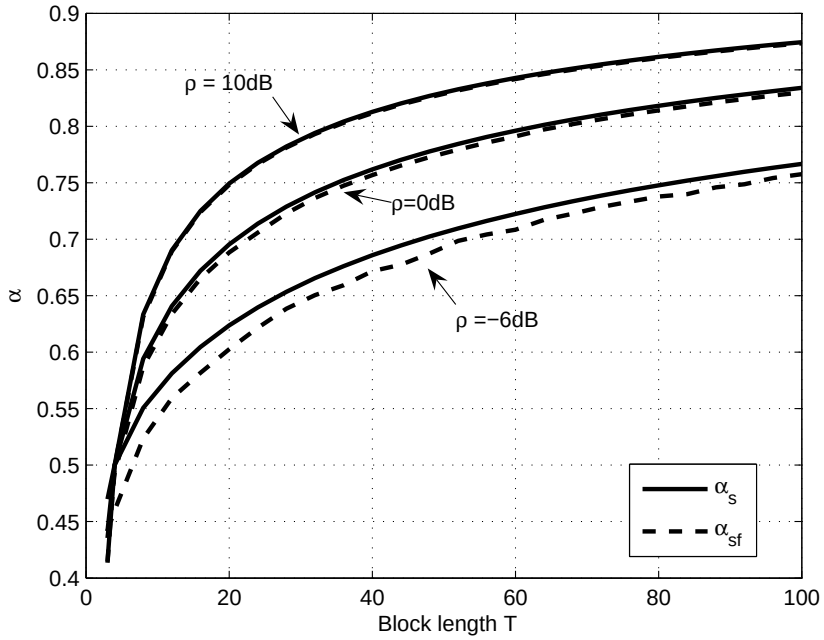

Figure 5.3: $\alpha_{\rm sf}$ vs. $\alpha_{\rm s}$, where $\alpha_{\rm sf}$ is obtained through numerical search ($M = N = 2$).

$C_{\rm s}$ can well approximate capacity $C_{\rm sf}$, and the scheme that employs only spatial power control is sufficient in the closed-loop MIMO systems to realize the feedback gain. This finding is of particular interest since in frequency division duplex (FDD) systems, the transmitter acquires channel information by feedback from the receiver. Since feedback channel is often bandwidth limited, the quantized covariance feedback scheme without fading power control [116][3] allows us to implement a simple codebook to realize most of capacity gain.

Fig. 5.3 and Fig. 5.4 show that the gap between $\alpha_{\rm s}$ in (5.23) and $\alpha_{\rm sf}$ decreases as SNR increases; the difference becomes negligible in high-SNR regime. We conjecture that $\alpha_{\rm sf}$ converges to $\alpha_{\rm s}$ when SNR approaches infinity, although an exact proof appears

[3]All quantized covariance matrices in the codebook from [116] have unit trace which means transmit power keeps constant for data transmission for any fading blocks.

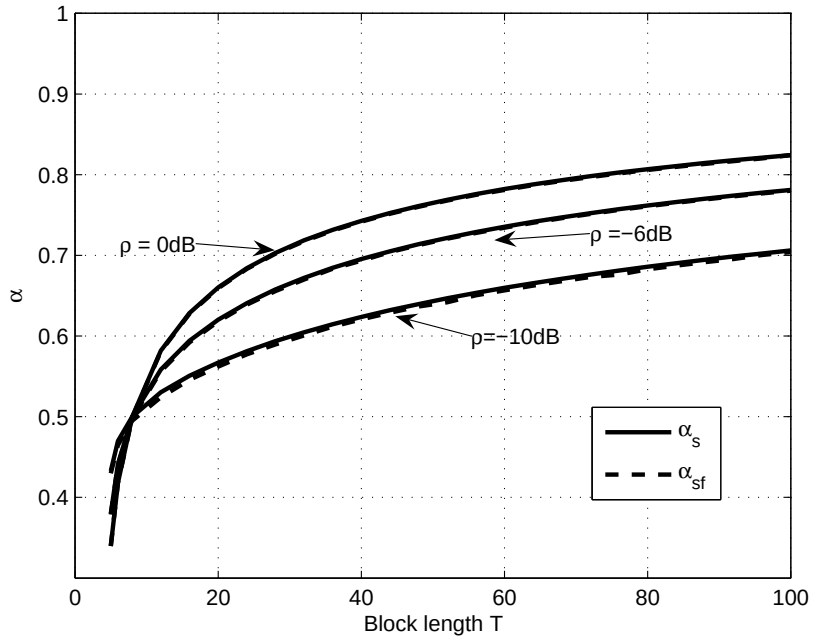

Figure 5.4: $\alpha_{\rm sf}$ vs. $\alpha_{\rm s}$, where $\alpha_{\rm sf}$ is obtained through numerical search ($M = N = 4$).

to be difficult.

Fig. 5.5 shows that the gap between $\alpha_{\rm s}$ and $\alpha_{\rm sf}$ reduces when $\min(M, N)$ increases. This confirms the conclusion we draw in Theorem 7 that $\alpha_{\rm s}$ is asymptotically optimal for the MIMO systems with spatial and fading power control when $\min(M, N) \to \infty$.

5.6 Conclusion

This chapter has explored power allocation between training phase and data transmission phase in closed-loop MIMO systems. Due to channel estimation errors, the exact channel capacity and capacity-achieving input distribution are unknown. We focus on input with a Gaussian distribution to derive the capacity lower bound and then max-

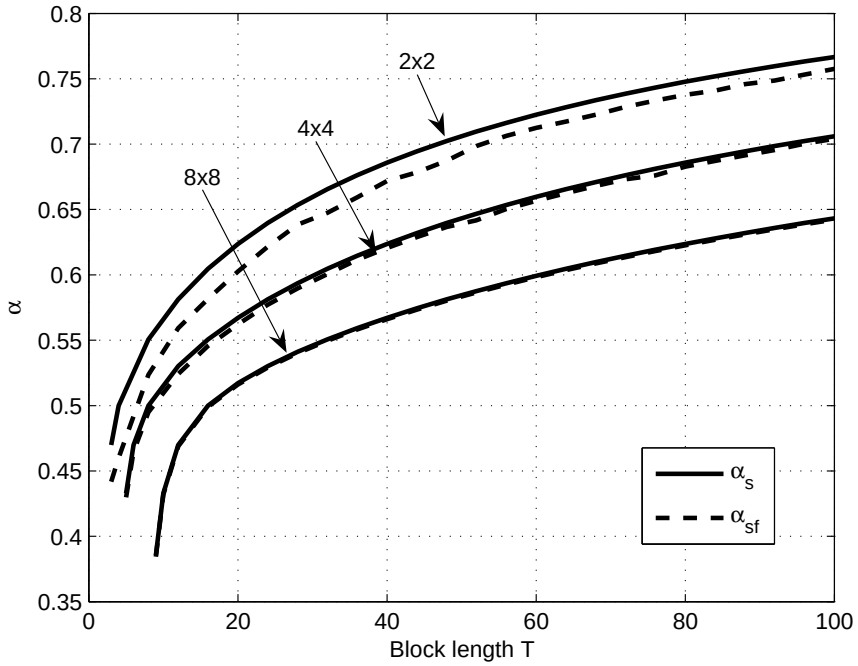

Figure 5.5: $\alpha_{\rm sf}$ vs. $\alpha_{\rm s}$, where $\alpha_{\rm sf}$ is obtained through numerical search for 2×2, 4×4, and 8×8 MIMO systems ($\rho = -6$ dB).

imize this lower bound by optimizing power allocation in training-based, closed-loop MIMO systems. When only spatial power control is used, the optimal power allocation $\alpha_{\rm s}$ is the same as that in the open-loop systems. When both spatial and fading power control is performed, we show, by using the theory of large random matrices, that optimal power allocation $\alpha_{\rm sf}$ converges to $\alpha_{\rm s}$, the case of spatial power control only when $\min(M, N) \to \infty$. This asymptotic result provides insights about the relationship of $\alpha_{\rm sf}$ and $\alpha_{\rm s}$. When the number of antennas is small, it is shown via simulations that the power allocation $\alpha_{\rm sf}$ is still very close to $\alpha_{\rm s}$ and the closed-loop capacity gap between $C_{\rm sf}$ using $\alpha_{\rm sf}$ and $C_{\rm sf}$ using suboptimal $\alpha_{\rm s}$ is very small. Therefore, suboptimal $\alpha_{\rm s}$, instead of $\alpha_{\rm sf}$, could be used for spatial and fading power control without causing a significant

capacity loss. We also find in the simulations that the capacity of the closed-loop MIMO systems under spatial power control using α_{s} is sufficient to approximate the capacity C_{sf} promised by spatial and fading power control using α_{sf}; thus complex fading power control may not be necessary in practice.

Chapter 6 – Diversity-Multiplexing Tradeoff of Multiple Beamforming

6.1 Introduction

The antenna elements of a MIMO system could be exploited for spatial multiplexing gain and/or diversity gain (to increase the link reliability). The fundamental diversity-multiplexing tradeoff (DMT) is addressed in [7] for i.i.d., Rayleigh-fading MIMO channels when CSI is available only at the receiver (CSIR). It is well known that CSI at the transmitter (CSIT) can further improve the system performance. For example, the single beamforming, which communicates through the largest channel eigenmode, results in an average error probability proportional to $1/\mathrm{SNR}^{MN}$ at high SNR [21] for any fixed data rate, where M is the number of transmit antennas and N is the number of receive antennas. For multiple beamforming, where two or more channel eigenmodes are used (say K, $1 < K \leq \min\{M, N\}$), the diversity order is derived in [21] and is shown to be $(M-K+1)(N-K+1)$ when independent K symbols are simultaneously transmitted through the K largest channel eigenmodes at a fixed rate. In [118], the DMT of MIMO systems with both CSIT and CSIR is investigated with independent data streams transmitted over the channel eigenmodes using water-filling power allocation with short term power constraint, but without coding over channel eigenmodes and time. The DMT of each of individual substreams is obtained and the result is used to derive an optimal rate allocation scheme to obtain the optimal DMT of spatial multiplexing.

In this chapter, unlike [21, 118] that assume perfect CSIT, we are interested in the optimal DMT of multiple beamforming systems where only the right singular vectors of

channel matrix corresponding to the first K largest singular values are available at the transmitter (i.e., partial CSIT). The resulting optimal DMT can provide a performance limit for limited feedback unitary precoding [32]. On the other hand, MIMO systems with perfect CSIT do not have a finite diversity order if CSIT are fully exploited [119].

6.2 System Model

Again, we consider a MIMO system with M transmit antennas and N receive antennas over a frequency-flat, block-fading channel. The $N \times M$ channel matrix $\boldsymbol{H}$ has i.i.d. entries, each of which is a circularly symmetric, complex Gaussian random variable with zero mean and unit variance, i.e., $[\boldsymbol{H}]_{ij} \sim \mathcal{CN}(0,1)$. $[\boldsymbol{H}]_{ij}$ (the element at the ith row and the jth column of $\boldsymbol{H}$) represents the path gain (channel coefficient) between the jth transmit antenna and the ith receive antenna. The channel matrix $\boldsymbol{H}$ remains constant within a block of T symbols, i.e., the block length is much smaller than the channel coherent time. The system model per transmission can be expressed as

$$\tilde{\boldsymbol{y}} = \boldsymbol{H}\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{w}} \tag{6.1}$$

where $\tilde{\boldsymbol{x}} \in \mathcal{C}^{M\times 1}$ and $\tilde{\boldsymbol{y}} \in \mathcal{C}^{N\times 1}$ are the transmitted and received signals, respectively. The $N \times 1$ vector $\tilde{\boldsymbol{w}}$ is the circularly symmetric, complex Gaussian noise and has i.i.d. entries, $\tilde{w}_i \sim \mathcal{CN}(0,1)$. The transmitted signal $\tilde{\boldsymbol{x}}$ is normalized so that the total transmit power per channel use satisfies

$$\frac{1}{T}\mathrm{E}[\left\|\tilde{\boldsymbol{X}}\right\|_F^2] \leq \mathrm{SNR}$$

where $\tilde{\boldsymbol{X}} = [\tilde{\boldsymbol{x}}_1, ..., \tilde{\boldsymbol{x}}_T]$ is a block of transmitted signal. SNR here refers to the average SNR at each receive antenna, and $\|\cdot\|_F^2$ is the Frobenius norm of a matrix.

We assume that the receiver knows the channel matrix $\boldsymbol{H}$ perfectly. Define

$$m \triangleq \max\{M, N\}, \qquad n \triangleq \min\{M, N\}.$$

The singular value decomposition (SVD) of $\boldsymbol{H}$ is written as

$$\boldsymbol{H} = \boldsymbol{U}\bar{\boldsymbol{\Lambda}}\boldsymbol{V}^H$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary matrices of size $N \times N$ and $M \times M$, and $\bar{\boldsymbol{\Lambda}}$ is an $N \times M$ diagonal matrix with n nonnegative singular values $\lambda_i^{\frac{1}{2}}$ of $\boldsymbol{H}$ on its main diagonal. In fact, the diagonal entries of $\bar{\boldsymbol{\Lambda}}$ are the nonnegative square roots of the eigenvalues of $\boldsymbol{H}\boldsymbol{H}^H$ if $N \leq M$ (or $\boldsymbol{H}^H\boldsymbol{H}$ if $N > M$); the columns of $\boldsymbol{U}$ are the eigenvectors of $\boldsymbol{H}\boldsymbol{H}^H$; and the columns of $\boldsymbol{V}$ are the eigenvectors of $\boldsymbol{H}^H\boldsymbol{H}$.

With SVD of $\boldsymbol{H}$, when $\boldsymbol{V}$ is known at the transmitter we can rewrite the channel model (6.1) as

$$\bar{\boldsymbol{y}} = \bar{\boldsymbol{\Lambda}}\bar{\boldsymbol{x}} + \bar{\boldsymbol{w}} \tag{6.2}$$

where $\bar{\boldsymbol{y}} = \boldsymbol{U}^H\tilde{\boldsymbol{y}}$, $\bar{\boldsymbol{x}} = \boldsymbol{V}^H\tilde{\boldsymbol{x}}$, and $\bar{\boldsymbol{w}} = \boldsymbol{U}^H\tilde{\boldsymbol{w}}$ contains i.i.d., circularly symmetric, complex Gaussian entries of zero mean and unit variance, i.e., $\bar{w}_i \sim \mathcal{CN}(0, 1)$, since unitary transformation does not change the distribution of $\tilde{\boldsymbol{w}}$. $\bar{\boldsymbol{x}}$ is the equivalent transmitted signal vector and can be completely constructed at the transmitter. Eq. (6.2) effectively decouples the channel (6.1) into n independent, parallel channels

$$\bar{y}_i = \lambda_i^{\frac{1}{2}}\bar{x}_i + \bar{w}_i, \qquad 1 \leq i \leq n \tag{6.3}$$

since the rest of components of $\bar{\boldsymbol{y}}$ (if any) are independent of the transmitted signal and the $\bar{x}_i$'s for $i > n$ (if any) don't have any contribution. We call these independent, parallel channels the channel eigenmodes (or eigenchannels), since the transmission and the reception are performed along the column vector directions of $\boldsymbol{V}$ and $\boldsymbol{U}$, respectively. Using the channel eigenmodes for data transmission (known as beamforming) requires linear precoding at the transmitter (see [22, and references therein]). Without of loss generality, we assume $\lambda_n \geq ... \geq \lambda_1 \geq 0$.

We consider multiple beamforming with size K $(1 < K \leq n)$ and assume that only K column vectors of $\boldsymbol{V}$ corresponding to the first K largest eigenvalues are known, and the instantaneous eigenvalues are not known at the transmitter. Therefore, the channel eigenmodes

$$\bar{y}_i = \lambda_i^{\frac{1}{2}} \bar{x}_i + \bar{w}_i, \qquad n - K + 1 \leq i \leq n, \tag{6.4}$$

are used. We use

$$\boldsymbol{y} = \boldsymbol{\Lambda x} + \boldsymbol{w} \tag{6.5}$$

to express (6.4) in a vector-matrix form, where $\boldsymbol{\Lambda}$ is a $K \times K$ diagonal matrix with $\lambda_{n-K+1}^{\frac{1}{2}}, ..., \lambda_n^{\frac{1}{2}}$ on its main diagonal. $K \times 1$ vectors $\boldsymbol{y}$, $\boldsymbol{x}$, and $\boldsymbol{w}$ are, respectively, the output, input, and noise of channel eigenmodes. $\mathrm{E}[\boldsymbol{w}\boldsymbol{w}^H] = \boldsymbol{I}$. In multiple beamforming, the transmit power still needs to satisfy the total power constraint

$$\frac{1}{T}\mathrm{E}[\|\boldsymbol{X}\|_F^2] \leq \mathrm{SNR}$$

where $\boldsymbol{X} = [\boldsymbol{x}_1 \ \boldsymbol{x}_2 \ \cdots \ \boldsymbol{x}_T] \in \mathcal{C}^{K \times T}$ and $\boldsymbol{x}_i$ is the transmitted signal vector of the ith channel use within a block.

Diversity gain and multiplexing gain are two fundamental concepts in DMT theory.

MIMO channels provide spatial diversity to improve link reliability. A MIMO channel with M transmit antennas and N receive antennas provides MN individual transmit-receive antenna pairs. The goal of the diversity gain is to provide the receiver with multiple independently faded replicas of the transmitted symbols in the spatial domain. Statistically, the probability that all independently faded replicas experience deep fading simultaneously is small. Thus, the link reliability is improved.

On the other hand, the MIMO channel ergodic capacity with CSIR in the high-SNR regime can be well approximated as [1, 2]

$$C(\mathrm{SNR}) \approx n \log \frac{\mathrm{SNR}}{M}.$$

The channel capacity increases with SNR as $n \log \mathrm{SNR}$. Compared with $\log \mathrm{SNR}$, the SISO ergodic capacity with CSIR in the high-SNR regime, n serves as the multiplexing gain. In order to achieve a certain fraction of the capacity in the high-SNR regime, we consider schemes that support a data rate which also increases with SNR. We need to define a scheme as a family of codes $\{\mathcal{C}(\mathrm{SNR})\}$ with coding over a block (with coding length T), one at each SNR level. Let $R(\mathrm{SNR})$ be the rate of the code $\mathcal{C}(\mathrm{SNR})$. If the supported data rate of this scheme meets

$$R(\mathrm{SNR}) \approx r \log \mathrm{SNR},$$

then we say that multiplexing gain r is achieved.

Definition: A coding scheme $\{\mathcal{C}(\mathrm{SNR})\}$ is said to achieve a spatial multiplexing gain

r and a diversity gain d if the data rate

$$\lim_{\mathrm{SNR}\to\infty} \frac{R(\mathrm{SNR})}{\log \mathrm{SNR}} = r$$

and the average error probability (averaged over the channel fading, the noise and the transmitted signals)

$$\lim_{\mathrm{SNR}\to\infty} \frac{\log P_e(\mathrm{SNR})}{\log \mathrm{SNR}} = -d.$$

For each r, define $d^*(r)$ to be the supremum of the diversity gain achieved over all schemes.

We define the symbol $\doteq$ to denote exponential equality, i.e., if

$$\lim_{\mathrm{SNR}\to\infty} \frac{\log f(\mathrm{SNR})}{\log \mathrm{SNR}} = c$$

we write $f(\mathrm{SNR}) \doteq \mathrm{SNR}^c$. The symbols $\dot{\leq}$ and $\dot{\geq}$ are similarly defined.

6.3 Optimal DMT of Multiple Beamforming

6.3.1 Outage Probability - A Lower Bound on Error Probability

In this subsection, we calculate the outage probability in the high-SNR regime, which provides a lower bound on error probability, an upper bound on diversity gain.

The outage probability of the channel described by (6.5) is defined as [7]

$$\begin{aligned} P_{\mathrm{out}}(R) &= \inf_{\boldsymbol{Q}\geq 0, \mathrm{Tr}\{\boldsymbol{Q}\}\leq \mathrm{SNR}} \Pr(\log\det(\boldsymbol{I} + \boldsymbol{\Lambda}\boldsymbol{Q}\boldsymbol{\Lambda}^H) < R) \\ &= \inf_{\boldsymbol{Q}\geq 0, \mathrm{Tr}\{\boldsymbol{Q}\}\leq \mathrm{SNR}} \Pr(\log\det(\boldsymbol{I} + \boldsymbol{Q}\boldsymbol{\Lambda}^2) < R) \end{aligned}$$

where $\boldsymbol{Q} = \mathrm{E}[\boldsymbol{x}\boldsymbol{x}^H]$. Since $P_{\mathrm{out}}(R)$ can be upper-bounded by taking $\boldsymbol{Q} = (\mathrm{SNR}/K)\boldsymbol{I}$ and lower-bounded by taking $\boldsymbol{Q} = (\mathrm{SNR})\boldsymbol{I}$, we have

$$\Pr(\log\det(\boldsymbol{I} + \frac{\mathrm{SNR}}{K}\boldsymbol{\Lambda}^2) < R) \geq P_{\mathrm{out}}(R) \geq \Pr(\log\det(\boldsymbol{I} + \mathrm{SNR}\boldsymbol{\Lambda}^2) < R).$$

At high SNR,

$$\begin{aligned}
&\lim_{\mathrm{SNR}\to\infty} \frac{\log\Pr(\log\det(\boldsymbol{I} + \frac{\mathrm{SNR}}{K}\boldsymbol{\Lambda}^2) < R)}{\log\mathrm{SNR}} \\
&= \lim_{\mathrm{SNR}\to\infty} \frac{\log\Pr(\log\det(\boldsymbol{I} + \frac{SNR}{K}\boldsymbol{\Lambda}^2) < R)}{\log\frac{\mathrm{SNR}}{K}} \\
&= \lim_{\mathrm{SNR}\to\infty} \frac{\log\Pr(\log\det(\boldsymbol{I} + \mathrm{SNR}\boldsymbol{\Lambda}^2) < R)}{\log\mathrm{SNR}}.
\end{aligned}$$

We have

$$\begin{aligned}
P_{\mathrm{out}}(R) &\doteq \Pr(\log\det(\boldsymbol{I} + \mathrm{SNR}\boldsymbol{\Lambda}^2) < R) \\
&= \Pr(\log\prod_{i=n-K+1}^{n}(1 + \mathrm{SNR}\lambda_i) < R).
\end{aligned}$$

In order to compute $P_{\mathrm{out}}(R)$, we need the joint probability density (pdf) function of $\{\lambda_{n-K+1}, ..., \lambda_n\}$. We first review the joint pdf of all eigenvalues of an uncorrelated central Wishart matrix and the joint pdf of transformations of these eigenvalues given in [7, Lemma 3].

Lemma 9 *Let $\boldsymbol{R}$ be an $m \times n$ random matrix with i.i.d. $\mathcal{CN}(0,1)$ entries. Suppose that $m \geq n$, $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ are the ordered nonzero eigenvalues of $\boldsymbol{R}^H\boldsymbol{R}$. Then the joint pdf of λ_i's is*

$$p(\lambda_1, ..., \lambda_n) = K_{m,n}^{-1}\prod_{i=1}^{n}\lambda_i^{m-n}\prod_{i<j}(\lambda_i - \lambda_j)^2 e^{-\sum_{i=1}^{n}\lambda_i}$$

where $K_{m,n}$ is a normalization constant. Define $\alpha_i := -\log \lambda_i / \log SNR$, for all i. The joint pdf of the random vector $\boldsymbol{\alpha} = [\alpha_1, ..., \alpha_n]$ is

$$p(\boldsymbol{\alpha}) = K_{m,n}^{-1} (\log SNR)^n \prod_{i=1}^{n} SNR^{-(m-n+1)\alpha_i} \prod_{i<j} (SNR^{-\alpha_i} - SNR^{-\alpha_j})^2 \exp\left[-\sum_{i=1}^{n} SNR^{-\alpha_i} \right]$$

by changing variable $\lambda_i = SNR^{-\alpha_i}$, for all i.

Let $R = r \log \mathrm{SNR}$, where $0 \leq r \leq K$ [118]. In the high-SNR regime, we have

$$P_{\text{out}}(r \log \mathrm{SNR}) \doteq \Pr(\log \prod_{i=n-K+1}^{n} (1 + \mathrm{SNR}\lambda_i) < r \log \mathrm{SNR}).$$

Let $\lambda_i = \mathrm{SNR}^{-\alpha_i}$. At high SNR, $(1 + \mathrm{SNR}\lambda_i) \doteq \mathrm{SNR}^{(1-\alpha_i)^+}$, where $(x)^+$ denotes $\max\{0, x\}$. We have

$$\begin{aligned} P_{\text{out}}(r \log \mathrm{SNR}) &\doteq \Pr\left[\prod_{n-K+1}^{n} \mathrm{SNR}^{(1-\alpha_i)^+} < \mathrm{SNR}^r \right] \\ &= \Pr\left[\sum_{n-K+1}^{n} (1-\alpha_i)^+ < r \right]. \end{aligned}$$

Define $\boldsymbol{\alpha}_1 = (\alpha_1, ..., \alpha_{n-K})$ and $\boldsymbol{\alpha}_2 = (\alpha_{n-K+1}, ..., \alpha_n)$. In order to calculate the probability of $\sum_{n-K+1}^{n} (1-\alpha_i)^+ < r$, we need the marginal pdf of $\boldsymbol{\alpha}_2$ which can be expressed as

$$p(\boldsymbol{\alpha}_2) = \int_{\alpha_{n-K+1}}^{\infty} ... \int_{\alpha_2}^{\infty} p(\boldsymbol{\alpha}) d\alpha_1 ... d\alpha_{n-K}.$$

By definition, for $i < j$, $\alpha_i \geq \alpha_j$. Denote the set $\mathcal{A}_2 := \{(\alpha_{n-K+1}, ..., \alpha_n) : \sum_{i=n-K+1}^{n} (1-$

$\alpha_i)^+ < r\}$. The outage probability is obtained as

$$
\begin{aligned}
&P_{\text{out}}(r\log \text{SNR}) \\
&\doteq \int_{\mathcal{A}_2} p(\boldsymbol{\alpha}_2) d\boldsymbol{\alpha}_2 \\
&= \int_{\mathcal{A}_2} \left(\int_{\alpha_{n-K+1}}^{\infty} ... \int_{\alpha_2}^{\infty} p(\boldsymbol{\alpha}) d\alpha_1 ... d\alpha_{n-K} \right) d\alpha_{n-K+1} ... \alpha_n \\
&= \int_{\mathcal{A}_2} \Bigg(\int_{\alpha_{n-K+1}}^{\infty} ... \int_{\alpha_2}^{\infty} K_{m,n}^{-1} (\log \text{SNR})^n \prod_{i=1}^{n} \text{SNR}^{-(m-n+1)\alpha_i} \prod_{i<j} (\text{SNR}^{-\alpha_i} - \text{SNR}^{-\alpha_j})^2 \\
&\cdot \exp\left[-\sum_{i=1}^{n} \text{SNR}^{-\alpha_i} \right] d\alpha_1 ... d\alpha_{n-K} \Bigg) d\alpha_{n-K+1} ... d\alpha_n
\end{aligned}
$$

We are interested in the SNR exponent of $P_{\text{out}}(r\log \text{SNR})$ in the high-SNR regime, i.e.,

$$\lim_{\text{SNR}\to\infty} \frac{\log P_{\text{out}}(r\log \text{SNR})}{\log \text{SNR}},$$

using the same arguments in [7]: 1) The term $K_{m,n}^{-1}(\log \text{SNR})^n$ has no effect on the SNR exponent because

$$\lim_{\text{SNR}\to\infty} \frac{\log[K_{m,n}^{-1}(\log \text{SNR})^n]}{\log \text{SNR}} = 0;$$

2) For any $\alpha_i < 0$, the term $\exp(-\text{SNR}^{-\alpha_i})$ decays with SNR exponentially. At high SNR, the integral over any $\alpha_i < 0$ can be ignored. Therefore, the integrals are only considered for any $\alpha_i \geq 0$. For any $\alpha_i \geq 0$, at high SNR, $\exp(-\text{SNR}^{-\alpha_i})$ approaches 1 when $\alpha_i > 0$, and approaches $1/e$ when $\alpha_i = 0$, this form has no effect on the SNR exponent and can be ignored. For any $i < j$, $\alpha_i \geq \alpha_j$. Let us define the following set:

$$\mathcal{A}_2^{'} := \{\boldsymbol{\alpha}_2 \in R^{K+} | \alpha_{n-K+1} \geq ... \geq \alpha_n \geq 0 \text{ and } \sum_{i=n-K+1}^{n} (1-\alpha_i)^+ < r\} \quad (6.6)$$

$$\mathcal{A}_1^{'} := \{\boldsymbol{\alpha}_1 \in R^{(n-K)+} | \alpha_1 \geq ... \geq \alpha_{n-K} \geq \alpha_{n-K+1}\} \quad (6.7)$$

where R^{m+} represents the set of $m \times 1$ vectors with non-negative elements. The integral becomes

$$P_{\text{out}}(r \log \text{SNR}) \doteq \int_{\mathcal{A}_2'} \left(\int_{\mathcal{A}_1'} \prod_{i=1}^{n} \text{SNR}^{-(m-n+1)\alpha_i} \prod_{i<j} (\text{SNR}^{-\alpha_i} - \text{SNR}^{-\alpha_j})^2 d\boldsymbol{\alpha}_1 \right) d\boldsymbol{\alpha}_2. \tag{6.8}$$

We have the following theorem about the outage probability when SNR$\to \infty$.

Theorem 10 *For the beamforming channel described by Eq.(6.5), let the data rate be $R = r \log SNR$(b/s/Hz), with $0 \leq r \leq K$. The outage probability satisfies*

$$P_{out}(r \log SNR) \doteq SNR^{-d_{out}(r)}$$

where

$$d_{out}(r) = \inf_{\boldsymbol{\alpha}_2 \in \mathcal{A}_2'} \left[(n-K+1)(m-K+1)\alpha_{n-K+1} + \sum_{n-K+2}^{n} (m-n+2i-1)\alpha_i \right] \tag{6.9}$$

and $\mathcal{A}_2'$ is defined in (6.6).

Proof. Please see Appendix C. ■

Given $0 \leq r \leq K$, $d_{\text{out}}(r)$ can be explicitly calculated as follows.

1. Denote the coefficient of α_i $(n-K+1 \leq i \leq n)$ as c_i. We have

$$\begin{aligned} c_{n-K+1} &= (n-K+1)(m-K+1) > 0 \\ c_i &= m-n+2i-1 > 0 \qquad\qquad i = n-K+2, ..., n \end{aligned} \tag{6.10}$$

2. Find the largest index i, denoted as i^*, $(n-K+1 \le i^* \le n)$ such that

$$i^* = \arg \min_{n-K+1 \le i_1 \le n} \frac{1}{i_1 - n + K} \sum_{i=n-K+1}^{i_1} c_i \tag{6.11}$$

3. When $n - i^* \le r \le K$,

$$\alpha_{n-K+1} = ... = \alpha_{i^*} = \frac{K-r}{i^* - n + K},$$
$$\alpha_{i^*+1} = ... = \alpha_n = 0;$$

and

$$d_{\text{out}}(r) = \frac{K-r}{i^* - n + K} \sum_{i=n-K+1}^{i^*} c_i$$

When $0 \le r \le n - i^*$,

$$\alpha_{n-K+1} = ... = \alpha_{n-\lceil r \rceil} = 1,$$
$$\alpha_{n-\lceil r \rceil+1} = \lceil r \rceil - r,$$
$$\alpha_{n-\lceil r \rceil+2} = ... = \alpha_n = 0;$$

and

$$d_{\text{out}}(r) = \sum_{n-K+1}^{n-\lceil r \rceil} c_i + \alpha_{n-\lceil r \rceil+1} c_{n-\lceil r \rceil+1}$$

where $\lceil x \rceil$ denotes the smallest integer not less than x.

This procedure can be well explained using Fig. 6.1. There are K identical containers of height 1, labeled α_{n-K+1} to α_n. When the faucet is turned on, the water level goes up. The water level in each container is the value of its corresponding α. The first $i^* - n + K$ containers from α_{n-K+1} to α_{i^*} are connected together at their bottom such

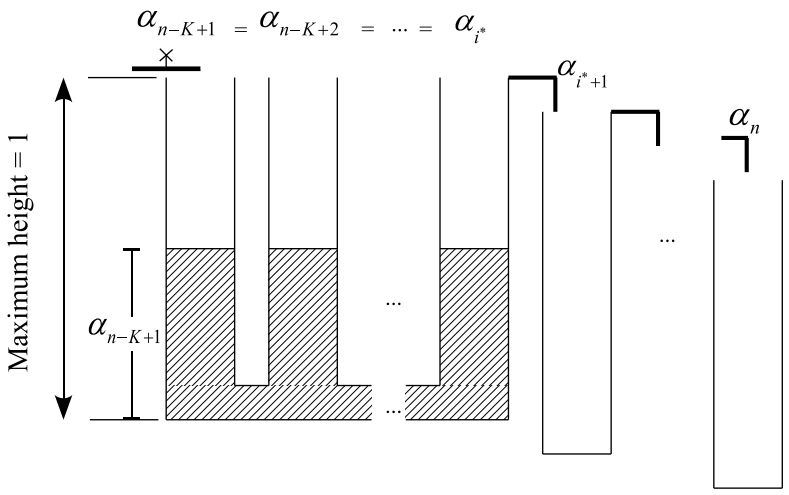

Figure 6.1: Solution of $\boldsymbol{\alpha}_2$ by water-pouring.

that they always have the same water level. When overflow occurs, water goes to the next containers (on the right, for $i \geq i^*+1$) one by one through the water pipes with each connecting two adjacent containers at their top. If water is not shut down, all containers are eventually full, i.e. $\alpha_{n-K+1} = ... = \alpha_n = 1$, and this is the case of $r = 0$. For the general case of $0 \leq r \leq K$, water is shut down when the condition $\sum_{i=n-K+1}^{n}(1-\alpha_i)^+ = r$ is met and the optimal values of α_i's to reach $d_{\text{out}}(r)$ are determined by the final water level.

Consider coding across K independent channel eigenmodes in the channel described by (6.5). The transmission rate per channel use is $R = r \log \text{SNR}$. We assume a Gaussian codebook $\mathcal{X}$ of size 2^{RT}. Each codeword $\boldsymbol{X}_i$ has a coding length of T, $1 \leq i \leq 2^{RT}$, which means that transmission of each codeword needs to use channel (6.5) T times, i.e.,

$\boldsymbol{X}_i \in \mathcal{C}^{K\times T}$. T is much less the channel coherent time such that during each codeword transmission, the channel is constant. Since the transmitter has no knowledge of the channel matrix $\boldsymbol{\Lambda}$, the construction of codebook $\mathcal{X}$ is independent of $\boldsymbol{\Lambda}$ such that each codeword is chosen from i.i.d. Gaussian ensemble. The transmission of codeword $\boldsymbol{X}_i$ is independent of channel matrix $\boldsymbol{\Lambda}$ and is uniformly drawn from the codebook. The channel with codeword transmission can be written as

$$\boldsymbol{Y} = \boldsymbol{\Lambda}\boldsymbol{X} + \boldsymbol{W} \tag{6.14}$$

where $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{W} \in \mathcal{C}^{K\times T}$. By Fano's inequality, the lower-bound of the average error probability, that is, the lower-bound of the probability the receiver detects a codeword other than the one the transmitter sends, is determined by the following lemma.

Lemma 11 *For the channel in (6.5), let the data rate be $R = r \log SNR$(b/s/Hz). For any coding scheme, the probability of a detection error is lower-bounded by*

$$P_e(SNR) \dot{\geq} SNR^{-d_{out}(r)} \tag{6.15}$$

where $d_{out}(r)$ is defined in (6.9).

Proof. Follow the proof of [7, Lemma 5] with $\boldsymbol{H}$ being replaced by $\boldsymbol{\Lambda}$, we have this result. ■

6.3.2 Gaussian Bound - An Upper Bound on Error Probability

Next we develop an upper-bound on $P_e(\text{SNR})$. We assume that the channel input $\boldsymbol{X}$ is drawn from codebook $\mathcal{X}$ with data rate $R = r \log \text{SNR(b/s/Hz)}$. The average error

probability is upper-bounded as [7]

$$\begin{aligned} P_e(\text{SNR}) &= P_{\text{out}}(R)P(\text{error} \mid \text{outage}) + P(\text{error, no outage}) \\ &\leq P_{\text{out}}(R) + P(\text{error, no outage}). \end{aligned}$$

Using the same technique in the proof of [7, Theorem 2], we can show that

$$P(\text{error, no outage}) \doteq \int_{\bar{\mathcal{A}}_2'} \text{SNR}^{-d_G(\boldsymbol{\alpha}_2, r)} d\boldsymbol{\alpha}_2$$

where

$$d_G(\boldsymbol{\alpha}_2, r) = (n-K+1)(m-K+1)\alpha_{n-K+1} + \sum_{i=n-K+2}^{n} (m-n+2i-1)\alpha_i + T[\sum_{i=n-K+1}^{n} (1-\alpha_i)^+ - r]$$

and

$$\bar{\mathcal{A}}_2' = \{\boldsymbol{\alpha}_2 \in R^{K+} | \alpha_{n-K+1} \geq ... \geq \alpha_n \geq 0 \text{ and } \sum_{i=n-K+1}^{n} (1-\alpha_i)^+ \geq r\}$$

represents the set of no-outage event, a complement of the set of outage event $\mathcal{A}_2'$. In the high-SNR regime, $\int_{\bar{\mathcal{A}}_2'} \text{SNR}^{-d_G(\boldsymbol{\alpha}_2, r)} d\boldsymbol{\alpha}_2$ is dominated by $\int_{\bar{\mathcal{A}}_2'} \text{SNR}^{-d_G(r)} d\boldsymbol{\alpha}_2$ with

$$d_G(r) = \min_{\boldsymbol{\alpha}_2 \in \bar{\mathcal{A}}_2'} d_G(\boldsymbol{\alpha}_2, r). \tag{6.16}$$

When the coding length T satisfies

$$T \geq T^* \triangleq \max \left\{ \left\lceil \frac{1}{i^* - n + K} \sum_{i=n-K+1}^{i^*} c_i \right\rceil, \; c_n \right\} \tag{6.17}$$

where $c_n = m-n+2n-1 = m+n-1$, the minimum is reached when $\sum_{i=n-K+1}^{n}(1-\alpha_i)^+ = r$. The term in (6.16) is expressed as

$$d_G(r) = \min_{\sum_{i=n-K+1}^{n}(1-\alpha_i)^+=r} \Bigg[(n-K+1)(m-K+1)\alpha_{n-K+1} + \sum_{i=n-K+2}^{n}(m-n+2i-1)\alpha_i\Bigg].$$

From (6.9), we can verify that $d_G(r) = d_{\text{out}}(r)$, for $0 \le r \le K$, and

$$\begin{aligned} P_e(\text{SNR}) &\le P_{\text{out}}(R) + P(\text{error, no outage}) \\ &\dot{\le} \text{SNR}^{-d_{\text{out}}(r)} + \text{SNR}^{-d_G(r)} \\ &\doteq \text{SNR}^{-d_{\text{out}}(r)}. \end{aligned} \tag{6.18}$$

Combining (6.15) with (6.18), we have

$$P_e(\text{SNR}) \doteq \text{SNR}^{-d_{\text{out}}(r)}.$$

Let $d^*_{\text{MB}}(r)$ denote the optimal DMT of multiple beamforming with $T \ge T^*$. The largest diversity gain obtained by any coding scheme, for $0 \le r \le K$, is

$$d^*_{\text{MB}}(r) = d_{\text{out}}(r).$$

Clearly, when $r = 0$, $d^*_{\text{MB}}(0) = d^*_{\max} = MN$; when $r = K$, $d^*_{\text{MB}}(K) = d^*_{\min} = 0$.

6.4 Optimal DMT of Multiple Beamforming with $T \geq T^*$ versus Fundamental Optimal DMT of Multiple Antenna System with CSIR with $T \geq M + N - 1$

In this section, we evaluate the optimal DMT of multiple beamforming to gain insights of the tradeoff. The fundamental optimal DMT of independently Rayleigh-faded multiple-antenna channels with CSIR, denoted as $d^*_{\text{CSIR}}(r)$, is given by the piecewise-linear function connecting the point $(k, d^*_{\text{CSIR}}(k))$, $k = 0, 1, ..., n$, where

$$d^*_{\text{CSIR}}(k) = (m-k)(n-k) \tag{6.19}$$

when the coding length $T \geq M + N - 1$ [7, Theorem 2].

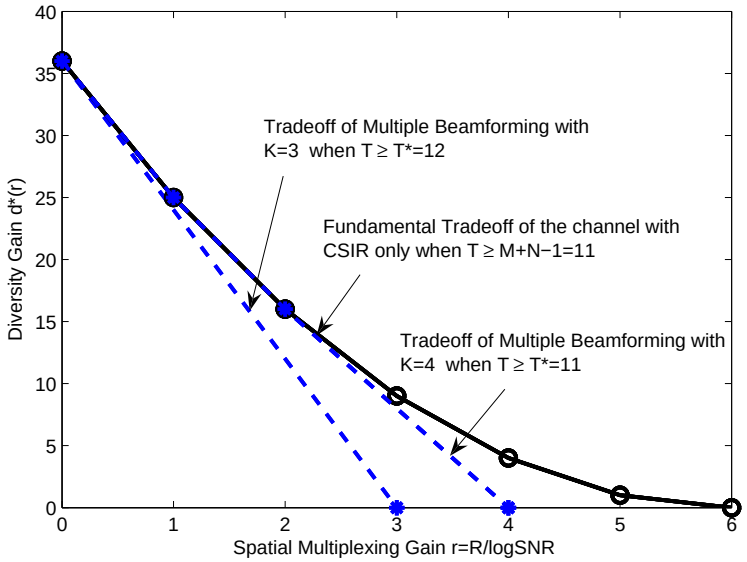

Figure 6.2: Diversity-multiplexing tradeoff: $d^*_{\text{MB}}(r)$ and $d^*_{\text{CSIR}}(r)$ ($M = N = 6$ and $K = 3, 4$).

Fig. 6.2 shows the tradeoff curves of multiple beamforming with $K = 3$ and $K = 4$ in a system with 6 transmit antennas and 6 receive antennas. The fundamental tradeoff curve of the same system with CSIR is also included. When $K = 3$, the tradeoff curve $d^*_{\mathrm{MB}}(r)$ is a straight line connecting the points $(0, 36)$ and $(3, 0)$ and $d^*_{\mathrm{MB}}(r) < d^*_{\mathrm{CSIR}}(r)$, for $0 < r \leq 3$; when $K = 4$, the tradeoff curve overlaps with the fundamental tradeoff curve for $0 \leq r \leq 2$ and $d^*_{\mathrm{MB}}(2) = d^*_{\mathrm{CSIR}}(2) = 16$. When $2 \leq r \leq 4$, $d^*_{\mathrm{MB}}(r)$ is a straight line connecting the points $(2, 16)$ and $(4, 0)$ and $d^*_{\mathrm{MB}}(r) < d^*_{\mathrm{CSIR}}(r)$ when $2 < r \leq 4$.

For Fig. 6.3, the system settings are $M = 5$, $N = 4$, and $K = 2, 3$. When $K = 2$, the tradeoff curve $d^*_{\mathrm{MB}}(r)$ is a straight line connecting the points $(0, 20)$ and $(2, 0)$. $d^*_{\mathrm{MB}}(r)$ is equal to $d^*_{\mathrm{CSIR}}(r)$ when $r = 0$ and is strictly less than $d^*_{\mathrm{CSIR}}(r)$ when $0 < r \leq 2$. When $K = 3$, $d^*_{\mathrm{MB}}(r) = d^*_{\mathrm{CSIR}}(r)$ when $0 \leq r \leq 2$ and $d^*_{\mathrm{MB}}(r) < d^*_{\mathrm{CSIR}}(r)$ when $2 < r \leq 3$.

In Fig. 6.2 and Fig. 6.3 we observe that $d^*_{\mathrm{MB}}(r) \leq d^*_{\mathrm{CSIR}}(r)$. The following Theorem summarizes this result.

Theorem 12 *Given any multiplexing gain r, $0 \leq r \leq K$ ($1 \leq K \leq n$), the optimal diversity gain of multiple beamforming of channel (6.5) $d^*_{MB}(r)$ with $T \geq T^*$ is less than or equal to the optimal diversity gain $d^*_{CSIR}(r)$ provided by the same multiple antenna system with CSIR when $T \geq m+n-1$, where T^* is given in (6.17). For the special case of $K = n$, $d^*_{MB}(r) = d^*_{CSIR}(r)$.*

Proof. Assume $T \geq T^*$ for $d^*_{\mathrm{MB}}(r)$ and $T \geq m + n - 1$ for $d^*_{\mathrm{CSIR}}(r)$. First we consider $K = n$. It's obvious that $d^*_{\mathrm{MB}}(r) = d^*_{\mathrm{CSIR}}(r)$ by comparing (6.9) with [7, Eq. 14]. Both DMT curves are identical and they are both piecewise-linear function connecting the points $(k, (m-k)(n-k))$, $k = 0, 1, ..., n$.

Now consider $K < n$. Please note $d^*_{\mathrm{MB}}(r) = 0$ when $r \geq K$, while $d^*_{\mathrm{CSIR}}(r) > 0$ when $K \leq r < n$ and $d^*_{\mathrm{CSIR}}(r) = 0$ when $r \geq n$. In order to finish the proof, we only need to

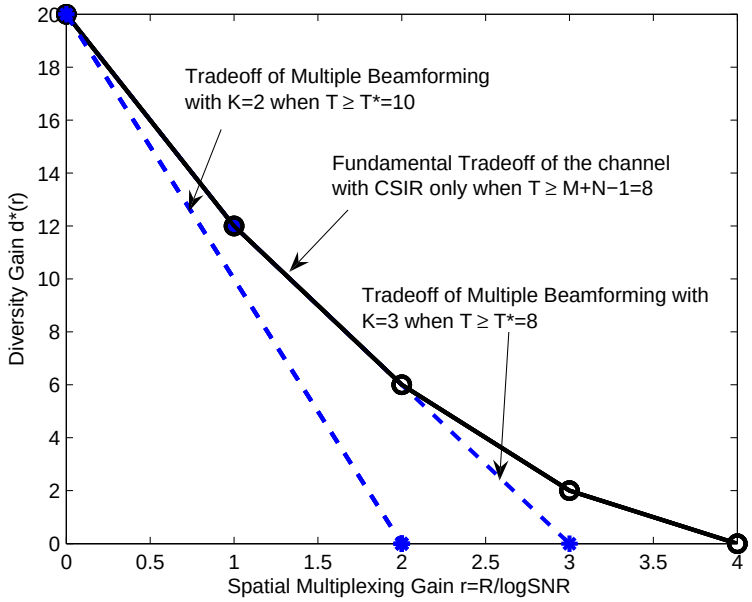

Figure 6.3: Diversity-multiplexing tradeoff: $d^*_{\mathrm{MB}}(r)$ and $d^*_{\mathrm{CSIR}}(r)$ ($M = 5$, $N = 4$ and $K = 2, 3$).

show $d^*_{\mathrm{MB}}(r) \leq d^*_{\mathrm{CSIR}}(r)$ when $0 \leq r \leq K$.

Between the point k and $k+1$ ($k = 0, ..., n-1$), $d^*_{\mathrm{CSIR}}(r)$ is a straight line and has the slope

$$(m-k-1)(n-k-1) - (m-k)(n-k) = -(m+n-2k-1). \tag{6.20}$$

Recall in step 3) of the calculation of $d_{\mathrm{out}}(r)$ in Section 6.3, there are two regions: $n - i^* \leq r \leq K$ and $0 \leq r \leq n - i^*$, where the optimal diversity is calculated differently.

By definition of i^* in (6.11):

1. When $n-K+2 \leq i^* \leq n$, we have

$$\frac{1}{(i^*-1)-n+K}\sum_{i=n-K+1}^{i^*-1} c_i \geq \frac{1}{i^*-n+K}\sum_{i=n-K+1}^{i^*} c_i$$

which implies

$$\frac{1}{i^*-n+K}\sum_{i=n-K+1}^{i^*} c_i \geq c_{i^*} \tag{6.21}$$

i) When $0 \leq r \leq n-i^*$,

$$d^*_{\mathrm{MB}}(r) = \sum_{n-K+1}^{n-\lceil r\rceil} c_i + \alpha_{n-\lceil r\rceil+1} c_{n-\lceil r\rceil+1}$$

where $\alpha_{n-\lceil r\rceil+1} = \lceil r\rceil - r$. When r takes integer k, $0 \leq k \leq n-i^*$, with c_i in (6.10) plugged in, $d^*_{\mathrm{MB}}(k) = \sum_{n-K+1}^{n-k} c_i = (n-k)(m-k)$. Comparing $d^*_{\mathrm{MB}}(k)$ with (6.19), we have $d^*_{\mathrm{MB}}(k) = d^*_{\mathrm{CSIR}}(k)$. When r is not an integer, say $k < r < k+1$ ($k = 0, ..., n-i^*-1$), $d^*_{\mathrm{MB}}(r)$ is a straight line connecting $(k, d^*_{\mathrm{MB}}(k))$ and $(k+1, d^*_{\mathrm{MB}}(k+1))$ with a slope of $-c_{n-\lceil r\rceil+1} = -c_{n-k} = -(m-n+2(n-k)-1) = -(m+n-2k-1)$, which is the same as (6.20). Therefore, we have $d^*_{\mathrm{MB}}(r) = d^*_{\mathrm{CSIR}}(r)$ for $0 \leq r \leq n-i^*$. The optimal DMT curve $d^*_{\mathrm{MB}}(r)$ exactly matches the fundamental DMT curve $d^*_{\mathrm{CSIR}}(r)$ in this region.

ii) When $n-i^* < r \leq K$,

$$d^*_{\mathrm{MB}}(r) = \frac{K-r}{i^*-n+K}\sum_{i=n-K+1}^{i^*} c_i,$$

is a straight line connecting $(n-i^*, \sum_{i=n-K+1}^{i^*} c_i)$ and $(K, 0)$ with a slope

$\frac{-1}{i^*-n+K}\sum_{i=n-K+1}^{i^*} c_i$. Using (6.20), we can verify that the slope of $d^*_{\text{CSIR}}(r)$ when $n-i^* \leq r \leq K$ is lower bounded by $-(m+n-2(n-i^*)-1) = -(m-n+2i^*-1) = -c_{i^*}$. By (6.21), $\frac{-1}{i^*-n+K}\sum_{i=n-K+1}^{i^*} c_i \leq -c_{i^*}$. Since at $r = n-i^*$, $d^*_{\text{MB}}(n-i^*) = d^*_{\text{CSIR}}(n-i^*)$ (both tradeoff curves start decaying from same point), we have $d^*_{\text{MB}}(r) \leq d^*_{\text{CSIR}}(r)$ for $n-i^* < r \leq K$.

2. When $i^* = n-K+1$:

   i) When $0 \leq r \leq n-i^* = K-1$, using the same arguments as in case i) of 1), we have $d^*_{\text{MB}}(r) = d^*_{\text{CSIR}}(r)$, for $0 \leq r \leq K-1$.

   ii) When $n-i^* = K-1 < r \leq K$, we use arguments slightly different from those in case ii) of 1). In this region,

$$d^*_{\text{MB}}(r) = (K-r)c_{n-K+1},$$

   is a straight line connecting $(K-1, c_{n-K+1})$ and $(K,0)$ with a slope $-c_{n-K+1} = -(n-K+1)(m-K+1)$. Using (6.20), we find the slope of $d^*_{\text{CSIR}}(r)$: $-(m+n-2K+1)$ when $K-1 \leq r \leq K$. Since $-c_{n-K+1} < -(m+n-2K+1)$ and $d^*_{\text{MB}}(K-1) = d^*_{\text{CSIR}}(K-1)$, we have $d^*_{\text{MB}}(r) < d^*_{\text{CSIR}}(r)$ for $K-1 < r \leq K$.

Combining 1) and 2), we complete the proof. ■

For any fixed-rate transmission, the multiplexing gain $r = 0$. The maximum diversity gain provided by multiple beamforming is $d^*_{\text{MB}}(0) = MN$. Comparing $d^*_{\text{MB}}(0)$ with the results in [21, 118], which show that the scheme transmitting K symbols independently drawn from the same constellation (equivalent to the same data rate on each channel eigenmode) with coding length $T = 1$ on the first K ($K > 1$) largest channel eigenmodes

has a maximum diversity order $(M-K+1)(N-K+1)$, we observe an increase in diversity gain of

$$MN-(M-K+1)(N-K+1)=(K-1)(M+N-K+1)>0.$$

This diversity increase is provided by coding over both channel eigenmodes and time.

In Fig. 6.2 and Fig. 6.3, the optimal DMT curve $d^*_{\mathrm{MB}}(r)$ decreases faster when K is small than the one with a large K. This can be explained as follows. As K decreases, more channel eigenmodes are discarded, losing protection that could have been provided by the channel. Thus, when r increases, the error rate increases faster when K is small.

6.5 Optimal DMT of Multiple Beamforming with $T \geq T^*$ versus Optimal DMT of Spatial Multiplexing with CSIT with $T=1$ in [118]

The optimal DMT of spatial multiplexing with CSIT is derived without coding across eigenmodes and time, i.e., independent n symbols with coding length $T=1$ are sent over n eigenchannels [118]. Since Full CSIT is available at the transmitter, the capacity-achieving water-filling power allocation is performed on the symbols on all available eigenmodes based on instantaneous eigenvalues λ_i's. The channel model (6.5) becomes

$$\boldsymbol{y}=\boldsymbol{\Lambda P x}+\boldsymbol{w}, \tag{6.22}$$

where $\boldsymbol{P}$ is a $K\times K$ diagonal matrix with diagonal entries $p_1,...,p_K$. p_k is the power for the symbols on the eigenmode corresponding to the kth largest eigenvalue λ_{n-k+1}. For notation simplicity, we denote the kth largest eigenvalue as $\tilde{\lambda}_k=\lambda_{n-k+1}$. $p_k=(\mu-\tilde{\lambda}_k^{-1})^+$ due to water-filling principle and μ is determined by the short term power constraint

$\sum_{k=1}^{K} p_k$ = SNR (i.e., no power adaptation to the fading in the time domain). The individual DMT $d_S^k(r)$ of the kth ($1 \le k \le n$) largest eigenchannel with water-filling power allocation for the channel model (6.22) without coding over eigenmodes and time is given by [118, Theorem 1]

$$d_S^{(k)}(r_k) = d_k(1 - r_k) \quad 0 \le r_k \le 1 \tag{6.23}$$

where $d_k = (m-k+1)(n-k+1)$. Based on this result, for the general case of $0 \le r \le n$, by imposing the optimal rate allocation ($\sum_{k=1}^{I} r_k = r$, $I \ge \lceil r \rceil$) that assures the same SNR exponent $d(I, r) = d_k(1 - r_k)$ for the active I substreams and further maximizing the resulting $d(I, r)$ over I ($\lceil r \rceil \le I \le n$), the optimal DMT curve of channel (6.22), denoted as $d^*_{\text{CSIT}}(r)$, can be obtained for the spatial multiplexing systems with full CSI at the transmitter without coding over eigenmodes (space) and time.

Now we compare $d^*_{\text{MB}}(r)$ and $d^*_{\text{CSIT}}(r)$. First, Theorem 12 shows that when $K = n$, $d^*_{\text{MB}}(r) = d^*_{\text{CSIR}}(r)$ for the entire range of r, i.e., $0 \le r \le n$. [118, Theorem 3] shows that $d^*_{\text{CSIT}}(r) = d^*_{\text{CSIR}}(r)$ only when $r = 0, n$ and $d^*_{\text{CSIT}}(r) < d^*_{\text{CSIR}}(r)$ when $0 < r < n$. Therefore, $d^*_{\text{CSIT}}(r) \le d^*_{\text{MB}}(r)$ for $0 \le r \le n$ when $K = n$. This result is not surprising since $d^*_{\text{CSIT}}(r)$ is derived without coding over space and time, i.e., coding independently over eigenmodes and time. On the other hand, $d^*_{\text{MB}}(r)$ is obtained with coding over space and time even with partial CSI[1]. This shows coding is more beneficial.

To gain more insights, consider equal power allocation $p_k = \frac{\text{SNR}}{K}$, instead of water-filling power allocation $p_k = (\mu - \tilde{\lambda}_k^{-1})^+$, and no coding over space and time for K eigenchannels, $K = 1, ..., n-1$. It is straightforward to go through the steps in [118, IV.A]

[1]Clearly water-filling power adaptation is not possible in the configuration of multiple beamforming since eigenvalue information is not available at the transmitter.

to obtain the

$$P_e^k(R_k) \doteq \Pr(\tilde{\lambda}_k \leq \mathrm{SNR}^{r_k-1})$$

where $R_k = r_k \log \mathrm{SNR}$ is the data rate of the kth largest eigenmode corresponding to $\tilde{\lambda}_k$. $P_e^k(R_k)$ is the symbol error rate of the corresponding eigenmode. Following the technique in the proof of [118, Theorem 1], we have same result $d_S^{(k)}(r_k) = d_k(1-r_k)$ for $0 \leq r_k \leq 1$ as in (6.23). This shows that water-filling power allocation with short term power constraint $\sum_{k=1}^{K} p_k = \mathrm{SNR}$ without coding over space and time has no advantage in terms of maximizing diversity gain over equal power allocation; both power allocation schemes indeed have exactly the same DMT. For fair comparison between $d_{\mathrm{MB}}^*(r)$ and $d_{\mathrm{CSIT}}^*(r)$ for $K = 1, ..., n-1$, we only use the full CSIT of the first K largest channel eigenmodes, denoted as CSIT_K, for spatial multiplexing. Since the optimal DMT of multiple beamforming for any K is obtained by using equal power allocation with coding over space and time ($T \geq T^*$), intuitively it should be better than the DMT of spatial multiplexing with CSIT_K using water-filling power allocation with short term power constraint (equivalent to equal power allocation) and the optimal rate allocation, but without coding over space and time.

Next, we derive $d_{\mathrm{CSIT}_K}^*(r)$, for $K = 1, ..., n-1$. This is a direct extension of [118, Theorem 3]. Following the same arguments, we can show that $d_{\mathrm{CSIT}_K}^*(r)$ is a piecewise-linear function connecting the points $(0, mn)$, $(r_1(k), d_S^*(k))$, and $(K, 0)$, where

$$\begin{aligned} r_1(k) &= k - d_{k+1}\left(\sum_{j=1}^{k} 1/d_j\right) \\ d_S^*(k) &= (m-k)(n-k) \qquad \text{for} \quad k = 1, ..., K-1 \end{aligned} \tag{6.24}$$

with $d_k = (m-k+1)(n-k+1)$. In Fig. 6.4, we plot the optimal DMT curves of $d_{\mathrm{MB}}^*(r)$,

$d^*_{\mathrm{CSIT}_3}(r)$ and $d^*_{\mathrm{CSIR}}(r)$ with $M = N = 5$ and $K = 3$. The plot shows $d^*_{\mathrm{CSIT}_3}(r) \leq d^*_{\mathrm{MB}}(r)$ for $0 \leq r \leq 3$.

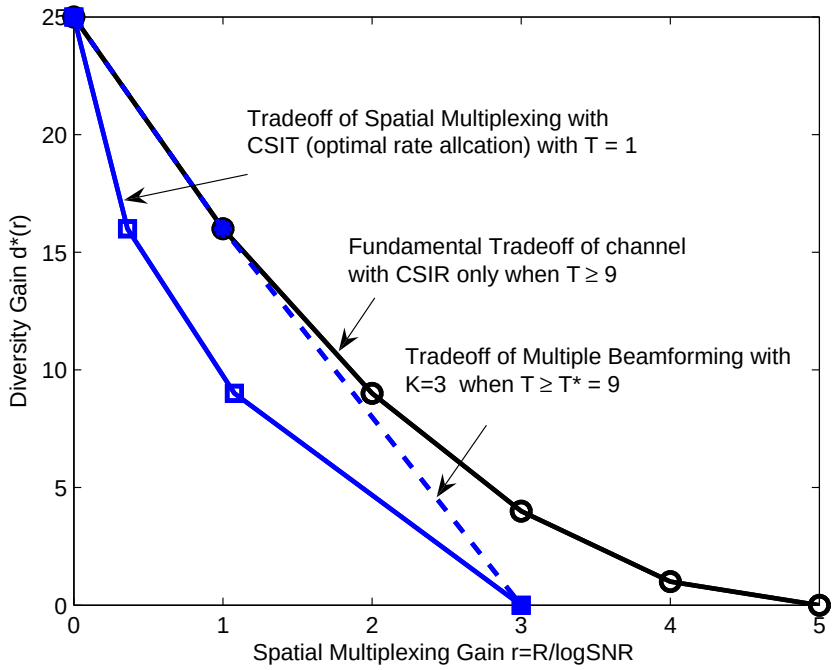

Figure 6.4: Diversity-multiplexing tradeoff: $d^*_{\mathrm{MB}}(r)$, $d^*_{\mathrm{CSIT}_3}(r)$ and $d^*_{\mathrm{CSIR}}(r)$ ($M = N = 5$ and $K = 3$).

In general, we have following result.

$$d^*_{\mathrm{CSIT}_K}(r) \leq d^*_{\mathrm{MB}}(r) \quad 0 \leq r \leq K, \tag{6.25}$$

for $K = 1, 2, ..., n$.

The proof is provided as follows:

Since the case of $K = n$ has been analyzed in the beginning of this section, we only

focus on $K = 1, ..., n-1$. In the proof of Theorem 12, we show that 1) The optimal DMT $d^*_{\mathrm{MB}}(r)$ is a piecewise-linear function connecting the points $(k, (m-k)(n-k))$, for $k = 0, 1, ..., n - i^*$, and the point $(K, 0)$ (See illustration in Fig. 6.5, where the point $C = (n-i^*, [m-(n-i^*)][n-(n-i^*)])$ and the point $B = (K, 0)$); 2) $d^*_{\mathrm{MB}}(r) = d^*_{\mathrm{CSIR}}(r)$, for $0 \leq r \leq n - i^*$.

When $k = 1, ..., n - i^*$, the performance degradation of $d^*_{\mathrm{CSIT}_K}(r)$ compared with optimal DMT curve $d^*_{\mathrm{MB}}(r)$ (equivalent to $d^*_{\mathrm{CSIR}}(r)$) can be characterized by multiplexing gain loss [118]

$$\Delta_r(k) = k - r_1(k) = d_{k+1}\left(\sum_{j=1}^{k} 1/d_j\right) > 0 \tag{6.26}$$

at the diversity gain $(m-k)(n-k)$. Thus, we have

$$d^*_{\mathrm{CSIT}_K}(r) \leq d^*_{\mathrm{MB}}(r) \qquad 0 \leq r \leq r_1(n-i^*) \tag{6.27}$$

and equality holds when $r = 0$.

Next, we need to show $d^*_{\mathrm{CSIT}_K}(r) \leq d^*_{\mathrm{MB}}(r)$ when $r_1(n-i^*) < r \leq K$. To facilitate the analysis, we first show that the function $d^*_{\mathrm{CSIT}_K}(r)$ with $r \in [0, K]$ is a convex function [2].

The curve of $d^*_{\mathrm{CSIT}_K}(r)$ when $0 \leq r \leq K$ has K line segments. The line represented by the kth segment from left to right, denoted as L_k, has a slope of (which can be derived

[2]Convex function [120, 3.1.1]: A function $f : R^n \to R$ is convex if $\boldsymbol{dom} f$ is a convex set and if for all $x, y \in \boldsymbol{dom} f$, and θ with $0 \leq \theta \leq 1$, we have $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$. Convex set [120, 2.14]: A set C is convex if the line segment between any two points in C lies in C, i.e., if for any $x_1, x_2 \in C$ and any θ with $0 \leq \theta \leq 1$, we have $\theta x_1 + (1-\theta)x_2 \in C$.

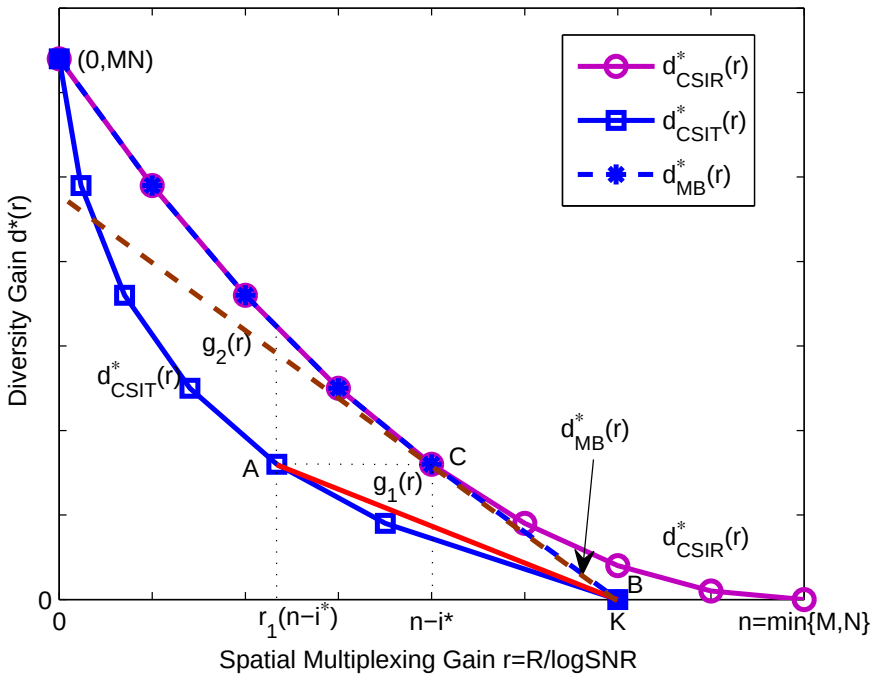

Figure 6.5: Diversity-multiplexing tradeoff: $d^*_{\mathrm{MB}}(r)$, $d^*_{\mathrm{CSIT}_K}(r)$ and $d^*_{\mathrm{CSIR}}(r)$.

using (6.24))

$$\rho_k = -\frac{1}{\sum_{j=1}^{k} 1/d_j} \qquad \text{and}$$

$$\rho_1 < \rho_2 < ... < \rho_K.$$

We can see $d^*_{\mathrm{CSIT}_K}(r) = \max\{L_1(r), ..., L_K(r)\}$ when $0 \le r \le K$. Since L_k is a convex function (a line is both convex and concave function), and it is well known that if $f_1, ..., f_m$ are convex, then their pointwise maximum $f(x) = \max\{f_1(x), ..., f_m(x)\}$ is also convex [120, 3.23], we readily have that $d^*_{\mathrm{CSIT}_K}(r)$ with $r \in [0, K]$ is a convex

function.

Define a linear function $g_1(r)$ representing the line segment connecting the points $A = (r_1(n-i^*), [m-(n-i^*)][n-(n-i^*)])$ and $B = (K, 0)$. See illustration in Fig. 6.5. By using the convex function definition, we have $d^*_{\mathrm{CSIT}_K}(r) \leq g_1(r)$ with $r_1(n-i^*) \leq r \leq K$. In order to show $d^*_{\mathrm{CSIT}_K}(r) \leq d^*_{\mathrm{MB}}(r)$ when $r_1(n-i^*) < r \leq K$, we only need to show $g_1(r) \leq d^*_{\mathrm{MB}}(r)$ when $r_1(n-i^*) < r \leq K$.

When $i^* = n$, $g_1(r) = d^*_{\mathrm{MB}}(r)$ with $0 \leq r \leq K$. Thus we have $g_1(r) \leq d^*_{\mathrm{MB}}(r)$ with $r_1(n-i^*) \leq r \leq K$.

When $i^* = n-K+1, ..., n-1$, by definition of i^* in (6.11), we have

$$\frac{1}{i^*-n+K}\sum_{i=n-K+1}^{i^*} c_i < \frac{1}{i^*+1-n+K}\sum_{i=n-K+1}^{i^*+1} c_i$$

which implies

$$\frac{1}{i^*-n+K}\sum_{i=n-K+1}^{i^*} c_i < c_{i^*+1}.$$

Using the arguments regarding the slope of each segment in $d^*_{\mathrm{MB}}(r)$ in the proof of Theorem 12, we can easily show that when $0 \leq r \leq n-i^*$, each segment of $d^*_{\mathrm{MB}}(r)$ has a slope smaller than $-\frac{1}{i^*-n+K}\sum_{i=n-K+1}^{i^*} c_i$ (i.e., the absolute value of the slope is greater than $\frac{1}{i^*-n+K}\sum_{i=n-K+1}^{i^*} c_i$), which is the slope of segment BC with points $B = (K, 0)$ and $C = (n-i^*, [m-(n-i^*)][n-(n-i^*)])$. Thus,

$$g_2(r) \leq d^*_{\mathrm{MB}}(r) \qquad 0 \leq r \leq K \tag{6.28}$$

where $g_2(r)$ is defined as a linear function representing the segment of BC and its extension between $r = 0$ and $r = K$. Since point A locates horizontally on the left side of

point C (due to multiplexing gain loss in (6.26)), we have

$$g_1(r) \leq g_2(r) \qquad r_1(n-i^*) \leq r \leq K. \tag{6.29}$$

Combining (6.28) and (6.29), we conclude that $g_1(r) \leq d^*_{\mathrm{MB}}(r)$ with $r_1(n-i^*) \leq r \leq K$. Thus

$$d^*_{\mathrm{CSIT}_K}(r) \leq d^*_{\mathrm{MB}}(r) \qquad r_1(n-i^*) \leq r \leq K. \tag{6.30}$$

Consider both (6.27) and (6.30), we have (6.25).

6.6 Conclusion

In this chapter, we have analyzed the optimal DMT of multiple beamforming in MIMO channels. We show that multiple beamforming with $T \geq T^*$ based on eigenvector information at the transmitter does not increase the diversity gain compared with the fundamental DMT of MIMO channels with CSIR and $T \geq M+N-1$. Compared with the optimal DMT of spatial multiplexing with full CSIT of the first K largest channel eigenmodes using the scheme without coding over space and time, however, the optimal DMT of multiple beamforming with $T \geq T^*$ is still better. Water-filling power allocation with short-term power constraint (i.e., water-filling over space but not over time) does not have any DMT advantage over equal power allocation if independent data streams are sent over channel eigenmodes in parallel with coding length of $T = 1$. Thus, space-time coding is important to explore the potentially achievable DMT.

Chapter 7 – Conclusion

7.1 Summary

In this thesis, we first study the problem of binary index assignment for beamforming codewords when the feedback channel is not error free. In a noisy feedback channel, feedback errors are inevitable in the feedback indices. They lead to incorrect beamforming vectors to be applied at the transmitter and thus degrade beamforming performance. An index-assignment algorithm that minimizes the impact of feedback errors are proposed. The proposed algorithm performs better than random index assignments. We also prove that when there exist feedback errors, for a beamforming system with M_t transmit antennas and M_r receive antennas, the diversity order is only M_r in an i.i.d. Rayleigh-fading channel.

Second, in the limited-feedback beamforming scheme, the receiver must determine the best codeword from the beamforming codebook and sends its index to the transmitter. Exhaustive codeword search for the large-size codebooks becomes impractical when the receiver has the limited computational power. The problem becomes worse for OFDM systems where multiple subcarriers or subcarrier groups have to select their own beamforming vectors simultaneously. We propose an ordering algorithm to reduce the number of codeword being searched with negligible performance loss, thus reducing codeword search complexity and search time.

Third, we compare angle feedback scheme and transmit antenna shuffling feedback scheme for double space-time transmit diversity (DSTTD) systems with four transmit

antennas and at least two receive antennas. Both schemes are simple and easy to implement. Transmit antenna shuffling feedback scheme for DSTTD systems has been adopted in IEEE 802.16e standard. It is of interest to have a performance comparison between these two schemes. First, we establish the equivalence between minimizing interference terms and minimizing MSE of linear MMSE or ZF receiver in angle feedback scheme. Then we show that angle rotation is sufficient to be applied on one transmit antenna, which can be chosen arbitrarily in advance. For transmit antenna shuffling scheme, we present a simplified general result about the selection of antenna shuffling matrices. We prove that 1-bit angle feedback scheme does not provide a better performance than the 1-bit transmit antenna shuffling feedback scheme. We also show by simulations that transmit antenna shuffling feedback is better than angle feedback with more than 1-bit feedback in an i.i.d. Rayleigh-fading channel.

Fourth, we consider training power allocation for a closed-loop MIMO system in i.i.d., Rayleigh flat-fading channels with power constraint. For a block fading model with block length T, if more intervals are dedicated to training, there will be fewer time slots left for data transmission; if more power is dedicated to training, there will be less power for data transmission. By maximizing the achievable rate, we first consider power allocation between the training phase and the data transmission phase when every fading block is assigned the same power so that only spatial power control is performed. In this case, we show that the optimal percentage of the power used for training (or data transmission) is the same as that in open-loop systems. We then consider power allocation between the training phase and the data transmission phase when the transmitter varies data power to adapt fading using both spatial and fading power control. We show that the optimal percentage of the average power used for training asymptotically converges to its counterpart in the first case where only spatial power control is performed.

Finally, we analyze the optimal DMT of multiple beamforming in MIMO channels. We show that multiple beamforming with $T \geq T^*$ based on eigenvector information at the transmitter does not increase the diversity gain compared with the fundamental DMT of MIMO channel with CSIR and $T \geq M + N - 1$. Compared with the optimal DMT of spatial multiplexing with full CSIT of the first K largest channel eigenmodes using the scheme without coding over channel eigenmodes (i.e., space) and time, however, the optimal DMT of multiple beamforming with $T \geq T^*$ provides a better diversity gain. This advantage is guaranteed by space-time coding.

7.2 Future work

In analyzing the problem of optimal power allocation between training and data transmission, we have assumed the symbol synchronization is ideal for both open-loop systems and closed-loop systems. This is not realistic in certain applications where each burst of the transmission packet must carry known symbols, known as a training sequence, for synchronization. Like training-based channel estimation, training-based synchronization has the same problem in training sequence design. In a block fading model with a block length of T symbol intervals and power constraint, if more symbol intervals are dedicated to synchronization, there will be fewer time slots for data transmission; if more power is dedicated to synchronization, there will be less power for data transmission. How to design the optimal training sequence for synchronization or how to jointly design training sequence for both synchronization and channel estimation in the sense of maximizing achievable rate is of great value.

Bibliography

[1] G. J. Foschini, Jr. and M. J. Gans,“ On limits of wireless communication in a fading environment when using multiple antennas,” *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[2] E. Telatar, “Capacity of multi-antenna Gaussian channels,”*Eur. Trans. Telecomm.*, vol. 10, no. 6, pp. 585-596, Nov. 1999.

[3] T. Marzetta and B. Hochwald,“Capacity of mobile multiple-antenna communication link in a Rayleigh flat-fading environment,” *IEEE Trans. Inform. Theory*, vol. 45, pp. 139-157, Jan. 1999.

[4] G. J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,” *Bell Labs Technical Journal*, pp. 41–59, Aug. 1996.

[5] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, “V- BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel,” in *Proc. URSI Int. Symp. Signals, Systems, and Electronics*, Pisa, Italy, pp. 295–300, Oct. 1998.

[6] B. Hochwald and T. Marzetta,“Unitary spacetime modulation for multiple- antenna communications in Rayleigh flat fading,” *IEEE Trans. Inform. Theory*, vol. 46, pp. 543-565, Mar. 2000.

[7] L. Zheng and D. Tse,“Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels,” *IEEE Trans. Inform. Theory*, vol. 49, pp. 1073-1096, MAY. 2003.

[8] W. C. Jakes, Jr., *Mobile Microwave Communication*. New York: Wiley, 1974.

[9] R. Heath, Jr. and A. Paulraj, “Switching between multiplexing and diversity based on constellation distance,” in *Proc. Allerton Conf. Communication, Control and Computing*, Oct. 2000.

[10] S. M. Alamouti, “A simple transmit diversity technique for wireless communications,” *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[11] V. Tarokh, N. Seshadri, and A. R. Calderbank,"Space-time codes for high data rate wireless communication: Performance criterion and code construction,"*IEEE Trans. Inform. Theory*, vol. 44, pp. 744-765, Mar. 1998.

[12] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456-1467, July 1999.

[13] H. E. Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels," *IEEE Trans. Inform. Theory*, vol. 50, pp. 968-985, June 2004.

[14] A. Goldsmith, *Wireless Communications*. NY: Cambridge University Press, 2005.

[15] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. UK: Cambridge University Press, 2003.

[16] H. Jafarkhani, " A quasi-orthogonal space-time block code," *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 1-4, Jan. 2001.

[17] N. Sharma and C. B. Papadias, "Improved quasi-orthogonal codes through constellation rotation,"*IEEE Trans. Commun.*, vol. 51, no. 3, pp. 332-335, Mar. 2003.

[18] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Combined array processing and space-time coding," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1121-1128, May 1999.

[19] Y. G. Li, J. H. Winters, and N. R. Sollenberger, "MIMO-OFDM for wireless communications: Signal detection with enhanced channel estimation," *IEEE Trans. Commun.*, vol. 50, no. 9, pp. 1471-1477, Sep. 2002.

[20] Titus K. Y. Lo,"Maximum ratio transmission," *IEEE Trans. Commun.* vol. 47, no. 10, pp. 1458-1461, Oct. 1999.

[21] E. Sengul, E. Akay, and E. Ayanoglu,"Diversity analysis of single and multiple beamforming," *IEEE Trans. Commun.*, vol. 54, no. 6, pp. 990-993, June. 2006.

[22] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders,"*IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1051-1064, May 2002.

[23] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198-2206, Dec. 2001.

[24] L. Collin, O. Berder, P. Rostaing, and G. Burel, "Optimal minimum distance-based precoder for MIMO spatial multiplexing systems," *IEEE Trans. Signal Process.*, vol. 52, no. 3, pp. 617-627, Mar. 2004.

[25] Z. Yan, K. M. Wong, and Z. Q. Luo, "Optimal diagonal precoder for multiantenna Communication Systems," *IEEE Trans. Signal Process.* vol. 53, no. 6, pp. 2089–2100, Jun. 2005.

[26] S. K. Jayaweera, and H. Vincent Poor, "Capacity of multiple-antenna systems with both receiver and transmitter channel state information," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2697-2709, Oct. 2003.

[27] A. Narula, M. J. Lopez, M. D. Trott, and G. W.Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 7, pp. 1423-1436, Oct. 1998.

[28] D. J. Love and R. W. Heath, "What is the value of limited feedback for MIMO channels," IEEE Commun. Mag., vol. 42, pp. 54-59, Oct. 2004.

[29] S. Zhou, Z.Wang, and G. B. Giannakis, "Performance analysis of transmit beamforming with finite-rate feedback," In *Proc. of 38th Conf. on Info. Sciences and Systems*, Princeton Univ., March 2004.

[30] B. Mondal and R. W. Heath, Jr.,"Performance analysis of quantized beamforming MIMO systems," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4753-4766, Dec. 2006.

[31] D. J. Love and R. W. Heath, "Limited feedback unitary precoding for orthogonal space-time block codes," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 64-73, Jan. 2005.

[32] D. J. Love and R. W. Heath, "Limited feedback unitary precoding for spatial multiplexing systems,"*IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967-2976, Aug. 2005.

[33] R.W. Heath Jr., S. Sandhu, and A. Paulraj, "Antenna selection for spatial multiplexing systems with linear receivers," *IEEE Commun. Lett.*, vol. 5, no. 4, pp. 142-144, Apr. 2001.

[34] V. Lau, Y. Liu, and T.-A. Chen, "On the design of MIMO blockfading channels with feedback-link capacity constraint,"*IEEE Trans. Commun.*, vol. 52, no. 1, pp. 62-70, Jan. 2004.

[35] W. Santipach and M. L. Honig, "Asymptotic performance of MIMO wireless channels with limited feedback," in *Proc. IEEE Mil. Comm. Conf.*, vol. 1, pp. 141146, Oct. 2003.

[36] D. A. Gore and A. J. Paulraj, "MIMO antenna subset selection with space-time coding," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2580-2588, Oct. 2002.

[37] H. Huang, G. Wu, and S. Li, "Optimized non-unitary linear precoding for orthogonal space-time block codes," *IEEE Commun. Lett.*, vol. 13, no. 6, pp. 414–416, Jun. 2009.

[38] M. Tao, Q. Li, and H. K. Garg, "Extended space-time block coding with transmit antenna selection over correlated fading channels," *IEEE Trans. Wireless Comm.*, vol. 6, no. 9, pp. 3137-3141, Sept. 2007.

[39] J. Kim and S. L. Ariyavisitakul, "Optimum 4-transmit-antenna STBC/SFBC with angle feedback and a near-optimum 1-bit feedback scheme," *IEEE Commun. Lett.*, vol. 11, no. 11, pp. 868870, Nov. 2007.

[40] J. Kim, S. L. Ariyavisitakul, and N. Seshadri, "STBC/SFBC for 4 transmit antennas with 1-bit feedback," in *Proc. ICC08, IEEE International Conference on Communications*, Beijing, China, pp. 3943-3947, May. 2008.

[41] J. Kim, "Interference suppression for high data rate STBC/SFBC with 1-bit feedback," *IEEE Commun. Lett.*, vol. 13, no. 1, pp. 13–15, Jan. 2009.

[42] P. Tan and N. C. Beaulieu, "Effect of channel estimation error on bit error probability in OFDM systems over Rayleigh and Ricean fading channels," *IEEE Trans. Commun.*, vol. 54, no. 6, pp. 675-685, Apr. 2008.

[43] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 686- 693, Nov. 1991.

[44] X. Tang, M. S. Alouini, and A. J. Goldsmith, "Effect of channel estimation error on M-QAM BER performance in Rayleigh fading," *IEEE Trans. Commun.*, vol. 47, pp. 1856-1864, Dec. 1999.

[45] P. Garg, R. K. Mallik, and H. M. Gupta,"Performance analysis of space-time coding with imperfect channel estimation," *IEEE Trans. Wireless Comm.*, vol. 4, no. 1, pp. 257-265, Jan. 2005.

[46] L. Cao and N. C. Beaulieu, "Exact error-rate analysis of diversity 16-QAM with channel estimation error," *IEEE Trans. Commun.*, vol. 52, pp. 1019-1029, June 2004.

[47] B. Xia and J. Wang, "Effect of channel-estimation error on QAM systems with antenna diversity," *IEEE Trans. Commun.*, vol. 53, pp. 481-488, Mar. 2005.

[48] R. Narasimhan, "Performance of diversity schemes for OFDM systems with frequency offset, phase noise, and channel estimation errors," *IEEE Trans. Commun.*, vol. 50, pp. 1561-1565, Oct. 2002.

[49] J. K. Tugnait and B. Huang, "Second-order statistics-based blind equalization of IIR single-input multiple-output channels with common zeros," *IEEE Trans. Signal Processing*, vol. 47, pp. 147157, July 1999.

[50] X. Zhuang, Z. Ding, and A. L. Swindlehurst, "A statistical subspace method for blind channel identification in OFDM communications," in *Proc. IEEE ICASSP*, Istanbul, Turkey, 2000.

[51] C. Tepedelenlioglu and G. B. Giannakis, "Transmitter redundancy for blind estimation and equalization of time- and frequency-selective channels," *IEEE Trans. Signal Processing*, vol. 48, pp. 2029-2043, July 2000.

[52] H. Blcskei, R.W. Heath, Jr., and A. J. Paulraj, "Blind channel identification and equalization in OFDM-based multi-antenna systems," *IEEE Trans. Signal Processing*, vol. 50, pp. 96-109, Jan. 2002.

[53] D. Reynolds, X. Wang, and H. V. Poor, "Blind adaptive space-time multiuser detection with multiple transmitter and receiver antennas," *IEEE Trans. Signal Processing*, vol. 50, pp. 1261-1276, Jun. 2002.

[54] C. Li and S. Roy, "Subspace-based blind channel estimation for OFDM by exploiting virtual carriers," *IEEE Trans. Wireless Commun.*, vol. 2, no. 1, pp. 141-150, Jan. 2003.

[55] A. Petropulu, R. Zhang, and R. Lin, "Blind OFDM channel estimation through simple linear precoding," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 647-655, Mar. 2004.

[56] S. Shahbazpanahi, A. B. Gershman, and J. H. Manton, "Closed-form blind MIMO channel estimation for orthogonal spacetime block codes," *IEEE Trans. Signal Process.* vol. 53, no. 12, Dec. 2005.

[57] E. Eidinger and A. Yeredor, "Blind MIMO identification using the second characteristic function," *IEEE Trans. Signal Process.* vol. 53, no. 11, pp. 4067–4079, Nov. 2005.

[58] G. Leus and M. Moonen, "Semi-blind channel estimation for block transmission with nonzero padding," in *Proc. Asilomar Conf. Signals, Systems, Computers*, vol. 1, Pacific Grove, CA, pp. 762-766, Nov. 2001.

[59] A. L. Swindlehurst and G. Leus, "Blind and semi-blind equalization for generalized spacetime block codes," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2489-2498, Oct. 2002.

[60] T. Cui, and C. Tellambura, "Semiblind channel estimation and data detection for OFDM systems with optimal pilot design," *IEEE Trans. Commun.*, vol. 55, no. 5, pp. 1053-1062, May 2007.

[61] M. Biguesh, and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.* vol. 54, no. 3, pp. 884–893, Mar. 2006.

[62] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consum. Electron.*, vol. 44, pp. 1112-1128, Aug. 1998.

[63] O. Rousseaux, G. Leus, P. Stoica, and M. Moonen, "Generalized training based channel identification," in *Proc. IEEE Global Conf. Communications*, San Francisco, CA, pp. 2432-2436, Dec. 2003.

[64] O. Rousseaux, G. Leus, P. Stoica, and M. Moonen, "Gaussian maximum likelihood channel estimation with short training sequences,"*IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2945-2955, Nov. 2005.

[65] P. Schniter, Low-complexity estimation of doubly-selective channels, in *Proc. Signal Processing Advances in Wireless Communications*, Rome, Italy, pp. 200-204, Jun. 2003.

[66] S. Ohno and G. B. Giannakis, "Capacity maximizing pilots for wireless OFDM over rapidly fading channels," in *Proc. Int. Symp. Signals, Syst., Electron.*, Tokyo, Japan, July 24-27, 2001, pp. 246-249.

[67] M. K. Tsatsanis and Z. Xu, "Pilot symbol assisted modulation in frequency selective fading wireless channels," *IEEE Trans. Signal Processing*, vol. 48, pp. 2353-2365, Aug. 2000.

[68] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Processing*, vol. 50, pp. 3055-3069, Dec. 2002.

[69] Y. G. Li, L. J. Cimini, and N. R. Sollegberger, "Robust channels estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 902-915, July 1998.

[70] Y. G. Li, N. Seshadri, and S. Ariyavisitakul, "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 461-471, Mar. 1999.

[71] Y. G. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, pp. 1207-1216, July 2000.

[72] Y. G. Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 67-75, Jan. 2002.

[73] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Processing*, vol. 52, pp. 546-557, Feb. 2004.

[74] M. C. Valenti and B. D. Woerner, "Iterative channel estimation and decoding of pilot assisted turbo codes over flat-fading channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 9, pp. 1697-1705, Sep. 2001.

[75] Q. Li, C. N. Georghiades, and X.Wang, "An iterative receiver for turbocoded pilot-assisted modulation in fading channels," *IEEE Commun. Lett.*, vol. 5, no. 4, pp. 145-147, Apr. 2001.

[76] M. Hsieh and C.Wei, "Channel estimation for OFDM systems based on comb-type pilot arrangement in frequency selective fading channels," *IEEE Trans. Consumer Electron.*, vol. 44, pp. 217-225, Feb. 1998.

[77] S. Coleri, M. Ergen, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Trans. Broadcast.*, vol. 48, pp. 223-229, Sept. 2002.

[78] C. Fragouli, N. Al-Dhahir, and W. Turin, " Training-based channel estimation for multile-antenna broadband transmissions," *IEEE Trans. on Wireless Commun.*, vol. 2, no. 2, pp. 384–391, Mar. 2003.

[79] D. J. Love, R. W. Heath Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[80] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple antenna systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2562-2579, Oct. 2003.

[81] P. Xia and G. B. Giannakis, "Design and analysis of transmit-beamforming based on limited-rate feedback," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1853–1863, May 2006.

[82] IEEE Std 802.16e-2005. "IEEE Standard for Local and Metropolitan Area Networks – Part 16, Amendment 2 and Corrigendum 1," New York, NY: IEEE, Feb. 2006.

[83] R. A. Horn and C. A. Johnson, *Matrix Analysis*. New York: Cambridge University Press, 1985.

[84] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, vol. 49, pp. 951–963, Apr. 2003.

[85] J. H. Conway, R. H. Hardin, and N. J. A. Sloane,"Packing lines, planes, etc.: Packings in Grassmannian spaces," *Experimental Math.*, vol. 5, no. 2, pp. 139-159, 1996.

[86] T. Strohmer and R. W. Heath Jr. "Grassmannian frames with applications to coding and communications, " *Applied and Compittational Harmonic Analysis*, 14(3): pp. 257–275, May 2003.

[87] M. A. Sustik, J. A. Tropp, I. Dhillon, and R. W. Heath, Jr., "On the existence of equiangular tight frames," *Linear Algebra and its Applications*. vol. 426, Issues 2-3, 15 October 2007, Pages 619–635.

[88] B. Mondal, R. Samanta, and R. W. Heath Jr., "Frame Theoretic Quantization for Limited Feedback MIMO Beamforming Systems," in *Proc. Int. Conf. Wireless Networks, Commun. and Mobile Comput.*, Maui, Hawaii, Jun. 2005, pp. 1065–1070.

[89] D. J. Love, R. W. Heath Jr., "Limited feedback diversity techniques for correlated channels," *IEEE Trans. Veh. Technol.*, vol. 55, no. 2, pp. 718–722, Mar. 2006.

[90] D. J. Love and R. W. Heath, Jr., "Necessary and sufficient conditions for full diversity order in correlated Rayleigh fading beamforming and combining systems," *IEEE Trans. Wireless Commun.*, vol. 4, no. 1, pp. 20-23, Jan. 2005.

[91] L. Liu and H. Jafarkhani, "Novel transmit beamforming schemes for time-selective fading multiantenna systems," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4767–4781, Dec. 2006.

[92] A. Ghaderipoor and C. Tellambura, "On the design, selection algorithm and performance analysis of limited feedback transmit beamforming," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4948–4957, Dec. 2008.

[93] S. Zhou and B. Li, "BER criterion and codebook construction for finite-rate precoded spatial multiplexing with linear receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1653–1665, May 2006.

[94] P. Xia, Collected Codebooks, Available on line : http//spincom.ece.umn.edu/pengfei/codebooks

[95] D. Love, Grassmanian subspace packing, Available on line : http//dynamo.ecn.purdue.edu/ djlove/grass.html

[96] T. Xu and H. Liu, " Index assignment for beamforming with limited-rate imperfect feedback," *IEEE Commun. Lett.*, vol. 11, no. 5, pp. 865–867, Nov. 2007.

[97] E. O. Onggosanusi, A. G. Dabak, and T. M. Schmidl, "High rate space-time block coded scheme: performance and improvement in correlated fading channels," in *Proc. IEEE WCNC'02*, vol. 1, Mar. 2002, pp. 194–199.

[98] S. Shim, K. Kim, and C. lee, "Efficient antenna shuffling scheme for a DSTTD system," *IEEE Commun. Lett.*, vol. 9, no. 2, pp 124–126, Feb. 2005.

[99] J. Joung, E. Jeong, and Y. H. Lee, "A computationally efficient criterion for antenna shuffling in DSTTD systems," *IEEE Commun. Lett.*, vol. 11, no. 9, pp. 732–734, Sep. 2007.

[100] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, pp. 1615–1624, Jun. 2003.

[101] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1158–1168, May 2006.

[102] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of training for frequency selective block-fading channels," *IEEE Trans. Inf. Theory*, vol. 48, pp. 2338–2353, Aug. 2002.

[103] S. Adireddy, L. Tong, "Optimal placement of known symbols for slowly varying frequency-selective channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1292–1296, Jul. 2005.

[104] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "On capacity of frequency-selective channels in training-based transmission schemes," *IEEE Trans. Signal Process.*, vol. 52, pp. 2572–2583, Sep. 2004.

[105] S. Ohno and G. B. Giannakis, "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2138–2145, Sep. 2004.

[106] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly selective wireless fading channels," *IEEE Trans. Signal Process.*, vol. 51, pp. 1531–1366, May 2003.

[107] I. Barhumi, G. Leus, and M. Moonen, "Time-varying FIR equalization of doubly selective channels,"*IEEE Trans. Wireless Commu.*, vol. 4, no. 1, pp. 202-214, Jan. 2005.

[108] G. Leus, I. Barhumi, and M. Moonen, "MMSE time-varying FIR equalization of doubly-selective channels," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICCASP 2003)*, Hong Kong, pp.485-488, Apr. 2003.

[109] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, pp. 933–946, May 2000.

[110] A. Lapidoth and S. Shamai, "Fading channels: How perfect need "perfect side information" be?," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1118–1134, May 2002.

[111] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, pp. 2203–2214, May 2006.

[112] A. Feiten, R. Mathar and S. Hanly, "Eigenvalue-based optimum-power allocation for Gaussian vector channels," *IEEE Trans. Inf. Theory*, vol. 53, pp. 2304–2309, Jun. 2007.

[113] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: Wiley, 1991

[114] J.W. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *J. of Multivariate Analysis*, vol. 55, pp. 331–339, 1995.

[115] P. A. Dighe, K. Mallik, and S. S. Jamuar,"Analysis of transmit-receive diversity in Rayleigh fading," *IEEE Trans. Commun*, vol. 51, no. 4, pp. 694–703, Apr. 2003.

[116] A. D. Dabbagh and D. J. Love, "Feedback rate-capacity loss tradeoff for limited feedback MIMO systems," *IEEE Trans. Inf. Theory*, vol. 52, pp. 2190–2202, May. 2006.

[117] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," *Ann. Math. Stat.*, vol.35, no. 2, pp. 475–501, Jun. 1964

[118] L. Garcia-Ordonez, A. Pages-Zamora, and J. R. Fonollosa,"Diversity and multiplexing tradeoff of spatial multiplexing MIMO systems with CSI," *IEEE Trans. Inform. Theory*, vol. 54, No. 7 pp. 2959-2975, Jul. 2008.

[119] A. Khoshnevis and A. Sabharwal, "On diversity and multiplexing gain of multiple antennas systems with transmitter channel information," in *Proc. Annu. Allerton Conf. Communication, Control, Computing*, Sept. 2004.

[120] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press 2004

APPENDICES

Appendix A – Proof of Lemma 3

Please note in transmit antenna shuffling scheme, the rotation factor $c_1 = c_2 = c_3 = c_4 = 1$. In the first step of the proof of the lemma, we claim that

$$\rho\mu - \eta = \sum_{m=1}^{M} (\rho_m \mu_m - \eta_m),$$

where $\rho_m = \sum_{i=1}^{2} \sum_{k=1}^{2} \left| h_{i\boldsymbol{\ell}_m(k)} \right|^2$, $\mu_m = \sum_{i=1}^{2} \sum_{k=1}^{2} \left| g_{i\boldsymbol{\ell}_m(k)} \right|^2$ and $\eta_m = |\delta_{1,m}|^2 + |\delta_{2,m}|^2$, with

$$\begin{aligned} \delta_{1,m} &= \sum_{k=1}^{2} (h^*_{1\boldsymbol{\ell}_m(k)} g_{1\boldsymbol{\ell}_m(k)} + h_{2\boldsymbol{\ell}_m(k)} g^*_{2\boldsymbol{\ell}_m(k)}), \\ \delta_{2,m} &= \sum_{k=1}^{2} (h^*_{1\boldsymbol{\ell}_m(k)} g_{2\boldsymbol{\ell}_m(k)} - h_{2\boldsymbol{\ell}_m(k)} g^*_{1\boldsymbol{\ell}_m(k)}). \end{aligned}$$

In this appendix, we give a detail proof of this claim using mathematical induction method:

a) when $N_r = 2$, there is only one combination set which is $\boldsymbol{\ell}_1 = [1\ 2]$, the claim can be verified easily.

b) Assume the claim is valid for $N_r = K$, where $K > 2$, we have

$$\begin{aligned}
\rho\mu - \eta &\triangleq \Xi_K \\
&= \Big(\sum_{i=1}^{2}\sum_{j=1}^{K}|h_{ij}|^2\Big)\Big(\sum_{i=1}^{2}\sum_{j=1}^{K}|g_{ij}|^2\Big) \\
&\quad -\Big|\sum_{j=1}^{K}h_{1j}^{*}g_{1j} + \sum_{j=1}^{K}h_{2j}g_{2j}^{*}\Big|^2 - \Big|\sum_{j=1}^{K}h_{1j}^{*}g_{2j} - \sum_{j=1}^{K}h_{2j}g_{1j}^{*}\Big|^2 \\
&= \sum_{m=1}^{M_K}(\rho_m\mu_m - \eta_m).
\end{aligned}$$

$M_K = \begin{pmatrix} K \\ 2 \end{pmatrix}$ combination sets are $\{[1\ 2], [1\ 3], ..., [1\ K], [2\ 3], [2\ 4], ..., [2\ K], ..., [(K-1)\ K]\}$ with two indices in each set.

By mathematical induction theory, we need to show for $N_r = K+1$ the claim is valid too under the assumption that the claim is valid when $N_r = K$. When $N_r = K+1$, we have

$$\begin{aligned}
\rho\mu - \eta &\triangleq \Xi_{K+1} \\
&= \Big(\sum_{i=1}^{2}\sum_{j=1}^{K+1}|h_{ij}|^2\Big)\Big(\sum_{i=1}^{2}\sum_{j=1}^{K+1}|g_{ij}|^2\Big) \\
&\quad -\Big|\sum_{j=1}^{K+1}h_{1j}^{*}g_{1j} + \sum_{j=1}^{K+1}h_{2j}g_{2j}^{*}\Big|^2 - \Big|\sum_{j=1}^{K+1}h_{1j}^{*}g_{2j} - \sum_{j=1}^{K+1}h_{2j}g_{1j}^{*}\Big|^2 \\
&= \Xi_K + \Delta.
\end{aligned}$$

where

$$\begin{aligned}\Delta &= \Big(\sum_{i=1}^{2}\sum_{j=1}^{K}|h_{ij}|^2\Big)\sum_{i=1}^{2}|g_{i(K+1)}|^2+\Big(\sum_{i=1}^{2}\sum_{j=1}^{K}|g_{ij}|^2\Big)\sum_{i=1}^{2}|h_{i(K+1)}|^2\\ &-\big(h_{1(K+1)}g^*_{1(K+1)}+h^*_{2(K+1)}g_{2(K+1)}\big)\sum_{j=1}^{K}\big(h^*_{1j}g_{1j}+h_{2j}g^*_{2j}\big)\\ &-\big(h^*_{1(K+1)}g_{1(K+1)}+h_{2(K+1)}g^*_{2(K+1)}\big)\sum_{j=1}^{K}\big(h_{1j}g^*_{1j}+h^*_{2j}g_{2j}\big)\\ &-\big(h_{1(K+1)}g^*_{2(K+1)}-h^*_{2(K+1)}g_{1(K+1)}\big)\sum_{j=1}^{K}\big(h^*_{1j}g_{2j}-h_{2j}g^*_{1j}\big)\\ &-\big(h^*_{1(K+1)}g_{2(K+1)}-h_{2(K+1)}g^*_{1(K+1)}\big)\sum_{j=1}^{K}\big(h_{1j}g^*_{2j}-h^*_{2j}g_{1j}\big)\end{aligned}$$

Please note when $N_r = K+1$, $M_{K+1} = \begin{pmatrix} K+1 \\ 2 \end{pmatrix}$ combination sets ($M_{K+1} = M_K + K$) contain all M_K sets in $N_r = K$ case and K new sets $[1\ (K+1)], [2\ (K+1)], ..., [K\ (K+1)]$. Without loss of generality, we let $\boldsymbol{\ell}_{M_K+l} = [l\ (K+1)]$ represent K new sets ($1 \le l \le K$) and let $\boldsymbol{\ell}_m$'s ($1 \le m \le M_K$) be the other M_K sets. We have

$$\sum_{m=M_K+1}^{M_K+K}(\rho_m\mu_m - \eta_m) = \sum_{m=M_K+1}^{M_K+K}\rho_m\mu_m - \sum_{m=M_K+1}^{M_K+K}\eta_m$$

Plug receive antenna indices contained in new sets into ρ_m, μ_m and η_m, we have

$$\begin{aligned}
\sum_{m=M_K+1}^{M_K+K} \rho_m\mu_m &= \Big(\sum_{i=1}^{2}\sum_{j=1}^{K}|h_{ij}|^2\Big)\sum_{i=1}^{2}|g_{i(K+1)}|^2 \\
&+\Big(\sum_{i=1}^{2}\sum_{j=1}^{K}|g_{ij}|^2\Big)\sum_{i=1}^{2}|h_{i(K+1)}|^2 \\
&+K\Big(\sum_{i=1}^{2}|h_{i(K+1)}|^2\Big)\sum_{i=1}^{2}|g_{i(K+1)}|^2 \\
&+\sum_{j=1}^{K}\Big((\sum_{i=1}^{2}|h_{ij}|^2)\sum_{i=1}^{2}|g_{ij}|^2\Big) \\
\sum_{m=M_K+1}^{M_K+K} \eta_m &= \big(h_{1(K+1)}g^*_{1(K+1)}+h^*_{2(K+1)}g_{2(K+1)}\big)\sum_{j=1}^{K}\big(h^*_{1j}g_{1j}+h_{2j}g^*_{2j}\big) \\
&+\big(h^*_{1(K+1)}g_{1(K+1)}+h_{2(K+1)}g^*_{2(K+1)}\big)\sum_{j=1}^{K}\big(h_{1j}g^*_{1j}+h^*_{2j}g_{2j}\big) \\
&+\big(h_{1(K+1)}g^*_{2(K+1)}-h^*_{2(K+1)}g_{1(K+1)}\big)\sum_{j=1}^{K}\big(h^*_{1j}g_{2j}-h_{2j}g^*_{1j}\big) \\
&+\big(h^*_{1(K+1)}g_{2(K+1)}-h_{2(K+1)}g^*_{1(K+1)}\big)\sum_{j=1}^{K}\big(h_{1j}g^*_{2j}-h^*_{2j}g_{1j}\big) \\
&+K\Big(\sum_{i=1}^{2}|h_{i(K+1)}|^2\Big)\sum_{i=1}^{2}|g_{i(K+1)}|^2 \\
&+\sum_{j=1}^{K}\Big((\sum_{i=1}^{2}|h_{ij}|^2)\sum_{i=1}^{2}|g_{ij}|^2\Big).
\end{aligned}$$

and we can verify

$$\sum_{m=M_K+1}^{M_K+K}(\rho_m\mu_m-\eta_m)=\Delta.$$

We complete the proof that when $N_r = K + 1$,

$$\begin{aligned}\rho\mu - \eta &= \Xi_K + \Delta = \sum_{m=1}^{M_K}(\rho_m\mu_m - \eta_m) + \sum_{m=M_K+1}^{M_{K+1}}(\rho_m\mu_m - \eta_m) \\ &= \sum_{m=1}^{M_{K+1}}(\rho_m\mu_m - \eta_m)\end{aligned}$$

In the second step of proof of the lemma, we claim that

$$\rho_m\mu_m - \eta_m = \Lambda_m - |\Gamma_m|^2$$

with

$$\begin{aligned}\Lambda_m &= \sum_{i=1}^{4}\sum_{j=1,j\neq i}^{4}|\hbar_{i\boldsymbol{\ell}_m(1)}\hbar_{j\boldsymbol{\ell}_m(2)}|^2 \\ &\quad -2\Re\left\{\sum_{i=1}^{4}\left(\hbar^*_{i\boldsymbol{\ell}_m(1)}\hbar_{i\boldsymbol{\ell}_m(2)}\sum_{j=i+1}^{4}\hbar_{j\boldsymbol{\ell}_m(1)}\hbar^*_{j\boldsymbol{\ell}_m(2)}\right)\right\}.\end{aligned}$$

We give a step by step derivation about this claim:

$$\begin{aligned}\rho_m\mu_m - \eta_m &= \Big(\sum_{i=1}^{2}\sum_{k=1}^{2}\left|h_{i\boldsymbol{\ell}_m(k)}\right|^2\Big)\Big(\sum_{i=1}^{2}\sum_{k=1}^{2}\left|g_{i\boldsymbol{\ell}_m(k)}\right|^2\Big) - \delta^*_{1,m}\delta_{1,m} - \delta^*_{2,m}\delta_{2,m} \\ &= A - B\end{aligned}$$

where

$$
\begin{aligned}
A &= |h_{1\boldsymbol{\ell}_{m(1)}}|^2|g_{1\boldsymbol{\ell}_{m(2)}}|^2 + |h_{1\boldsymbol{\ell}_{m(1)}}|^2|g_{2\boldsymbol{\ell}_{m(2)}}|^2 \\
&\quad + |h_{1\boldsymbol{\ell}_{m(2)}}|^2|g_{1\boldsymbol{\ell}_{m(1)}}|^2 + |h_{1\boldsymbol{\ell}_{m(2)}}|^2|g_{2\boldsymbol{\ell}_{m(1)}}|^2 \\
&\quad + |h_{2\boldsymbol{\ell}_{m(1)}}|^2|g_{1\boldsymbol{\ell}_{m(2)}}|^2 + |h_{2\boldsymbol{\ell}_{m(1)}}|^2|g_{2\boldsymbol{\ell}_{m(2)}}|^2 \\
&\quad + |h_{2\boldsymbol{\ell}_{m(2)}}|^2|g_{1\boldsymbol{\ell}_{m(1)}}|^2 + |h_{2\boldsymbol{\ell}_{m(2)}}|^2|g_{2\boldsymbol{\ell}_{m(1)}}|^2 \\
B &= 2\Re\Big\{ h^*_{1\boldsymbol{\ell}_{m(1)}} h_{1\boldsymbol{\ell}_{m(2)}} \left(g_{1\boldsymbol{\ell}_{m(1)}} g^*_{1\boldsymbol{\ell}_{m(2)}} + g_{2\boldsymbol{\ell}_{m(1)}} g^*_{2\boldsymbol{\ell}_{m(2)}} \right) \\
&\quad + h^*_{2\boldsymbol{\ell}_{m(1)}} h_{2\boldsymbol{\ell}_{m(2)}} \left(g_{1\boldsymbol{\ell}_{m(1)}} g^*_{1\boldsymbol{\ell}_{m(2)}} + g_{2\boldsymbol{\ell}_{m(1)}} g^*_{2\boldsymbol{\ell}_{m(2)}} \right) \\
&\quad + \left(h_{1\boldsymbol{\ell}_{m(1)}} h_{2\boldsymbol{\ell}_{m(2)}} - h_{1\boldsymbol{\ell}_{m(2)}} h_{2\boldsymbol{\ell}_{m(1)}} \right) \left(g^*_{1\boldsymbol{\ell}_{m(1)}} g^*_{2\boldsymbol{\ell}_{m(2)}} - g^*_{1\boldsymbol{\ell}_{m(2)}} g^*_{2\boldsymbol{\ell}_{m(1)}} \right) \Big\}
\end{aligned}
$$

Let

$$
\begin{aligned}
A' &= |h_{1\boldsymbol{\ell}_{m(1)}}|^2|h_{2\boldsymbol{\ell}_{m(2)}}|^2 + |h_{1\boldsymbol{\ell}_{m(2)}}|^2|h_{2\boldsymbol{\ell}_{m(1)}}|^2 \\
&\quad + |g_{1\boldsymbol{\ell}_{m(1)}}|^2|g_{2\boldsymbol{\ell}_{m(2)}}|^2 + |g_{1\boldsymbol{\ell}_{m(2)}}|^2|g_{2\boldsymbol{\ell}_{m(1)}}|^2 \\
B' &= 2\Re\Big\{ h^*_{1\boldsymbol{\ell}_{m(1)}} h_{1\boldsymbol{\ell}_{m(2)}} h_{2\boldsymbol{\ell}_{m(1)}} h^*_{2\boldsymbol{\ell}_{m(2)}} + g^*_{1\boldsymbol{\ell}_{m(1)}} g_{1\boldsymbol{\ell}_{m(2)}} g_{2\boldsymbol{\ell}_{m(1)}} g^*_{2\boldsymbol{\ell}_{m(2)}} \\
&\quad - \left(h_{1\boldsymbol{\ell}_{m(1)}} h_{2\boldsymbol{\ell}_{m(2)}} - h_{1\boldsymbol{\ell}_{m(2)}} h_{2\boldsymbol{\ell}_{m(1)}} \right) \left(g^*_{1\boldsymbol{\ell}_{m(1)}} g^*_{2\boldsymbol{\ell}_{m(2)}} - g^*_{1\boldsymbol{\ell}_{m(2)}} g^*_{2\boldsymbol{\ell}_{m(1)}} \right) \Big\}
\end{aligned}
$$

Use the fact that $|a|^2 + |b|^2 - 2\Re(a^*b) = |a-b|^2$, where a and b are complex numbers.

We have

$$\begin{aligned} A' - B' &= |h_{1\boldsymbol{\ell}_{m(1)}} h_{2\boldsymbol{\ell}_{m(2)}} - h_{1\boldsymbol{\ell}_{m(2)}} h_{2\boldsymbol{\ell}_{m(1)}}|^2 \\ &\quad + |g_{1\boldsymbol{\ell}_{m(1)}} g_{2\boldsymbol{\ell}_{m(2)}} - g_{1\boldsymbol{\ell}_{m(2)}} g_{2\boldsymbol{\ell}_{m(1)}}|^2 \\ &\quad + 2\Re\Big\{ \left(h_{1\boldsymbol{\ell}_{m(1)}} h_{2\boldsymbol{\ell}_{m(2)}} - h_{1\boldsymbol{\ell}_{m(2)}} h_{2\boldsymbol{\ell}_{m(1)}} \right) \\ &\quad \times \left(g^*_{1\boldsymbol{\ell}_{m(1)}} g^*_{2\boldsymbol{\ell}_{m(2)}} - g^*_{1\boldsymbol{\ell}_{m(2)}} g^*_{2\boldsymbol{\ell}_{m(1)}} \right) \Big\} \\ &= |\Gamma_m|^2 \end{aligned}$$

Note that

$$\begin{aligned} A + A' = \quad & |h_{1\boldsymbol{\ell}_{m(1)}}|^2 \left\{ |h_{2\boldsymbol{\ell}_{m(2)}}|^2 + |g_{1\boldsymbol{\ell}_{m(2)}}|^2 + |g_{2\boldsymbol{\ell}_{m(2)}}|^2 \right\} \\ & + |h_{2\boldsymbol{\ell}_{m(1)}}|^2 \left\{ |h_{1\boldsymbol{\ell}_{m(2)}}|^2 + |g_{1\boldsymbol{\ell}_{m(2)}}|^2 + |g_{2\boldsymbol{\ell}_{m(2)}}|^2 \right\} \\ & + |g_{1\boldsymbol{\ell}_{m(1)}}|^2 \left\{ |h_{1\boldsymbol{\ell}_{m(2)}}|^2 + |h_{2\boldsymbol{\ell}_{m(2)}}|^2 + |g_{2\boldsymbol{\ell}_{m(2)}}|^2 \right\} \\ & + |g_{2\boldsymbol{\ell}_{m(1)}}|^2 \left\{ |h_{1\boldsymbol{\ell}_{m(2)}}|^2 + |h_{2\boldsymbol{\ell}_{m(2)}}|^2 + |g_{1\boldsymbol{\ell}_{m(2)}}|^2 \right\} \end{aligned}$$

has a unique structure which shows $A + A'$ is the sum of four terms, each of which (in each line on the right-hand side of equation) represents the product of the channel power from one logical transmit antenna to the $\boldsymbol{\ell}_{m(1)}$th receive antenna and the sum of channel powers from the other three logical transmit antennas to the $\boldsymbol{\ell}_{m(2)}$th receive antenna. Due to symmetry, we have

$$A + A' = \sum_{i=1}^{4} \sum_{\substack{j=1 \\ j \neq i}}^{4} |\hbar_{i\boldsymbol{\ell}_{m(1)}} \hbar_{j\boldsymbol{\ell}_{m(2)}}|^2.$$

On the other hand,

$$\begin{aligned} B + B' &= 2\Re\Big\{ h^*_{1\boldsymbol{\ell}_{m(1)}} h_{1\boldsymbol{\ell}_{m(2)}} \left(h_{2\boldsymbol{\ell}_{m(1)}} h^*_{2\boldsymbol{\ell}_{m(2)}} + g_{1\boldsymbol{\ell}_{m(1)}} g^*_{1\boldsymbol{\ell}_{m(2)}} + g_{2\boldsymbol{\ell}_{m(1)}} g^*_{2\boldsymbol{\ell}_{m(2)}} \right) \\ &+ h^*_{2\boldsymbol{\ell}_{m(1)}} h_{2\boldsymbol{\ell}_{m(2)}} \left(g_{1\boldsymbol{\ell}_{m(1)}} g^*_{1\boldsymbol{\ell}_{m(2)}} + g_{2\boldsymbol{\ell}_{m(1)}} g^*_{2\boldsymbol{\ell}_{m(2)}} \right) \\ &+ g^*_{1\boldsymbol{\ell}_{m(1)}} g_{1\boldsymbol{\ell}_{m(2)}} g_{2\boldsymbol{\ell}_{m(1)}} g^*_{2\boldsymbol{\ell}_{m(2)}} \Big\} \\ &= 2\Re\Big\{ \sum_{i=1}^{4} \big(\hbar^*_{i\boldsymbol{\ell}_{m(1)}} \hbar_{i\boldsymbol{\ell}_{m(2)}} \sum_{j=i+1}^{4} \hbar_{j\boldsymbol{\ell}_{m(1)}} \hbar^*_{j\boldsymbol{\ell}_{m(2)}} \big) \Big\} \end{aligned}$$

The second equality comes from the fact that the value of $B + B'$ is constant no matter which permutation matrix is applied at transmitter. In fact, the right-hand side of first equality is nothing but the mutual products of six terms of $\hbar^*_{i\boldsymbol{\ell}_{m(1)}} \hbar_{i\boldsymbol{\ell}_{m(2)}}$'s, $(1 \le i \le 4)$, under any permutation matrices.

Now, we can show

$$\rho_m \mu_m - \eta_m = A - B = (A + A') - (B + B') - (A' - B') = \Lambda_m - (A' - B') = \Lambda_m - |\Gamma_m|^2$$

where Λ_m is independent of $\boldsymbol{W}$. Combining the above two steps, we have

$$\begin{aligned} \arg\max_{\boldsymbol{W} \in \mathrm{S}_{\boldsymbol{W}}} (\rho\mu - \eta) &= \arg\max_{\boldsymbol{W} \in \mathrm{S}_{\boldsymbol{W}}} \sum_{m=1}^{M} (\rho_m \mu_m - \eta_m) \\ &= \arg\max_{\boldsymbol{W} \in \mathrm{S}_{\boldsymbol{W}}} \sum_{m=1}^{M} (\Lambda_m - |\Gamma_m|^2) \\ &= \arg\min_{\boldsymbol{W} \in \mathrm{S}_{\boldsymbol{W}}} \sum_{m=1}^{M} |\Gamma_m|^2 . \end{aligned}$$

This completes the proof of Lemma 3.

Appendix B – Proof of Theorem 7

Clearly the entries of matrix $\bar{\mathbf{H}}^H$ are still i.i.d. Gaussian random variables with unit variance. The identity matrix $\boldsymbol{I}_M$ is a special case of an $M \times M$ random Hermitian non-negative matrix whose eigenvalue pdf is $\delta(x-1)$, where $\delta(\cdot)$ is the Dirac-delta function.

Based on Lemma 6, and consider $\mathbf{X}_M = \bar{\mathbf{H}}^H$ and $\boldsymbol{T}_M = \boldsymbol{I}_M$, the distribution of the eigenvalues of $(1/N)\bar{\mathbf{H}}^H\bar{\mathbf{H}}$ converges in distribution to a non-random pdf f_1 as $M \to \infty$. Let $K_1 = \min(M,N)$ and $K_2 = \max(M,N)$, $\bar{\mathbf{H}}^H\bar{\mathbf{H}}$ has at most K_1 non-zero eigenvalues. We scale $(1/N)\bar{\mathbf{H}}^H\bar{\mathbf{H}}$ by N/K_1. The distribution of the eigenvalues of $(1/K_1)\bar{\mathbf{H}}^H\bar{\mathbf{H}}$ still converges in distribution to a non-random pdf f_λ as $M \to \infty$ and we have $f_\lambda(x) = f_1(x \cdot K_1/N)$. This shows that the eigenvalue distributions for any channel realizations tend to be the same when the number of transmit and receive antennas approach infinity and their ratio is fixed. Furthermore, $f_\lambda(x)$ is expressed as [2]

$$f_\lambda(x) = \begin{cases} \frac{1}{2\pi}\sqrt{(\frac{x_+}{x}-1)(1-\frac{x_-}{x})} & \text{for } x \in [x_-, x_+]; \\ 0 & \text{otherwise} \end{cases}$$

with $x_\pm = \left(\sqrt{K_2/K_1} \pm 1\right)^2$.

For simplicity, whenever the condition that $K_1 \to \infty$ is applied in this proof, it means that both K_1 and K_2 approach infinity and their ratio is a constant. From (5.15) we

have

$$\begin{aligned} C_{\text{inst}}(\bar{\mathbf{H}}, \sigma_{\tilde{H}}^2, p) &= \sum_{k=1}^{K_1} [\log_2(\lambda_k \nu)]^+ \\ &= K_1 \left(\frac{1}{K_1} \sum_{k=1}^{K_1} \left[\log_2(\frac{\lambda_k}{K_1} \cdot K_1 \nu) \right]^+ \right). \end{aligned}$$

When $K_1 \to \infty$, $\frac{\lambda_k}{K_1} \to x \in [x_-, x_+]$ in eigenvalue spectrum with pdf $f_\lambda(x)$, $K_1 \nu \to \mu$, μ is the water level of the effective power poured in inverse eigenvalue spectrum. Thus, when $K_1 \to \infty$, we obtain

$$\begin{aligned} \frac{C_{\text{inst}}(\bar{\mathbf{H}}, \sigma_{\tilde{H}}^2, p)}{K_1} &= \frac{1}{K_1} \sum_{k=1}^{K_1} \left[\log_2(\frac{\lambda_k}{K_1} \cdot K_1 \nu) \right]^+ \\ &\to \int_{x_-}^{x_+} [\log_2(x\mu)]^+ f_\lambda(x) dx \end{aligned} \tag{B.1}$$

where μ must be finite and can be shown as follow. From (5.14), we have

$$\begin{aligned} p_{\text{eff}} &= \sum_{k=1}^{m} \left(\nu - \frac{1}{\lambda_k} \right)^+ \\ &= \frac{1}{K_1} \sum_{k=1}^{K_1} \left(K_1 \nu - \frac{K_1}{\lambda_k} \right)^+. \end{aligned} \tag{B.2}$$

When $K_1 \to \infty$, (B.2) becomes

$$\int_{x_-}^{x_+} (\mu - \frac{1}{x})^+ f_\lambda(x) dx = p_{\text{eff}} = \frac{\sigma_{\hat{H}}^2 p}{1 + \sigma_{\tilde{H}}^2 p}. \tag{B.3}$$

Since $\bar{P}/T > 0$, $P_p/M > P_p/T = (1-\alpha)\bar{P}/T > 0$ for $0 < \alpha < 1$. Using (5.6), we have $\sigma_{\tilde{H}}^2 < 1$ and $\sigma_{\hat{H}}^2 = 1 - \sigma_{\tilde{H}}^2 > 0$. When p is finite, μ must be finite.

From (B.1), we can see that for any channel realization $\bar{\mathbf{H}}$, the instantaneous capacity

lower bound normalized by K_1, when $K_1 \to \infty$, is determined by μ and $f_\lambda(x)$ only.

Let us define

$$\overrightarrow{C}_{\text{inst}} = \int_{x_-}^{x_+} [\log_2(x\mu)]^+ f_\lambda(x) dx. \tag{B.4}$$

We will compute $\frac{\partial \overrightarrow{C}_{\text{inst}}}{\partial p}$ to evaluate how $\overrightarrow{C}_{\text{inst}}$ changes as p changes. Note that

$$\frac{\partial \overrightarrow{C}_{\text{inst}}}{\partial p} = \frac{\partial \overrightarrow{C}_{\text{inst}}}{\partial \mu} \frac{\partial \mu}{\partial p_{\text{eff}}} \frac{\partial p_{\text{eff}}}{\partial p}. \tag{B.5}$$

From (B.4) and the first and the second equality of (B.3), we have

$$\begin{aligned}
\frac{\partial \overrightarrow{C}_{\text{inst}}}{\partial \mu} &= \frac{1}{\ln 2} \frac{1}{\mu} \int_{\frac{1}{\mu}}^{x_+} f_\lambda(x) dx, \\
\frac{\partial \mu}{\partial p_{\text{eff}}} &= \left(\int_{\frac{1}{\mu}}^{x_+} f_\lambda(x) dx \right)^{-1}, \\
\frac{\partial p_{\text{eff}}}{\partial p} &= \frac{\sigma_{\hat{H}}^2}{\left(1 + \sigma_{\hat{H}}^2 p\right)^2}.
\end{aligned}$$

Therefore, we obtain

$$\frac{\partial \overrightarrow{C}_{\text{inst}}}{\partial p} = \frac{1}{\ln 2} \frac{1}{\mu} \frac{\sigma_{\hat{H}}^2}{\left(1 + \sigma_{\hat{H}}^2 p\right)^2} > 0. \tag{B.6}$$

In order to maximize capacity, using Lagrange multipliers we need $\frac{\partial \overrightarrow{C}_{\text{inst}}}{\partial p} = \gamma_\infty$. γ_∞ is determined by the average data power $\bar{P}_d$. Because when $K_1 \to \infty$ the eigenvalue spectrum follows pdf $f_\lambda(x)$ which is independent of channel realization $\bar{\mathbf{H}}$, (B.3) shows that μ is a positive increasing function of p and is uniquely determined by p. Thus, (B.6) shows that $\overrightarrow{C}_{\text{inst}}$ is uniquely determined by p. It is easily to show in order to make $\frac{\partial \overrightarrow{C}_{\text{inst}}}{\partial p} = \gamma_\infty$ hold for any data transmission intervals, p should be the same at any data

transmission intervals. We readily have following results:

$$P_d(\bar{\mathbf{H}}) = \bar{P}_d.$$

Thus, when $K_1 \to \infty$, fading power control reduces to fixed power assignment. The optimal power allocation parameter $\alpha_{\mathrm{rf}} \to \alpha_{\mathrm{s}}$, and the following equalities almost surely hold:

$$P_d(\bar{\mathbf{H}}) = \bar{P}_d = \alpha_{\mathrm{sf}}\bar{P} = \alpha_{\mathrm{s}}\bar{P}$$

which proves Theorem 7.

Appendix C – Proof of Theorem 10

This proof follows the line of [7, proof of Theorem 4]. Starting from (6.8), we need to prove

$$F(\mathrm{SNR}) \stackrel{\Delta}{=} \int_{\mathcal{A}_2'} \left(\int_{\mathcal{A}_1'} \prod_{i=1}^{n} \mathrm{SNR}^{-(m-n+1)\alpha_i} \prod_{i<j} (\mathrm{SNR}^{-\alpha_i} - \mathrm{SNR}^{-\alpha_j})^2 d\boldsymbol{\alpha}_1 \right) d\boldsymbol{\alpha}_2 \doteq d_{\mathrm{out}}(r),$$

where

$$\begin{aligned} \mathcal{A}_2' &= \{\boldsymbol{\alpha}_2 \in R^{K+} | \alpha_{n-K+1} \geq ... \geq \alpha_n \geq 0 \text{ and } \sum_{i=n-K+1}^{n} (1-\alpha_i)^+ < r\} \\ \mathcal{A}_1' &= \{\boldsymbol{\alpha}_1 \in R^{(n-K)+} | \alpha_1 \geq ... \geq \alpha_{n-K} \geq \alpha_{n-K+1}\}. \end{aligned}$$

Since when $\alpha_i = \alpha_j$ the integral is zero, only distinct α_i's are considered in the integral. By definition $\alpha_i > \alpha_j$ for $i < j$ and $(\mathrm{SNR}^{-\alpha_i} - \mathrm{SNR}^{-\alpha_j})^2 \leq (0 - \mathrm{SNR}^{-\alpha_j})^2$ for $i < j$.

We have

$$
\begin{aligned}
F(\mathrm{SNR}) &\leq \bar{F}(\mathrm{SNR}) \\
&\triangleq \int_{\mathcal{A}_2'} \left(\int_{\mathcal{A}_1'} \prod_{i=1}^{n} \mathrm{SNR}^{-(m-n+1)\alpha_i} \prod_{i<j} (0 - \mathrm{SNR}^{-\alpha_j})^2 d\boldsymbol{\alpha}_1 \right) d\boldsymbol{\alpha}_2 \\
&= \int_{\mathcal{A}_2'} \left(\int_{\mathcal{A}_1'} \prod_{i=1}^{n} \mathrm{SNR}^{-(m-n+2i-1)\alpha_i} d\boldsymbol{\alpha}_1 \right) d\boldsymbol{\alpha}_2 \\
&= \int_{\mathcal{A}_2'} \left(\int_{\mathcal{A}_1'} \prod_{i=1}^{n-K} \mathrm{SNR}^{-(m-n+2i-1)\alpha_i} d\boldsymbol{\alpha}_1 \right) \prod_{i=n-K+1}^{n} \mathrm{SNR}^{-(m-n+2i-1)\alpha_i} d\boldsymbol{\alpha}_2 \\
&= C_1(m,n,K)(\ln \mathrm{SNR})^{(K-n)} \cdot \\
&\qquad \int_{\mathcal{A}_2'} \mathrm{SNR}^{-(n-K)(m-K)\alpha_{n-K+1}} \prod_{i=n-K+1}^{n} \mathrm{SNR}^{-(m-n+2i-1)\alpha_i} d\boldsymbol{\alpha}_2 \\
&= C_1(m,n,K)(\ln \mathrm{SNR})^{(K-n)} \cdot \\
&\qquad \int_{\mathcal{A}_2'} \mathrm{SNR}^{-(n-K+1)(m-K+1)\alpha_{n-K+1} + \sum_{i=n-K+2}^{n} -(m-n+2i-1)\alpha_i} d\boldsymbol{\alpha}_2 \\
&\doteq \int_{\mathcal{A}_2'} \mathrm{SNR}^{-(n-K+1)(m-K+1)\alpha_{n-K+1} - \sum_{i=n-K+2}^{n} (m-n+2i-1)\alpha_i} d\boldsymbol{\alpha}_2 \\
&= \int_{\mathcal{A}_2'} \mathrm{SNR}^{-f(\boldsymbol{\alpha}_2)} d\boldsymbol{\alpha}_2
\end{aligned}
$$

The first exponential equality uses the fact that

$$
\lim_{\mathrm{SNR}\to\infty} \frac{\log[C_1(m,n,K)(\ln \mathrm{SNR})^{(K-n)}]}{\log \mathrm{SNR}} = 0
$$

where $C_1(m,n,K) = \prod_{i=1}^{n-K} [i(m-n)+i^2]$ is a finite positive real coefficient, which is a function of m, n and K, and is independent of SNR. and

$$f(\boldsymbol{\alpha}_2) = (n-K+1)(m-K+1)\alpha_{n-K+1} + \sum_{i=n-K+2}^{n} (m-n+2i-1)\alpha_i$$

Denote $\boldsymbol{\alpha}_2^* = \arg\inf_{\mathcal{A}_2'} f(\boldsymbol{\alpha}_2)$. We can follow the exact same technique in [7] to prove that

$$\bar{F}(\mathrm{SNR}) \doteq \mathrm{SNR}^{-f(\boldsymbol{\alpha}_2^*)}$$

by replacing $\mathcal{A}'$ and $I = [0, mn]^n$ in [7, eq. (45)] with $\mathcal{A}_2'$ and $I = [0, mn]^K$, respectively. We have

$$F(\mathrm{SNR}) \dot{\leq} \mathrm{SNR}^{-f(\boldsymbol{\alpha}_2^*)} \tag{C.1}$$

In order to finish the proof, we develop a lower bound on $F(\mathrm{SNR})$. For any $\delta > 0$, define the set

$$S_\delta = \{\boldsymbol{\alpha} : \alpha_i - \alpha_j \geq \delta, \forall\ i < j\}$$

$$
\begin{aligned}
F(\mathrm{SNR}) &= \int_{\mathcal{A}_2'} \left(\int_{\mathcal{A}_1'} \prod_{i=1}^{n} \mathrm{SNR}^{-(m-n+1)\alpha_i} \prod_{i<j} (\mathrm{SNR}^{-\alpha_i} - \mathrm{SNR}^{-\alpha_j})^2 d\boldsymbol{\alpha}_1 \right) d\boldsymbol{\alpha}_2 \\
&\geq \int_{\mathcal{A}_2' \cap S_\delta} \left(\int_{\mathcal{A}_1' \cap S_\delta} \prod_{i=1}^{n} \mathrm{SNR}^{-(m-n+1)\alpha_i} \prod_{i<j} (\mathrm{SNR}^{-\alpha_i} - \mathrm{SNR}^{-\alpha_j})^2 d\boldsymbol{\alpha}_1 \right) d\boldsymbol{\alpha}_2 \\
&\geq \int_{\mathcal{A}_2' \cap S_\delta} \left(\int_{\mathcal{A}_1' \cap S_\delta} \prod_{i=1}^{n} \mathrm{SNR}^{-(m-n+1)\alpha_i} \prod_{i<j} [(1 - \mathrm{SNR}^{-\delta})\mathrm{SNR}^{-\alpha_j}]^2 d\boldsymbol{\alpha}_1 \right) d\boldsymbol{\alpha}_2 \\
&= (1 - \mathrm{SNR}^{-\delta})^{n(n-1)} \int_{\mathcal{A}_2' \cap S_\delta} \left(\int_{\alpha_{n-K+1}+\delta}^{\infty} \cdots \int_{\alpha_2+\delta}^{\infty} \right. \\
&\qquad \left. \prod_{i=1}^{n} \mathrm{SNR}^{-(m-n+2i-1)\alpha_i} d\alpha_1 \ldots d\alpha_{n-K} \right) d\boldsymbol{\alpha}_2 \\
&= (1 - \mathrm{SNR}^{-\delta})^{n(n-1)} C_1(m,n,K) (\ln \mathrm{SNR})^{(K-n)} \\
&\qquad \cdot \int_{\mathcal{A}_2' \cap S_\delta} \mathrm{SNR}^{-f(\boldsymbol{\alpha}_2) - C_2(m,n,K)\delta} d\boldsymbol{\alpha}_2 \\
&\doteq (1 - \mathrm{SNR}^{-\delta})^{n(n-1)} \int_{\mathcal{A}_2' \cap S_\delta} \mathrm{SNR}^{-f(\boldsymbol{\alpha}_2) - C_2(m,n,K)\delta} d\boldsymbol{\alpha}_2
\end{aligned}
$$

where $C_2(m,n,K) = \sum_{i=1}^{n-K} [i(m-n) + i^2]$. Given $\delta > 0$, the last equality has the dominated SNR exponent

$$\inf_{\boldsymbol{\alpha}_2 \in \mathcal{A}_2' \cap S_\delta} f(\boldsymbol{\alpha}_2) + C_2(m,n,K)\delta$$

which approaches $f(\boldsymbol{\alpha}_2^*)$ when $\delta \to 0$, since $f(\boldsymbol{\alpha}_2)$ is a continuous function. Thus we have

$$F(\mathrm{SNR}) \dot{\geq} \mathrm{SNR}^{-f(\boldsymbol{\alpha}_2^*)}. \tag{C.2}$$

Combining (C.1) and (C.2), we finish the proof.