Modulation Division for Multiuser Wireless Communication Networks

MODULATION DIVISION FOR MULTIUSER WIRELESS COMMUNICATION NETWORKS

BY

ZHENG DONG, B.Sc., (Electronic Information Science and Technology)

M. Eng., (Communication and Information Systems)

Shandong University, Jinan, China

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Doctor of Philosophy (2016) McMaster University

(Electrical & Computer Engineering) Hamilton, Ontario, Canada

TITLE:	Modulation Division for Multiuser Wireless Communication Networks
AUTHOR:	Zheng Dong B.Sc., (Electronic Information Science and Technology) M. Eng., (Communication and Information Systems) Shandong University, Jinan, China
SUPERVISOR:	Dr. Jian-Kang Zhang
NUMBER OF PAGES:	xxi, 228

Dedicated to my parents for their endless love and support.

Abstract

This thesis considers the modulation division based on the concept of uniquely factorable constellation pair (UFCP) and uniquely decodable constellation group (UDCG) in multiuser wireless communication networks.

We first consider a two-hop relay network consisting of two single-antenna users and a two-antenna relay node, for which a novel distributed concatenated Alamouti code is devised. This new design allows the source and relay nodes to transmit their own information to the destination node concurrently at the symbol level with the aid of the UFCP generated from both PSK and square QAM constellations as well as by jointly processing the noisy signals received at the relay node. Moreover, an asymptotic symbol error probability (SEP) formula is derived for the ML receiver, showing that the maximum diversity gain function is achieved, which is proportional to $\ln \mathtt{SNR}/\mathtt{SNR}^2$.

Then, we concentrate on the point-to-point correlated multiple-input and multiple-output (MIMO) communication systems where full knowledge of channel state information (CSI) is available at the receiver and only the first- and second-order statistics of the channels are available at the transmitter. When the number of antenna elements of both ends goes to infinity while keeping their ratio constant, the asymptotic SEP analysis is carried out for either optimally precoded or uniformly

precoded correlated large MIMO fading channels using the zero-forcing (ZF) detector with equally likely PAM, PSK or square QAM constellations. For such systems, we reveal some very nice structures which inspire us to explore two very useful mathematical tools (i.e., the Szegö's theorem on large Hermitian Toeplitz matrices and the well-known limit: $\lim_{x\to\infty}(1+1/x)^x = e$), for the systematic study of asymptotic behaviors on their error performance. This new approach enables us to attain a very simple expression for the SEP limit as the number of the available antenna elements goes to infinity. In what follows, the problem of precoder design using a zero-forcing decision-feedback (ZF-DF) detector is also addressed. For such a MIMO system, our principal goal is to efficiently design an optimal precoder that minimizes the asymptotic SEP of the ZF-DF detector under a perfect decision feedback. By fully taking advantage of the product majorization relationship among eigenvalues, singular-values and Cholesky values of the precoded channel matrix parameters, a necessary condition for the optimal solution to satisfy is first developed and then the structure of the optimal solution is characterized. With these results, the original non-convex problem is reformulated into a convex one that can be efficiently solved by using an interior-point method. In addition, by scaling up the antenna array size of both terminals without bound for such a network, we propose a novel method as we did for the ZF receiver scenario to analyze the asymptotic SEP performance of an equal-diagonal QRS precoded large MIMO system when employing an abstract Toeplitz correlation model for the transmitter antenna array. This new approach has a simple expression with a fast convergence rate and thus, is efficient and effective for error performance evaluation.

For multiuser communication networks, we first consider a discrete-time multiple-input single-output (MISO) Gaussian broadcast channel (BC) where perfect CSI is available at both the transmitter and all the receivers. We propose a flexible and explicit design of a uniquely decomposable constellation group (UDCG) based on PAM and rectangular QAM constellations. With this new concept, a modulation division (MD) transmission scheme is developed for the considered MISO BC. The proposed MD scheme enables each receiver to uniquely and efficiently recover their desired signals from the superposition of mutually interfering cochannel signals in the absence of noise. Using max-min fairness as a design criterion, the optimal transmitter beamforming problem is solved in a closed-form for two-user MISO BC. Then, for a general case with more than two receivers, a user-grouping based beamforming scheme is developed, where the grouping method, beamforming vector design and power allocation problems are addressed by employing weighted max-min fairness.

Then, we consider an uplink massive single-input and multiple-output (SIMO) network consisting of a base station (BS) and several single-antenna users. To recover the transmitted signal matrix of all the users when the antenna array size is large, a novel multi-user space-time modulation (MUSTM) scheme is proposed for the considered network based on the explicit construction of QAM uniquely-decomposable constellation groups (QAM-UDCGs). In addition, we also develop a sub-constellation allocation method at the transmitter side to ensure the signal matrix is always invertible. In the meanwhile, an efficient training correlation receiver (TCR) is proposed which calculates the correlation between the received sum training signal vector and the sum information carrying vector. Moreover, the optimal power allocation problems are addressed by maximizing the coding gain or minimizing the average SEP

of the received sum signal under both average and peak power constraints on each user. The proposed transmission scheme not only allows the transmitted signals with strong mutual interference to be decoded by a simple TCR but it also enables the CSI of all the users to be estimated within a minimum number of time slots equal to that of the users.

Comprehensive computer simulations are carried out to verify the effectiveness of the proposed uniquely decomposable space-time modulation method in various network topologies and configurations. Our modulation division method will be one of the promising technologies for the fifth generation (5G) communication systems.

Acknowledgements

I am grateful to work with all those extraordinary people during the past four years.

First and foremost, I would like to show my deepest gratitude to my supervisor Dr. Jian-Kang Zhang, a respectable, responsible and earnest scholar who has influenced me in many aspects of my life, not only for his invaluable guidance in every stage of my research but also for his encouragement, impressive patience and genuine concern for students. Since the first time I met him, Dr. Zhang has introduced me so many things on how to think critically, to work in earnest and to appreciate the beauty of good ideas. I feel extremely fortunate and proud of being his student and I will try my best to pass his generous help down to other people in my future life.

I am also deeply indebted to Dr. Jun Chen for his knowledge, continuous encouragement, support and stimulating discussions. Dr. Chen is alway willing to help and acts as both a great mentor and a close friend. I would also like to thank Dr. Sorina Dumitrescu for her invaluable guidance and inspiration. Besides, my sincere gratitude goes to Dr. Kon Max Wong who has taught me so much both for the academic competence and how to do research. My sincere appreciation also goes to all my friends in McMaster University, who have made my life a story worth telling.

Most importantly, I would like to thank my parents, Shuqing Dong and Shiying Zheng, who have provided me the best they can, for their endless love and support.

Notations and Abbreviations

e.g., $\mathbf{a}$	column vectors are denoted by lowercase boldface characters
e.g., $\mathbf{A}$	matrices are denoted by uppercase boldface characters
$\mathbf{A}^T$	transpose of matrix $\mathbf{A}$
$\mathbf{A}^*$	complex conjugate of matrix $\mathbf{A}$
$\mathbf{A}^H$	Hermitian (conjugate transpose) of matrix $\mathbf{A}$
$\mathrm{tr}(\mathbf{A})$	trace of square matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	inverse of square matrix $\mathbf{A}$
$[\mathbf{A}]_{k,\ell}$ or $a_{k,\ell}$	the (k,ℓ)-th element of matrix $\mathbf{A}$
$[\mathbf{A}]_k$	the k-th diagonal entry of matrix $\mathbf{A}$
$\det(\mathbf{A})$	determinant of matrix $\mathbf{A}$
$\mathrm{diag}(\cdot)$	diagonal matrix
$\|\mathbf{b}\|$	the Euclidean norm of $\mathbf{b}$
$\|\mathbf{A}\|_F$	Frobenius norm of matrix $\mathbf{A}$
$\mathbf{I}_N$	$N \times N$ identity matrix
$\mathbf{A} \preceq \mathbf{B}$	$\mathbf{A}$ and $\mathbf{B}$ are positive semidefinite, $\mathbf{B} - \mathbf{A}$ is also positive semidefinite
$\otimes$	Kronecker product
$\mathbb{E}[\cdot]$	expectation operator
$\mathrm{vec}(\cdot)$	the operator stacking the columns of a matrix on top of one another

$\approx$	approximately equal to
$\Re(\cdot)$	the real part of a vector or a matrix
$\Im(\cdot)$	the imaginary part of a vector or a matrix
$\varnothing$	empty set
$\mathbb{Z}$	ring of integers
$\mathbb{C}$	field of complex numbers
ln	natural logarithm
j	$\sqrt{-1}$
$\binom{n}{k} = \frac{n!}{k!(n-k)!}$	the binomial coefficient
$\lfloor a \rfloor$	the floor function which represents the largest integer not larger than a
$\lceil b \rceil$	the ceiling function that returns the smallest integer not smaller than b
$\gcd(a, b)$	the greatest common divisor of a and b
$a \vert b$	a divides b
$a \nmid b$	a does not divide b
$a \equiv b \mod m$	$m \vert (a - b)$
$f(x) = o(g(x))$	$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$
$f(x) = O(g(x))$	There exists a positive real number M and a real number x_0, such that $\vert f(x) \vert \leq M \vert g(x) \vert$ for all $x \geq x_0$

AWGN	additive white Gaussian noise
BC	broadcast channel
BER	bit error rate
BS	base station
CLT	central limit theorem
CSCG	circularly symmetric complex Gaussian
CSI	channel state information
CSIR	CSI at the receiver
CSIT	CSI at the transmitter
DFE	decision-feedback equalization
DFT	discrete Fourier transform
DPC	dirty paper coding
DPSK	differential phase shift keying
LoS	line of sight
MD	modulation division
MIMO	multiple-input and multiple-output
MISO	multiple-input and single-output
ML	maximum likelihood
MMSE	minimum mean square error
MRC	maximum-ratio combining
MUSTM	multi-user space-time modulation
NOMA	non-orthogonal multiple access
PAM	pulse amplitude modulation
PEP	pairwise error probability

PSD	power spectral density
PSK	phase-shift keying
QAM	quadrature amplitude modulation
QoS	quality of service
SEP	symbol error probability
SIMO	single-input and multiple-output
SINR	signal-to-interference-plus-noise ratio
SLNR	signal-to-leakage-and-noise ratio
SNR	signal-to-noise ratio
STBC	space-time block code
TCR	training correlation receiver
TD	time division
UDCG	uniquely decomposable constellation group
UFCP	uniquely factorable constellation pair
ZF	zero-forcing

Contents

List of Figures

Chapter 1

Introduction

In the past decade, we have witnessed tremendous advances and a shifted research paradigm in physical-layer technology and theory of wireless communications. The information theoretical research has brought about new perspectives on communication system design (e.g., the effect of channel fading, the tradeoff between energy efficiency and spectral efficiency, the method of interference management), and meanwhile, the rapid evolution of wireless communication systems driven by breakthrough in radio frequency and computer science removes many obstacles of the original technologies (e.g., the emergence of machine-to-machine communication, millimeter wave communication, massive multiple-input and multiple-output (MIMO)). However, there are still many technical challenges which must be addressed to meet the ever-increasing rate requirement and network connectivity demand. Among all the promising technologies, multi-antenna technology and ad-hoc network receive intensive attention in the past decade. In this thesis, we concentrate on the uniquely decomposable space-time modulation and its application in multi-hop relay networks and single-hop multi-antenna networks. Based on the difference in modulation division method

and system model, this thesis can be basically divided into two main parts.

The first part of this thesis considers the modulation division by employing the uniquely factorable constellation pair (UFCP) in two-hop relay networks. The wireless ad-hoc network that does not rely on the existing infrastructure is highly appealing for its robustness to node failure and flexibility in network topology. As signal processing tends to be performed by network nodes in a distributed manner, making energy efficiency and computational complexity challenging design constraints. In an ad-hoc networks, the communication between different nodes typically takes place over several hops by adopting amplify-and-forward, compress-and-forward or decode-and-forward protocols in a half duplex mode. In this thesis, a new relay protocol based on the concept of UFCP is proposed that allows the source and relay nodes to transmit their individual messages to the destination node concurrently at the symbol level. This new method has a low average delay compared to the conventional relay strategy where different users are scheduled in the packet level.

The second part of this thesis focuses on the modulation division by using the uniquely decomposable constellation group (UDCG) and signal processing technologies in modern multi-antenna systems where the additional spatial dimension enables much higher data rates and reliability in fading channels. We will consider the precoder design problem when employing both zero-forcing (ZF) and zero-forcing decision-feedback (ZF-DF) receivers in the point-to-point correlated fading MIMO channels where full channel state information (CSI) is known to the receiver. In particular, new mathematical tools are developed for the systematical analysis of the symbol error probability (SEP) when the number of available antennas is unbounded. Then, for the multiuser multiple-input and single-output (MISO) broadcast channel

in the downlink with full CSI at both the transmitter and all the receivers, a modulation division (MD) transmission scheme is developed by the flexible and explicit design of a UDCG, which outperforms the ZF method significantly when the number of users is very close to that of transmitter antennas. The last chapter deals with the multiuser massive single-input and multiple-output (SIMO) channel in the uplink with no CSI known to both the transmitters and the base station (BS) where a QAM division method is proposed that allows all the users to transmit their messages to the BS concurrently at the same frequency band. In our design, only one time slot is needed for the training process and hence the pilot contamination can be greatly mitigated or even be eliminated.

In the following part of this chapter, comprehensive introduction to the design motivation, system model and technical contributions of the thesis is given as follows.

1.1 UFCG in Relay Networks

Cooperative diversity is widely recognized as a promising tool which can be used for enhancing network connectivity and improving reliability beyond physical size and complexity constraint of wireless communication terminals [1–11]. The in-cell mobile users are allowed to share the use of their antennas to form a virtual array through collaborative transmission and distributed signal processing. Therefore, the existing diversity techniques for the conventional MIMO systems have been modified and generalized to such relay networks for the design of the distributed space-time block coding (STBC) [7, 9, 12–16]. In practical wireless relay network communication systems, it is desirable to allow relay nodes to send their own information to terminal nodes (e.g., transmitting CSI or control signaling) which helps to build the network

topology or to manage connectivity. However, the relay nodes in currently-available relay networks with distributed STBCs just passively forward whatever they have received from the source node to the destination without being able to transmit their own information. The conventional strategy to realize information sharing is accomplished by allocating orthogonal subchannels to relay such as time slots or frequency bands, operating virtually at a packet level and this leads to significant delay. Despite the fact that this method may work well for some networks with long coherence time, it will not be applicable to certain strictly-constrained relay systems. Motivated by all the aforementioned facts, in Chapter 3 of this thesis, we consider a one-way relay network consisting of two single-antenna terminals and one relay node having two antennas. Our main objective is to design a new distributed STBC for the network that allows the source node and relay node to transmit information simultaneously at the symbol level. The primary idea of achieving this goal is to properly make use of the recently developed concept of UFCP [17, 18] defined in Chapter 2 generated from both square-QAM and PSK constellations, as well as the Alamouti coding scheme at the relay node. In particular, it should be mentioned that the unique factorization based on PSK constellation is closely related to that of coprime PSK constellations, which was originally proposed in [19–21] for the design of full diversity noncoherent STBC. Our code design is also closely related to those in [22, 23]. Two significant advantages of this novel code design will be revealed: (a) The optimal ML detector is equivalent to a symbol-by-symbol detector; (b) The maximum diversity gain function is achieved.

1.2 Point-to-Point MIMO with ZF and ZF-DF Receivers

Multiple antenna systems have been proved to be a very promising technology in wireless communications, where the channel capacity and reliability can be substantially improved by exploiting multi-path scattering from antenna arrays equipped at both the transmitter and the receiver [24–26]. With recent advances in radio frequency (RF) chains and integrated circuit designs, MIMO systems with a large number of antennas (commonly known as Large/Massive MIMO systems) have emerged as an important breakthrough, drawing tremendous attention from both academia and industry [27–31]. In recent years, information-theoretic research has shown that many significant advantages are brought about in the large MIMO architectures by arguably unbounded number (typically tens or hundreds) of available antennas, such as even higher data rate, increased diversity, and better energy efficiency than conventional MIMO systems [28, 31–36].

Unfortunately, several significant issues arise meanwhile in the large MIMO fading channels. Thus far, it has been well understood that channel correlation between neighboring antennas, which is mainly determined by antenna spacing and the location of scatterers, is a major factor to affect information rate and reliability in a general MIMO system [37]. However, the effect of fading correlation on the large MIMO systems still needs to be better understood, in which, to accommodate a large number of antennas within affordable size and cost, fading correlation is virtually inevitable. In Chapter 4, to show the effect of fading correlation explicitly, we specifically consider an exponential correlation model [33, 38]. This is a simplified one-parameter

model in practical environment which can capture the main phenomenon of spatial correlation between antennas, especially for a uniform linear array. Another important issue arising in the large MIMO fading channels is that the drastically increased detection complexity restricts the widespread deployment of a large MIMO array in realistic application. Despite the fact that the maximum likelihood (ML) detector is universally optimal, its complexity is prohibitively high even for a moderate size of linear array. The ML receiver for the large MIMO systems is arguably impractical. Therefore, for the sake of practicality, in Chapter 4 we consider the use of the simple linear ZF detector [39] at the receiver side.

In general, for a MIMO system in which the first- and second-order statistics are completely available at the transmitter, it is known that linear precoding is a very efficient and effective scheme to significantly alleviate performance loss suffering from the channel fading correlation [40–43]. Presently, for the large MIMO channels, the most existing works mainly concentrate on the information-theoretical capacity analysis of un-precoded transmission models [33, 35] (generally assuming a uniform power allocation strategy). In Chapter 4, we are interested in the systematic analysis on asymptotic error performance for either optimally precoded or uniformly precoded large correlated MIMO fading channels using the ZF detector. Our primary goal is to attain a simple expression with a very fast convergence rate for the SEP limit for PAM, PSK and square QAM constellations when the number of the transmitter antennas goes to infinity. The main technical approach proposed in this chapter to reaching our goal is to fully take advantage of the characteristic of the MIMO channels, the structure of the transmitter as well as of the ZF receiver, the Szegö's theorem [44] on large Hermitian Toeplitz matrices, and the well known limit: $\lim_{x\to\infty}(1+1/x)^x = e$.

On the other hand, in terms of a trade-off between system performance and implementation complexity, it is known that the decision-feedback (DF) receiver is an attractive alternative detection scheme [45] to linear ZF and minimum mean square error (MMSE) receivers [40]. Recently, the authors in [46] proposed an optimal precoder design minimizing the average arithmetic mean-squared error (MSE) for correlated MIMO channels using the ZF-DF detector. In spite of the fact that its performance is better than that of the optimal linear receivers [40], the design is restricted to the case when $N > M$ for a MIMO system equipped with M transmitter antennas and N receiver antennas. In addition, minimizing the MSE does not necessarily minimize detection error probability. It is known that the ultimate goal in the transmitter design from the viewpoint of detection theory for communication systems is to optimize error performance with a maximum reliable transmission rate. Therefore, in Chapter 5 we are interested in the design of an optimal transmitter for the correlated MIMO communication system using the DF receiver, but instead of minimizing the MSE, our purpose here is to minimize the average SEP over all random channel coefficients under a perfect decision feedback. However, it should be explicitly pointed out that despite the fact that our design problem is actually an extension of the design problem in [40], we need to confront more technical difficulties in solving this problem due to the following two major factors: 1) Unlike the problem in [40] in which each subchannel for the ZF receiver has the same diversity gain $(N - M + 1)$ and thus, a closed-form optimal solution in a high signal-noise ratio (SNR) region can be derived using the Jessen's inequality, each subchannel for the ZF-DF receiver in our problem has different diversity gains, making the problem more difficult to be handled. 2) The linear ZF receiver does not depend on with the detection order.

However, the detection order for the ZF-DF receiver substantially affects the error performance, resulting in a major hurdle for us to transform our original non-convex design problem into a convex one.

In order to overcome these two technical difficulties, in Chapter 5, we first develop a necessary condition for the optimal solution to satisfy and then, characterize the structure of the optimal solution by carefully utilizing the product majorization relationship among eigenvalues, singular-values and Cholesky values of the design matrix parameters. Therefore, our original non-convex problem is reformulated into a convex one which can be efficiently solved using an interior-point method [47]. Computer simulations show that the error performance of our optimal precoder proposed in this chapter outperforms those of all existing designs. Particularly, our design obtains a significant SNR gain over the MIMO system with the V-BALST detector when $N = M$. In the following part of Chapter 5, we are also interested in the asymptotic SEP analysis for a specific transmitter using the QRS decomposition [48] for the correlated large MIMO fading channels using the ZF-DF receiver. Normally, the exact SEP analysis for the large MIMO systems results in enormously computational complexity. Fortunately, a careful observation on the average SEP expression for such QRS precoder reveals a nice structure which naturally leads us to propose the use of two useful mathematical tools for the asymptotic analysis on the error performance of the large MIMO systems as we did for the ZF case. Likewise, the first is the Szegö's theorem on large Hermitian Toeplitz matrices and the second is the well-known limit: $\lim_{x\to\infty}(1+1/x)^x = e$ in real analysis. It is these two powerful tools that also enable us to successfully attain a simple expression for the SEP limit when the number of the transmitter antennas goes to infinity. This new approach enjoys a fast rate of

convergence and hence, is effective and efficient for error performance evaluation for the large MIMO systems.

1.3 QAM Division for Multiuser MISO Broadcast Channels

Multiuser systems in a downlink, also known as broadcast channels, have long been the main building block of modern wireless communication systems such as cellular system, telephone conference and digital TV broadcasting. In Chapter 6, we concentrate on MISO BC, where one multi-antenna access point serves several single-antenna receivers at the same time. Hence, the design of the receiver can be significantly simplified, since it has only one antenna and the system still enjoys the multi-antenna diversity and multiuser diversity. BC was first introduced by Cover [49], who demonstrated the idea of superposition coding for both binary-symmetric and Gaussian BC. Since then, great efforts have been devoted to obtaining the capacity regions of different BCs and to seeking for the optimal transmission strategies under various constraints. Although the capacity region for the general discrete memoryless broadcast channel (DM-BC) is still unknown, much progress has been made since [49]. In particular, the achievability and converse of the capacity region for the degraded DM-BC were proved by Bergmans [50] and Gallager [51], respectively. Surveys of the literature on the BC can be found in [52–54]. On the other hand, if the transmitter and/or receiver nodes are allowed to have more than a single antenna, there will be a Gaussian vector channel in which much higher spectral efficiency (through spatial multiplexing) and reliability (by multi-path diversity) can be achieved by exploiting

the scattering medium between the transmitter and receiver antenna arrays [24, 55]. Specifically for the MISO BCs, the achievable throughput was developed in [56] based on Costa's dirty paper coding approach [57] that achieves the sum-capacity for a two-user case with a single transmitting antenna. Then, the sum-capacity for a general multiuser MISO BC was established in [58–61] by exploiting the uplink-downlink duality between the multiaccess channel (MAC) and the corresponding BC. By using more practical finite-alphabet constellations rather than Gaussian input signals, the transmission schemes that maximize the mutual information between the BS and all the receivers of the multi-user BC are considered in [62–65].

The aforementioned information-theoretic analyses serve as a guideline for a general system design. For transmitter design, non-linear precoding techniques such as the dirty paper coding (DPC) method can be used to approach the sum rate of the MISO BC [59,60]. It was shown in [56] that a successive interference cancellation procedure, namely the ZF-DPC, can be performed at the transmitter side to completely remove mutual interference between receivers. Given the complexity of DPC, the Tomlinson-Harashima precoding (THP) method [66–69] serves as a suboptimal but practical approximation of DPC by introducing a modulo operation to transmitted symbols. Despite the fact that there is a modulo loss [70], the transmitted symbols are guaranteed to have a finite dynamic range. All these precoding methods were primarily devoted to improving the sum rate of multiuser MISO BCs. On the other hand, practical transmitter designs may also aim to improve signal quality at the receiver side. Among such transmitter designs, linear precoding techniques receive tremendous attention because of their potential and simplicity. Using signal-to-noise ratio (SNR) as a design criterion, it was shown that transmitter beamforming can

increase the received SNR in the multiuser MISO channel by performing optimization on the beamformer design and the power allocation scheme [71–73]. In addition, by employing MMSE as a performance measure, an optimal precoder was proposed in [74] with regularized channel inversion, which outperforms the ZF scheme when the channel condition number is large. By maximizing SLNR for all users simultaneously, a closed-form beamforming method was given in [75]. Even so, however, it was demonstrated that the ZF beamforming technique, which is simple to implement, can achieve most of the capacity in moderdate and high SNR regimes [76, 77]. Comprehensive comparisons of these schemes can be found in [78].

As we can see, interference has long been the central focus for meeting the ever increasing requirements on quality of service (QoS) in modern and future wireless communication systems. The key to the understanding of multiuser communications is the understanding of interference. Traditional approaches, which treat interference as a detrimental phenomenon are, therefore, to suppress or eliminate it. The classical information-theoretic study on the two-user Gaussian interference channel [79] suggests us that we should treat the interference as noise when it is weak and that the optimal strategy is to decode the interference when it is very strong. In addition, when the level of interference is of the same order of the power of a desired signal, one good strategy is to suppress all the undesired interferences into a smaller space that has no overlap with the signal space [80–82]. However, some recent innovative approaches, which consider interference as a useful resource, are, thus, to make use of it for developing energy efficient and secure 5G communication systems [83–85]. For example, interference can be used for boosting up the desired signal [86–88] for energy harvesting [89–96] or to deteriorate the signal of the eavesdropper for secure

communication [97].

Inspired by [81, 98], in Chapter 6, we consider the management of interference for BC by carefully designing communication signals. To better elaborate on our idea, we would like to revisit some early seminal work [99,100] of how to strategically take advantage of the finite alphabet properties of digital communication signals for managing interference for a two-user access binary channel. Essentially, the Kasami and Lin's main idea is to carefully design such two finite-length codes for the two users that when any sum binary signal of the two user codewords is received in a noiseless environment, each individual user codeword can be uniquely decoded, as well as in a noisy case, the resulting error is able to be correctable. Specifically, such uniquely decodable code (UDC) was explicitly constructed for a two-user binary ensure channel [101, 102]. Then, this important concept was extended to the design of UDC based on trellis modulation for a multiuser binary multi-access channel [103], which allows a number of users to access a common receiver simultaneously and outperforms the time sharing method in terms of the probability of error. Furthermore, the design of trellis-coded UDC was investigated in a complex number domain to extract the desired signal from the superposition of the signal and cochannel interferences [104–106]. In addition, the concept of UDC was also exploited to design varieties of multi-resolution modulation schemes for BC and it was shown that they not only outperform the frequency division scheme by properly designing the resulting constellation [107–110], but also reduce the transmission delay of the network at the cost of increased transmitting power for fading channels [109]. Recently, a pair of uniquely decodable constellations was designed to study the capacity region of a two user Gaussian multi-access channel [111].

Indeed, it is the above aforementioned factors that greatly motivate and enlighten us to look into interference from the perspective of signal processing. In Chapter 6, we are interested in exploring a novel signal processing technique to manage interference for BC, which allows strongly interfered user signals to cooperate each other as a common desired sum signal from which each individual user signal is able to be uniquely and efficiently decoded. Specifically, our main contributions of Chapter 6 can be summarized as follows:

1. An explicit construction of UDCG, which can be considered as a UDC in the complex domain for a multiuser case, for general PAM and rectangular QAM constellations for *any* number of users is proposed. The main difference between our UDCG design and all currently available UFC designs in literature is that in our UDCG design, the sum-constellation and all the user constellations are PAM and QAM constellations with different scales. It is because of this nice geometric structure that once the sum signal is received, each individual user signal can be efficiently and uniquely decoded (see Algorithms 1 and 2).

2. Using the newly developed UDCG, we propose a novel non-orthogonal multiple access (NOMA) transmission scheme called QAM-modulation division for the multiuser MISO BC. First, an optimal beamformer that maximizes the minimum SNR between the two receivers is obtained in a closed-form for a two-user case. Then, for a general network topology with more than two receivers, a grouping-based transmitter design problem is also investigated, with ZF eliminating the inter-group interference, where the grouping policy, the beamformer design and power allocation strategy are addressed. It is demonstrated that when the Hermitian angle of the two channel vectors is small, our proposed

division method has a much lower probability of error, which confirms that the NOMA method with proper interference control is a promising technology for 5G communications.

Our work in Chapter 6 can be considered as a concrete, simple and systematic design of the constellation for NOMA [112, 113], serving different users with different power levels, and it has a considerable spectral gain over the traditional methods.

1.4 Additive UDCG in the Uplink Multiuser Massive SIMO

As we have mentioned above, as one of the most promising technologies to meet the ever-increasing bandwidth requirement and connectivity of 5G communications, massive MIMO technique receives considerable attention recently [27, 28, 114–118]. Due to the large number of available antennas, extremely high energy efficiency and spectral efficiency can be achieved compared with most of the available systems off-the-shelf [115]. More importantly, for the uplink multiuser massive single-input and multiple-output (SIMO) network considered in this thesis, as the number of BS antennas tends to infinity, the cross-correlation between the channel vectors of different users vanishes in rich scattering environment (and it is also true for line-of-sight channel with enough angular separation) which will result in negligible mutual interference if the CSI is perfectly known at the BS. Meanwhile, the overall system design for massive MIMO systems can be greatly simplified [27,114] (e.g., the maximum-ratio combining (MRC) receiver would be sufficient to approach the sum-capacity).

In Chapter 7, we consider an uplink multiuser massive SIMO system where a

multi-antenna BS serves several single-antenna users simultaneously in the same time-frequency band and we attempt to minimize the training overhead. This is motivated by the fact that, with the rise of machine-to-machine (M2M) communications and internet of things (IoT), the number of users needed to be supported by a macrocell BS increases dramatically nowadays [119–121]. To accommodate more terminals, spatial multiplexing technique that allows time and frequency reuse must be adopted which relies on the availability of the CSI and it is commonly acquired by training methods [122, 123]. In general, the minimum number of time slots needed for the training process is equal to the number of users in our model [124] while it is further limited by the channel coherence time and delay spread of wireless channels. As a consequence, the pilot contamination due to using non-orthogonal pilot sequences arises as a key-limiting factor that will introduce a non-vanishing interference term for the intended signal [114, 117, 125]. To alleviate such constraint and to save the overhead caused by channel estimation in massive MIMO systems, in Chapter 7, a new transmission scheme is proposed which can be used to estimate the CSI for all the users in a short coherence time. The proposed new transmission framework is based on the concept of quadrature amplitude modulation division (QAMD).

Now, let us review some conventional resource division methods in wireless communications. It is widely accepted that dividing scarce spectral resources among multiple users to allow them to access the wireless media simultaneously is one of the core issues in wireless communications. The multiple access methods can be divided into two basic categories [126], i.e., reservation-based multiple access by using time, frequency, code and space dimensions of the physical channels (e.g., TDMA, FDMA, CDMA and SDMA) and random multiple access through competition on the virtual

channels (e.g., ALOHA and CSMA). The later case actually builds on the former case to further increase the number of terminals that can be served. Here, we concentrate on the reservation-based multiple access where the main idea is to allocate different users into orthogonal channel subspaces causing no mutual interference to each other and different signal scales picked from a finite alphabet can be sent to convey information in the absence of interference in every dimension. These methods are relatively simple to realize and are able to approach the capacity [127], especially when the number of users is smaller than the available degrees of freedom (DoF) of the channel, orthogonality between all the sub-channels can be maintained. However, one drawback of these schemes is that the maximum number of terminals that can be served simultaneously is limited by the available subspaces (i.e., the DoF) of the physical channels. Moreover, in low and moderate SNR regimes, some sub-channels are too poor to be of practical use (e.g., the multiplexing gain should be traded for increased diversity gain and power gain). In addition, these methods typically require global CSI over all the terminals which is a rather strong assumption and might be too complicated to implement in practice [80, 128].

In Chapter 7, we propose a novel multiple access method, namely the multi-user space-time modulation (MUSTM) scheme which can support multiple access in the modulation level by explicit factorization of QAM constellations into a uniquely-decomposable constellation group (UDCG) with predetermined data rate as defined in Chapter 2. The minimum training time required is as short as one time slot. As we have mentioned above, interference is the central topic of multiuser network information theory [54, 129] where conventional approaches are to suppress or to eliminate interference. Multiple access is, essentially, a way to manage interference to maintain

the decodability of the transmitted signals at the intended terminals. In the considered multiuser massive SIMO systems, to improve the spectral efficiency and alleviate the pilot contamination, all the training signals and the information for all the users are transmitted concurrently, which therefore will cause strong interference to each other. Inspired by the classic works on interference channels [81,98], in Chapter 7, we consider the management of interference for multiple access channel (MAC) by carefully designing communication signals. To enable the receiver to recover the desired transmitted signal from the received signal mixture, we consider a new interference management scheme through interference collaboration, or more specifically, QAMD method by letting some smaller constellations to form a QAM-UDCG. The main idea is to take advantage of channel statistics of massive MIMO networks and the finite alphabet of digital signals. More specifically, in massive MIMO communications, the fluctuation caused by the local scattering vanishes and then we can let multiple transmitters to transmit at proper power levels to form a constellation with a good geometric structure that would be uniquely decodable at the receiver side.

The above factors motivate and enlighten us to perform interference management from the perspective of signal processing. Our contributions in Chapter 7 can be stated as follows:

1. For our proposed method with N single-antenna users, the minimum number of training slots is as short as one while it equals to N for conventional training method such as [124]. Our scheme is especially suited for the 5G communication systems with a extremely short coherence time where the mobility support at speed $\geq$500 km/h for ground transportation [130]. Also, the overhead for the training procedure is $1/N$ the length of conventional method and thus the

average delay is also reduced considerably.

2. A sub-constellation allocation method is developed which ensures the transmitted signal matrix to be invertible. As a result, the channel vector can be uniquely determined in the absence of noise, which can be used for the downlink precoding. In addition, the coding gain for the received sum signal in each time slot is derived and the optimal power allocation methods are also addressed under both peak and average power constraints.

In this chapter, we have discussed the motivations and technical contributions of this thesis. Also the literature related to the thesis is surveyed. In the following part of this thesis, the system models and technical details of our work will be given explicitly.

Chapter 2

Uniquely Factorable Constellation Pair (UFCP) and Uniquely Decomposable Constellation Group (UDCG)

In this chapter, we first briefly review the concept of uniquely factorable constellation pair (UFCP) proposed in [17,18] and then introduce the definition of a novel concept called uniquely decomposable constellation group (UDCG) devised in [131].

2.1 Uniquely Factorable Constellation Pair (UFCP)

As one of the key concepts in this thesis, the formal definition of UFCP is given as follows:

Definition 1 (UFCP) *[18] A pair of constellations $\mathcal{X}$ and $\mathcal{Y}$ is said to form a UFCP, which is denoted by $\mathcal{Y} \sim \mathcal{X}$, if for any $x, \tilde{x} \in \mathcal{X}$ and $y, \tilde{y} \in \mathcal{Y}$ such that $x\tilde{y} = \tilde{x}y$, then $x = \tilde{x}, y = \tilde{y}$.* ■

By the definition, the following example provides us with a trivial UFCP.

Example 1 *For any set $\mathcal{Y}$, if we take $\mathcal{X} = \{1\}$, then, $\mathcal{X}$ and $\mathcal{Y}$ form a UFCP.*

The above Example 1 actually shows that if $\mathcal{X}$ contains only a non-zero element, then it will form a UFCP with any set $\mathcal{Y}$. However, constellation $\mathcal{X}$ in Example 1 can not be used for information transmission since there is only one element in it. The following example provides us a non-trivial UFCP.

Example 2 *If we let $\mathcal{X} = \{1, j\}$ and $\mathcal{Y} = \{1, -1\}$, then, such a pair of $\mathcal{X}$ and $\mathcal{Y}$ constitutes a UFCP.*

In this thesis, we are interested in the design of non-trival UFCPs each element of which is a Gaussian integer. To that end, the necessary condition which a UFCP must satisfy is developed.

Definition 2 *[18] Given a UFCP $\mathcal{Y} \sim \mathcal{X}$, $0 \notin \mathcal{X}$ and a fixed $x \in \mathcal{X}$, a set generated from x, denoted by $\mathcal{Z}_x$,*

$$\mathcal{Z}_x = \left\{ z : z = \frac{y}{x}, y \in \mathcal{Y} \right\},$$

is called a Group-x. ■

Proposition 1 *[18]*

- *Let $\mathcal{X}$ and $\mathcal{Y}$ form a UFCP. If $|\mathcal{Y}| \geq 2$, then, $0 \notin \mathcal{X}$.*
- *For a pair of given constellations $\mathcal{X}$ and $\mathcal{Y}$ with each having finite size and $0 \notin \mathcal{X}$, if a new constellation $\mathcal{Z}$ is defined as*

$$\mathcal{Z} = \left\{ z : z = \frac{y}{x}, x \in \mathcal{X}, y \in \mathcal{Y} \right\}, \tag{2.1}$$

then such a pair of $\mathcal{X}$ and $\mathcal{Y}$ constitutes a UFCP if and only if

$$|\mathcal{Z}| = |\mathcal{X}||\mathcal{Y}|. \tag{2.2}$$

■

2.1.1 Unique Factorization of PSK Constellation

We first consider the UFCP constructed based on the commonly used PSK constelaltion as follows.

Proposition 2 (PSK-UFCP) *If we let two sets $\mathcal{X}$ and $\mathcal{Y}$ be*

$$\mathcal{X} = \left\{ \exp\left(\frac{j2\pi m}{2^p} \right) \right\}_{m=0}^{2^p-1}, \tag{2.3a}$$

$$\mathcal{Y} = \left\{ \exp\left(\frac{j2\pi n(2^p-1)}{2^r} \right) \right\}_{n=0}^{2^q-1}, \tag{2.3b}$$

where $r = p + q$ in which p and q are positive integers, then such a pair of $\mathcal{X}$ and $\mathcal{Y}$ constitutes a UFCP. ■

Proof: Let $x = \exp\left(j\frac{2\pi m}{2^p}\right)$, $\tilde{x} = \exp\left(j\frac{2\pi \tilde{m}}{2^p}\right)$, $y = \exp\left(j\frac{2\pi n(2^p-1)}{2^r}\right)$, and $\tilde{y} = \exp\left(j\frac{2\pi \tilde{n}(2^p-1)}{2^r}\right)$, where $0 \leq m, \tilde{m} \leq 2^p - 1$, $0 \leq n, \tilde{n} \leq 2^q - 1$. By the property of the PSK constellation, $xy = \tilde{x}\tilde{y}$ is equivalent to $m2^q + n(2^p - 1) \equiv \tilde{m}2^q + \tilde{n}(2^p - 1) \mod 2^r$, or equivalently

$$(m - \tilde{m})2^q + (n - \tilde{n})(2^p - 1) \equiv 0 \mod 2^r. \tag{2.4}$$

Since $2^q | 2^r$, we attain that $(n - \tilde{n})(2^p - 1) \equiv 0 \mod 2^q$. Notice that $(2^p - 1, 2^q) = 1$. Hence, we have $2^q | (n - \tilde{n})$. Since $0 \leq n, \tilde{n} \leq 2^q - 1$, we obtain $n = \tilde{n}$ and as a result, (2.4) reduces to

$$(m - \tilde{m})2^q \equiv 0 \mod 2^r. \tag{2.5}$$

Dividing both sides by 2^q yields

$$m - \tilde{m} \equiv 0 \mod 2^p. \tag{2.6}$$

In other words, $2^p | (m - \tilde{m})$. Once noticing that $0 \leq m, \tilde{m} \leq 2^p - 1$, we can immediately deduce that $m = \tilde{m}$. Therefore, $x = \tilde{x}$ and $y = \tilde{y}$, such a pair of $\mathcal{X}$ and $\mathcal{Y}$ constitutes a UFCP. This completes the proof of Proposition 2. □

2.1.2 Fast Factorization of PSK Constellation

Now consider a PSK-UFCP $\mathcal{Y} \sim \mathcal{X}$ defined in Proposition 2 where $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and $z \in \mathcal{Z}$. By the definition of UFCP, once z has been observed, x, y can be uniquely determined by resorting to an exhaustive search. However, when the size of

constellation $\mathcal{Z}$ is large, it is not computationally efficient. To resolve this problem, we develop a closed-form solution for the fast factorization of the PSK constellation.

Proposition 3 (PSK-Factorization) *Let p, q and r be positive integers such that $r = p + q$, and $\mathcal{Z}$ denote a 2^r-PSK constellation, i.e.,*

$$\mathcal{Z} = \left\{ \exp\left(\frac{j2\pi k}{2^r}\right) \right\}_{k=0}^{2^r-1}. \tag{2.7}$$

Then, for any $z \in \mathcal{Z}$, there exists a pair of $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ such that $xy = z$. Furthermore, x and y are uniquely and explicitly determined by $x = \exp\left(j\frac{2\pi m}{2^p}\right)$, $y = \exp\left(j\frac{2\pi n(2^p-1)}{2^r}\right)$, where $n \equiv k(2^p-1)^{2^{q-1}-1} \mod 2^q$ and $m \equiv \frac{k-(2^p-1)n}{2^q} \mod 2^p$. ■

Proof: Let $x = \exp\left(j\frac{2\pi m}{2^p}\right), y = \exp\left(j\frac{2\pi n(2^p-1)}{2^r}\right)$ and $z = \exp\left(j\frac{2\pi k}{2^r}\right)$. Then, equation $xy = z$ is equivalent to

$$m2^q + n(2^p - 1) \equiv k \mod 2^r. \tag{2.8}$$

Since $2^q | 2^r$, we have

$$n(2^p - 1) \equiv k \mod 2^q. \tag{2.9}$$

With the help of the Euler theorem in [132], we can attain

$$(2^p - 1)^{2^{q-1}} \equiv 1 \mod 2^q, \tag{2.10}$$

Now, combining (2.9) with (2.10) results in

$$n \equiv k(2^p - 1)^{2^{q-1}-1} \mod 2^q. \tag{2.11}$$

There is only one solution to (2.11) such that $0 \leq n \leq 2^q - 1$. In other words, the solution to n is unique. On the other hand, from (2.9), we know that $2^q | \big(k - n(2^p - 1)\big)$. Then, according to (2.8) and noticing that $2^q | 2^r$, we can arrive at

$$m \equiv \frac{k - (2^p - 1)n}{2^q} \mod 2^p. \tag{2.12}$$

Hence, m can also be uniquely determined for $0 \leq m \leq 2^p - 1$. This completes the proof of Proposition 3. □

Proposition 3 tells us that any 2^r-PSK symbol can be efficiently factorized into the product of a 2^p-PSK symbol and a 2^q-PSK symbol with $r = p + q$.

2.1.3 Unique Factorizations of Square-QAM Constellation

Now, we consider factorizing a 2^K-QAM constellation rather than PSK constellation. For a given constellation $\mathcal{Z}$, a UFCP $\mathcal{Y} \sim \mathcal{X}$ is said to be generated from $\mathcal{Z}$, which is denoted by $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, if $\mathcal{Z} = \{z : z = xy, x \in \mathcal{X}, y \in \mathcal{Y}\}$ and $|\mathcal{Z}| = |\mathcal{X}| \times |\mathcal{Y}|$. Here, we are particularly interested in the case when constellation $\mathcal{Y}$ has unit energy, i.e., $\mathcal{Y} = \{1, j\}, \{1, j, -1, -j\}$. For the sake of completeness, we recast the result of [18] in Proposition 4.

Proposition 4 (Square-QAM UFCP) *Let $\mathcal{Z}$ be a given square 2^K-QAM constellation such that, where $K \geq 4$ is even. Then, for any $\mathcal{Y} \subseteq \{1, -1, j, -j\}$ with a fixed*

size of $\mathcal{Y}$ greater than one, one of the solutions to the following optimization problem

$$\{\mathcal{X}_{\text{opt}}, \mathcal{Y}_{\text{opt}}\} = \arg \max_{\mathcal{X}\times\mathcal{Y}=\mathcal{Z}} \min_{x_1\neq x_2\in\mathcal{X}} |x_1 - x_2| \tag{2.13}$$

is given as follows:

1. *If $|\mathcal{Y}| = 2$ and $K \geq 4$, then, $\mathcal{Y}_{\text{opt}}^{(1)} = \{1, j\}$ and, $\mathcal{X}_{\text{opt}}^{(1)} = \big\{(2^{\frac{K}{2}} - 1 - 4m) + (2^{\frac{K}{2}} - 1 - 4n)j : 0 \leq m, n \leq 2^{\frac{K-2}{2}} - 1\big\} \bigcup \big\{(2^{\frac{K}{2}} - 3 - 4m) + (2^{\frac{K}{2}} - 3 - 4n)j : 0 \leq m, n \leq 2^{\frac{K-2}{2}} - 1\big\}$.*

2. *If $|\mathcal{Y}| = 4$ and $K \geq 4$, then, $\mathcal{Y}_{\text{opt}}^{(2)} = \{1, -1, j, -j\}$ and, $\mathcal{X}_{\text{opt}}^{(2)} = \big\{(4m - 2^{\frac{K}{2}} + 3) + (2^{\frac{K}{2}} - 1 - 4n)j : 0 \leq m, n \leq 2^{\frac{K-2}{2}} - 1\big\}$*

■

As a direct application of Proposition 4, we give the following example.

Example 3 *Let $\mathcal{Z}$ be the 16-QAM constellation. By Proposition 4, a UFCP $\mathcal{Y} \sim \mathcal{X}$ can be obtained by factorizing $\mathcal{Z}$:*

$$\mathcal{Y} = \{1, j\},$$
$$\mathcal{X} = \{3 + 3j, 3 + j, 1 + 3j, 1 + j, -1 - j, -3 - 3j, -3 - j, -1 - 3j\}.$$

Example 4 *Likewise, by Proposition 4, a UFCP $\mathcal{Y} \sim \mathcal{X}$ can be obtained by factorizing 16-QAM constellation $\mathcal{Z}$ in the following way:*

$$\mathcal{Y} = \{1, -1, j, -j\},$$
$$\mathcal{X} = \{3 + 3j, 3 + j, 1 + 3j, 1 + j\}.$$

2.1.4 Fast Factorization of Square-QAM Constellation

Similar to the PSK-UFCP, a closed-form solution is provided for the factorization procedure of the Square-QAM-UFCP.

Proposition 5 *Let $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and $z \in \mathcal{Z}$ be such that $xy = z$ where $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ are defined in Proposition 4. Also we denote $z = p + qj$ and then, x, y can be determined as follows:*

1. *If $|\mathcal{Y}| = 2$ and $K \geq 4$, then $y = 1 - \frac{\mathrm{mod}(p-q,4)}{2} + \frac{\mathrm{mod}(p-q,4)}{2} j, x = \frac{p}{y} + \frac{q}{y} j$.*
2. *If $|\mathcal{Y}| = 4$ and $K \geq 4$, then $y = -2 + \frac{\mathrm{mod}(p,4)+\mathrm{mod}(q,4)}{2} + \frac{\mathrm{mod}(q,4)-\mathrm{mod}(p,4)}{2} j, x = \frac{p}{y} + \frac{q}{y} j$.*

■

Proof: We first notice that after y has been obatined, we have $x = \frac{z}{y}$, since $z = x\,y$. Hence, we only need to show how to determine y based on z. We consider the following two cases.

1. If $|\mathcal{Y}| = 2$, then, from Proposition 4, we know that, if $z \in \mathcal{X}_{\mathrm{opt}}^{(1)}$, i.e., $y = 1$, we have $\mathrm{mod}(p-q,4) = 0$, and that if $z \in j\mathcal{X}_{\mathrm{opt}}^{(1)}$, i.e., $y = j$, we have $\mathrm{mod}(p-q,4) = 2$.

2. If $|\mathcal{Y}| = 4$ and $K \geq 4$, then, we have the following observations summarized in Table 2. This completes the proof of Statement 2 of Proposition 5.

□

Estimated signals	y	$\mathrm{mod}(p,4)$	$\mathrm{mod}(q,4)$
$z \in \mathcal{X}_{\mathrm{opt}}^{(2)}$	1	3	3
$z \in -\mathcal{X}_{\mathrm{opt}}^{(2)}$	-1	1	1
$z \in j\mathcal{X}_{\mathrm{opt}}^{(2)}$	j	1	3
$z \in -j\mathcal{X}_{\mathrm{opt}}^{(2)}$	-j	3	1

Table 2.1: The numerical relationship of the UFCP

2.2 Uniquely Decomposable Constellation Group

In this section, we will introduce the definition of the UDCG and then, provide the flexible and the explicit construction of the UDCG using commonly-used PAM and QAM constellations and the corresponding efficient decoding algorithms.

Definition 3 (UDCG) *A group of constellations* $\{\mathcal{X}_i\}_{i=1}^N$ *is said to form a UDCG, denoted by* $\{\sum_{i=1}^N x_i : x_i \in \mathcal{X}_i\} = \uplus_{i=1}^N \mathcal{X}_i = \mathcal{X}_1 \uplus \mathcal{X}_2 \uplus \ldots \uplus \mathcal{X}_N$*, for any groups of* $x_i, \tilde{x}_i \in \mathcal{X}_i$ *for* $i = 1, 2, \cdots, N$ *such that* $\sum_{i=1}^N x_i = \sum_{i=1}^N \tilde{x}_i$*, then, we have* $x_i = \tilde{x}_i$ *for* $i = 1, 2, \cdots, N$. ■

For presentation convenience, constellation $\uplus_{i=1}^N \mathcal{X}_i$ in Definition 3 is called the *sum-constellation* of all $\mathcal{X}_i$ and each $\mathcal{X}_i$ is called the *i-th sub-constellation* of $\uplus_{i=1}^N \mathcal{X}_i$ or *i-th user constellation*. The concept of UDCG can be considered as an extension of UDC in binary field to the complex number domain for N-users [99–111]. However, we would like to emphasize a major difference between the definition of UDCG and the traditional concept of UDC. In our Definition 3, we are interested in each analysis component of the decomposition, i.e., the geometrical structure of each user-constellation, as well as in the synthesis component of the decomposition, i.e., the geometrical structure of the sum constellation.

The following property reveals such a fact that checking whether or not a group of

constellations forms a UDCG is equivalent to checking whether or not the cardinality of the sum constellation is equal to the product of the cardinalities of the user-constellations.

Property 1 (Unambiguity) *Given a group of constellations* $\{\mathcal{X}_i\}_{i=1}^N$ *with each having a finite size, if we let* $\mathcal{G} = \{\sum_{i=1}^N x_i : x_i \in \mathcal{X}_i\}$*, then,* $\mathcal{G} = \uplus_{i=1}^N \mathcal{X}_i$ *if and only if* $|\mathcal{G}| = \prod_{i=1}^N |\mathcal{X}_i|$. ■

Proof: Let $\mathcal{Y}$ denote a set of N-tuples $\mathcal{Y} = \{(x_1, x_2, \ldots, x_N) : x_i \in \mathcal{X}_i\}$. Then, $|\mathcal{Y}| = \prod_{i=1}^N |\mathcal{X}_i|$ by the combinatorial rule of product and $\mathcal{Y}$ is a finite set, since each $\mathcal{X}_i$ is a finite set.

If $|\mathcal{G}| = \prod_{i=1}^N |\mathcal{X}_i|$, then, we have $|\mathcal{G}| = |\mathcal{Y}|$. Since $\mathcal{G}$ and $\mathcal{Y}$ are finite sets, there exists a bijection map between these two sets, which is denoted by $f_{\text{bij}} : \mathcal{G} \to \mathcal{Y}$ [133]. Without loss of generality, we let $f_{\text{bij}}(\sum_{i=1}^N x_i) = (x_1, x_2, \ldots, x_N)$. As $f_{\text{bij}} : \mathcal{G} \to \mathcal{Y}$ is a bijective map, then, if $\sum_{i=1}^N x_i = \sum_{i=1}^N \tilde{x}_i$, we must have $(x_1, x_2, \ldots, x_N) = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_N)$ and hence, $x_i = \tilde{x}_i$ for $i = 1, 2, \cdots, N$. Then, by Definition 3, we have $\mathcal{G} = \uplus_{i=1}^N \mathcal{X}_i$.

If $\mathcal{G} = \uplus_{i=1}^N \mathcal{X}_i$, by Definition 3, for any $(x_1, x_2, \ldots, x_N), (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_N) \in \mathcal{Y}$ satisfying $\sum_{i=1}^N x_i = \sum_{i=1}^N \tilde{x}_i$, we have $x_i = \tilde{x}_i$ for $i = 1, 2, \cdots, N$, or equivalently $(x_1, x_2, \ldots, x_N) = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_N)$. Therefore, there exists an injective function $f_{\text{inj}} : \mathcal{Y} \to \mathcal{G}$ such that $f_{\text{inj}}\big((x_1, x_2, \ldots, x_N)\big) = \sum_{i=1}^N x_i$. Hence, $|G| \geq |\mathcal{Y}| = \prod_{i=1}^N |\mathcal{X}_i|$. From the construction of $\mathcal{G}$, we know that $|\mathcal{G}| \leq \prod_{i=1}^N |\mathcal{X}_i|$. As a result, $|\mathcal{G}| = \prod_{i=1}^N |\mathcal{X}_i|$. This completes the proof of the property. □

Since PAM and QAM constellations are commonly used in modern digital communication systems, in this chapter we are interested in uniquely decomposing them into the sum of a group of scaled PAM or QAM constellations.

Theorem 1 (PAM) *Given two positive integers K and N, let K_i be any N non-negative integers satisfying $\sum_{i=1}^{N} K_i = K$. Then, 2^K-ary PAM constellation $\mathcal{G} = \{\pm(m - \frac{1}{2}) : m = 1, 2, \ldots, 2^{K-1}\}$ can be uniquely decomposed into the sum of N sub-constellations $\mathcal{X}_i$ for $i = 1, 2, \cdots, N$, i.e., $\mathcal{G} = \uplus_{i=1}^{N} \mathcal{X}_i$, where*

$$\mathcal{X}_1 = \begin{cases} \{0\} & K_1 = 0 \\ \{\pm(m - \frac{1}{2}) : m = 1, 2, \ldots, 2^{K_1 - 1}\} & K_1 \geq 1 \end{cases}$$

and for $i \geq 2$,

$$\mathcal{X}_i = \begin{cases} \{0\} & K_i = 0 \\ \{\pm(m - \frac{1}{2}) \times 2^{\sum_{n=1}^{i-1} K_n} : m = 1, 2, \ldots, 2^{K_i - 1}\} & K_i \geq 1 \end{cases}.$$

■

Proof: On one hand, we notice that $\sum_{i=1}^{N} x_i \in \mathcal{G}$, for any $x_i \in \mathcal{X}_i, \forall i \in \{1, 2, \ldots, N\}$. On the other hand, we also note that $|\mathcal{X}_i| = \begin{cases} 1 & K_i = 0 \\ 2^{K_i} & K_i \geq 1 \end{cases}, \forall i \in \{1, 2, \ldots, N\}$ and $|\mathcal{G}| = 2^K$. Since $K = \sum_{i=1}^{N} K_i$, we have $|\mathcal{G}| = \prod_{i=1}^{N} |\mathcal{X}_i|$. By Property 1, we attain $\mathcal{G} = \uplus_{i=1}^{N} \mathcal{X}_i$. This completes the proof of Theorem 1. □

Theorem 1 reveals a significant property on the PAM constellation that any PAM constellation of large size can be uniquely decomposed into the sum of a group of the scaled version of the PAM constellations of variety of small sizes. Furthermore, the following algorithm proceeds to uncover an important advantage of such unique decomposition.

Algorithm 1 (Fast detection of PAM UDCGs) *Given a UDCG* $\mathcal{G} = \uplus_{i=1}^{N}\mathcal{X}_i$ *generated from a PAM constellation by Theorem 1. For an observed noisy real signal* $y = \sum_{i=1}^{N} x_i + \xi$*, where* $x_i \in \mathcal{X}_i$ *and* $\xi \sim \mathcal{N}(0, \sigma^2/2)$ *is a real additive white Gaussian noise. Then, we have a fast detection method for estimating all user-signals stated as follows:*

1. *Quantization of the sum signal: Given* y*, the optimal estimate of* $g = \sum_{i=1}^{N} x_i$ *is given as follows:*

$$\begin{aligned}\hat{g} &= \arg_{g\in\mathcal{G}} \min |y-g| \\ &= \begin{cases} -\frac{2^K-1}{2}, & y \le -\frac{2^K}{2}; \\ \lfloor y + \frac{2^K}{2} \rfloor - \frac{2^K-1}{2}, & -\frac{2^K}{2} < y \le \frac{2^K}{2}; \\ \frac{2^K-1}{2}, & y > \frac{2^K}{2}. \end{cases}\end{aligned} \tag{2.14}$$

2. *Decoding of the user-signals: Let* $\hat{g}$ *be defined by* (2.14)*. Then, the estimates of all the original user-signals* $\hat{x}_i$*, satisfying* $\sum_{i=1}^{N} \hat{x}_i = \hat{g}$*, are uniquely determined as*

$$\hat{x}_1 = \begin{cases} 0 & K_1 = 0 \\ (\hat{g} + \frac{2^K-1}{2}) \bmod 2^{K_1} - \frac{2^{K_1}-1}{2} & K_1 \ge 1 \end{cases} \tag{2.15}$$

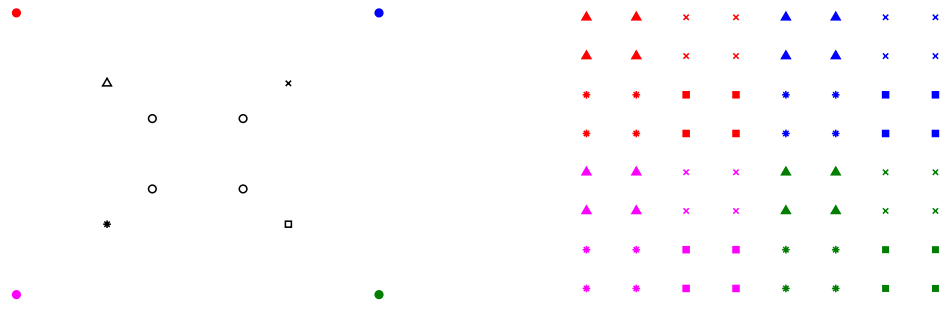
and for $i \ge 2$*,*

$$\hat{x}_i = \begin{cases} 0 & K_i = 0 \\ \left(\frac{\hat{g} + \frac{2^K-1}{2} - (\hat{g} + \frac{2^K-1}{2}) \mod 2^{\sum_{\ell=1}^{i-1} K_\ell}}{2^{\sum_{\ell=1}^{i-1} K_\ell}} \mod 2^{K_i} - \frac{2^{K_i}-1}{2} \right) 2^{\sum_{\ell=1}^{i-1} K_\ell} & K_i \ge 1 \end{cases}. \tag{2.16}$$

■

The proof of Algorithm 1 is given in Appendix A.1. As we know, a rectangular QAM constellation is generated from a pair of the PAM constellations. Hence, Theorem 1 can be extended to the PAM and QAM mixed case in a straightforward manner, whose proof, therefore, is omitted.

Theorem 2 (QAM) *For two positive integers N and $K = K^{(c)} + K^{(s)}$, with $K^{(c)}$ and $K^{(s)}$ being nonnegative integers, let $K_i^{(c)}$ and $K_i^{(s)}$ for $i = 1, 2, \cdots, N$ denote any two given nonnegative integer sequences satisfying $K^{(c)} = \sum_{i=1}^{N} K_i^{(c)}$ and $K^{(s)} = \sum_{i=1}^{N} K_i^{(s)}$ with $K_i^{(c)} + K_i^{(s)} > 0$. Then, there exists a PAM and QAM mixed constellation $\mathcal{Q}$ such that $\mathcal{Q} = \uplus_{i=1}^{N} \mathcal{X}_i$, where $\mathcal{X}_i = \mathcal{X}_i^{(c)} \uplus j\mathcal{X}_i^{(s)}$, with $j\mathcal{X}_i^{(s)} = \{jx : x \in \mathcal{X}_i^{(s)}\}$. In addition, $\mathcal{Q}^{(c)} = \uplus_{i=1}^{N} \mathcal{X}_i^{(c)}$ and $\mathcal{Q}^{(s)} = \uplus_{i=1}^{N} \mathcal{X}_i^{(s)}$ are two PAM UDCGs given in Theorem 1 according to the rate-allocation $K_i^{(c)}$ and $K_i^{(s)}$, respectively.* ■



(a) Three 4-QAM constellation $\mathcal{X}_1$, $\mathcal{X}_2$ and $\mathcal{X}_3$.

(b) The sum-constellation $\mathcal{G} = \mathcal{X}_1 \uplus \mathcal{X}_2 \uplus \mathcal{X}_3$.

Figure 2.1: An example of the UDCG with three sub-constellations.

Similar to Algorithm 1, we also have an efficient detection method for a UDCG based on the QAM constellation below:

Algorithm 2 (Fast detection of QAM UDCG) *Let a UDCG* $\mathcal{G} = \uplus_{i=1}^{N} \mathcal{X}_i$ *be generated from the QAM constellation by Theorem 2. Then, for an observed noisy complex signal* $y = \sum_{i=1}^{N} x_i + \xi$, *where* $x_i \in \mathcal{X}_i$ *and* $\xi \sim \mathcal{CN}(0, \sigma^2)$ *is an additive circularly-symmetric complex Gaussian noise, all the user-signals are efficiently estimated using the following two successive steps:*

1. *Quantization of the sum signal: Let* $y = y^{(c)} + jy^{(s)}$. *Find the quantized PAM signal* $\hat{g}^{(c)} \in \mathcal{Q}^{(c)}$ *and* $\hat{g}^{(s)} \in \mathcal{Q}^{(s)}$ *of* $y^{(c)}$ *and* $y^{(s)}$, *respectively by solving the following optimization problems*

$$\hat{g}^{(c)} = \arg_{g \in \mathcal{Q}^{(c)}} \min |y^{(c)} - g|,$$
$$\hat{g}^{(s)} = \arg_{g \in \mathcal{Q}^{(s)}} \min |y^{(s)} - g|.$$

2. *Decoding of the user-signals: By Algorithm 1, the estimates of all the user real signals* $\hat{x}_i^{(c)} \in \mathcal{X}_i^{(c)}$ *and* $\hat{x}_i^{(s)} \in \mathcal{X}_i^{(s)}$ *for* $i = 1, 2, \cdots, N$ *can be efficiently obtained such that* $\hat{g}^{(c)} = \sum_{i=1}^{N} \hat{x}_i^{(c)}$ *and* $\hat{g}^{(s)} = \sum_{i=1}^{N} \hat{x}_i^{(s)}$ *, and thus,* $\hat{g}_i = \hat{x}_i^{(c)} + j\hat{x}_i^{(s)}$. ■

Example 5 *Consider a UDCG such that* $\mathcal{X}_1 = \{\frac{1}{2} + \frac{j}{2}, \frac{1}{2} - \frac{j}{2}, -\frac{1}{2} + \frac{j}{2}, -\frac{1}{2} - \frac{j}{2}\}$, $\mathcal{X}_2 = \{1 + j, 1 - j, -1 + j, -1 - j\}$ *and* $\mathcal{X}_3 = \{2 + 2j, 2 - 2j, -2 + 2j, -2 - 2j\}$ *as illustrated in Fig. 2.1(a) and the sum-constellation* $\mathcal{G} = \mathcal{X}_1 \uplus \mathcal{X}_2 \uplus \mathcal{X}_3$ *as depicted in Fig. 2.1(b). For example, once a red triangle constellation point in the northwest corner of* $\mathcal{G}$ *is observed, then we can claim that* $-\frac{1}{2} + \frac{j}{2} \in \mathcal{X}_1$, $-1 + j \in \mathcal{X}_2$ *and* $-2 + 2j \in \mathcal{X}_3$ *are transmitted in the absence of noise.*

Chapter 3

Distributed Concatenated Alamouti Code Design for Relay Networks with PSK/QAM UFCP

In this chapter, a novel distributed concatenated Alamouti code is devised for a one-way relay network consisting of two end nodes each with a single antenna, and one relay node equipped with two antennas. With the aid of the newly developed uniquely-factorable constellation pair (UFCP) generated from both phase-shift keying (PSK) and square quadrature amplitude modulation (QAM) constellations as well as by jointly processing the noisy signals received at the relay node, such a design allows the terminal nodes and the relay node to transmit their own information concurrently at the symbol level, and turns the equivalent channel between the two end nodes into a product of two Alamouti channels, thus, called UFCP concatenated Alamouti space-time block code (STBC) while maintaining the equivalent noise being still white Gaussian, thereby, leading to a symbol-by-symbol decodable optimal

maximum-likelihood (ML) receiver. In addition, an asymptotic symbol error probability (SEP) formula is derived with the ML detector, showing that the maximum diversity gain function is achieved, which is proportional to $\ln \mathtt{SNR}/\mathtt{SNR}^2$.

3.1 System Model

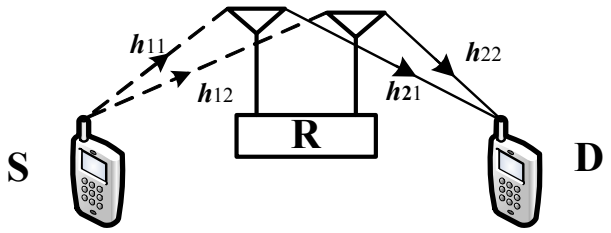


Figure 3.2: One-way dual-hop relay networks with uniquely-factorable distributed concatenated Alamouti codes

Let us consider a one-way dual-hop relay network as depicted in Fig. 3.2, where the communication between the two single-antenna terminals S and D is assisted by a relay node R equipped with two antennas. Our transmission scheme for such network is briefly described as follows. There are two different phases with each covering two consecutive time slots. The channel gain of the i-th phase of each communication associated with the j-th antenna of the relay R is denoted by h_{ij} for $i, j = 1, 2$, which are assumed to be independent and circularly symmetric complex Gaussian (CSCG) distributed with each having zero mean and the variances of which are assumed to be Ω_i, i.e., $\mathrm{E}[|h_{ij}|^2] = \Omega_i$. Let $x_\ell \in \mathcal{X}$, $y_\ell \in \mathcal{Y}, \ell = 1, 2$ be symbols to be transmitted from the source and relay, respectively, which are randomly, independently and equally likely chosen from the UFCP $\mathcal{X} \sim \mathcal{Y}$ generated from PSK or QAM constellations mentioned in Chapter 2 (see Propositions 2 and 4).

During the first and second time slots in the first communication phase, source node S transmits its message x_1 and x_2^* to R respectively, all along the channel link $\mathbf{h}_1 = [h_{11}, h_{12}]^T$. Therefore, within these two time slots, the relay R receives two signal vectors $\mathbf{r}_\ell$ for $\ell = 1, 2$, given by

$$\mathbf{r}_1 = \sqrt{P_1}\mathbf{h}_1 x_1 + \mathbf{n}_1, \tag{3.17a}$$

$$\mathbf{r}_2 = \sqrt{P_1}\mathbf{h}_1 x_2^* + \mathbf{n}_2, \tag{3.17b}$$

where P_1 is the transmission power of S in each time slot, $\mathbf{r}_\ell = [r_{\ell 1}, r_{\ell 2}]^T$, $\mathbf{n}_\ell = [n_{\ell 1}, n_{\ell 2}]^T$ denotes complex Gaussian noise vectors with each having zero mean and covariance matrix $\sigma^2\mathbf{I}$.

In the second phase, in order to enable the relay to transmit its own information together with S, the received signals $\mathbf{r}_1$ and $\mathbf{r}_2$ are multiplied by y_1 and y_2^* , respectively, to generate the composite signal z_ℓ, where $z_\ell = x_\ell y_\ell, \ell = 1, 2$, which are given by

$$\mathbf{s}_1 = \sqrt{P_1}\mathbf{h}_1 z_1 + y_1\mathbf{n}_1, \tag{3.18a}$$

$$\mathbf{s}_2 = \sqrt{P_1}\mathbf{h}_1 z_2^* + y_2^*\mathbf{n}_2. \tag{3.18b}$$

In what follows, R first properly combines these four received signals using the Alamouti coding scheme, i.e., $t_1 = s_{11} + s_{22}^*$ and $t_2 = s_{12} - s_{21}^*$, which can be re-expressed in a two-by-one vector,

$$\mathbf{t} = \sqrt{P_1}\mathbf{H}_1^T\mathbf{z} + \boldsymbol{\mu} \tag{3.19}$$

with $\mathbf{t} = [t_1, t_2]^T$, $\mathbf{z} = [z_1, z_2]^T$, and

$$\mathbf{H}_1 = \begin{bmatrix} h_{11} & h_{12} \\ h_{12}^* & -h_{11}^* \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} y_1 n_{11} + y_2 n_{22}^* \\ y_1 n_{12} - y_2 n_{21}^* \end{bmatrix}.$$

It is worth noticing that $\boldsymbol{\mu}$ is a two-by-one complex Gaussian noise vector with zero mean and covariance matrix $2\sigma^2\mathbf{I}$ for given $y_\ell, \ell = 1, 2$.

Then, R spends another consecutive two time slots on broadcasting the scaled versions of t_1 and t_2 through the two antennas to D also using the Alamouti coding scheme, or more specifically, during the following two time slots, it transmits two signal vectors $\mathbf{u}_1 = \beta[t_1, \ -t_2^*]^T$ and $\mathbf{u}_2 = \beta[t_2, \ t_1^*]^T$ to D. The scale β is a fixed value and determined to satisfy the average power constraint at the relay, i.e., $\beta^2 = \frac{1}{4(P_1\Omega_1+\sigma^2)} \approx \frac{1}{4P_1\Omega_1}$ in high SNR regime.

Hence, the signal received at D is represented by

$$\mathbf{v} = \sqrt{P_2}h_{21}\mathbf{u}_1 + \sqrt{P_2}h_{22}\mathbf{u}_2 + \boldsymbol{\eta}, \tag{3.20}$$

where P_2 is the transmission power of R in each time slot. Equation (3.20) is equivalent to

$$\bar{\mathbf{v}} = \beta\sqrt{P_2}\mathbf{H}_2\mathbf{t} + \bar{\boldsymbol{\eta}}, \tag{3.21}$$

where $\bar{\mathbf{v}} = [v_1, v_2^*]^T$ and $\bar{\boldsymbol{\eta}} = [\eta_1, \eta_2^*]^T$ denotes the two by one complex Gaussian noise vector with zero mean and covariance matrix $\sigma^2\mathbf{I}_2$, and $\mathbf{H}_2$ is an Alamouti matrix,

i.e.,

$$\mathbf{H}_2 = \begin{bmatrix} h_{21} & h_{22} \\ h_{22}^* & -h_{21}^* \end{bmatrix}.$$

Now, substituting (3.19) into (3.21) yields

$$\bar{\mathbf{v}} = \beta\sqrt{P_2 P_1}\mathbf{H}_2\mathbf{H}_1^T\mathbf{z} + \boldsymbol{\xi}, \tag{3.22}$$

where $\boldsymbol{\xi} = [\xi_1, \xi_2]^T$, in which

$$\begin{aligned} \xi_1 &= \beta\sqrt{P_2}\left[h_{21}\left(y_1 n_{11} + y_2 n_{22}^*\right) + h_{22}\left(y_1 n_{12} - y_2 n_{21}^*\right)\right] + \eta_1, \\ \xi_2 &= \beta\sqrt{P_2}\left[h_{22}^*\left(y_1 n_{11} + y_2 n_{22}^*\right) - h_{21}^*\left(y_1 n_{12} - y_2 n_{21}^*\right)\right] + \eta_2^*. \end{aligned}$$

We can now clearly see that, by carefully using the Alamouti scheme twice at the relay, the equivalent channel between S and D, i.e., $\mathbf{H}_2\mathbf{H}_1^T$ in (3.22) is essentially a product of two Alamouti matrices, thereby, still being Alamouti matrices. This is the reason why our code can be called distributed concatenated Alamouti STBC. In addition, it would be important to make the following remark:

Remark 1 *If $y \in \mathcal{Y}$ all have unit energy (e.g., $\mathcal{Y} \in \{1, -1, j, -j\}$ or $\mathcal{Y}$ taking PSK constellations), then for given h_{ij}, the noise vector $\boldsymbol{\xi}$ in* (3.22) *is Gaussian distributed with zero mean and covariance $\sigma^2[1 + 2\beta^2 P_2(|h_{21}|^2 + |h_{22}|^2)]\mathbf{I}_2$.* ■

Therefore, when the transmitted signals are randomly, independently and equally likely chosen from the constellation and channel state information is perfectly available at D, the optimal detector for estimation is the ML detector, which aims to solve

the following optimization problem,

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}} \|\bar{\mathbf{v}} - \beta\sqrt{P_2 P_1}\mathbf{H}_2\mathbf{H}_1^T\mathbf{z}\|^2. \tag{3.23}$$

After $\hat{\mathbf{z}}$ has been obtained, $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ can be uniquely determined by the property of unique factorization. The factorization process can be efficiently performed with the help of Proposition 3 and 5.

3.2 SEP Analysis for Distributed Concatenated Alamouti Codes

The primary goal of this section is to derive the asymptotic formulas of SEP in high SNR regime for the distributed concatenated Alamouti coded relay networks with UFCP generated from PSK and square QAM constellations and using the ML receiver proposed in the previous section.

For notational simplicity, we assume that $\mathbf{H}_1 = \sqrt{\Omega_1}\tilde{\mathbf{H}}_1$, $\mathbf{H}_2 = \sqrt{\Omega_2}\tilde{\mathbf{H}}_2$, where all the entries of $\tilde{\mathbf{H}}_1, \tilde{\mathbf{H}}_2$ are normalized CSCG random variables (or say Rayleigh fading). Let us recall that, the channel matrices $\mathbf{H}_2\mathbf{H}_1^T$ in (3.22) are unitary up to some scale, i.e., $(\mathbf{H}_2\mathbf{H}_1^T)^H(\mathbf{H}_2\mathbf{H}_1^T) = \Omega_2\Omega_1(|\tilde{h}_{11}|^2 + |\tilde{h}_{12}|^2)(|\tilde{h}_{21}|^2 + |\tilde{h}_{22}|^2)\mathbf{I}$, and each noise vector $\boldsymbol{\xi}$ is white Gaussian for the given h_{ij}, with the covariance matrix given by $\sigma^2[1 + 2\beta^2 P_2\Omega_2(|\tilde{h}_{21}|^2 + |\tilde{h}_{22}|^2]\mathbf{I}$. Notice that the received SNR at the terminal D is

$$\gamma = \frac{\Omega_2\Omega_1(|\tilde{h}_{11}|^2 + |\tilde{h}_{12}|^2)(|\tilde{h}_{21}|^2 + |\tilde{h}_{22}|^2)\rho}{1 + 2\beta^2 P_2\Omega_2(|\tilde{h}_{21}|^2 + |\tilde{h}_{22}|^2)}, \tag{3.24}$$

where $\rho = \frac{\beta^2 P_2 P_1}{\sigma^2}$. Hence, the optimal ML detection for (3.23) is equivalently reduced to a symbol-by-symbol detection and its arithmetic average SEP for the composite received signal with the given channel realization for PSK constellation and square-QAM constellation are:

- PSK constellation

$$\mathrm{P}_{e_{\mathrm{psk}}|(\tilde{h}_{11},\tilde{h}_{12},\tilde{h}_{21},\tilde{h}_{22})} = \frac{1}{\pi}\int_0^{\frac{(M-1)\pi}{M}} \exp\Big(-\frac{\gamma \sin^2\frac{\pi}{M}}{\sin^2\theta}\Big)\, d\theta. \tag{3.25}$$

- QAM constellation

$$\mathrm{P}_{e_{\mathrm{qam}}|(\tilde{h}_{11},\tilde{h}_{12},\tilde{h}_{21},\tilde{h}_{22})} = 4\Big(1-\frac{1}{\sqrt{M}}\Big)Q\Big(\sqrt{\frac{3\gamma}{M-1}}\Big) - 4\Big(1-\frac{1}{\sqrt{M}}\Big)^2 Q^2\Big(\sqrt{\frac{3\gamma}{M-1}}\Big). \tag{3.26}$$

The average SEP can be calculated by averaging over all the channel realizations, that is $\bar{\mathrm{P}}_e = \mathbb{E}[\mathrm{P}_{e|\tilde{h}_{11},\tilde{h}_{12},\tilde{h}_{21},\tilde{h}_{22}}]$, or equivalently:

- PSK constellation

$$\bar{\mathrm{P}}_{e_{\mathrm{psk}}} = \frac{1}{\pi}\int_0^{\frac{(M-1)\pi}{M}} J_1(\theta)d\theta, \tag{3.27}$$

where $J_1(\theta) = \mathbb{E}_{\tilde{h}_{11},\tilde{h}_{12},\tilde{h}_{21},\tilde{h}_{22}}\big[\exp\big(-\frac{\gamma\sin^2\frac{\pi}{M}}{\sin^2\theta}\big)\big]$.

- QAM constellation

$$\bar{\mathrm{P}}_{e_{\mathrm{qam}}} = 4\Big(1-\frac{1}{\sqrt{M}}\Big)\frac{1}{\pi}\int_0^{\frac{\pi}{2}} J_2(\theta)d\theta - 4\Big(1-\frac{1}{\sqrt{M}}\Big)^2\frac{1}{\pi}\int_0^{\frac{\pi}{4}} J_2(\theta)d\theta, \tag{3.28}$$

where $J_2(\theta) = \mathbb{E}_{\tilde{h}_{11},\tilde{h}_{12},\tilde{h}_{21},\tilde{h}_{22}}\big[\exp\big(\frac{-3\gamma}{2(M-1)\sin^2\theta}\big)\big]$.

Let us consider the error performance analysis of the PSK constellation and the case using QAM constellation is similar and hence the derivation process is omitted. Based on the assumption that the channel coefficients are independent, if we let $t_i = |\tilde{h}_{i1}|^2 + |\tilde{h}_{i2}|^2$, then t_1 and t_2 are independent, with each being χ_4^2-distributed, i.e., the probabilistic density function of t_i is $t_i e^{-t_i}$, $i = 1, 2$. Thus, $J_1(\theta)$ can be simply calculated by first taking the expectation over t_1 and then, t_2 such that

$$J_1(\theta) = \mathbb{E}_{t_2}\left[\mathbb{E}_{t_1}\left[\exp\left(-\frac{\tau(\theta)\rho t_1 t_2}{1+2\beta^2 P_2\Omega_2 t_2}\right)\right]\right] \tag{3.29a}$$

$$= \int_0^\infty \left(1+\frac{\tau(\theta)\rho t_2}{1+2\beta^2 P_2\Omega_2 t_2}\right)^{-2} t_2 e^{-t_2} dt_2 \tag{3.29b}$$

where $\tau(\theta) = \frac{\Omega_2\Omega_1 \sin^2\frac{\pi}{M}}{\sin^2\theta}$. To further simplify (3.29b), let $a = 2\beta^2 P_2\Omega_2, b = 2\beta^2 P_2\Omega_2 + \tau(\theta)\rho$ and rewrite it as

$$J_1(\theta) = \frac{a^2}{b^2}\int_0^\infty \frac{(t_2+a^{-1})^2}{(t_2+b^{-1})^2} \times t_2 e^{-t_2}\, dt_2. \tag{3.30}$$

Following a strategy similar to that in [23], now performing a partial fraction expansion of $\frac{(t_2+a^{-1})^2 t_2}{(t_2+b^{-1})^2}$ gives

$$\frac{(t_2+a^{-1})^2 t_2}{(t_2+b^{-1})^2} = t_2 + 2(a^{-1}-b^{-1}) + \frac{(a^{-1}-b^{-1})(a^{-1}-3b^{-1})}{t_2+b^{-1}} - \frac{(a^{-1}-b^{-1})^2 b^{-1}}{(t_2+b^{-1})^2}. \tag{3.31}$$

Substituting (3.31) into (3.30) and using a partial integral, we have

$$J_1(\theta) = \frac{a^2}{b^2}\Big\{1 + (a^{-1} - b^{-1})(2 - a^{-1} + b^{-1}) - \big[(a^{-1} - b^{-1}) (a^{-1} - 3b^{-1}) - (a^{-1} - b^{-1})^2 b^{-1}\big] e^{b^{-1}} \mathrm{Ei}(-b^{-1})\Big\}, \quad (3.32)$$

where we have used the fact [134] that $\int_0^\infty \frac{e^{-t_2}}{t_2 + b^{-1}} dt_2 = -e^{b^{-1}} \mathrm{Ei}(-b^{-1})$ with $\mathrm{Ei}(t)$ is the exponential integral function. Notice $b^{-1} = \frac{1}{\tau(\theta)\rho} + \mathcal{O}(\rho^{-2}), e^{b^{-1}} = 1 + \mathcal{O}(\rho^{-1})$ and $\mathrm{Ei}(-b^{-1}) = \mathrm{E} - \ln b + \sum_{k=1}^{\infty} (-1)^k \frac{b^{-k}}{k k!} = \mathrm{E} - \ln(\tau(\theta)\rho) + \mathcal{O}(\rho^{-2})$, where E is the Euler constant. Combining these with (3.32), and noticing that $\beta^2 \approx \frac{1}{4P_1\Omega_1}$, we can attain the following asymptotic formula for $J_1(\theta)$:

$$\begin{aligned} J_1(\theta) &= \frac{a^2 + 2a - 1 - \mathrm{E} + \ln(\tau(\theta)\rho)}{\tau^2(\theta)\rho^2} + \mathcal{O}(\rho^{-3}), \\ &= K_1(\theta)\rho^{-2} \ln \rho + K_2(\theta)\rho^{-2} + \mathcal{O}\left(\frac{\ln \rho}{\rho^3}\right), \end{aligned} \quad (3.33)$$

where $K_1(\theta) = \tau^{-2}(\theta) = \frac{\sin^4 \theta}{\Omega_2^2 \Omega_1^2 \sin^4 \frac{\pi}{M}}$, $K_2(\theta) = \tau^{-2}(\theta)(a^2 + 2a - 1 - \mathrm{E} + \ln \tau(\theta)) = K_1(\theta)\big(\frac{P_2\Omega_2(P_2\Omega_2 + 4P_1\Omega_1)}{4P_1^2\Omega_1^2} - 1 - \mathrm{E} + \ln \Omega_2 + \ln \Omega_1 + 2 \ln \sin \frac{\pi}{M} - 2 \ln \sin \theta\big)$ in which E is the Euler constant. Now, with the help of [134] and substituting (3.33) into (3.28) yields Theorem 3 below.

Theorem 3 (PSK-SEP) *The average SEP for the one way relay network with the PSK distributed concatenated Alamouti code has the following asymptotic formula:*

$$\bar{\mathrm{P}}_{e_{\mathrm{psk}}} = C_1 \rho^{-2} \ln \rho + C_2 \rho^{-2} + \mathcal{O}\left(\frac{\ln \rho}{\rho^3}\right), \quad (3.34)$$

where

$$C_1 = \frac{1}{\pi\Omega_2^2\Omega_1^2\sin^4\frac{\pi}{M}}\Big(\frac{3(M-1)\pi}{8M} - \frac{1}{4}\sin\frac{2(M-1)\pi}{M} + \frac{1}{32}\sin\frac{4(M-1)\pi}{M}\Big)$$
$$C_2 = C_1\Big[\frac{P_2\Omega_2(P_2\Omega_2+4P_1\Omega_1)}{4P_1^2\Omega_1^2} - 1 - \mathrm{E} + \ln\Omega_2 + \ln\Omega_1 + 2\ln\sin\frac{\pi}{M}\Big]$$
$$- \frac{2}{\pi\Omega_2^2\Omega_1^2\sin^4\frac{\pi}{M}}\underbrace{\int_0^{\frac{(M-1)\pi}{M}}\sin^4\theta\ln\sin\theta\,d\theta}_{\mathrm{T}_1},$$

in which

$$\mathrm{T}_1 = \Big(\frac{3(M-1)\pi}{8M} - \frac{1}{4}\sin\frac{2(M-1)\pi}{M} + \frac{1}{32}\sin\frac{4(M-1)\pi}{M}\Big)\ln\sin\frac{(M-1)\pi}{M}$$
$$+ \frac{3}{32}\sin\frac{2(M-1)\pi}{M} + \frac{3(M-1)\pi}{16M} - \frac{1}{128}\sin\frac{4(M-1)\pi}{M} + \frac{(M-1)\pi}{32M}$$
$$- \frac{3}{8}\sum_{k=0}^{\infty}(-1)^k\frac{2^{2k}B_{2k}}{(1+2k)(2k)!}\Big(\frac{(M-1)\pi}{M}\Big)^{1+2k}$$

with B_n being the Bernoulli numbers. ■

Following a similar strategy, the average SEP for QAM constellation is given by:

Theorem 4 (QAM-SEP) *The average SEP for the proposed one way relay network with the QAM distributed concatenated Alamouti codes has the following asymptotic formula:*

$$\mathrm{P}_{e_{\mathrm{qam}}} = D_1\rho^{-2}\ln\rho + D_2\rho^{-2} + \mathcal{O}\Big(\frac{\ln\rho}{\rho^3}\Big), \tag{3.35}$$

where

$$D_1 = \frac{8(\sqrt{M}-1)(M-1)^2}{9\pi M P_2^2 \Omega_2^2}(3\pi\sqrt{M}+8\sqrt{M}+3\pi-8)$$

$$D_2 = \frac{8(\sqrt{M}-1)(M-1)^2}{9\pi M P_2^2 \Omega_2^2}\Big[(3\pi\sqrt{M}+8\sqrt{M}+3\pi-8)$$
$$\times\big(\frac{P_2^2\Omega_2^2}{4P_1^2\Omega_1^2}+\frac{P_2\Omega_2}{P_1\Omega_1}+\ln P_2+\ln\Omega_2-\ln(M-1)+\ln 3-3\ln 2-\mathrm{E}-1\big)$$
$$+(6\pi\ln 2-\frac{7}{2}\pi)(\sqrt{M}+1)+(8\ln 2+6-12\mathrm{G})(\sqrt{M}-1)\Big]$$

with G *being the Catalan's constant.* ■

From Theorem 3 and 4, we can see that when SNR is large, the SEP for the one-way relay network with the proposed distributed concatenated Alamouti STBC decays as fast as $\ln\rho/\rho^2$.

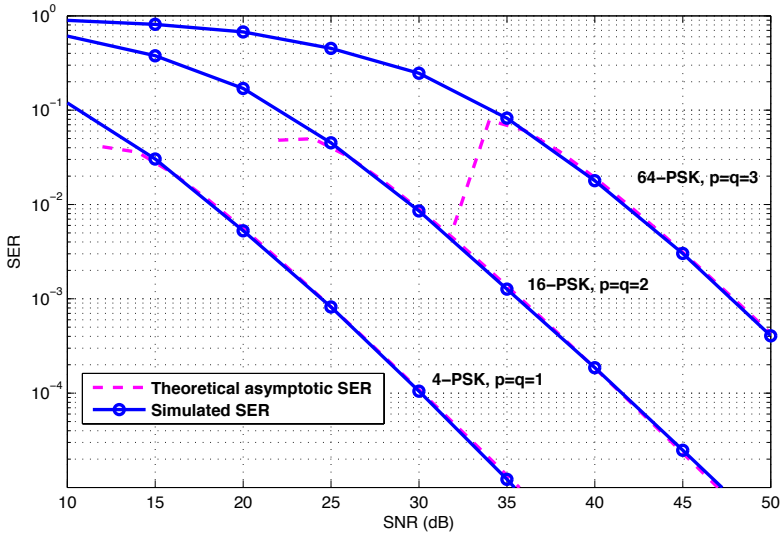


Figure 3.3: SER performance of the relay networks with uniquely-factorable distributed concatenated Alamouti codes.

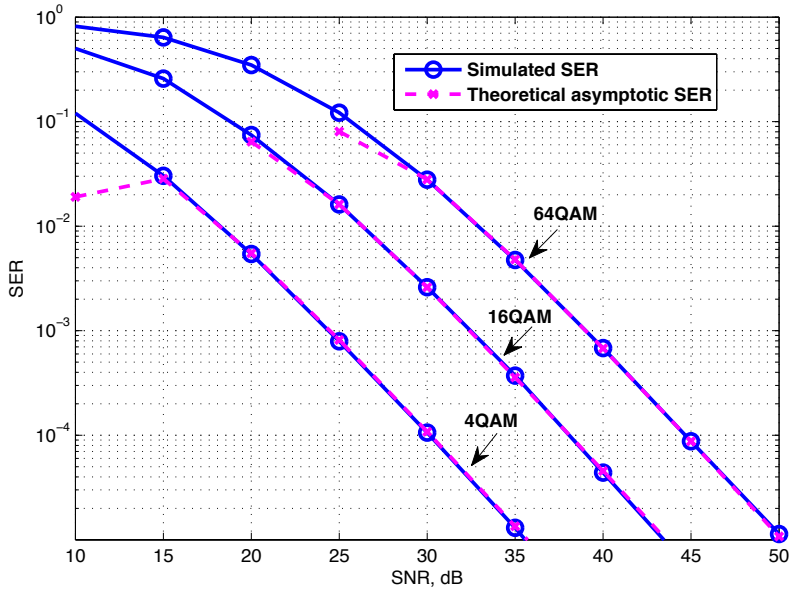


Figure 3.4: SER performance of the relay networks with uniquely-factorable distributed concatenated Alamouti codes.

3.3 Numerical Simulations

In this section, computer simulations are carried out to verify the error performance of our relay scheme. Throughout the simulation, we assume that the destination node D knows the perfect channel state information, whereas only first- and second-order channel statistics are available at the relay node. The distances between S and R, and between R and D are denoted by d_{S-R} and d_{R-D}, respectively. Therefore, the pathloss is $\Omega_1 = d_{S-R}^{-\alpha}$ and $\Omega_2 = d_{R-D}^{-\alpha}$, where $\alpha \in [2, 6]$ in most practical wireless communication systems [135].

Fig. 3.3 shows the simulated symbol error rate (SER) and the dominant theoretical SEP of the proposed relay network against the total SNR of the network with a equal power distribution between source and relay. Without loss of generality, we

set $p = q = 1, 2, 3$, which implies that $4, 16, 64$-PSK constellation are received at destination node. It can be observed that asymptotic and simulated error performance curves match very well when SNR is relatively high, which verifies the accuracy of our asymptotic SEP expression given by (3.34). In addition, the slopes of the SER curves for different PSK constellations are identical in the high SNR regime, which further affirms the conclusion that the full diversity gain function for the network is proportional to $\rho^{-2} \ln^2 \rho$.

Similarly, Fig. 3.4 shows the simulated SER and the dominant theoretical SEP of the proposed relay network versus the aggregate SNR of the network, with a equal power allocation between the source and relay node. For simulation simplicity, we choose $(|\mathcal{X}|, |\mathcal{Y}|) = (2, 2), (4, 4), (16, 4)$, which corresponds to $4, 16, 64$-QAM constellations. It can be observed that asymptotic and simulated error performance curves match perfectly when SNR is relatively high, which verifies the correctness of our asymptotic SEP expression given by (3.35). In addition, the slopes of the SER curves for different QAM constellations are identical in the high SNR regime, which also affirms the conclusion that the full diversity gain function for the proposed network is proportional to $\rho^{-2} \ln \rho$.

3.4 Conclusion

In this chapter, by using the novel concept called UFCP for both PSK and QAM constellations, we have developed a new distributed concatenated Alamouti code for a relay network consisting of two single-antenna terminals and one relay node having two antennas. This newly-designed code allows the relay to transmit its own information. By taking advantage of the Alamouti coding scheme twice and jointly

processing the signals from the two antennas at the relay node, the equivalent channel between source and destination becomes a product of the two Alamouti channels, thus, called distributed concatenated Alamouti STBC. In addition, the asymptotic symbol error probability (SEP) formula is attained for the maximum likelihood (ML), showing that the maximum diversity gain is achieved and proportional to $\ln \mathtt{SNR}/\mathtt{SNR}^2$.

Chapter 4

Optimally Precoded Large MIMO Fading Channels and Asymptotic SEP Analysis with Zero-Forcing Detection

This chapter considers the asymptotic analysis of symbol error probability (SEP) for either optimally precoded or uniformly precoded large correlated MIMO fading channels using the zero-forcing (ZF) detector and equally likely PAM, PSK or square QAM signalling points. For such systems, we reveal some very nice structures which naturally lead us to the exploration of two very strong and very useful mathematical tools for the systematic study of asymptotic behaviors on their error performance. The first tool is the Szegö's theorem on large Hermitian Toeplitz matrices and the second tool is the well known limit: $\lim_{x\to\infty}(1+1/x)^x = e$. This new approach

enables us to attain a very simple expression for the SEP limit as the number of the transmitter antennas goes to infinity. One of the major advantages for this method is that its convergence rate is very fast. Hence, this expression is very efficient and effective SEP approximation for the large MIMO systems. Due to the constraint on allowable space and limited (sparse) multi-path scattering, fading correlation between neighbouring antenna elements is almost inevitable in a large MIMO architecture. By specifically examining an exponential correlation matrix model, we show that the optimal precoding technique can yield substantial power gain over the uniform power allocation strategy.

4.1 Precoded Transmission Model with Zero-Forcing Detection

4.1.1 System Model

Consider the complex baseband-equivalent model of a narrow-band MIMO communication system, with N_t transmitting antennas and N_r receiving antennas. The receiver is assumed to have more antennas than the transmitter ($N_r > N_t$) just like that in the V-BLAST system [136]. The sequence of serially transmitted symbols is first de-multiplexed into a vector signal $\mathbf{s} = [s_1, s_2, \cdots, s_{N_t}]^T$, and then this vector signal $\mathbf{s}$ is transformed by an $N_t \times N_t$ full-rank square precoding matrix $\mathbf{F}$ into another vector signal $\mathbf{x} = [x_1, x_2, \cdots, x_{N_t}]^T = \mathbf{Fs}$. Then, each element x_k of $\mathbf{x}$ is fed to the k-th transmitter antenna for transmission. In the space-time communication, N_t transmitter each can transmit uncoded M-ary PAM, PSK or QAM symbols using

the same waveform during the same time interval, but we assume that the same constellation is used for each antenna. At the receiver array, the discrete received $N_r \times 1$ signal vector $\mathbf{r}$ can be written as

$$\mathbf{r} = \mathbf{HFs} + \boldsymbol{\xi}, \tag{4.36}$$

where $\mathbf{H}$ is the $N_r \times N_t$ complex channel matrix and $\boldsymbol{\xi}$ is the $N_r \times 1$ additive complex noise vector. Throughout this chapter, we adopt the following assumptions:

1. *The perfect channel estimates are available at the receiver to allow coherent detection*;
2. *The channel* $\mathbf{H}$ *is complex Gaussian distributed, with zero-mean, and covariance matrix* $\mathbf{I} \otimes \boldsymbol{\Sigma}$;
3. $\boldsymbol{\xi}$ *is circularly-symmetric complex Gaussian noise with covariance* $\sigma^2\mathbf{I}$;
4. *Each element of* $\mathbf{s}$ *is independently and equally likely chosen from PAM, PSK or QAM constellations of the same size with the covariance matrix of* $\mathbf{s}$ *being* $\mathbf{I}$;
5. *The total power budget of the transmitter array is unified to one, and as a consequence the system SNR is defined as* $\eta \triangleq 1/\sigma^2$.

4.1.2 Zero-Forcing Equalization

Suppose that we use zero-forcing equalization to recover the information symbols. To this end, first we obtain the pseudo inverse of the super-channel matrix $\mathbf{HF}$; i.e.

$$(\mathbf{HF})^{+} = \left(\mathbf{F}^H\mathbf{H}^H\mathbf{HF}\right)^{-1}\mathbf{F}^H\mathbf{H}^H. \tag{4.37}$$

Here we need to explain why the inverse in (4.37) exists. Under Assumption 2 above, the matrix $\mathbf{H}^H\mathbf{H}$ is the Wishart distribution and as result, $\mathbf{F}^H\mathbf{H}^H\mathbf{H}\mathbf{F}$ is also subject to the Wishart distribution [137, 138]. Therefore, the inverse (4.37) exists with probability one if the number of receiving antennas is not less than that of transmitting antennas [137, 138].

The ZF detection is captured by the following two steps:

1. Perform ZF equalization. Multiplying both sides of equation (4.36) by the pseudo-inverse $(\mathbf{HF})^+$ of $\mathbf{HF}$, we get

$$\mathbf{r}' = \mathbf{s} + \boldsymbol{\xi}', \tag{4.38}$$

   where $\mathbf{r}' = (\mathbf{HF})^+\mathbf{r}$ and $\boldsymbol{\xi}' = (\mathbf{HF})^+\boldsymbol{\xi}$. Under Assumption 3, $\boldsymbol{\xi}'$ is the circularly-symmetric complex Gaussian noise with covariance $\sigma^2(\mathbf{F}^H\mathbf{H}^H\mathbf{H}\mathbf{F})^{-1}$;

2. Perform a hard decision to obtain an estimate $\hat{s}_k$ of s_k, i.e.,

$$\hat{s}_k = \arg \min_{s_k \in \mathcal{S}} |r'_k - s_k|^2.$$

Since ZF equalizer performs a memoryless detection, i.e., the decision on the current symbol does not affect the decision on the next symbol. Therefore, the average SEP over one vector signal $\mathbf{s}$ just is the arithmetic mean of all SEPs.

4.2 Optimal Precoders Minimizing the Average Symbol Error Probability

Our primary purpose of this section is to first give an explicit convex region for which the optimally precoding matrix $\mathbf{F}$ that minimizes the average SEP of the ZF detector can be obtained and then, to uncover some nice structures for the optimally precoded system, which naturally leads us to taking full advantage of the Szegö's theorem for the systematic study of the asymptotic behaviour on the resulting large MIMO systems.

4.2.1 SEP Expressions for M-ary PAM, PSK and QAM signals

PAM signals

The SEP for M-ary PAM signal s_k is

$$\mathrm{P}_{\mathrm{PAM}}(\mathbf{H},\mathbf{F},s_k)=\frac{2(M-1)}{M}Q\left(\sqrt{\frac{6\eta}{(M^2-1)[(\mathbf{F}^H\mathbf{H}^H\mathbf{H}\mathbf{F})^{-1}]_k}}\right).$$

Therefore, the arithmetic average of all SEPs in one block is

$$\mathrm{P}_{\mathrm{PAM}}(\mathbf{H},\mathbf{F})=\frac{2(M-1)}{MN_t}\sum_{k=1}^{N_t}Q\left(\sqrt{\frac{6\eta}{(M^2-1)[(\mathbf{F}^H\mathbf{H}^H\mathbf{H}\mathbf{F})^{-1}]_k}}\right). \tag{4.39}$$

For our purpose, we now prefer to use another expression for the Gaussian Q-function [139, 140] i.e.,

$$Q(t) = \frac{1}{\pi}\int_0^{\pi/2} \exp\left(-\frac{t^2}{2\sin^2\theta}\right) d\theta, \qquad t \geq 0. \tag{4.40}$$

Substituting (4.40) into (4.39) yields

$$\mathrm{P}_{\mathrm{PAM}}(\mathbf{H},\mathbf{F}) = \frac{2(M-1)}{MN_t\pi}\sum_{k=1}^{N_t}\int_0^{\pi/2} \exp\left(-\frac{3\eta}{(M^2-1)[(\mathbf{F}^H\mathbf{H}^H\mathbf{H}\mathbf{F})^{-1}]_k \sin^2\theta}\right) d\theta. \tag{4.41}$$

It is known [137,138] that $\gamma = \frac{[(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_k}{[(\mathbf{F}^H\mathbf{H}^H\mathbf{H}\mathbf{F})^{-1}]_k}$ is subject to $\mathcal{X}^2_{2(N_r-N_t+1)}$, i.e., its density function is

$$f(\gamma) = \frac{1}{\Gamma(N_r-N_t+1)} e^{-\gamma}\gamma^{N_r-N_t} \qquad \text{for } \gamma > 0,$$

where $\Gamma(t)$ denotes the gamma function. Now, taking the expectation in (4.41) over random channel $\mathbf{H}$ yields

$$\begin{aligned}
\mathrm{P}_{\mathrm{PAM}}(\mathbf{F}) &= \frac{2(M-1)}{MN_t\pi}\sum_{k=1}^{N_t}\int_0^{\pi/2} \mathbb{E}_{\mathbf{H}}\exp\left(-\frac{3\eta}{(M^2-1)[(\mathbf{F}^H\mathbf{H}^H\mathbf{H}\mathbf{F})^{-1}]_k \sin^2\theta}\right) d\theta \\
&= \frac{2(M-1)}{MN_t\pi}\sum_{k=1}^{N_t}\int_0^{\pi/2} \mathbb{E}_{\gamma}\exp\left(-\frac{3\eta\gamma}{(M^2-1)[(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_k \sin^2\theta}\right) d\theta \\
&= \frac{2(M-1)}{MN_t\pi}\sum_{k=1}^{N_t}\int_0^{\pi/2} \left(1+\frac{3\eta}{(M^2-1)[(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_k \sin^2\theta}\right)^{-(N_r-N_t+1)} d\theta \\
&= \frac{1}{N_t}\sum_{k=1}^{N_t} G_{\mathrm{PAM}}\left([(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_k\right),
\end{aligned} \tag{4.42}$$

where function $G_{\mathrm{PAM}}(t)$ is defined as

$$G_{\mathrm{PAM}}(t) = \frac{2(M-1)}{M\pi} \int_0^{\pi/2} \left(1 + \frac{3\eta}{(M^2-1)t\sin^2\theta}\right)^{-(N_r-N_t+1)} d\theta.$$

PSK signals

The SEP for M-ary PSK signal s_k is

$$\mathrm{P}_{\mathrm{PSK}}(\mathbf{H}, \mathbf{F}, s_k) = \frac{1}{\pi} \int_0^{(M-1)\pi/M} \exp\left(-\frac{\eta \sin^2(\pi/M)}{[(\mathbf{F}^H\mathbf{H}^H\mathbf{H}\mathbf{F})^{-1}]_k \sin^2\theta}\right) d\theta.$$

Therefore, the arithmetic mean of all SEPs is

$$\mathrm{P}_{\mathrm{PSK}}(\mathbf{H}, \mathbf{F}) = \frac{1}{N_t\pi} \sum_{k=1}^{N_t} \int_0^{(M-1)\pi/M} \exp\left(-\frac{\eta \sin^2(\pi/M)}{[(\mathbf{F}^H\mathbf{H}^H\mathbf{H}\mathbf{F})^{-1}]_k \sin^2\theta}\right) d\theta. \quad (4.43)$$

Similarly, taking the expectation of (4.43) over random channel $\mathbf{H}$ produces

$$\begin{aligned} \mathrm{P}_{\mathrm{PSK}}(\mathbf{F}) &= \frac{1}{N_t\pi} \sum_{k=1}^{N_t} \int_0^{(M-1)\pi/M} \left(1 + \frac{\eta \sin^2(\pi/M)}{[(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_k \sin^2\theta}\right)^{-(N_r-N_t+1)} d\theta \\ &= \frac{1}{N_t} \sum_{k=1}^{N_t} G_{\mathrm{PSK}}\left([(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_k\right), \end{aligned}$$

where function $G_{\mathrm{PSK}}(t)$ is defined as

$$G_{\mathrm{PSK}}(t) = \frac{1}{\pi} \int_0^{(M-1)\pi/M} \left(1 + \frac{\eta \sin^2(\pi/M)}{t\sin^2\theta}\right)^{-(N_r-N_t+1)} d\theta.$$

QAM signals

The SEP for M-ary QAM signal s_k is

$$\begin{aligned}\mathrm{P}_{\mathrm{QAM}}(\mathbf{H},\mathbf{F},s_k) = 4\left(1-1/\sqrt{M}\right)Q\left(\sqrt{\frac{3\eta}{(M-1)[(\mathbf{F}^H\mathbf{H}^H\mathbf{H}\mathbf{F})^{-1}]_k}}\right)\\ -4\left(1-1/\sqrt{M}\right)^2 Q^2\left(\sqrt{\frac{3\eta}{(M-1)[(\mathbf{F}^H\mathbf{H}^H\mathbf{H}\mathbf{F})^{-1}]_k}}\right) \qquad (4.44)\end{aligned}$$

The first term in (4.44) can be replaced by (4.40). Similarly, $Q^2(\cdot)$ function also has a very nice formula [140],

$$Q^2(t) = \frac{1}{\pi}\int_0^{\pi/4}\exp\left(-\frac{t^2}{2\sin^2\theta}\right)d\theta. \qquad (4.45)$$

Substituting (4.40) and (4.45) into (4.44) and then, taking the expectation over the random channel matrix, we can obtain

$$\begin{aligned}\mathrm{P}_{\mathrm{QAM}}(\mathbf{F}) &= \frac{4(\sqrt{M}-1)}{\sqrt{M}N_t\pi}\sum_{k=1}^{N_t}\int_0^{\pi/2}\left(1+\frac{3\eta}{2(M-1)[(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_k\sin^2\theta}\right)^{-(N_r-N_t+1)}d\theta\\ &\quad -\frac{4(\sqrt{M}-1)^2}{MN_t\pi}\sum_{k=1}^{N_t}\int_0^{\pi/4}\left(1+\frac{3\eta}{2(M-1)[(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_k\sin^2\theta}\right)^{-(N_r-N_t+1)}d\theta\\ &= \frac{4(\sqrt{M}-1)}{\sqrt{M}N_t\pi}\sum_{k=1}^{N_t}\int_{\pi/4}^{\pi/2}\left(1+\frac{3\eta}{2(M-1)[(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_k\sin^2\theta}\right)^{-(N_r-N_t+1)}d\theta\\ &\quad +\frac{4(\sqrt{M}-1)}{MN_t\pi}\sum_{k=1}^{N_t}\int_0^{\pi/4}\left(1+\frac{3\eta}{2(M-1)[(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_k\sin^2\theta}\right)^{-(N_r-N_t+1)}d\theta\\ &= \frac{1}{N_t}\sum_{k=1}^{N_t}G_{\mathrm{QAM}}\left([(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_k\right), \qquad (4.46)\end{aligned}$$

where function $G_{\mathrm{QAM}}(t)$ is defined as

$$\begin{aligned} G_{\mathrm{QAM}}(t) &= \frac{4(\sqrt{M}-1)}{\sqrt{M}\pi}\int_{\pi/4}^{\pi/2}\left(1+\frac{3\eta}{2(M-1)t\sin^2\theta}\right)^{-(N_r-N_t+1)}d\theta \\ &\quad +\frac{4(\sqrt{M}-1)}{M\pi}\int_{0}^{\pi/4}\left(1+\frac{3\eta}{2(M-1)t\sin^2\theta}\right)^{-(N_r-N_t+1)}d\theta. \end{aligned}$$

4.2.2 Convexity of Objective Functions

In order to obtain the optimal precoder, let us discuss the convexity of function $G(t)$.

1. For PAM signals, the second-order derivative of $G_{\mathrm{PAM}}(t)$ is given by

$$\begin{aligned} \frac{d^2G_{\mathrm{PAM}}(t)}{dt^2} &= \frac{2(M-1)}{M\pi}\int_{0}^{\pi/2}\left(1+\frac{3\eta}{(M^2-1)t\sin^2\theta}\right)^{-(N_r-N_t+3)} \\ &\quad \times\frac{3(N_r-N_t+1)}{(M^2-1)\sigma^2t^4\sin^2\theta}\left(\frac{3\eta(N_r-N_t)}{(M^2-1)\sin^2\theta}-2t\right)d\theta. \end{aligned}$$

   Since $3\eta(N_r-N_t)/\big((M^2-1)\sin^2\theta\big)-2t \geq 3\eta(N_r-N_t)/((M^2-1))-2t$, then if the following condition is satisfied, i.e.,

$$0 < t \leq \frac{3\eta(N_r-N_t)}{2(M^2-1)} = T_{\mathrm{PAM}},$$

   we have $d^2G_{\mathrm{PAM}}(t)/dt^2 \geq 0$. This implies that $G_{\mathrm{PAM}}(t)$ is convex in this interval.

2. For PSK signals, we have

$$\begin{aligned} \frac{d^2G_{\mathrm{PSK}}(t)}{dt^2} &= \frac{1}{\pi}\int_{0}^{(M-1)\pi/M}\left(1+\frac{\eta\sin^2(\pi/M)}{t\sin^2\theta}\right)^{-(N_r-N_t+3)} \\ &\quad \times\frac{\eta(N_r-N_t+1)\sin^2(\pi/M)}{t^4\sin^2\theta}\left(\frac{\eta(N_r-N_t)\sin^2(\pi/M)}{\sin^2\theta}-2t\right)d\theta. \end{aligned}$$

Since $\eta(N_r - N_t)\sin^2(\pi/M)/(\sin^2\theta) - 2t \geq \eta(N_r - N_t)\sin^2(\pi/M) - 2t \geq 0$, we have that if

$$0 < t \leq \frac{\eta(N_r - N_t)\sin^2(\pi/M)}{2} = T_{\text{PSK}},$$

then, in this interval, $d^2G_{\text{PSK}}(t)/dt^2 \geq 0$. This shows that $G_{\text{PSK}}(t)$ is convex in this range.

3. For QAM signals, we have

$$\begin{aligned}\frac{d^2G_{\text{QAM}}(t)}{dt^2} = 4(1-1/\sqrt{M})/\pi \int_{\pi/4}^{\pi/2} \left(1 + \frac{3\eta}{2(M-1)t\sin^2\theta}\right)^{-(N_r-N_t+3)} \\ \times \frac{3\eta(N_r - N_t + 1)}{2(M-1)t^4\sin^2\theta}\left(\frac{3\eta(N_r-N_t)}{2(M-1)\sin^2\theta} - 2t\right)d\theta \\ +4(1-1/\sqrt{M})/(\sqrt{M}\pi)\int_0^{\pi/4}\left(1 + \frac{3\eta}{2(M-1)t\sin^2\theta}\right)^{-(N_r-N_t+3)} \\ \times \frac{3\eta(N_r - N_t + 1)}{2(M-1)t^4\sin^2\theta}\left(\frac{3\eta(N_r-N_t)}{2(M-1)\sin^2\theta} - 2t\right)d\theta.\end{aligned}$$

Since $3\eta(N_r - N_t)/(2(M-1)\sin^2\theta) - 2t \geq 3\eta(N_r - N_t)/(2(M-1)) - 2t \geq 0$, if the following condition meets,

$$0 < t \leq \frac{3\eta(N_r - N_t)}{4(M-1)} = T_{\text{QAM}},$$

then, $d^2G_{\text{QAM}}(t)/dt^2 \geq 0$ and as a result, $G_{\text{QAM}}(t)$ is a convex function in this interval.

To obtain a unified expression, we drop the subscripts of both $G(t)$ and T. Hence, $G(t)$ is convex if $0 < t \leq T$. Correspondingly, the noise power associated with $\mathbf{F}$

satisfies the following condition,

$$[(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_k \;\leq\; T \qquad \text{for } k = 1, 2, \cdots, N_t. \tag{4.47}$$

4.2.3 Explicit Convex Regions and Optimal Precoders

To develop an explicit constraint from (4.47), we need to introduce the following two lemmas.

Lemma 1 (Rayleigh-Ritz) *[141] Let* $\mathbf{A} \in \mathbb{C}^{K\times K}$ *be Hermitian and let* $\zeta_1(\mathbf{A}) \leq \zeta_2(\mathbf{A}) \leq \ldots \leq \zeta_K(\mathbf{A})$ *be the eigenvalues of* $\mathbf{A}$*. Then*

$$\zeta_1(\mathbf{A}) = \min_{\|\mathbf{x}\|\neq 0} \frac{\mathbf{x}^H\mathbf{A}\mathbf{x}}{\mathbf{x}^H\mathbf{x}}, \quad \zeta_K(\mathbf{A}) = \max_{\|\mathbf{x}\|\neq 0} \frac{\mathbf{x}^H\mathbf{A}\mathbf{x}}{\mathbf{x}^H\mathbf{x}}.$$

■

Since $(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}$ is Hermitian and the k-th diagonal value $[(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_k = \mathbf{e}_k^H(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}\mathbf{e}_k$, where $\mathbf{e}_k = [0, \cdots, 0, 1, 0, \cdots, 0]^H$ has 1 only in the k-th entry, by Lemma 1 we have

$$\zeta_1\big((\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}\big) \leq [(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_k \leq \zeta_{N_t}\big((\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}\big), \text{ for } k = 1, 2, \ldots, N_t,$$

where $\zeta_1\big((\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}\big)$ and $\zeta_{N_t}\big((\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}\big)$ are the minimum and maximum eigenvalues of $(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}$, respectively and the equality is attainable when $\mathbf{F}$ diagonalizes $\mathbf{\Sigma}$. Therefore, if $\mathbf{F}$ satisfies

$$\zeta_{N_t}\big((\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}\big) \leq T, \tag{4.48}$$

then, such $\mathbf{F}$ also satisfies (4.47). To further simplify the constraint, we need another lemma.

Lemma 2 (Ostrowski) *Let $\mathbf{A} \in \mathbb{C}^{K\times K}$ be Hermitian and $\mathbf{S} \in \mathbb{C}^{K\times K}$ be nonsingular. If we let the eigenvalues of $\mathbf{A}$ and $\mathbf{SS}^H$ be given by $\zeta_1(\mathbf{A}) \leq \zeta_2(\mathbf{A}) \leq \ldots \leq \zeta_K(\mathbf{A})$ and $\zeta_1(\mathbf{SS}^H) \leq \zeta_2(\mathbf{SS}^H) \leq \ldots \leq \zeta_K(\mathbf{SS}^H)$, respectively, then, for each $i = 1, 2, \ldots, K$, there exists a positive real number κ_i such that $\zeta_1(\mathbf{SS}^H) \leq \kappa_i \leq \zeta_K(\mathbf{SS}^H)$ and $\zeta_i(\mathbf{SAS}^H) = \kappa_i\zeta_i(\mathbf{A})$.* ■

Let the eigenvalue decomposition of $\mathbf{\Sigma}$ be $\mathbf{\Sigma} = \mathbf{W\Lambda W}^H$, where $\mathbf{W}$ is a unitary matrix, and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_{N_t})$ with $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{N_t}$. Let the singular value decomposition (SVD) of $\mathbf{F}$ be $\mathbf{F} = \mathbf{UDV}$, and $\mathbf{D} = \text{diag}(\sqrt{d_1}, \sqrt{d_2}, \cdots, \sqrt{d_{N_t}})$, where $0 < d_1 \leq d_2 \leq \cdots \leq d_{N_t}$, since $\mathbf{F}$ is assumed to be of full-rank (nonsingular). Then, by Lemma 2 with $K = N_t$, $\mathbf{S} = \mathbf{F}^{-1}$ and $\mathbf{A} = \mathbf{\Sigma}^{-1}$, we have

$$\zeta_{N_t}\big((\mathbf{F}^H\mathbf{\Sigma F})^{-1}\big) = \zeta_{N_t}\big(\mathbf{F}^{-1}\mathbf{\Sigma}^{-1}\mathbf{F}^{-H}\big) \leq \zeta_{N_t}(\mathbf{\Sigma}^{-1})\zeta_{N_t}\big((\mathbf{F}^H\mathbf{F})^{-1}\big) = \frac{1}{\lambda_1 d_1}, \quad (4.49)$$

where the equality is also attainable when $\mathbf{U} = \mathbf{W}$. Therefore, if

$$d_1 \geq \frac{1}{\lambda_1 T}, \quad (4.50)$$

then, $\mathbf{F}$ satisfies constraint (4.47). Constraint (4.50) requires that the minimal average transmitting power of the subchannels, d_1 must be larger than certain predefined threshold that is related to the modulation signals, system SNR η and channel statistics. Since T is proportional to system SNR η, constraint (4.50) is valid in slightly high SNR regime. Now, following the same way as [40,142,143] and applying Jensen's

inequality [144] to function $G(t)$ under the constraint (4.50) result in

$$\frac{1}{N_t}\sum_{k=1,d_1\geq\frac{1}{\lambda_1 T}}^{N_t} G\left([(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_k\right) \geq G\left(\frac{1}{N_t}\sum_{k=1}^{N_t}[(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_k\right), \tag{4.51}$$

where the equality in (4.51) holds if and only if

$$[(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_1 = [(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_2 = \cdots = [(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_{N_t}. \tag{4.52}$$

On the other hand, by a well known trace-inequality [145], we have

$$\mathrm{tr}\left((\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}\right) \;=\; \mathrm{tr}\left((\mathbf{F}\mathbf{F}^H)^{-1}\mathbf{\Sigma}^{-1}\right) \geq \sum_{k=1}^{N_t} d_{N_t+1-k}^{-1}\lambda_k^{-1}, \tag{4.53}$$

where the equality in (4.53) holds if $\mathbf{U} = \mathbf{W}\mathbf{P}$, where $\mathbf{P}$ is an anti-diagonal permutation matrix given by

$$\mathbf{P} = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 1 & \cdots & 0 & 0 \end{bmatrix}.$$

Then, using the Cauchy-Schwarz inequality, we can attain

$$\sum_{k=1}^{N_t}\sqrt{d_{N_t+1-k}}\cdot\frac{1}{\sqrt{d_{N_t+1-k}\lambda_k}} \leq \sqrt{\sum_{\ell=1}^{N_t} d_\ell}\cdot\sqrt{\sum_{k=1}^{N_t}\frac{1}{d_{N_t+1-k}\lambda_k}}.$$

Combining this with the power constraint $\mathrm{tr}(\mathbf{F}^H\mathbf{F}) = 1$ gives us

$$\sum_{k=1}^{N_t} d_{N_t+1-k}^{-1}\lambda_k^{-1} \geq \left(\sum_{k=1}^{N_t}\lambda_k^{-1/2}\right)^2. \tag{4.54}$$

The equality in (4.54) holds if and only if

$$d_{N_t+1-k} = \frac{\lambda_k^{-1/2}}{\sum_{\ell=1}^{N_t}\lambda_\ell^{-1/2}}, \qquad k = 1, 2, \ldots, N_t. \tag{4.55}$$

Since $G(t)$ monotonically increases, combining (4.55) with (4.51) leads to

$$\frac{1}{N_t}\sum_{k=1, d_1 \geq \frac{1}{\lambda_1 T}}^{N_t} G\left([(\mathbf{F}^H\mathbf{\Sigma}\mathbf{F})^{-1}]_k\right) \geq G\bigg(\Big(\sum_{k=1}^{N_t}\lambda_k^{-1/2}\Big)^2/N_t\bigg),$$

where the equality holds if $\mathbf{U} = \mathbf{WP}$, the square of the singularvalues of $\mathbf{F}$ meets (4.55) and $\mathbf{V}$ is chosen as the normalized DFT matrix. For presentation clarity, all the above discussions can be summarized as the following theorem.

Theorem 5 *Let the eigenvalue decomposition of* $\mathbf{\Sigma}$ *be* $\mathbf{\Sigma} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^H$*, where* $\mathbf{W}$ *is a unitary matrix, and* $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_{N_t})$ *with* $\lambda_1 \leq \lambda_2 \leq \cdots \lambda_{N_t}$*. If* $\mathbf{F}$ *is restricted to be in set* $\left\{\mathbf{F} : \zeta_1(\mathbf{F}^H\mathbf{F}) \geq \frac{1}{\lambda_1 T}\right\}$*, where* T *is the threshold in* (4.47) *and* $\zeta_1(\mathbf{F}^H\mathbf{F})$ *is the minimum eigenvalues of* $\mathbf{F}^H\mathbf{F}$*. Then, the optimal precoder minimizing the average SEP is given by*

$$\widetilde{\mathbf{F}} = \frac{1}{\sqrt{\mathrm{tr}(\mathbf{\Lambda}^{-1/2})}}\mathbf{W}\mathbf{\Lambda}^{-1/4}\widetilde{\mathbf{V}}_2, \tag{4.56}$$

where $\widetilde{\mathbf{V}}_2$ *is the* $N_t \times N_t$ *normalized DFT matrix, and the resulting minimum average*

SEP is determined by

$$P_{min}(\widetilde{\mathbf{F}}) = G\bigg(\Big(\sum_{k=1}^{N_t}\lambda_k^{-1/2}\Big)^2/N_t\bigg). \tag{4.57}$$

■

We would like to make the following two comments on Theorem 5.

1. The optimal precoder design problems with ZF detection were also considered in [40, 142]. However, our Theorem 5 provides an explicitly sufficient condition that guarantees the optimality of the proposed precoder.

2. Here, it is highly worth pointing out that the resulting SEP for the optimal precoder exposes a very interesting structure which motivates us to systematically study the asymptotic SEP performance in large MIMO systems.

4.3 Asymptotic SEP Analysis for Optimally Precoded Large MIMO Systems

In this section, our main purpose is to investigate the asymptotic behavior of SEP for the optimally precoded correlated large MIMO systems equipped with the ZF receiver. The array size of both the transmitter and the receiver is assumed to go unbounded while maintaining a constant ratio between them.

4.3.1 Array Correlation Model

In general, MIMO techniques can yield linear increasing in the data rate against the minimum number of the transmitter and the receiver antennas in rich scattering

environment, particularly when the array elements are uncorrelated [24]. However, in a practical radio propagation process, correlation is almost inevitable, especially for a large MIMO architecture. In this chapter, we assume that the transmitter array is arbitrarily correlated and that the correlation between each element of the receiver array is negligible. This case can be considered as a MIMO system in the up-link where the transmitter is a mobile terminal with correlated array and the receiver is a base station, where the distance between adjacent antenna elements can be made as large as desired to eliminate correlation. To facilitate our analysis, we also assume that the correlation matrix is a Hermitian Toeplitz matrix [38]. This is a simplified model of measurement in practical environment, but it can capture the main phenomenon of spatial correlation between antennas (see, e.g., [37] for other models). This model enables us to completely take advantage of the structure provided by the optimal system as well as of the Szegö's theorem on large Hermitian Toeplitz matrices so that we can attain a simple closed-form solution in terms of the correlation coefficients, from which some important insightful information on the effect of correlation can be extracted.

4.3.2 Asymptotic Behaviour of Large Toeplitz Matrices

To fully make use of the optimal structure provided by (4.57) for our analysis on the asymptotic behavior of the statistical average SEP, let us review an important property on a sequence of large Hermitian Toeplitz matrices $\{\mathbf{T}_K\}_{K=1}^{\infty}$. Without loss

of generality, we let

$$\mathbf{T}_K = \begin{bmatrix} t(0) & t(-1) & \cdots & t(-(K-1)) \\ t(1) & t(0) & \cdots & t(-(K-2)) \\ \vdots & \vdots & \ddots & \vdots \\ t(K-1) & t(K-2) & \cdots & t(0) \end{bmatrix}. \tag{4.58}$$

where $t(k) = t^*(-k)$ and $t(k)$ are assumed to be absolutely square-summable, i.e., $\sum_{k=-\infty}^{\infty} |t(k)|^2 < \infty$. Thus, the following pair of discrete-time Fourier transforms exists,

$$\begin{aligned} s_{\mathbf{T}}(\omega) &= \sum_{k=-\infty}^{\infty} t(k) e^{-jk\omega}, \\ t(k) &= \frac{1}{2\pi} \int_0^{2\pi} s_{\mathbf{T}}(\omega) e^{jk\omega} d\omega. \end{aligned}$$

It is worth noting that the function $s_{\mathbf{T}}(w)$ is real, since $\mathbf{T}$ is Hermitian and $s_T(w)$ is also known as the power spectral density (PSD) function. The above relationship is also known as the Wiener-Khinchin theorem of a discrete-time process.

Lemma 3 (Szegö's theorem) *[44] Let $\{\mathbf{T}_K\}_{K=1}^{\infty}$ be a sequence of Hermitian Toeplitz matrices with K eigenvalues of $\mathbf{T}_K$ given by $\mu_{K,1} \leq \mu_{K,2} \leq \cdots \leq \mu_{K,K}$, and $\sum_{k=0}^{\infty} |t(k)|^2$ being convergent. Then for any function $\mathtt{F}(x)$ that is continuous on $[L_{s_{\mathbf{T}}}, U_{s_{\mathbf{T}}}]$, we have*

$$\lim_{K\to\infty} \frac{1}{K} \sum_{\ell=1}^{K} \mathtt{F}(\mu_{K,\ell}) = \frac{1}{2\pi} \int_0^{2\pi} \mathtt{F}(s_{\mathbf{T}}(\omega)) d\omega, \tag{4.59}$$

where $L_{s_{\mathbf{T}}} = ess\ inf\ s_{\mathbf{T}}(\omega)$ is the essential infimum of $s_{\mathbf{T}}(\omega)$ and defined to be the

largest value of c for which $s_{\mathbf{T}}(\omega) \geq c$ except on a set of measure 0, and $U_{s_{\mathbf{T}}} =$ ess sup $s_{\mathbf{T}}(\omega)$ is the smallest number d for which $s_{\mathbf{T}}(\omega) \leq d$ except for a set of measure 0 ■

This lemma plays a vital role in our chapter. From now on, we assume that the ratio of the number of the receiver antennas to that of the transmitter antennas is fixed, i.e., $N_r/N_t = \beta > 1$ is constant. Now, our main result of this chapter can be formally stated as the following theorem.

Theorem 6 *Let us consider large MIMO systems using the optimal precoder in* (4.56)*, the ZF detector and the M-ary PAM, PSK or QAM constellations. If the entries of the channel covariance matrix $\mathbf{\Sigma}$ are absolutely square-summable and the resulting $L_{s_\Sigma} > 0$, then, $\lim_{N_t\to\infty} \mathrm{P}_{N_t}(\widetilde{\mathbf{F}}) = \bar{\mathrm{P}}_{\mathrm{opt}}$ exists and*

- $\bar{\mathrm{P}}_{\mathrm{opt,PAM}} = \frac{2(M-1)}{M} Q\left(\sqrt{\frac{6\eta(\beta-1)}{(M^2-1)\Lambda^2}}\right)$;
- $\bar{\mathrm{P}}_{\mathrm{opt,PSK}} = \frac{1}{\pi}\int_0^{(M-1)\pi/M} \exp\left(-\frac{\eta(\beta-1)\sin^2(\pi/M)}{\Lambda^2\sin^2\theta}\right) d\theta$;
- $\bar{\mathrm{P}}_{\mathrm{opt,QAM}} = \frac{4(\sqrt{M}-1)}{\sqrt{M}} Q\left(\sqrt{\frac{3\eta(\beta-1)}{(M-1)\Lambda^2}}\right) - \frac{4(\sqrt{M}-1)^2}{M} Q^2\left(\sqrt{\frac{3\eta(\beta-1)}{(M-1)\Lambda^2}}\right)$.

where Λ is defined by $\Lambda = \frac{1}{2\pi}\int_0^{2\pi} \frac{d\omega}{\sqrt{s_\Sigma(\omega)}}$ with $s_\Sigma(\omega) = \sum_{k=-\infty}^{\infty} \sigma(k) e^{-jk\omega}$. ■

Before proving this theorem, we would like to make the following two comments:

1. From Theorem 5 we can see that the diversity gain for the optimally precoded MIMO system for a fixed N_t with the ZF receiver is $N_r - N_t + 1$. However, when N_t tends to infinity, Theorem 6 reveals that the limiting SEP of the optimally precoded MIMO system equipped with the ZF detector decays exponentially.

2. Despite the fact that the assumption of Theorem 6 requires that the correlation matrix is Toeplitz so that we can make use of the Szegö's theorem, we can infer from the following proof that the assumption can be actually relaxed to any invertible correlation matrix $\mathbf{\Sigma}$ with the condition that $\lim_{N_t\to\infty}\frac{1}{N_t}\sum_{n=1}^{N_t}\lambda_k^{-1/2}$ exists, where $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{N_t}$ are the eigenvalues of $\mathbf{\Sigma}$.

Proof: Using Lemma 3 with $K = N_t, \mathbf{T}_{N_t} = \mathbf{\Sigma}$ and $\mathtt{F}(x) = 1/\sqrt{x}$, we have

$$\Lambda = \lim_{N_t\to\infty}\frac{\sum_{k=1}^{N_t}\lambda_k^{-1/2}}{N_t} = \frac{1}{2\pi}\int_0^{2\pi}\frac{1}{\sqrt{s_\Sigma(\omega)}}d\omega. \tag{4.60}$$

where $\lambda_1, \lambda_2 \ldots, \lambda_{N_t}$ are the eigenvalues of $\mathbf{\Sigma}$. For notational simplicity, let $\bar{\lambda}_{N_t} = \big(\sum_{k=1}^{N_t}\lambda_k^{-1/2}\big)/N_t$. Now, using the optimal precoder given in Theorem 5, the resulting minimum average SEP is

$$\lim_{N_t\to\infty}\mathrm{P}_{\min}(\widetilde{\mathbf{F}}) = \lim_{N_t\to\infty}G\bigg(\Big(\sum_{k=1}^{N_t}\lambda_k^{-1/2}\Big)^2/N_t\bigg). \tag{4.61}$$

Correspondingly, for PAM signal, we obtain

$$\begin{aligned}\bar{\mathrm{P}}_{\text{opt,PAM}} &\overset{(a)}{=} \frac{2(M-1)}{M\pi}\int_0^{\pi/2}\lim_{N_t\to\infty}\Big(1+\frac{3\eta}{(M^2-1)N_t\bar{\lambda}_{N_t}^2\sin^2\theta}\Big)^{-(\beta-1)N_t-1}d\theta \\ &\overset{(b)}{=} \frac{2(M-1)}{M\pi}\int_0^{\pi/2}\exp\Big(-\frac{3\eta(\beta-1)}{(M^2-1)\Lambda^2\sin^2\theta}\Big)d\theta \\ &= \frac{2(M-1)}{M}Q\bigg(\sqrt{\frac{6\eta(\beta-1)}{(M^2-1)\Lambda^2}}\bigg),\end{aligned} \tag{4.62}$$

where equality (a) follows from the fact that $\Big(1+\frac{3\eta}{(M^2-1)N_t\bar{\lambda}_{N_t}^2\sin^2\theta}\Big)^{-(\beta-1)N_t-1} < 1$ for all $N_t \geq 1$ and $\theta \in [0, \pi/2]$ and thus, by the Lebesgue's Dominated Convergence Theorem [146], we can change the order of the limit and integration. The equality (b)

is due to the well-known limit of the Euler's number. Following a similar argument, for PSK signal, we have

$$\begin{aligned}\bar{\mathrm{P}}_{\mathrm{opt,PSK}} &= \frac{1}{\pi}\int_0^{(M-1)\pi/M} \lim_{N_t\to\infty}\left(1+\frac{\eta\sin^2(\pi/M)}{N_t\bar{\lambda}^2_{N_t}\sin^2\theta}\right)^{-(\beta-1)N_t-1} d\theta \\ &= \frac{1}{\pi}\int_0^{(M-1)\pi/M} \exp\left(-\frac{\eta(\beta-1)\sin^2(\pi/M)}{\Lambda^2\sin^2\theta}\right) d\theta. \end{aligned} \tag{4.63}$$

and for QAM signal, we can attain

$$\begin{aligned}\bar{\mathrm{P}}_{\mathrm{opt,QAM}} &= \frac{4(\sqrt{M}-1)}{\sqrt{M}\pi}\int_0^{\pi/2} \lim_{N_t\to\infty}\left(1+\frac{3\eta}{2(M-1)N_t\bar{\lambda}^2_{N_t}\sin^2\theta}\right)^{-(\beta-1)N_t-1} d\theta \\ &\quad - \frac{4(\sqrt{M}-1)^2}{M\pi}\int_0^{\pi/4} \lim_{N_t\to\infty}\left(1+\frac{3\eta}{2(M-1)N_t\bar{\lambda}^2_{N_t}\sin^2\theta}\right)^{-(\beta-1)N_t-1} d\theta \\ &= \frac{4(\sqrt{M}-1)}{\sqrt{M}}Q\left(\sqrt{\frac{3\eta(\beta-1)}{(M-1)\Lambda^2}}\right) - \frac{4(\sqrt{M}-1)^2}{M}Q^2\left(\sqrt{\frac{3\eta(\beta-1)}{(M-1)\Lambda^2}}\right). \end{aligned} \tag{4.64}$$

This completes the proof of Theorem 2. □

In particular, when the channel covariance matrix $\mathbf{\Sigma}$ is the commonly-used non-symmetric Kac-Murdock-Szegö (KMS) matrix [147, 148], i.e.,

$$[\mathbf{\Sigma}]_{mn} = \begin{cases} \rho^{n-m} & m \le n \\ [\mathbf{\Sigma}]^*_{nm} & m > n, \end{cases} \tag{4.65}$$

where $0 < |\rho| < 1$ indicates the degree of correlation, we have the following corollary.

Corollary 1 *Consider large MIMO systems with the optimal precoder* (4.56)*, ZF detector and the M-ary PAM, PSK and QAM constellation. If* $0 < |\rho| < 1$*, then,*

$\lim_{N_t\to\infty} \mathrm{P}_{N_t}(\widetilde{\mathbf{F}}) = \bar{\mathrm{P}}_{\mathrm{KMS}}$ *exists and*

$$\bar{P}_{\mathrm{KMS,PAM}} = \frac{2(M-1)}{M} Q\left(\sqrt{\frac{3\pi^2\eta(1-|\rho|)(\beta-1)}{2(1+|\rho|)(M^2-1)\mathtt{E}^2\left(\frac{2\sqrt{|\rho|}}{1+|\rho|}\right)}}\right),$$

$$\bar{P}_{\mathrm{KMS,PSK}} = \frac{1}{\pi}\int_0^{(M-1)\pi/M} \exp\left(-\frac{\pi^2\eta(1-|\rho|)(\beta-1)\sin^2(\pi/M)}{4(1+|\rho|)\mathtt{E}^2\left(\frac{2\sqrt{|\rho|}}{1+|\rho|}\right)\sin^2\theta}\right) d\theta,$$

$$\begin{aligned}\bar{P}_{\mathrm{KMS,QAM}} = {} & \frac{4(\sqrt{M}-1)}{\sqrt{M}} Q\left(\sqrt{\frac{3\pi^2\eta(1-|\rho|)(\beta-1)}{4(1+|\rho|)(M-1)\mathtt{E}^2\left(\frac{2\sqrt{|\rho|}}{1+|\rho|}\right)}}\right) \\ & -\frac{4(\sqrt{M}-1)^2}{M} Q^2\left(\sqrt{\frac{3\pi^2\eta(1-|\rho|)(\beta-1)}{4(1+|\rho|)(M-1)\mathtt{E}^2\left(\frac{2\sqrt{|\rho|}}{1+|\rho|}\right)}}\right),\end{aligned}$$

where $\mathtt{E}(k) = \int_0^{\pi/2}\sqrt{1-k^2\sin^2\theta}\,d\theta$ *denotes the complete elliptic integral of the second kind [134].* ■

Proof: Since

$$s_\Sigma(\omega) = \sum_{k=-\infty}^{\infty} \sigma(k) e^{-jk\omega} = \frac{1-|\rho|^2}{1+|\rho|^2-2\mathtt{Re}[\rho e^{j\omega}]}.$$

for $0 < |\rho| < 1$, we have $s_\Sigma(\omega) \geq \frac{1-|\rho|}{1+|\rho|}$ and as a result, $L_{s_\Sigma} \geq \frac{1-|\rho|}{1+|\rho|} > 0$. Now by

Lemma 3 with $\mathtt{F}(x) = 1/\sqrt{x}$, we attain

$$\begin{aligned}\Lambda_{KMS} &= \frac{1}{\pi\sqrt{1-|\rho|^2}} \int_0^{\pi} \sqrt{1+|\rho|^2-2|\rho|\cos\omega} d\omega \\ &= \frac{2}{\pi}\sqrt{\frac{1+|\rho|}{1-|\rho|}}\mathtt{E}\left(\frac{2\sqrt{|\rho|}}{1+|\rho|}\right).\end{aligned}$$

Combining this with Theorem 1 completes the proof of Corollary 1. □

Theorem 6 requires that the first-order and second-order channel statistics are known at the transmitter. However, in practice, it is not easy to obtain the perfect channel estimate at the transmitter. In this case, we can consider asymptotic SEP for uniformly precoded large MIMO channels, i.e., $\bar{\mathbf{F}} = \frac{1}{\sqrt{N_t}}\mathbf{I}$.

Theorem 7 *Consider large MIMO systems using the uniform precoder, ZF detector and the M-ary square QAM constellation. If the channel covariance matrix* $\mathbf{\Sigma}$ *is the KMS matrix in* (6.139)*, then,* $\lim_{N_t\to\infty} \mathrm{P}(\bar{\mathbf{F}}) = \bar{\mathrm{P}}_{\mathrm{U}}$ *exists and*

- $\bar{\mathrm{P}}_{\mathrm{U,PAM}} = \frac{2(M-1)}{M} Q\left(\sqrt{\frac{6\eta(\beta-1)(1-\rho^2)}{(M^2-1)(1+\rho^2)}}\right)$;
- $\bar{\mathrm{P}}_{\mathrm{U,PSK}} = \frac{1}{\pi}\int_0^{(M-1)\pi/M} \exp\left(-\frac{\eta(\beta-1)(1-\rho^2)\sin^2(\pi/M)}{(1+\rho^2)\sin^2\theta}\right) d\theta$;
- $\bar{\mathrm{P}}_{\mathrm{U,QAM}} = \frac{4(\sqrt{M}-1)}{\sqrt{M}} Q\left(\sqrt{\frac{3\eta(\beta-1)(1-|\rho|^2)}{(M-1)(1+|\rho|^2)}}\right) - \frac{4(\sqrt{M}-1)^2}{M} Q^2\left(\sqrt{\frac{3\eta(\beta-1)(1-|\rho|^2)}{(M-1)(1+|\rho|^2)}}\right)$.

■

Proof: Note that in this case, matrix $\mathbf{\Sigma}$ has a simple tridiagonal inverse [148], given

by

$$\mathbf{\Sigma}^{-1} = \frac{1}{1-|\rho|^2}\begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho^* & 1+|\rho|^2 & -\rho & \cdots & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & \cdots & -\rho^* & 1+|\rho|^2 & -\rho \\ 0 & \cdots & 0 & -\rho^* & 1 \end{bmatrix}.$$

Combining this with (4.46) and the uniform precoder yields

$$\mathrm{P}_{\mathrm{QAM}}(\bar{\mathbf{F}}) = \frac{2}{N_t} G_{\mathrm{QAM}}\Big(\frac{N_t}{1-|\rho|^2}\Big) + \frac{N_t-2}{N_t} G_{\mathrm{QAM}}\Big(\frac{N_t(1+|\rho|^2)}{1-|\rho|^2}\Big).$$

Since

$$\lim_{N_t\to\infty}\left(1+\frac{3\eta(1-|\rho|^2)}{2(M-1)N_t\Phi\sin^2\theta}\right)^{-(N_r-N_t+1)} = \exp\left(-\frac{3\eta(\beta-1)(1-|\rho|^2)}{2(M-1)\Phi\sin^2\theta}\right),$$

where $\Phi = 1$ or $1+|\rho|^2$, we have

$$\begin{aligned}\bar{\mathrm{P}}_{\mathrm{U,QAM}} &= \lim_{N_t\to\infty} G_{\mathrm{QAM}}\Big(\frac{N_t(1+|\rho|^2)}{1-|\rho|^2}\Big) \\ &= \frac{4(\sqrt{M}-1)}{\sqrt{M}} Q\left(\sqrt{\frac{3\eta(\beta-1)(1-|\rho|^2)}{(M-1)(1+|\rho|^2)}}\right) - \frac{4(\sqrt{M}-1)^2}{M} Q^2\left(\sqrt{\frac{3\eta(\beta-1)(1-|\rho|^2)}{(M-1)(1+|\rho|^2)}}\right).\end{aligned}$$

Similarly, we can have $\bar{\mathrm{P}}_{\mathrm{U,PAM}}$ and $\bar{\mathrm{P}}_{\mathrm{U,PSK}}$ as desired. This completes the proof of Theorem 7. □

4.3.3 Convex Region for KMS Matrices

Let $\zeta_1(\mathbf{\Sigma}) \leq \zeta_2(\mathbf{\Sigma}) \leq \ldots \leq \zeta_{N_t}(\mathbf{\Sigma})$ be the eigenvalues of the KMS matrix and then from [149], we have

$$\zeta_k(\mathbf{\Sigma}) = \frac{1-|\rho|^2}{1+|\rho|^2+2|\rho|\cos\theta_k}, \text{ for } k = 1, 2, \ldots, N_t,$$

where $\cos\theta_1 \geq \cos\theta_2 \geq \ldots \geq \cos\theta_{N_t}$, in which θ_k is the solution to

$$|\rho|^2\Big(\sin(N_t+1)\theta_k + 2|\rho|\sin N_t\theta_k + |\rho|^2\sin(N_t-1)\theta_k\Big) = 0.$$

Since $|\cos\theta_k| \leq 1$ and $0 < |\rho| < 1$, then for the KMS matrix,

$$\zeta_1(\mathbf{\Sigma}) \geq \frac{1-|\rho|^2}{(1+|\rho|)^2} = \frac{1-|\rho|}{1+|\rho|} > 0.$$

Recall that the optimality condition for the precoder is

$$\zeta_1(\mathbf{F}^H\mathbf{F}) \geq \frac{1}{\zeta_1(\mathbf{\Sigma})T}, \tag{4.66}$$

where $\zeta_1(\mathbf{F}^H\mathbf{F})$ is the minimum eigenvalue of $\mathbf{F}^H\mathbf{F}$. Finally, for KMS covariance matrix, the constraint (4.50) have the following simple sufficient form

$$\zeta_1(\mathbf{F}^H\mathbf{F}) \geq \frac{1+|\rho|}{T(1-|\rho|)}.$$

4.3.4 Asymptotic Behaviour on Individual SNR for Each Subchannel

To deeply appreciate the asymptotic SEP properties derived for the optimally precoded large MIMO systems in the previous subsection, we would like here to also study the asymptotic distribution of the SNR for each sub-channel when the array size is large. Notice that at the output of the ZF receiver for each sub-channel, the average signal power is $\mathbb{E}[|s_k|^2] = 1$, and the power of the equalized noise is $\sigma^2 \left[(\widetilde{\mathbf{F}}^H \mathbf{H}^H \mathbf{H} \widetilde{\mathbf{F}})^{-1}\right]_k$. Therefore, the instantaneous SNR of each sub-channel as a function of random channel realization is

$$\tau_k = \frac{\eta}{\left[(\widetilde{\mathbf{F}}^H \mathbf{H}^H \mathbf{H} \widetilde{\mathbf{F}})^{-1}\right]_k} = \frac{N_t \eta \widetilde{\gamma}_k}{\left(\sum_{m=1}^{N_t} \lambda_m^{-1/2}\right)^2}, \quad \text{for } k = 1, 2, \ldots, N_t,$$

where $\widetilde{\gamma}_k = \frac{[(\widetilde{\mathbf{F}}^H \boldsymbol{\Sigma} \widetilde{\mathbf{F}})^{-1}]_k}{[(\widetilde{\mathbf{F}}^H \mathbf{H}^H \mathbf{H} \widetilde{\mathbf{F}})^{-1}]_k}$ and $\widetilde{\mathbf{F}}$ is the optimal precoder given in (4.56). Therefore, the mean and variance of τ_k can be determined as follows:

$$\mathbb{E}[\tau_k] = \frac{N_t(N_r - N_t + 1)\eta}{\left(\sum_{m=1}^{N_t} \lambda_m^{-1/2}\right)^2}, \quad \text{var}[\tau_k] = \frac{N_t^2(N_r - N_t + 1)\eta^2}{\left(\sum_{m=1}^{N_t} \lambda_m^{-1/2}\right)^4}.$$

When scaling up the array size, and with the help of (4.60), we have $\lim_{N_t \to \infty} \mathbb{E}[\tau_k] = \frac{\eta(\beta-1)}{\Lambda^2}$ and $\lim_{N_t \to \infty} \text{var}[\tau_k] = 0$. Then, by the law of large numbers (LLN), we have $\lim_{N_t \to \infty} \tau_k \xrightarrow{\text{a.s.}} \frac{\eta(\beta-1)}{\Lambda^2}$ for $k = 1, 2, \ldots, N_t$. This suggests that when the size of the antenna array goes to infinity, the instantaneous SNR of each sub-channel becomes stable, i.e., it converges to a fixed value. This verifies Theorem 6.

Here, it should be pointed out that in the above discussion, the exact convergence

requires that the array size goes to infinity and it does not necessarily work well when the array size is small. In what follows, we give an intuitive approximation to the distribution of the SNR for each receiver branch, which is very accurate when the array size is moderately large. From the convergence of (4.60), we know that

$$\lim_{N_t\to\infty}\tau_k=\lim_{N_t\to\infty}\frac{N_t\eta\widetilde{\gamma}_k}{\left(\sum_{m=1}^{N_t}\lambda_m^{-1/2}\right)^2}\approx\frac{\eta\widetilde{\gamma}_k}{N_t\Lambda^2} \tag{4.67}$$

for a large N_t. Now letting $\tilde{\tau}_k=\frac{\eta\widetilde{\gamma}_k}{N_t\Lambda^2}$ and, as the Szegö's theorem converges very fast for the considered correlation matrix, $\tilde{\tau}_k\approx\tau_k$ when N_t is reasonably large. Since $f(\widetilde{\gamma}_k)=\frac{1}{\Gamma(N_r-N_t+1)}e^{-\widetilde{\gamma}_k}\widetilde{\gamma}_k^{N_r-N_t}$, $\tilde{\tau}_k$ is subject to the Gamma distribution with mean $\frac{(N_r-N_t+1)\eta}{N_t\Lambda^2}\approx\frac{\eta(\beta-1)}{\Lambda^2}$ and variance $\frac{(N_r-N_t+1)\eta^2}{N_t^2\Lambda^4}\approx\frac{\eta^2(\beta-1)}{N_t\Lambda^4}$ when N_t is large. Now, by the well-known central limit theorem, we have

$$\tilde{\tau}_k\dot{\sim}\mathcal{N}\left(\frac{\eta(\beta-1)}{\Lambda^2},\frac{\eta^2(\beta-1)}{N_t\Lambda^4}\right) \tag{4.68}$$

where $\dot{\sim}$ means approximately with the same distribution when the array size is large. Hence,

$$\begin{aligned}\lim_{N_t\to\infty}\tau_k=\lim_{N_t\to\infty}\tilde{\tau}_k\sim\ &\mathcal{N}\left(\frac{\eta(\beta-1)}{\Lambda^2},\lim_{N_t\to\infty}\frac{\eta^2(\beta-1)}{N_t\Lambda^4}\right)\\ &\xrightarrow{\text{a.s.}}\frac{\eta(\beta-1)}{\Lambda^2}.\end{aligned}$$

It is worth pointing out that the approximation in (4.68) is pretty accurate when the array size is relatively small, say, $N_t=10$, as can be seen in Fig. 4.8.

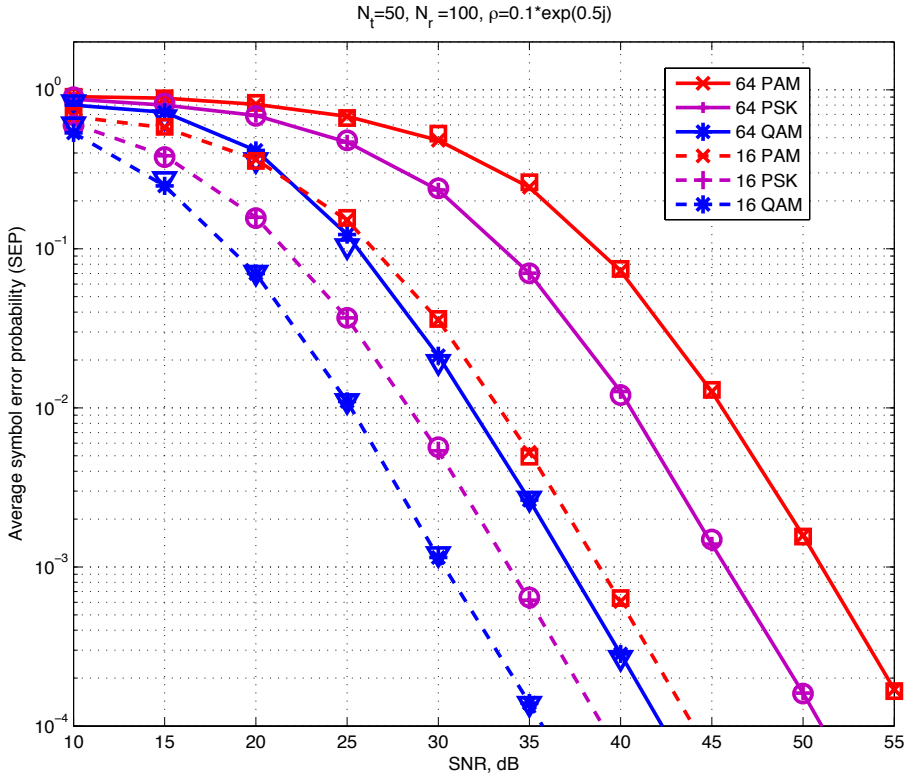


Figure 4.5: Average SEP performance against SNR η, where the correlation coefficient $\rho = 0.1 * \exp(0.5j)$. All the lines represent the theoretical values while the *squares*, *circles* and *triangles* denote the corresponding simulated SEP.

4.4 Numerical Simulations

In this section, we verify our theoretical results through computer simulations. In order to validate the theoretical SEP expression, Monte Carlo simulations are carried out. Let us first consider a uniform linear array with $N_t = 50$ transmitting antennas and $N_r = 100$ receiving antennas, where the receiver knows the CSI perfectly and the transmitter knows only the correlation matrix $\mathbf{\Sigma}$. In this simulation, the correlation matrix $\mathbf{\Sigma}$ is taken as the Kac-Murdock-Szegö matrix. The theoretical and the

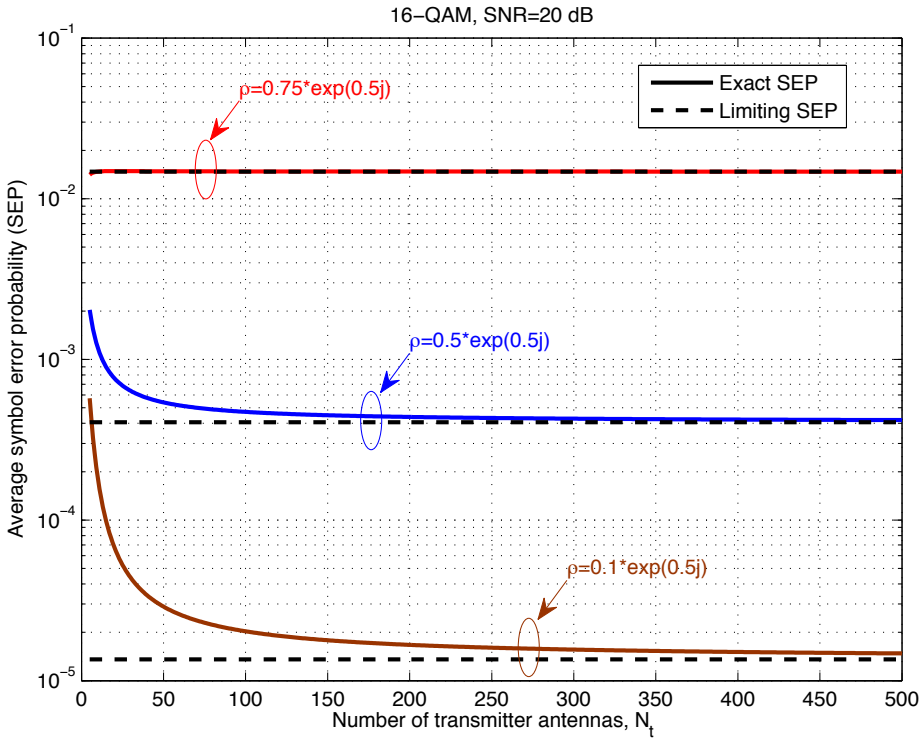


Figure 4.6: Average SEP performance against the number of transmitter antennas, with 16-QAM, $\beta = 2$, and SNR = 20dB.

simulated SEP results for the optimal precoder are plotted in Fig. 4.5 against the SNR η with the PAM, PSK and square QAM constellations. It can be observed that the simulated result matches with the theoretical expression very well, which verifies the correctness of our analysis. Therefore, in the following, we would like to use the theoretical result to examine some asymptotic properties.

To demonstrate the convergence rate in terms of the number of the transmitter antennas, the exact theoretical SEP and its limit are depicted versus the number of transmitting antennas in Fig. 4.6. Without loss of generality, 16-QAM constellation is adopted and the theoretical SEP is in solid line while its limit is represented by

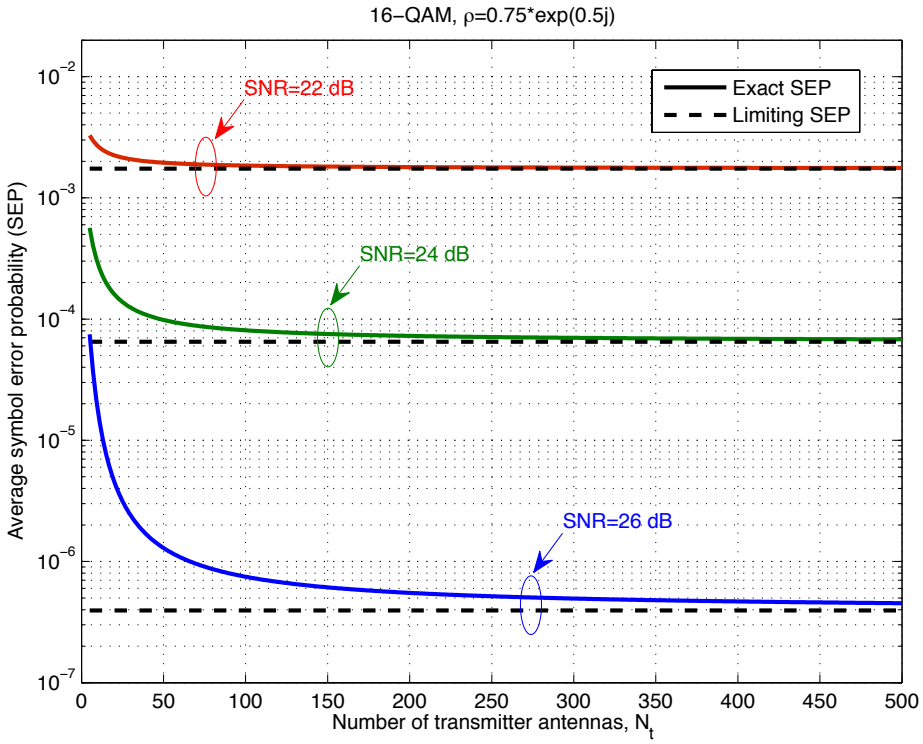


Figure 4.7: Average SEP performance against the number of transmitter antennas N_t, with 16-QAM, $\beta = 2$, and $\rho = 0.75 * \exp(0.5j)$.

dash line. Three different correlation matrices are generated according to ρ. It can be noticed that as the magnitude of the correlation coefficient ρ decreases, the correlation between the adjacent antennas reduces, and as a consequence, the corresponding SEP reduces substantially. In Fig. 4.7, the limiting SEP is also plotted against N_t but for different SNRs. It is expected that the SEPs drop as SNR increases. In both figures, it can be seen clearly that for given SNR, as the array size is scaled up, the theoretical SEP and the asymptotic result gradually meet together. The approximation is accurate for moderate and large number of antennas. Note that the mean of SNR for each sub-channel is a decreasing function of $|\rho|$ and an increasing

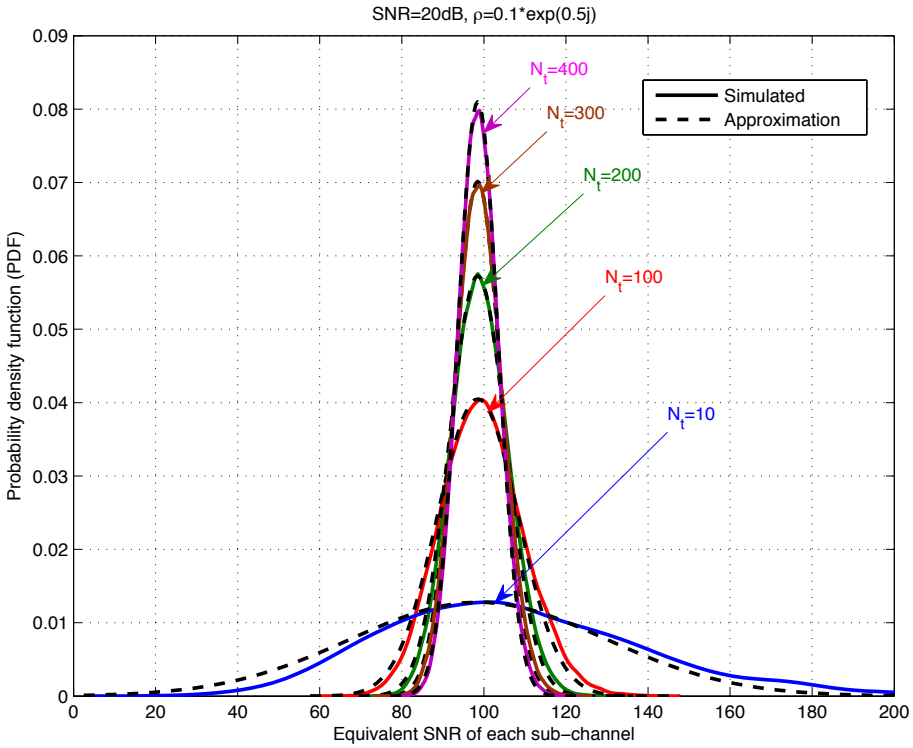


Figure 4.8: The distribution of the equivalent SNRs of each sub-channel with $\beta = 2$, and $\rho = 0.1 * \exp(0.5j)$, SNR= 20dB.

function of system SNR η. Hence, either decreasing $|\rho|$ or increasing η will eventually increase the mean of SNR for each sub-channel, and thus, result in a lower convergence rate for theoretic SEP approach to its limit expression against N_t. This phenomenon is also observed in [38] in the scenario of the approximation of channel capacity.

In addition, we would also like to show the convergence characteristic of the approximated distribution of individual SNR in each receiver branch for the optimally precoded large MIMO systems. Both the approximated PDF and the simulated PDF are given in Fig. 4.8, from which it can be seen clearly that the Gaussian approximation is very accurate even when we only have a very small number of antennas, say,

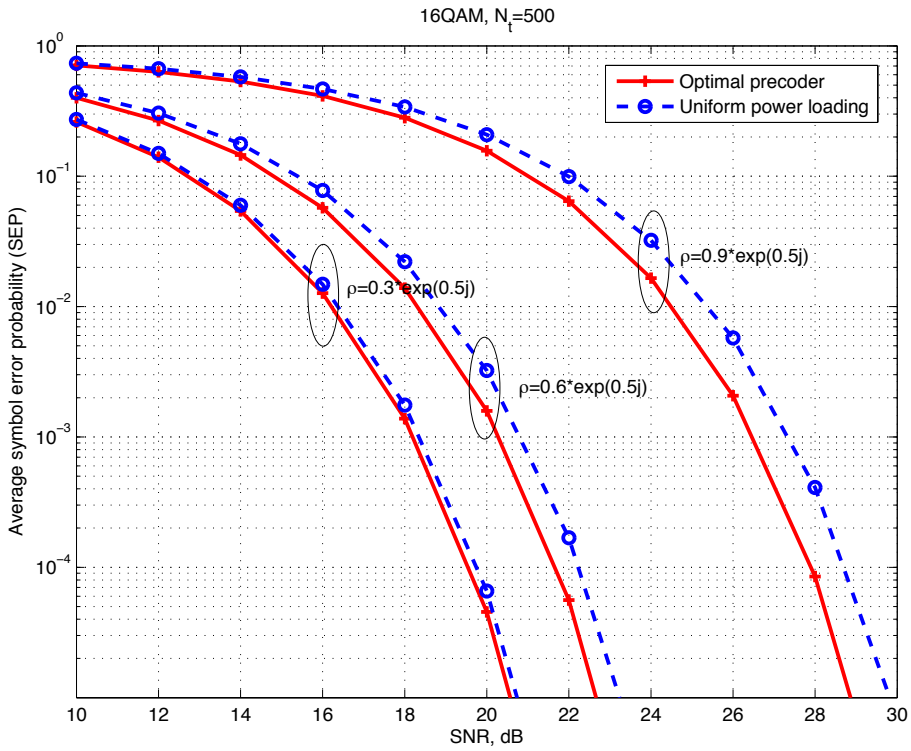


Figure 4.9: Average SEP performance against SNR, with 16-QAM, $N_t = 500$, $\beta = 2$ and different ρ.

10 transmitter antennas.

On the other hand, we also would like to compare the error performance of the optimal precoder with a uniform power allocation scheme. Consider the case where $N_t = 500$, $\beta = 2$, and using a 16-QAM constellation. The average SEPs are given for different correlation coefficient ρ in Fig. 4.9. Again, we can find that as $|\rho|$ increases, the SEP is becoming significantly worse. The optimal precoder always leads to better error performance. The gap between the optimal precoder and the uniform power allocation strategy becomes larger when $|\rho|$ increases. The reason is that when $|\rho|$ is very small, $\mathbf{\Sigma}$ is very close to a diagonal matrix with equal diagonal entries. Then, the

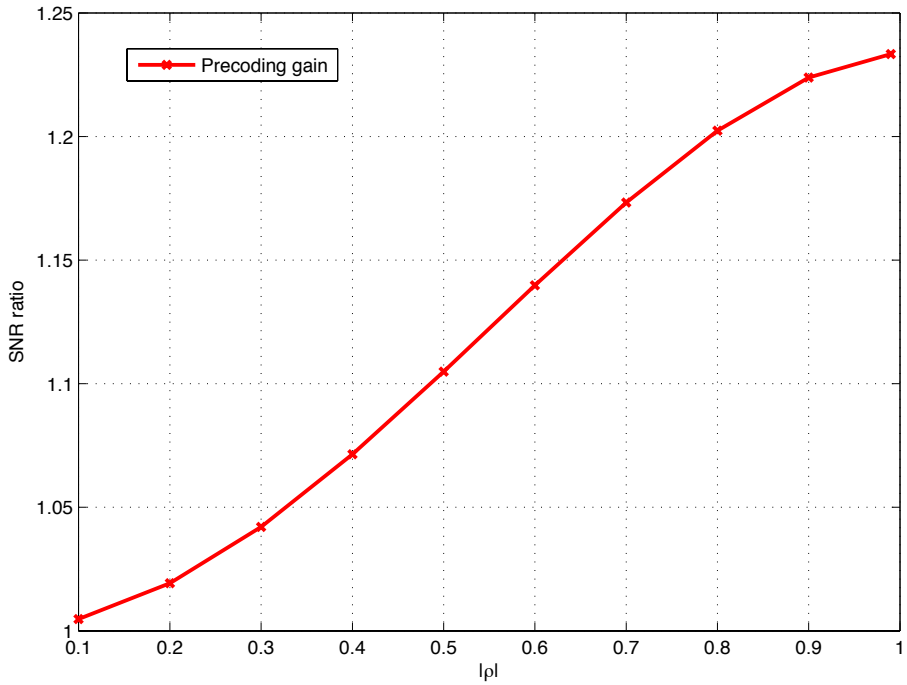


Figure 4.10: The precoding gain compared with uniform power allocation versus $|\rho|$.

optimal precoder will degrade into the uniform power allocation case. However, for general $\mathbf{\Sigma}$, the performance gap is non-negligible. To show this phenomenon clearly, the ratio of the SNR of individual sub-channel between optimal precoder and uniform power allocation transmitter are given in Fig. 4.10. The SNR ratio is a monotonic increasing function of $|\rho|$, which verifies the results in Fig. 4.9. Therefore, precoding at the transmitter side can yields much better performance over the uniform power allocation strategy.

4.5 Conclusion

In this chapter, we have derived an explicit convex region in terms of the modulated signals, system SNR and channel statistics for the optimal precoder minimizing the average SEP of the ZF detector. A simple expression with a very fast convergence rate for the SEP limit of the large MIMO systems with the PAM, PSK and square QAM constellations and the ZF receiver is obtained. An intuitive understanding of this convergent process has also been provided in terms of the approximation to the distribution of the individual SNR for each sub-channel. The main technical approach proposed in this chapter to deriving our results is to fully take advantage of the characteristic of the MIMO channels, the structure of the transmitter as well as of the ZF receiver, the Szegö's theorem [44] on large Hermitian Toeplitz matrices, and the well known limit: $\lim_{x\to\infty}(1+1/x)^x = e$.

Here, we would like to emphasize the fact that the Szegö's theorem [44] is a right and strong mathematical tool for the deep and systematic study of asymptotic behaviors on the large MIMO systems from both information-theoretic and detection viewpoints.

Chapter 5

Optimal Precoder Design and Asymptotic SEP Analysis for Correlated MIMO Channels using ZF-DF Detection

In this chapter, the problem of designing a precoder using a zero-forcing (ZF) decision-feedback (DF) detector is addressed for correlated multiple-input multiple-output (MIMO) communication systems having M transmitter antennas and N receiver antennas ($M \leq N$). It is assumed that full knowledge of channel state information (CSI) is available at the receiver and only the zero-mean and second-order statistics of the channels are available at the transmitter. For such a MIMO system, the principal goal of this chapter is to efficiently design an optimal precoder that minimizes asymptotic symbol error rate (SER) of the ZF-DF detector under a perfect decision

feedback. By fully taking advantage of the product majorization relationship among eigenvalues, singular-values and Cholesky values of the design matrix parameters, a necessary condition for the optimal solution to satisfy is first developed and then, the structure of the optimal solution is characterized. With these results, the original non-convex problem is reformulated into a convex one that can be efficiently solved using an interior-point method. Computer simulations show that the error performance of the optimal precoder proposed in this chapter outperforms those of all existing designs. Particularly, the optimal system obtains a significant SNR gain over the unprecoded MIMO system with the V-BLAST detector when $N = M$.

On the other hand, by scaling up the antenna array size of both terminals without bound for such network to form a large MIMO system, we propose a novel method based on the Szegö's theorem and the well-known limit $\lim_{x\to\infty}(1 + 1/x)^x = e$ to analyze the asymptotic behavior on the error performance of an equal-diagonal QRS precoded large MIMO system when employing an abstract Toeplitz correlation model. This new approach bears a simple expression with a fast convergence rate and thus, is efficient and effective for error performance evaluation. Then, the impact of channel correlation on the error performance is studied for different correlation coefficients. In addition, an explanation of this approach in terms of the entropy power of the channel is also provided. Finally, computer simulations are also carried out for the considered large MIMO case to verify our analysis in comparison with a uniform power allocation strategy.

5.1 System Description and Design Problem

5.1.1 Precoded MIMO Channel Model

In this chapter, we are interested in a precoded MIMO [150] communication system with M transmitter antennas and N receiver antennas ($N \geq M$). The channel model can be described by

$$\mathbf{y} = \mathbf{HFx} + \mathbf{z}, \tag{5.69}$$

where $\mathbf{y}$ is an $N \times 1$ received signal vector, $\mathbf{H}$ is an $N \times M$ channel matrix, $\mathbf{F}$ is an $M \times M$ precoding matrix, $\mathbf{x}$ is an $M \times 1$ transmitting signal vector and $\mathbf{z}$ is an $N \times 1$ complex noise vector. We assume that the channel noise is circularly-symmetric complex Gaussian distributed with the covariance matrix being $\mathbf{R_{zz}} = 2\sigma^2\mathbf{I}_N$, and the transmitted symbols in $\mathbf{x}$ are uncorrelated with each other and uncorrelated with the channel noise. The channel matrix $\mathbf{H}$ includes the subchannels h_{nm} connecting the m-th transmitter antenna with the n-th receiver antenna. Each of the h_{nm} is a zero-mean, circularly-symmetric complex Gaussian distributed random variable with unit variance. Let $\mathbf{h}_n^T = [h_{n1} \cdots h_{nM}]$ denote the n-th row of $\mathbf{H}$. We assume that the channels linking to the same receiver antenna are correlated among themselves, but are uncorrelated with the channel linking to the different receiver antenna. That is to say,

$$\mathrm{E}[(\mathbf{h}_\ell^T)^H \mathbf{h}_n^T] = \begin{cases} \mathbf{\Sigma} & \ell, \\ \mathbf{0} & \ell \neq n. \end{cases} \tag{5.70}$$

It is well-known [138] that $\mathbf{H}^H\mathbf{H}$ follows the Wishart distribution denoted by $\mathcal{W}_M(N, \mathbf{\Sigma})$. If we consider the precoder matrix $\mathbf{F}$ as a part of the channel such that $\mathbf{C} = \mathbf{HF}$, then we have

$$\mathrm{E}[(\mathbf{c}_\ell^T)^H \mathbf{c}_n^T] = \begin{cases} \mathbf{F}^H\mathbf{\Sigma}\mathbf{F} & \ell = n \\ \mathbf{0} & \ell \neq n \end{cases} \tag{5.71}$$

and $\mathbf{C}^H\mathbf{C} = (\mathbf{HF})^H\mathbf{HF}$ is also of Wishart distribution denoted by $\mathcal{W}_M(N, \mathbf{F}^H\mathbf{\Sigma}\mathbf{F})$. The quantity $\mathbf{F}^H\mathbf{\Sigma}\mathbf{F}$ is of fundamental importance in our application and we will examine its properties in greater detail in the ensuing sections.

5.1.2 The ZF-DF Receiver using QR Decomposition

In this subsection, we briefly review the implementation of the ZF-DF detector for our channel model (5.69) using the QR decomposition [48]. Particularly for block data transmission, presenting the ZF-DF receiver in the way of backward successive symbol-by-symbol cancellation detection based on the QR decomposition, helps us more naturally understand this detection procedure itself, as well as more easily and clearly formulate and state our design problem. Let each symbol x_m of the transmitted signal vector $\mathbf{x}$ in (5.69) be independently and equally likely chosen from a finite-size alphabet set $\mathcal{X}$. We denote the estimate of $\mathbf{x}$ by $\hat{\mathbf{x}} = [\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_M]^T$. Then, the QR-decomposition-based ZF-DF detector is described by the following three steps.

Algorithm 1 *(The QR-decomposition-based ZF-DF detector)*

1. QR-decomposition. *Perform the QR decomposition [151],* $\mathbf{C} = \mathbf{HF} = \mathbf{QR}$*, where* $\mathbf{Q}$ *is an* $N \times M$ *conlumn-wise orthonormal matrix and* $\mathbf{R}$ *is an* $M \times M$

upper triangular matrix

$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1M} \\ 0 & r_{22} & \cdots & r_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{MM} \end{pmatrix}$$

with $r_{ii} > 0$. *Then, left-multiplying (5.69) by* $\mathbf{Q}^H$ *yields*

$$\widetilde{\mathbf{y}} = \mathbf{Q}^H \mathbf{y} = \mathbf{R}\mathbf{x} + \widetilde{\mathbf{z}} \tag{5.72}$$

where $\widetilde{\mathbf{y}} = \mathbf{Q}^H \mathbf{y}$ *and* $\widetilde{\mathbf{z}} = \mathbf{Q}^H \mathbf{z}$. *Equation (5.72) can be equivalently rewritten as*

$$\widetilde{y}_m = r_{mm} x_m + \sum_{k=m+1}^{M} r_{mk} x_k + \widetilde{z}_m, \quad m = 1, \cdots, M. \tag{5.73}$$

2. Hard decision. *Now, we employ (5.72) to first estimate the* M*-th transmitted symbol* x_M *by making the hard decision, i.e.,* $\hat{x}_M = \mathcal{Q}\big[\frac{\widetilde{y}_M}{r_{MM}}\big]$, *where the function* $q = \mathcal{Q}[u]$ *is the quantization operation that sets* q *to the element of* $\mathcal{X}$ *closest to* u *in terms of the Euclidean distance measure.*

3. Backward successive cancellation. *Assume that the estimate of* x_M *is perfect, i.e.,* $\hat{x}_M = x_M$. *Plug* $\hat{x}_M$ *back into the* $(M-1)$*-th row in (5.72) so as to completely cancel the interference term involving in* x_M *in* $\widetilde{y}_{M-1}$ *and then ready to detect* x_{M-1}. *Proceed this process until the first symbol* x_1 *is already detected. The above whole procedure can be simply summarized as the following backward*

recursive algorithm:

$$\begin{aligned} \hat{x}_M &= \mathcal{Q}\left[\frac{\widetilde{y}_M}{r_{MM}}\right], \\ \hat{x}_m &= \mathcal{Q}\left[\frac{\widetilde{y}_m - \sum_{k=m+1}^{M} r_{mk}\hat{x}_k}{r_{mm}}\right] \qquad \textit{for } m = M, \cdots, 1. \end{aligned}$$

■

5.1.3 Statement of the Design Problem

Suppose that the transmitter and receiver antennas transmit symbols from a K-ary Quadrature Amplitude Modulation (K-QAM) square constellation. Then, under the assumption that the previous symbols have been perfectly detected, the average symbol error probability of the ZF-DF detector is given by

$$P_e(\mathbf{H}) = \frac{1}{M}\sum_{m=1}^{M}\left(4\Big(1-\frac{1}{\sqrt{K}}\Big)Q\Big(\sqrt{\frac{3\rho}{K-1}}r_{mm}\Big) - 4\Big(1-\frac{1}{\sqrt{K}}\Big)^2 Q^2\Big(\sqrt{\frac{3\rho}{K-1}}r_{mm}\Big)\right) \tag{5.74}$$

where the Q-function is defined by $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{t^2}{2}}dt$ and ρ is signal to noise ratio per QAM symbol. Our goal in this chapter is to propose an efficient technique for designing a precoder matrix $\mathbf{F}$ that minimizes the probability of symbol error of the ZF-DF receiver. Our problem can be formally stated as:

Problem 1 *Design a precoding matrix* $\mathbf{F}$ *such that*

$$\mathbf{F}_{\text{opt}} = \arg \min_{\mathbf{F}} \mathbb{E}_{\mathbf{H}}\left[P_e(\mathbf{H})\right] \tag{5.75}$$

subject to the total transmitting power constraint

$$\mathrm{tr}(\mathbf{F}^H\mathbf{F}) \leq p_0$$

where p_0 is the total transmitting power and the notation $\mathbb{E}_{\mathbf{H}}(\cdot)$ denotes the expectation taken over all random channel realizations. ■

5.2 Reformulations of the Design Problem

The main purpose of this section is to reformulate Problem 1 stated in Section 5.1.3 into a convex problem that can be more efficiently solved using an interior-point method.

5.2.1 Simplification of the Objective Function

Let us first simplify the objection function. To do that, we need to know the probability density function (PDF) of the diagonal entries of the R-factor in the QR decomposition of the precoded channel matrix. Let $\mathbf{C} = \mathbf{HF} = \mathbf{QR}$. In order to represent r_{mm}^2 for $m = 1, 2, \cdots, M$ in terms of the determinants of the submatrices of $\mathbf{C}$, we use notation $\mathbf{A}_k$ to denote the $N \times k$ matrix consisting of the first k columns of an $N \times M$ matrix $\mathbf{A}$, while $\mathbf{A}_{kk}$ denotes the $k \times k$ matrix consisting of the first k rows and columns of $\mathbf{A}$. Now, extracting the diagonal elements r_{mm} for $m = 1, 2, \cdots, M$ as a diagonal matrix such that

$$\mathbf{R} = \mathrm{diag}\,(r_{11}, \cdots, r_{MM}) \begin{bmatrix} 1 & \frac{r_{12}}{r_{11}} & \cdots & \frac{r_{1M}}{r_{11}} \\ 0 & 1 & \ddots & \frac{r_{2M}}{r_{22}} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \triangleq \mathrm{diag}(r_{11}, \cdots, r_{MM})\mathbf{L}^H \qquad (5.76)$$

with $\mathbf{L}^H$ being the upper triangular matrix defined in (5.76), we obtain $\mathbf{C}_m = \mathbf{Q}\text{diag}(r_{11}, r_{22}, \cdots, r_{MM})[\mathbf{L}^H]_m$ and thus,

$$\begin{aligned}\mathbf{C}_m^H\mathbf{C}_m = \big[\mathbf{HF}\big]_m^H\big[\mathbf{HF}\big]_m = \big[\mathbf{F}^H\mathbf{H}^H\mathbf{HF}\big]_{mm} = [\mathbf{L}]_m\ \text{diag}(r_{11}^2, \cdots, r_{MM}^2)\ [\mathbf{L}^H]_m \\ = [\mathbf{L}]_{mm}\ \text{diag}(r_{11}^2, \cdots, r_{mm}^2)\ [\mathbf{L}^H]_{mm},\end{aligned} \tag{5.77}$$

from which we can derive

$$\det(\mathbf{C}_m^H\mathbf{C}_m) \ = \ \prod_{i=1}^{m} r_{ii}^2. \tag{5.78}$$

Therefore, a general formula for the evaluation of r_{mm}^2 in terms of the determinants of the submatrices of $\mathbf{C}$ is given by

$$r_{mm}^2 = \frac{\det\big(\mathbf{C}_m^H\mathbf{C}_m\big)}{\det\big(\mathbf{C}_{m-1}^H\mathbf{C}_{m-1}\big)} \tag{5.79}$$

for $m = 1, 2, \cdots, M$, where we make the stipulation that $\mathbf{C}_{00} = 1$. Now, let us define $\widetilde{\mathbf{C}} \triangleq \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{F}$, where $\mathbf{\Sigma}$ is the covariance matrix of the Wishart distributed $\mathbf{H}^H\mathbf{H}$. If we apply the QR-decomposition to $\widetilde{\mathbf{C}}$ and then, follow similar steps leading to Eq. (5.79), we can derive

$$\tilde{r}_{mm}^2 = \frac{\det\big(\widetilde{\mathbf{C}}_m^H\widetilde{\mathbf{C}}_m\big)}{\det\big(\widetilde{\mathbf{C}}_{m-1}^H\widetilde{\mathbf{C}}_{m-1}\big)}. \tag{5.80}$$

Since $\mathbf{F}^H\mathbf{\Sigma}\mathbf{F}$ is a positive definite matrix, we can apply the Cholesky decomposition [151] such that

$$\widetilde{\mathbf{C}}^H\widetilde{\mathbf{C}} = \mathbf{F}^H\mathbf{\Sigma}\mathbf{F} = \widetilde{\mathbf{L}}\mathbf{D}\widetilde{\mathbf{L}}^H \tag{5.81}$$

where $\widetilde{\mathbf{L}}$ is a lower triangular matrix having unity diagonal elements, and $\mathbf{D} = \text{diag}(d_1, d_2, \cdots, d_M)$. On the other hand, we know that

$$\widetilde{\mathbf{C}}^H\widetilde{\mathbf{C}} = \widetilde{\mathbf{R}}^H\widetilde{\mathbf{R}} = \widetilde{\mathbf{L}}\,\text{diag}(\widetilde{r}_{11}^2, \cdots, \widetilde{r}_{mm}^2)\,\widetilde{\mathbf{L}}^H.$$

Therefore, we have

$$d_m = \tilde{r}_{mm}^2 \;\; = \;\; \frac{\det\big(\widetilde{\mathbf{C}}_m^H\widetilde{\mathbf{C}}_m\big)}{\det\big(\widetilde{\mathbf{C}}_{m-1}^H\widetilde{\mathbf{C}}_{m-1}\big)} \tag{5.82}$$

for $m = 1, 2, \cdots, M$. The sequence $\{d_m\}$ is called the Cholesky values. Now, we are ready to state the probability density function of the scaled version of random variables r_{mm}^2

Property 2 *If we let*

$$\tau_m = \frac{det([\mathbf{F}^H\mathbf{H}^H\mathbf{H}\mathbf{F}]_{mm})}{det([\mathbf{F}^H\mathbf{H}^H\mathbf{H}\mathbf{F}]_{(m-1)(m-1)})} \cdot \frac{det([\mathbf{F}^H\mathbf{\Sigma}\mathbf{F}]_{(m-1)(m-1)})}{det([\mathbf{F}^H\mathbf{\Sigma}\mathbf{F}]_{mm})} = r_{mm}^2/d_m, \tag{5.83}$$

then, the probability density function of τ_m is given by [138]

$$p(\tau_m) = \frac{1}{\Gamma(N-m+1)}\;\tau_m^{N-m}e^{-\tau_m}, \qquad \tau_m > 0, \qquad m = 1, \cdots, M, \tag{5.84}$$

i.e., $\tau_1, \tau_2, \cdots, \tau_M$ are independent χ^2-distributed *with $2(N-m+1)$ degrees of freedom.* ■

By Property 2, we can significantly simplify the objective function of Problem 1. To this end, we use two alternative formulae of the Q-function in Eq. (4.40) and

Eq. (4.45), to represent our objective function such that

$$P_e(\mathbf{H}) = \frac{1}{M}\sum_{m=1}^{M}\left(\frac{4}{\pi}\Big(1-\frac{1}{\sqrt{K}}\Big)\int_0^{\frac{\pi}{2}} e^{-\frac{3\rho r_{mm}^2}{2(K-1)\sin^2\theta}}d\theta - \frac{4}{\pi}\Big(1-\frac{1}{\sqrt{K}}\Big)^2\int_0^{\frac{\pi}{4}} e^{-\frac{3\rho r_{mm}^2}{2(K-1)\sin^2\theta}}d\theta\right)$$
$$= \frac{1}{M}\sum_{m=1}^{M}\left(\frac{4}{\pi}\Big(1-\frac{1}{\sqrt{K}}\Big)\frac{1}{\sqrt{K}}\int_0^{\frac{\pi}{4}} e^{-\frac{3\rho r_{mm}^2}{2(K-1)\sin^2\theta}}d\theta + \frac{4}{\pi}\Big(1-\frac{1}{\sqrt{K}}\Big)\int_{\frac{\pi}{4}}^{\frac{\pi}{2}} e^{-\frac{3\rho r_{mm}^2}{2(K-1)\sin^2\theta}}d\theta\right). \tag{5.85}$$

Using Property 2 and taking the expectation on both sides of (5.85) yields

$$\mathbb{E}_{\mathbf{H}}\left[P_e(\mathbf{H})\right] = \frac{1}{M}\sum_{m=1}^{M}\frac{4}{\pi}\Big(1-\frac{1}{\sqrt{K}}\Big)\frac{1}{\sqrt{K}}\frac{1}{W(N-m+1)}\int_0^{\pi/4}\int_0^{\infty} e^{-\tau_m}e^{-\frac{3\rho d_m}{3(K-1)\sin^2\theta}\tau_m}\tau_m^{N-m}$$
$$\times d\tau_m d\theta + \frac{1}{M}\sum_{m=1}^{M}\frac{4}{\pi}\Big(1-\frac{1}{\sqrt{K}}\Big)\frac{1}{W(N-m+1)}\int_{\pi/4}^{\pi/2}d\theta\int_0^{\infty} e^{-\tau_m}e^{-\frac{3\rho d_m}{2(K-1)\sin^2\theta}\tau_m}\tau_m^{N-m}d\tau_m$$
$$= \frac{1}{M}\sum_{m=1}^{M}\frac{4}{\pi}\Big(1-\frac{1}{\sqrt{K}}\Big)\frac{1}{\sqrt{K}}\int_0^{\pi/4}\frac{1}{\big(1+\frac{3\rho d_m}{2(K-1)\sin^2\theta}\big)^{N-m+1}}d\theta \quad +$$
$$\frac{1}{M}\sum_{m=1}^{M}\frac{4}{\pi}\Big(1-\frac{1}{\sqrt{K}}\Big)\int_{\pi/4}^{\pi/2}\frac{1}{\big(1+\frac{3\rho d_m}{2(K-1)\sin^2\theta}\big)^{N-m+1}}d\theta. \tag{5.86}$$

When SNR is high, by replacing $\sin\theta$ in (5.86) by one, it can be approximated by

$$\mathbb{E}_{\mathbf{H}}\left[P_e(\mathbf{H})\right] \approx \frac{1}{M}\Big(1-\frac{1}{K}\Big)\sum_{m=1}^{M}\frac{1}{\Big(1+\frac{3\rho d_m}{2(K-1)}\Big)^{N-m+1}}. \tag{5.87}$$

Hence, the original design Problem 1 becomes:

Formulation 1 *Find a precoder matrix* $\mathbf{F}$ *such that*

$$\mathbf{F}_{\text{opt}} = \text{argmin}_{\mathbf{F}}\sum_{m=1}^{M}\frac{1}{\big(1+Cd_m\big)^{N-m+1}} \tag{5.88a}$$
$$\text{s.t.} \quad \text{tr}(\mathbf{F}^H\mathbf{F}) \le p_0, \tag{5.88b}$$

where $C = \frac{3\rho}{2(K-1)}$. ■

5.2.2 Structure of the Optimal Solution

In order to further simplify the objective function, we need to characterize the structure of the optimal solution. Therefore, we introduce the mathematical concept of majorization [145, 152], which now becomes a powerful tool in the optimal design of precoders for communication systems in which both the transmitter and the receiver know perfect channel state information [143, 153–156].

Let $\mathbf{a} = [a_1, a_2, \cdots, a_M]^T$ and $\mathbf{b} = [b_1, b_2, \cdots, b_M]^T$ be two M-dimensional real-valued sequences satisfying $a_{[1]} \geq a_{[2]} \geq \cdots \geq a_{[M]}$ and $b_{[1]} \geq b_{[2]} \geq \cdots \geq b_{[M]}$.

Definition 4 *A sequence* $\mathbf{a}$ *is said to be* additively majorized *by a sequence* $\mathbf{b}$*, denoted by* $\mathbf{a} \prec_+ \mathbf{b}$*, if*

$$\sum_{m=1}^{K} a_{[m]} \leq \sum_{m=1}^{K} b_{[m]}, \qquad 1 \leq K < M \tag{5.89a}$$

$$\sum_{m=1}^{M} a_{[m]} = \sum_{m=1}^{M} b_{[m]}. \tag{5.89b}$$

■

Definition 5 *If* $a_1, a_2, \cdots, a_M$ *and* $b_1, b_2, \cdots, b_M$ *are positive numbers. The sequence*

a *is said to be* multiplicatively majorized *by sequence* **b**, *denoted by* $\mathbf{a} \prec_{\times} \mathbf{b}$, *if*

$$\prod_{m=1}^{K} a_{[m]} \leq \prod_{m=1}^{K} b_{[m]}, \qquad 1 \leq K < M \tag{5.90a}$$

$$\prod_{m=1}^{M} a_{[m]} = \prod_{m=1}^{M} b_{[m]}. \tag{5.90b}$$

■

We now present three important properties of the Cholesky values of $\mathbf{F}^H\mathbf{\Sigma}\mathbf{F}$:

Property 3 *[145, 152] Let* $\{\alpha_m\}$, $m = 1, 2, \cdots, M$ *be the eigenvalues of* $\mathbf{F}^H\mathbf{\Sigma}\mathbf{F}$. *Then,*

$$\tilde{\mathbf{d}} \prec_{\times} \boldsymbol{\alpha} \tag{5.91}$$

where $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_M]^T$ *and* $\tilde{\mathbf{d}} = [d_1, \cdots, d_M]^T$ *is defined in Eq.* (5.82). *Conversely, if* $\{\alpha_m\}_{m=1}^M$ *majorizes* $\{d_m\}_{m=1}^M$ *multiplicatively as described in Eq. (5.91), then, for an arbitrarily given desired permutation,* $\{d_{j_1}, d_{j_2}, \cdots, d_{j_M}\}$, *of the sequence* $\{d_m\}_{i=m}^M$, *there exists a positive definite matrix* $\mathbf{F}^H\mathbf{\Sigma}\mathbf{F}$ *such that* $\{\alpha_m\}_{i=m}^M$ *and* $\{d_{j_m}\}_{m=1}^M$ *are the eigenvalues and the Cholesky values of* $\mathbf{F}^H\mathbf{\Sigma}\mathbf{F}$, *respectively.* ■

The following theorem and its corollaries on the parameters of the matrix $\mathbf{F}^H\mathbf{\Sigma}\mathbf{F}$ plays a crucial role in characterizing the structure of the optimal solution as well as in transforming the original non-convex optimization problem in Formulation 1 into a convex one.

Theorem 8*[H.1.a] [145] Let* $\{\alpha_m\}$, $\{\beta_m\}$, *and* $\{\gamma_m\}$, *where* $m = 1, 2, \cdots, M$, *be the*

eigenvalues of the positive-definite matrices $\mathbf{F}^H\mathbf{\Sigma}\mathbf{F}$, $\mathbf{F}^H\mathbf{F}$ *and* $\mathbf{\Sigma}$ *respectively. Then*

$$\prod_{m=1}^{k}\alpha_{[m]} \leq \prod_{i=1}^{k}\beta_{[m]}\gamma_{[m]}, \quad for \quad 1 \leq k < M \tag{5.92a}$$

$$\prod_{m=1}^{M}\alpha_{[m]} = \prod_{m=1}^{M}\beta_{[m]}\gamma_{[m]} \tag{5.92b}$$

■

Now, combining Property 3 with Theorem 8, by transitivity, we have the following corollary.

Corollary 2

$$\prod_{m=1}^{k}d_{[m]} \leq \prod_{i=1}^{k}\beta_{[m]}\gamma_{[m]}, \quad for \quad 1 \leq k < M \tag{5.93a}$$

$$\prod_{m=1}^{M}d_{[m]} = \prod_{m=1}^{M}\beta_{[m]}\gamma_{[m]} \tag{5.93b}$$

■

As a direct consequence of Property 3 and Corollary 2, we have

Corollary 3 *For any arbitrarily given desired permutation,* $\{d_{j_1}, d_{j_2}, \cdots, d_{j_M}\}$*, of* $\{d_m\}_{m=1}^{M}$*, there exists a positive definite matrix* $(\bar{\mathbf{F}}^H\mathbf{\Sigma}\bar{\mathbf{F}})$ *such that* $\{(\beta_m\gamma_m)\}_{m=1}^{M}$ *and* $\{d_m\}_{m=1}^{M}$ *are the eigenvalues and the Cholesky values of* $\bar{\mathbf{F}}^H\mathbf{\Sigma}\bar{\mathbf{F}}$*, respectively.* ■

Since the objective function in the Formulation 1 depends only on the Cholesky values of $\mathbf{F}^H\mathbf{\Sigma}\mathbf{F}$. It can be inferred from Corollary 3 that there exists another matrix $\bar{\mathbf{F}}$ such that

1. it diagonalizes the covariance matrix $\mathbf{\Sigma}$, i.e., if we let the eigenvalue decomposition of $\mathbf{\Sigma}$ be $\mathbf{\Sigma} = \mathbf{U}\text{diag}(\gamma_{[1]}, \gamma_{[2]}, \cdots, \gamma_{[M]})\mathbf{U}^H$, then, the singular value decomposition of $\bar{\mathbf{F}}$ is can be written in the form of $\bar{\mathbf{F}} = \mathbf{U}\text{diag}(\sqrt{\beta_{[1]}}, \sqrt{\beta_{[2]}}, \cdots, \sqrt{\beta_{[M]}})\mathbf{V}^H$, where $\mathbf{U}$ and $\mathbf{V}$ are $M \times M$ unitary matrices, and

2. the Cholesky values of $\bar{\mathbf{F}}^H\mathbf{\Sigma}\bar{\mathbf{F}}$ are equal to those given by $\mathbf{F}^H\mathbf{\Sigma}\mathbf{F}$.

In other words, if the feasible set of the optimization problem in Formulation 1 is constrained on a family of all such $\bar{\mathbf{F}}$, we will not lose its optimality. Therefore, in the following we maintain this kind of the optimal structure and only need to consider how to jointly and optimally distribute two power sequences $\{\beta_m\}$ and $\{d_m\}$ into each eigen-channel and Cholesky-channel, respectively, and then, to optimize the unitary matrix $\mathbf{V}$.

5.2.3 Optimal Order of the Cholesky Values

As we have discussed in Subsection 5.2.2, the optimization problem in Formulation 1 is reduced to finding the optimum β_m and d_m. To this end, we must solve the optimal order issue on the Cholesky values of $\mathbf{F}^H\mathbf{\Sigma}\mathbf{F}$ required in Corollary 3 for the product majorization. Now, a deep investigation of the objective function in (5.88a) reveals the following two features:

1. The objective function in (5.88a) is of the form $s = \sum_{m=1}^{M} a_m^{-b_{[m]}}$, where $a_m = 1 + Cd_m$, and $\{b_{[m]}\}$ is a decreasing sequence.

2. If we make an arrangement of the sequence $\{d_m\}$ in ascending order $\{d_{(m)}\}$ such that $d_{(1)} \leq d_{(2)} \leq \cdots \leq d_{(M)}$, then $\{d_{(m)}\}$ is an increasing sequence, and $\{a_{(m)}\}$

is also an increasing sequence.

The above observations lead us to find the optimal order of the Colesky values in the objective function in (5.88a), which will play an essential role in developing a numerical optimization algorithm for our design problem.

Lemma 4 *Let* $s = a_{(1)}^{-b_{[1]}} + a_{(2)}^{-b_{[2]}} + \cdots + a_{(M)}^{-b_{[M]}}$. *If we interchange the position of* $a_{(i)}$ *with* $a_{(j)}$ *in* s *to form* $s_{ij} = a_{(1)}^{-b_{[1]}} + \cdots + a_{(j)}^{-b_{[i]}} + \cdots + a_{(i)}^{-b_{[j]}} + \cdots + a_{(M)}^{-b_{[M]}}$, *then,* $s < s_{ij}$ *for* $j > i$ *if the following condition is satisfied:*

$$d_{(k)} \geq \frac{\left(\frac{N-k+1}{N-M+1}\right)^{1/(M-k)} - 1}{C} \quad \textit{for} \quad 1 \leq k < M \tag{5.94}$$

■

Proof: Lemma 4 can be shown by simply taking the difference between s and s_{ij} such that

$$\begin{aligned} s - s_{ij} &= a_{(i)}^{-b_{[i]}} + a_{(j)}^{-b_{[j]}} - a_{(j)}^{-b_{[i]}} - a_{(i)}^{-b_{[j]}} \\ &= \frac{1}{\left(1+Cd_{(i)}\right)^{N-i+1}} + \frac{1}{\left(1+Cd_{(j)}\right)^{N-j+1}} - \frac{1}{\left(1+Cd_{(j)}\right)^{N-i+1}} - \frac{1}{\left(1+Cd_{(i)}\right)^{N-j+1}} \\ &= \frac{\left(1+Cd_{(j)}\right)^{j-i} - 1}{\left(1+Cd_{(j)}\right)^{N-i+1}} - \frac{\left(1+Cd_{(i)}\right)^{j-i} - 1}{\left(1+Cd_{(i)}\right)^{N-i+1}} \end{aligned} \tag{5.95}$$

This leads us to considering the monotonicity of a function $f(x) = \frac{(1+x)^{j-i}-1}{(1+x)^{N-i+1}}$. Notice that the first-order derivative of $f(x)$ is given by

$$f'(x) = (N-j+1)\Big(\frac{1}{1+x}\Big)^{N-j+2}\Bigg(\frac{N-i+1}{N-j+1}\Big(\frac{1}{1+x}\Big)^{j-i} - 1\Bigg). \tag{5.96}$$

When $x = x_0 = \left(\frac{N-i+1}{N-j+1}\right)^{1/(j-i)} - 1$, $f'(x_0) = 0$. Since $f'(x) \leq 0$ for $x \geq x_0$, $f(x)$ is a decreasing function if $x \geq x_0$. Therefore, if we are able to prove that

$$Cd_{(i)} \geq \left(\frac{N-i+1}{N-m+1}\right)^{1/(m-i)} - 1, \quad m = i+1, i+2, \cdots, M, \tag{5.97}$$

then, for $j > i$ and $m = j$ we have

$$Cd_{(j)} \geq Cd_{(i)} \geq \left(\frac{N-i+1}{N-j+1}\right)^{1/(j-i)} - 1. \tag{5.98}$$

Now, utilizing the monotonic deceasing property of $f(x)$ for $x > x_0$ we can obtain that $s - s_{ij} \leq 0$. In the following we will show that the inequality (5.97) is indeed true for $i = 1, 2, \cdots, M-1$ when the condition (5.94) is satisfied. To do that, let a function $g(x)$ be defined as $g(x) = \left(\frac{N-i+1}{N-x+1}\right)^{1/(x-i)}$ for $i < x \leq M$. Since $N \geq M \geq x > i$, $\frac{N-i+1}{N-x+1} > 1$ and $g(x) > 1$. In addition, the first order derivative of $g(x)$ is given by

$$g'(x) = \frac{1}{x-i}\left(\frac{1}{N-x+1} - \frac{1}{x-i}\ln\left(\frac{N-i+1}{N-x+1}\right)\right) g(x). \tag{5.99}$$

On the other hand, using the fact that $\ln x < x - 1$ for $x > 1$ and $\frac{N-i+1}{N-x+1} > 1$, we have $\ln\left(\frac{N-i+1}{N-x+1}\right) < \frac{N-i+1}{N-x+1} - 1$. Now, applying this inequality to the left hand side of Eq. (5.99) yields

$$g'(x) > \frac{1}{x-i}\left(\frac{1}{N-x+1} - \frac{1}{x-i}\left(\frac{N-i+1}{N-x+1} - 1\right)\right) g(x) = 0. \tag{5.100}$$

Hence, $g(x)$ is a increasing function for $i < x \leq M$. As a result, for any $m = i+1, i+2, \cdots, M$, we have $\left(\frac{N-i+1}{N-m+1}\right)^{1/(m-i)} - 1 = g(m) \leq g(M) = \left(\frac{N-i+1}{N-M+1}\right)^{1/(M-i)} - 1$.

Therefore, if the condition (5.94) is met, then,

$$Cd_{(i)} \geq \Big(\frac{N-i+1}{N-M+1}\Big)^{1/(M-i)} - 1 \geq \Big(\frac{N-i+1}{N-m+1}\Big)^{1/(m-i)} - 1 \tag{5.101}$$

for $m = i+1, i+2, \cdots, M$. This completes the proof of Lemma 4. ■

Lemma 4 tells us that for any sum of sequence of the form $s = \sum_{m=1}^{M} a_m^{-b_{[m]}}$, the minimum value s is attained if $\{d_m\}$ is arranged in an ascending order and (5.94) is satisfied. Thus, *to achieve the minimum of the objective in Problem 1, we must keep the sequence* $\{d_m\}$ *in an ascending order of* $\{d_{(m)}\}$ such that $d_{(1)} \leq d_{(2)} \leq \cdots \leq d_{(M)}$, and $d_{(k)} \geq \frac{\left((N-k+1)/(N-M+1)\right)^{1/(M-k)}-1}{C}, 1 \leq k < M$. It is this lemma that helps us successfully transform the origonal non-convex optimization problem in Fomulation 1 into a convex problem, whose detail will be developed in the ensuing subsection.

5.2.4 Reformulations

We now transform Formulation 1 into a convex optimization problem. First, let the eigenvalue decomposition of $\mathbf{\Sigma}$ be $\mathbf{\Sigma} = \mathbf{U}\ \mathrm{diag}(\gamma_{[1]}, \cdots, \gamma_{[M]})\ \mathbf{U}^H$ and then, let

$$\begin{aligned} & & u_m &= \ln\ d_{[m]} = \ln\ d_{(M-m+1)} & \\ \text{such that} & & e^{u_{M-m+1}} &= d_{(m)}, & (5.102a) \\ & & v_m &= \ln\ \beta_{[m]}, & (5.102b) \\ & & w_m &= \ln\ \gamma_{[m]},, & (5.102c) \end{aligned}$$

for $m = 1, 2, \ldots, M$, where $\beta_{[m]}$ and $\gamma_{[m]}$ are the m-th largest eigenvalues of $\mathbf{F}^H\mathbf{F}$ and $\mathbf{\Sigma}$ respectively. Applying Theorem 8, Corollaries 2 and 3 to Formulation 1, we can

now reformulate the design problem as the following optimization problem:

Formulation 2 *Find two sequences* $\{u_m\}$ *and* $\{v_m\}$ *such that*

$$\{\boldsymbol{u}_{\text{opt}}, \boldsymbol{v}_{\text{opt}}\} = \arg\min_{\{u_m\}} \quad \sum_{m=1}^{M} \left(1 + Ce^{u_{M-m+1}}\right)^{-(N-m+1)} \tag{5.103a}$$

$$\text{s.t.} \quad \sum_{m=1}^{k} u_m, \leq \sum_{m=1}^{k} v_m + \sum_{m=1}^{k} w_m \quad 1 \leq k < M, \tag{5.103b}$$

$$\sum_{m=1}^{M} u_m = \sum_{m=1}^{M} v_m + \sum_{m=1}^{M} w_m, \tag{5.103c}$$

$$v_1 \geq v_2 \geq \cdots \geq v_M, \tag{5.103d}$$

$$u_m \geq \ln\left(\frac{\sqrt[M-m]{(N-m+1)/(N-M+1)} - 1}{C}\right), \quad 1 \leq m \leq M \tag{5.103e}$$

$$\sum_{m=1}^{M} e^{v_m} \leq p_0 \quad \text{(power constraint)}. \tag{5.103f}$$

■

Equation (5.103a) is the objective function in terms of u_m. Constraints (5.103b) and (5.103c) result from applying logarithm to (5.93). Constraint (5.103d) follows the order of $\tilde{\boldsymbol{\beta}}$ for majorization, the constraint (5.103e) is required by Lemma 4 in the equivalent logarithm domain, which can be satisfied when SNR is slightly large, and constraint (5.103f) is the power constraint represented in terms of v_m. All these constraints (5.103b)-(5.103f) are convex constraints. Now, we need to check whether or not the objective function in Formulation 2 is convex. Let $f(x) = \left(1 + Ce^x\right)^{-\tau}$, where τ is a given positive number. Then, the first order and second order derivatives of $f(x)$ with respect to x is given, respectively, by $f'(x) = -\tau(1 + Ce^x)^{-\tau-1}Ce^x$ and $f''(x) = -\tau Ce^x(1 + Ce^x)^{-\tau-1}\big((-\tau - 1)(1 + Ce^x)^{-1}Ce^x + 1\big)$. Now, we can see that if

$x \geq \ln \frac{1}{\tau C}$, then, $f''(x) \geq 0$ and thus, $f(x)$ is a convex function. Therefore, in order to make the optimization problem in Formulation 2 is a convex optimization problem, we must add M extra constraints: $u_{M-m+1} \geq \ln \frac{1}{(N-m+1)C}$ for $m = 1, 2, \cdots, M$, which is equivalent to the fact that $u_m \geq \ln \frac{1}{(N-M+m)C}$ for $m = 1, 2, \cdots, M$. The resulting convex problem is stated as follows:

Formulation 3 *Find two sequences* $\{u_m\}$ *and* $\{v_m\}$ *such that*

$$\{\boldsymbol{u}_{\text{opt}}, \boldsymbol{v}_{\text{opt}}\} = \arg\min_{\{u_m\}} \quad \sum_{m=1}^{M} \left(1 + Ce^{u_{M-m+1}}\right)^{-(N-m+1)} \tag{5.104a}$$

$$\text{s.t.} \quad \sum_{m=1}^{k} u_m \leq \sum_{m=1}^{k} v_m + \sum_{m=1}^{k} w_m \quad 1 \leq k < M \tag{5.104b}$$

$$\sum_{m=1}^{M} u_m = \sum_{m=1}^{M} v_m + \sum_{m=1}^{M} w_m \tag{5.104c}$$

$$v_1 \geq v_2 \geq \cdots \geq v_M \tag{5.104d}$$

$$u_m \geq \ln\left(\frac{\sqrt[M-m]{(N-m+1)/(N-M+1)} - 1}{C}\right), \quad 1 \leq m \leq M \tag{5.104e}$$

$$u_m \geq \ln \frac{1}{(N-M+m)C}, \quad 1 \leq m \leq M \tag{5.104f}$$

$$\sum_{m=1}^{M} e^{v_m} \leq p_0 \qquad \text{(power constraint)} \tag{5.104g}$$

■

The optimization problem in Formulation 3 is now convex and can be efficiently solved using the interior point method [47]. Here, we should point out the fact that the problem in Formulation 3 is not theoretically equivalent to that in Formulation 1 because of the extra constraints (5.104e) and (5.104f). However, they are practically equivalent when SNR is large.

5.3 Optimum Precoder Designs

As we have shown in Section 5.2, the optimization problem in Formulation 3 is convex and its solution can be efficiently found employing interior point methods. The solution produces numerically the optimum values of $\{u_m\}$ and $\{v_m\}$. Therefore, from the optimum values of $\{u_m\}$, utilizing (5.102a), we can attain the optimum $\{d_{(m)}\}$, the Cholesky values of $\mathbf{F}^H\mathbf{\Sigma}\mathbf{F}$ and thus, $\{\tilde{r}^2_{(mm)}\}$, since $\{d_{(m)}\} = \{\tilde{r}^2_{(mm)}\}$. In addition, from the optimum values of $\{v_m\}$, we can attain the optimum $\{\beta_{[m]}\}$, the eigenvalues of $\mathbf{F}^H\mathbf{F}$. Our algorithm for obtaining an optimum precoder based on the optimum values of $\{d_{(m)}\}$ and $\{\beta_{[m]}\}$ is now summarized in the following three successive steps:

1. Perform an eigen-decomposition on the given channel covariance, i.e., $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Delta}_\Sigma\mathbf{U}^H$, where $\mathbf{U}$ is the eigenvector matrix of $\mathbf{\Sigma}$ and $\mathbf{\Delta}_\Sigma = \text{diag}(\gamma_{[1]}, \cdots, \gamma_{[M]})$ is its eigenvalue matrix.

2. Using the eigenvector matrix $\mathbf{U}$ and the optimum eigenvalues $\{\beta_{[m]}\}$, constitute the optimum precoder:

$$\mathbf{F}_{\text{opt}} = \mathbf{U}\ \text{diag}(\sqrt{\beta_{[1]}}, \cdots, \sqrt{\beta_{[M]}})\ \mathbf{S}_{\text{opt}} = \mathbf{U}\ \mathbf{\Delta}_F^{1/2}\ \mathbf{S}_{\text{opt}} \tag{5.105}$$

   where $\mathbf{S}_{\text{opt}}$ is to be determined, and $\mathbf{\Delta}_F^{1/2} = \text{diag}(\sqrt{\beta_{[1]}}, \cdots, \sqrt{\beta_{[M]}})$. This results in

$$\mathbf{F}^H_{\text{opt}}\mathbf{\Sigma}\mathbf{F}_{\text{opt}} = \mathbf{S}^H_{\text{opt}}\mathbf{\Delta}_F^{1/2}\mathbf{\Delta}_\Sigma\mathbf{\Delta}_F^{1/2}\mathbf{S}_{\text{opt}} = \mathbf{S}^H_{\text{op}}\mathbf{\Theta}\mathbf{S}_{\text{opt}} \tag{5.106}$$

   where $\mathbf{\Theta} = \mathbf{\Delta}_F^{1/2}\mathbf{\Delta}_\Sigma\mathbf{\Delta}_F^{1/2}$.

3. To obtain $\mathbf{S}_{\text{opt}}$, we apply the closed-form QRS decomposition algorithm ([46,

152, 157, 158]) to $\mathbf{\Theta}^{1/2}$ with $r_{mm} = \sqrt{d_m}$ and r_{mm} being in *ascending order* (so that $\{d_{(m)}^{-1}\}$ is in descending order to minimize the objective function as shown in Lemma 4). This optimal $\mathbf{S}_{\text{opt}}$, together with the eigenvector matrix $\mathbf{U}$ of $\mathbf{\Sigma}$, and the optimum eigenvalue matrix $\mathbf{\Delta}_F$ of $\mathbf{F}^H\mathbf{F}$, forms the optimum precoder $\mathbf{F}_{\text{opt}}$ determined by (5.105).

It is known from [48,159] that for a system employing the ZF-DF detector, when CSI is perfectly available at both the transmitter and receiver, the optimal precoder is a QRS decomposition of $\mathbf{HF}$ having equal diagonal elements in the R-factor. However, in the case where CSI is attainable only at the receiver and channel statistics known at the transmitter, the above design tells us that the optimal precoder structure is the QRS decomposition of $\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{F}$ in which the diagonal entries of the R-factor form an increasing sequence. This optimal structure roots essentially in the fact that $\tau_m = r_{mm}^2/d_m$ for $m = 1, \cdots, M$, are χ^2 distributed with decreasing values of degrees of freedoms. As a result, estimation of x_{m-1} turns out more reliable than that of x_m for $1 < m \leq M$. Therefore, when the ZF-DF receiver is employed with a given detection order, to attain a good average error performance, more power ought be distributed to x_m than to x_{m-1} and increasingly so to x_M, leading to this non-decreasing diagonal of the R-factor in the QRS decomposition.

5.4 Asymptotic Error Performance in Large MIMO Systems

In this section, we will scale up the array size of both the transmitter and the receiver sides to form a large MIMO system. Then, we propose a novel asymptotic SEP analysis approach to a specific precoder $\tilde{\mathbf{F}} = \frac{1}{\sqrt{M}}\mathbf{S}$ based on the QRS decomposition [48], where $\mathbf{S}$ is a unitary matrix such that $\mathbf{\Sigma}^{1/2}\mathbf{S} = \bar{\mathbf{Q}}\bar{\mathbf{R}}$, with each diagonal entry of $\bar{\mathbf{R}}$ being $\det(\mathbf{\Sigma})^{1/(2M)}$.

Suppose that the ratio of the number of the receiver antennas to that of the transmitter antennas is a constant, i.e., $N/M = \beta$ is fixed. Then, our main results can be formally stated as the following theorem.

Theorem 9 *Consider the QRS-precoded large MIMO system. If the entries of the channel covariance matrix* $\mathbf{\Sigma}$ *are absolutely square summable, i.e.,* $\sum_{\ell=0}^{\infty}|\alpha(\ell)|^2$ *is convergent, and* $0 < \mathsf{L}_{S_\Sigma} \leq \mathsf{U}_{S_\Sigma} < \infty$*, then, we have*

$$\lim_{M\to\infty} P_e(\tilde{\mathbf{F}}) = \frac{4(\sqrt{K}-1)}{\pi K}\int_0^{\frac{\pi}{4}} \Psi(\theta)d\theta + \frac{4(\sqrt{K}-1)}{\pi\sqrt{K}}\int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \Psi(\theta)d\theta,$$

where $\Psi(\theta) = \frac{2(K-1)\sin^2\theta}{3\rho\gamma}\left(e^{-\frac{3\rho(\beta-1)\gamma}{2(K-1)\sin^2\theta}} - e^{-\frac{3\rho\beta\gamma}{2(K-1)\sin^2\theta}}\right)$ *and* $\gamma = e^{\frac{1}{2\pi}\int_0^{2\pi}\ln(s_\Sigma(\omega))d\omega}$. ■

Proof: First, we notice that for the QRS precoder, all the diagonal entries of $\tilde{\mathbf{R}}$ are equal to each other, i.e.,

$$\tilde{r}_{mm} = \frac{1}{\sqrt{M}}\bar{r}_{mm} = \frac{1}{\sqrt{M}}(\det\mathbf{\Sigma})^{1/(2M)}, \tag{5.107}$$

for $m = 1, 2, \ldots, M$. Let $\mathsf{F}(x) \triangleq \ln(x)$ which is continuous on $(0, \infty)$ and $\mu_{M,1} \leq \mu_{M,2} \leq \cdots \leq \mu_{M,M}$ be the eigenvalues of $\mathbf{\Sigma}$. Since $\sum_{\ell=0}^{\infty}|\alpha(\ell)|^2 < \infty$ and $0 < \mathsf{L}_{s_\Sigma} \leq$

$\mathsf{U}_{s_\Sigma} < \infty$, by Lemma 3 in Chapter 4, we have

$$\begin{aligned}\lim_{M\to\infty} (\det \mathbf{\Sigma})^{1/M} &= e^{\lim_{M\to\infty} \frac{1}{M}\sum_{\ell=1}^{M} \ln \mu_{M,\ell}} \\ &= e^{\frac{1}{2\pi}\int_0^{2\pi} \ln(s_\Sigma(\omega))d\omega} = \gamma. \end{aligned} \tag{5.108}$$

On the other hand, under the equal diagonal QRS precoder, equation (5.86) can now be formulated by

$$\begin{aligned}\lim_{M\to\infty} P_e(\tilde{\mathbf{F}}) = &\frac{4(\sqrt{K}-1)}{\pi K}\int_0^{\frac{\pi}{4}} \lim_{M\to\infty} \frac{\tau^{-N}(1-\tau^M)}{M(1-\tau)} d\theta \\ &+ \frac{4(\sqrt{K}-1)}{\pi\sqrt{K}}\int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \lim_{M\to\infty} \frac{\tau^{-N}(1-\tau^M)}{M(1-\tau)} d\theta, \end{aligned} \tag{5.109}$$

where $\tau \triangleq 1 + \frac{3\rho\tilde{r}_{mm}^2}{2(K-1)\sin^2\theta}$. Since

$$\begin{aligned}\lim_{M\to\infty} \frac{\tau^{-N}(1-\tau^M)}{M(1-\tau)} &= \lim_{M\to\infty} -\frac{2(K-1)\sin^2\theta}{3\rho M\tilde{r}_{mm}^2} \times \\ &\quad \Big(\big(1 + \frac{3\rho\tilde{r}_{mm}^2}{2(K-1)\sin^2\theta}\big)^{-N} - \big(1 + \frac{3\rho\tilde{r}_{mm}^2}{2(K-1)\sin^2\theta}\big)^{-(N-M)} \Big) \\ &= \frac{2(K-1)\sin^2\theta}{3\rho\gamma} \Big(e^{-\frac{3\rho(\beta-1)\gamma}{2(K-1)\sin^2\theta}} - e^{-\frac{3\rho\beta\gamma}{2(K-1)\sin^2\theta}} \Big). \end{aligned} \tag{5.110}$$

Substituting Eq. (5.110) into Eq. (5.109) completes the proof of Theorem 9. □

We would like to make the following two comments on Theorem 9.

1. If $\beta = 1$, i.e., $M = N$, then, the diversity gain of the limiting SEP is one with respect to SNR ρ.

2. If $\beta > 1$, i.e., $N > M$, the limiting SEP $\lim_{M\to\infty} P_e(\tilde{\mathbf{F}})$ decays exponentially in terms of SNR ρ.

5.4.1 A Case Study on Exponential Correlation Model

As an important application of Theorem 9, we consider an exponential correlation model [38], where the (m,n)-th entry of $\mathbf{\Sigma} = \mathbf{\Sigma}_{KMS}$ is given by

$$\alpha(m-n) = \begin{cases} \eta^{n-m} & m \leq n \\ \alpha^*(n-m) & m > n \end{cases} \tag{5.111}$$

with η being the (complex) correlation coefficients of two adjacent antennas. This matrix is sometimes known as (non-symmetric) Kac-Murdock-Szegö matrix. In this case, we have

$$s_{KMS}(\omega) = \sum_{\ell=-\infty}^{\infty} \alpha(\ell) e^{-j\ell\omega} = \frac{1-|\eta|^2}{1+|\eta|^2 - 2\mathrm{Re}[\eta e^{j\omega}]}$$

and thus,

$$\gamma = \gamma_{KMS} = e^{\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln\left(\frac{1-|\eta|^2}{1+|\eta|^2-2\mathrm{Re}[\eta e^{j\omega}]}\right) d\omega\right)} = 1-|\eta|^2.$$

Then we can evaluate the limiting SEP by Theorem 9.

5.4.2 Entropy Power of Channel

From Theorem 9, we know that the asymptotic error performance of our considered system is determined by γ, which is the geometric mean of the eigenvalues of the channel covariance matrix $\mathbf{\Sigma}$. This relationship is also known in estimation and information theory, but with different application, e.g., see [144, 160] and references therein.

Consider a stationary Gaussian process $\{X_\ell\}_{\ell=1}^{M}$ with a $M \times M$ Toeplitz covariance matrix equal to $\mathbf{\Sigma}$. As $M \to \infty$, the differential entropy rate can be expressed by

$$\begin{aligned} h(\mathcal{X}) &= \lim_{M\to\infty} \frac{h(X_1, X_2, \ldots, X_M)}{M} \\ &= \lim_{M\to\infty} \frac{\log(2\pi e)^M \det \mathbf{\Sigma}}{2M} \\ &= \frac{1}{2}\log 2\pi e + \frac{1}{2}\lim_{M\to\infty} \log\left(\det \mathbf{\Sigma}\right)^{1/M} \end{aligned} \tag{5.112}$$

For the Gaussian process [144], Kolmogorov showed that

$$h(\mathcal{X}) = \frac{1}{2}\log 2\pi e + \frac{1}{4\pi}\int_0^{2\pi} \log s_\Sigma(\omega) d\omega \tag{5.113}$$

where $s_\Sigma(\omega)$ is given in (4.59). Substituting Eq. (5.113) into Eq. (5.112), we have

$$\lim_{M\to\infty} \left(\det \mathbf{\Sigma}\right)^{1/M} = 2^{\frac{1}{2\pi}\int_0^{2\pi} \log s_\Sigma(\omega) d\omega} = e^{\frac{1}{2\pi}\int_0^{2\pi} \ln s_\Sigma(\omega) d\omega} = \gamma.$$

This verifies Eq. (5.108).

5.5 Numerical Simulations

In this section, computer simulations are carried out to verify our method for both MIMO system with finite and a large number of available antennas.

5.5.1 MIMO with Finite Number of Antennas

In this subsection, we examine the performance of the optimally precoded system equipped with the ZF-DF detector with finite number of antennas. We compare its

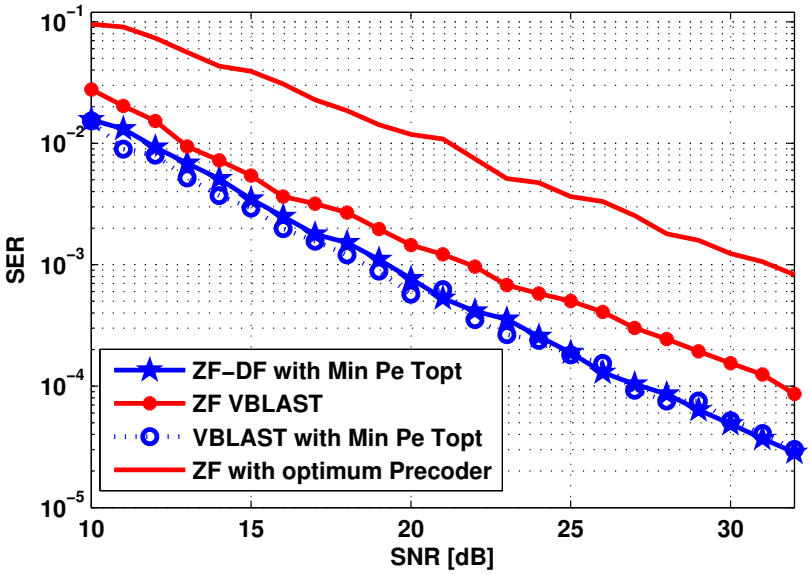


Figure 5.11: Simulation results when $\eta = 0.5e^{0.5j}$ $(N = M)$

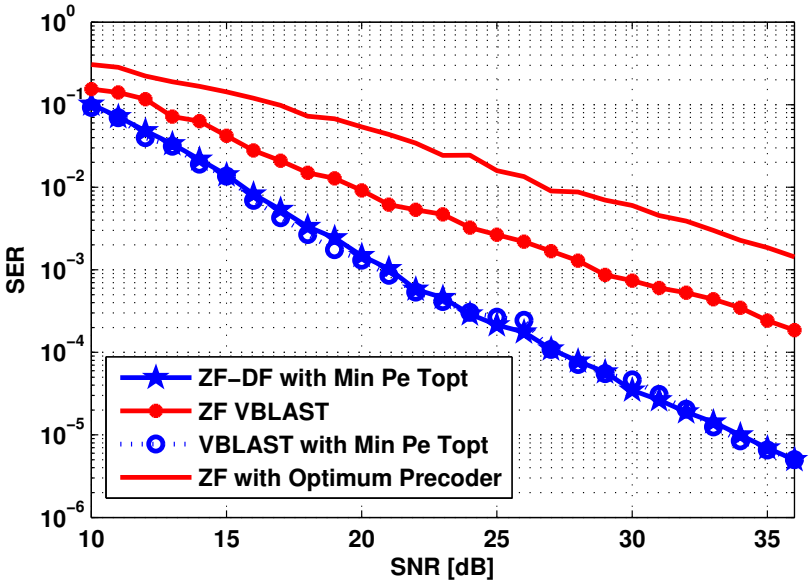


Figure 5.12: Simulation results when $\eta = 0.9e^{0.5j}$ $(N = M)$

performance with other schemes in the literatures which can be used in the scenario where channel state information is not completely available at the transmitter. The first scheme that we would like to compare is the optimally precoded system with the linear ZF receiver developed in [40], which first employed majorization theory as a major mathematical tool in the precoder design. With help of the same method, we actually extend the problem with the linear receiver to that with the non-linear ZF-DF receiver in this chapter. The comparison also covers the precoded system minimizing the average arithmetic mean-squared error (MSE) using the same ZF-DF receiver [46]. What is more, our main focus here is on the comparison with the celebrated V-BLAST scheme based on the ZF receiver [150]. In particular, we compare the scheme which combines our proposed optimal precoder for the ZF-DF receiver with the optimally ordered ZF-DF detector.

In the following examples, simulations are carried out for one MIMO system with 6 transmitter antennas and 10 receiver antennas ($N > M$) and another MIMO system with 6 transmitter antennas and 6 receiver antennas ($N = M$) transmitting symbols from a 4-QAM constellation.

Random realizations of correlated channels are generated following the exponential correlation model given in (5.111).

Example 1 $N = M$

We first investigate the symbol error rate (SER) performances of two MIMO systems where $N = M = 6$ under a moderately correlated channel fading environment with $\eta = 0.5e^{0.5j}$ and a highly correlated channel fading environment with $\eta = 0.9e^{0.5j}$. In this case, the precoded system [46] minimizing the average arithmetic MSE cannot

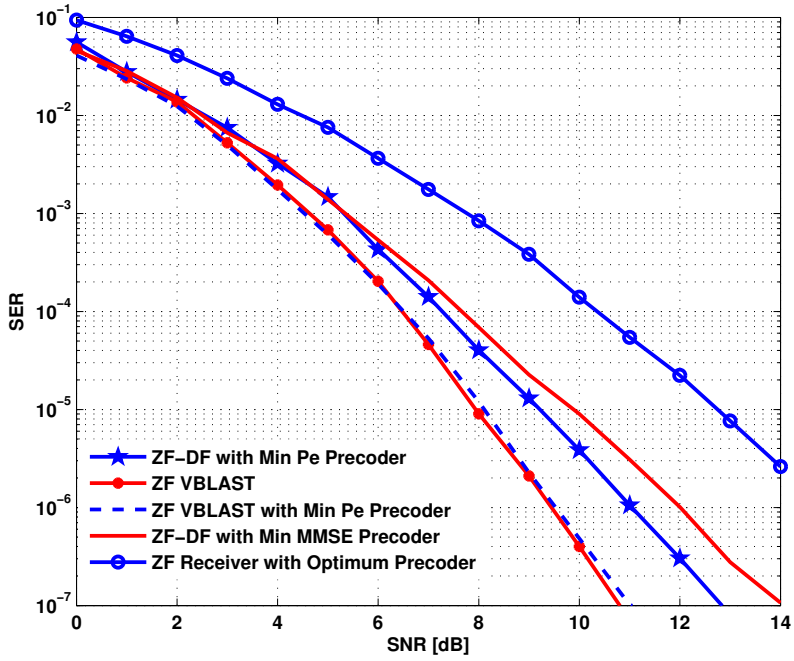


Figure 5.13: Simulation results when $\eta = 0.5e^{0.5j}$ $(N > M)$

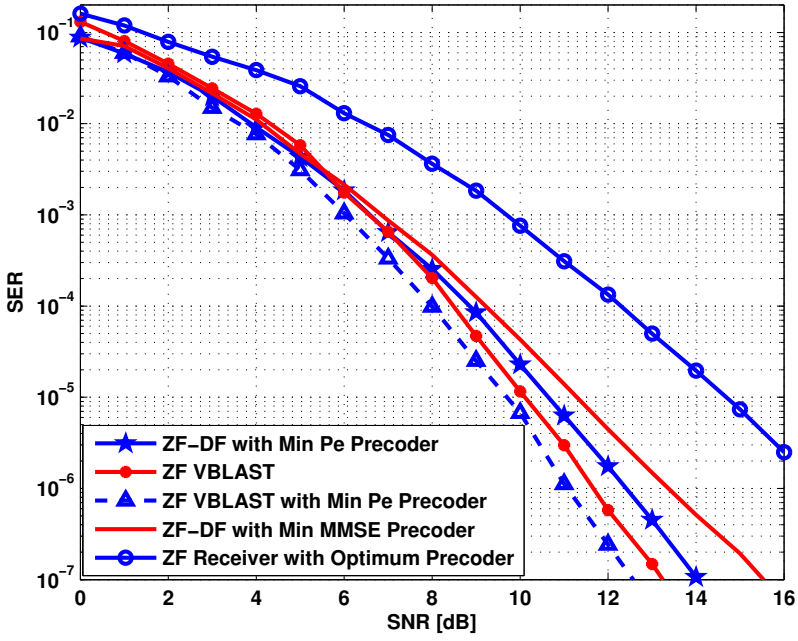


Figure 5.14: Simulation results when $\eta = 0.7e^{0.5j}$ $(N > M)$

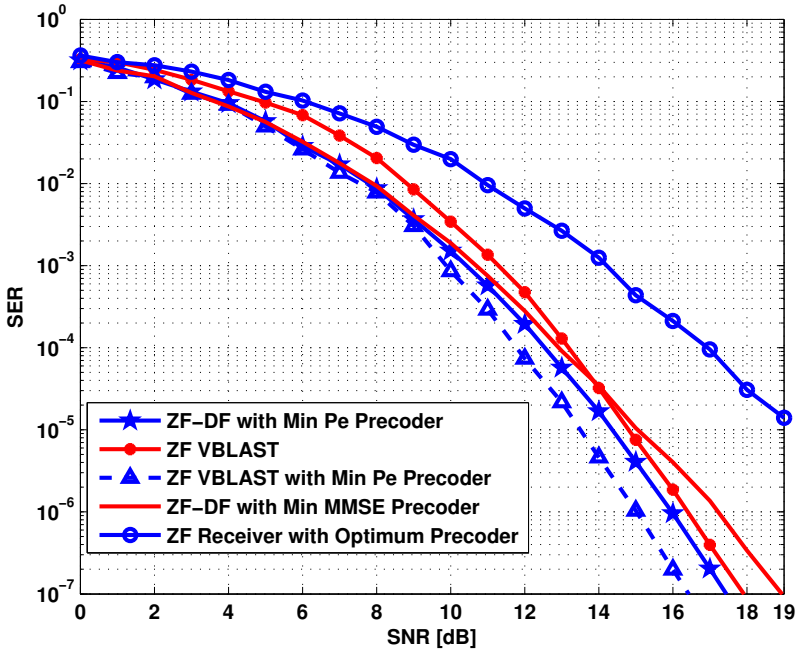


Figure 5.15: Simulation results when $\eta = 0.9e^{0.5j}$ $(N > M)$

be used anymore, since it requires that the number of the receiver antennas is larger than that of the transmitter antennas, i.e., $N > M$.

The SER performances are shown in Fig. 5.11 and Fig. 5.12, including our proposed optimally precoded system employing the ZF-DF detector, denoted by the star and the V-BLAST detector, denoted by the circle, the originally unprecoded system with the V-BLAST receiver, denoted by the dot, and the optimally precoded system using the linear ZF receiver [40], denoted by the solid line. It can be observed that there is a considerable deterioration of performance in the unprecoded ZF V-BLAST system and the optimally precoded system using the linear ZF receiver when the amplitude of correlation coefficient $|\eta|$ increases. However, the performances of the scheme equipped with the optimum precoder design developed in this chapter are both less sensitive to the correlation of the channels, especially when SNR is high.

What surprises us most is that the performance of our optimal scheme with the correlation coefficient $\eta = 0.9e^{0.5j}$ using the ZF-DF receiver is significantly better than that of the unprecoded ZF V-BLAST scheme. In particular, our system obtains about 10 dB SNR gain at the SER of 10^{-4} over the unprecoded V-BLAST system and even more than 15 dB gain over the optimal system using the linear ZF receiver [40]. In addition, our optimal system designed for the fixed detection order with the ZF-DF receiver has almost the same error performance as that of same scheme with the optimally ordered ZF-DF receiver, i.e., the ZF V-BLAST receiver. In other words, optimal ordering does not significantly affect the SER performance on our optimal system when the number of the transmitter antennas is equal to that of the receiver antennas.

Example 2 $N > M$

To make our investigation fair and complete, in this example we examine the case of $N > M$. Here, $N = 10$, $M = 6$, and the channel correlation coefficients are set to $\eta = 0.5e^{0.5j}$, $\eta = 0.7e^{0.5j}$ and $\eta = 0.9e^{0.5j}$ respectively. Figs. 5.13-5.15 show the performances of the various schemes with these correlation coefficients, from which we can see that there is severe error performance deterioration as the correlation coefficient becomes large. However, the system utilizing the proposed optimum precoder in this chapter outperforms all the other systems as the correlation coefficient is increasing. Specifically, at the SER of 10^{-7}, our scheme obtains a SNR gain of about 1.5 dB over the scheme equipped with the optimum precoder in [46]. We also observe from the Fig. 5.13 that when $\eta = 0.5e^{0.5j}$, our optimally precoded system and the original unprecoded system with the same ZF-VBLAST detector has almost the same

performance. However, as η increases, our scheme has the best performance among the other schemes. The larger $|\eta|$, the much better SER performance our scheme will obtain.

5.5.2 MIMO with A Large Number of Antennas

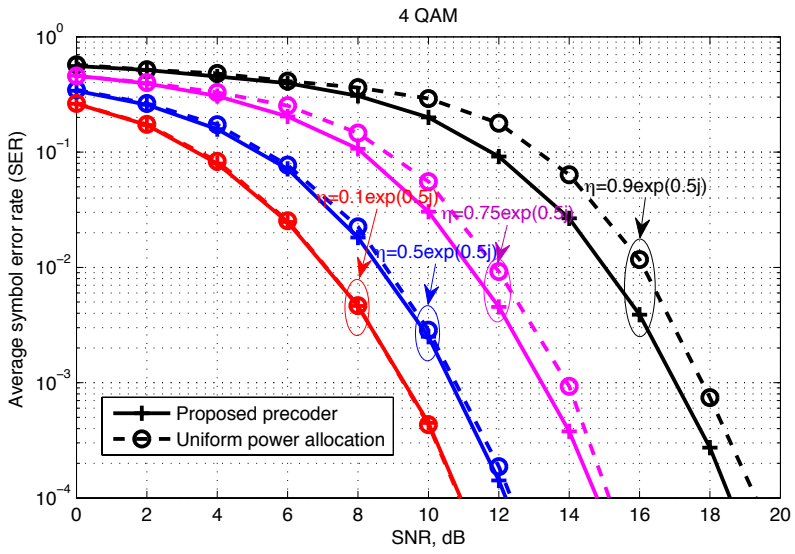


Figure 5.16: Simulated average SER against SNR η with $M = 50$ transmitter antennas and $N = 100$ receiver antennas.

In this subsection, computer simulations are carried out to verify our theoretical analysis with a large number of antennas. The error performance of the proposed precoder is plotted in Fig. 5.16 compared with a uniform power allocation method given by $\mathbf{F}_{\mathrm{U}} = \frac{1}{\sqrt{M}}\mathbf{I}_{M\times M}$ for different η. We can observe that the proposed precoder outperforms the uniform one, especially in a strong correlation case. In addition, when the channel is nearly uncorrelated (i.e., the curve where $\eta = 0.1\exp(0.5j)$), the error performance of the proposed approach is almost the same as that the uniform

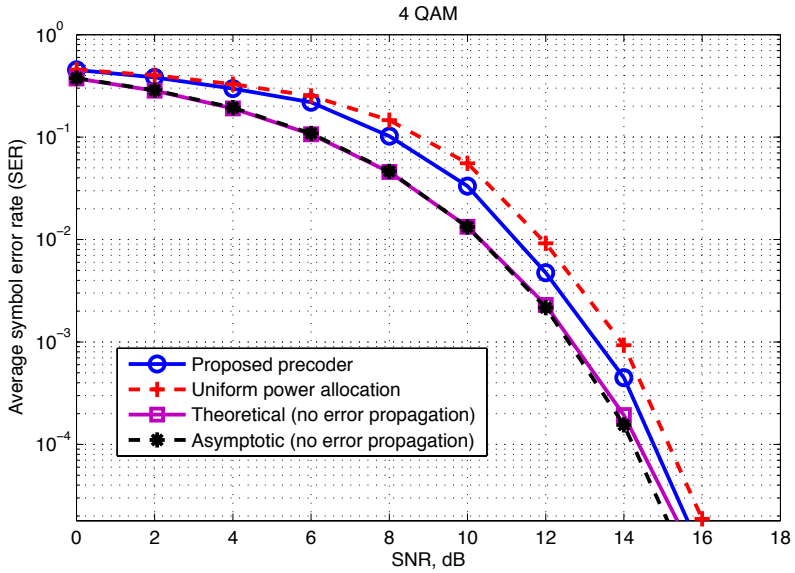


Figure 5.17: Comparison of simulated average SER to theoretical SEP against SNR, $\eta = 0.75\exp(0.5j)$ and $M = 50$ transmitter antennas and $N = 100$ receiver antennas.

power allocation precoder.

The theoretical SEP with no error propagation and the asymptotic SEP are illustrated in Fig. 5.17 along with the simulated symbol error rate (SER) of the proposed precoder and the uniform precoding strategy with $M = 50$ transmitter antennas and $N = 100$ receiver antennas. It can be seen that the asymptotic SEP is very close to the theoretical SEP, which verifies the accuracy of Theorem 9. It can be also expected that the simulated SER is larger than the theoretical one due to error propagation in practice.

To further show the asymptotic property of Theorem 9, we plot the theoretical and the asymptotic SEP against the number of transmitter antennas under various correlation coefficients and SNR in both Fig. 5.18 and Fig. 5.19. We can see that as the number of antennas increases, the asymptotic SEP and the theoretical SEP

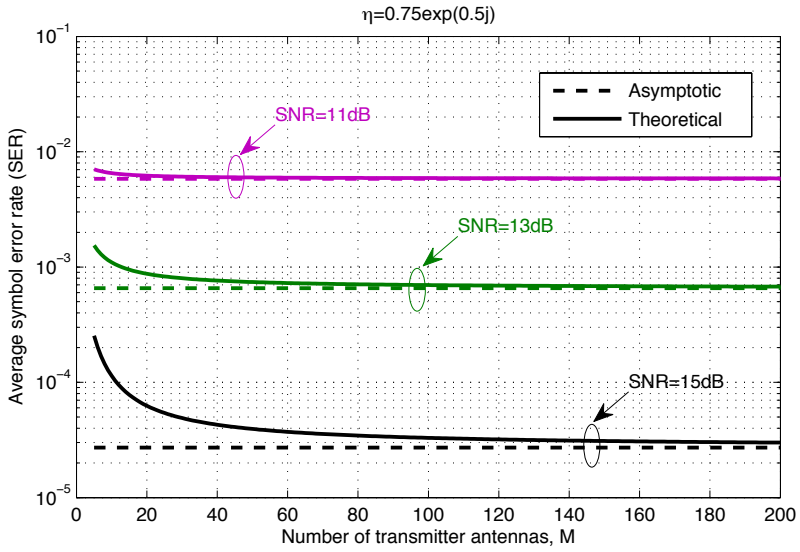


Figure 5.18: Average SEP against number of transmitter antennas, M, with different SNR.

gradually coincide with each other in both figures, which verifies our limiting results. It can be also observed that increasing the SNR or decreasing the channel correlation will improve the error performance, since in both cases, the received SNR will be increased [38]. However, this will results in a lower convergence rate for the limit of Euler's number $\lim_{x\to\infty}(1+1/x)^x = e$ in Eq. (5.110) with the moderate number of antennas.

5.6 Conclusion

In this chapter we developed an efficient technique for the design of an optimal precoder that minimizes the SER of the ZF-DF receiver for the correlated MIMO system in which channel state information is fully available at the receiver, but only the zero-mean and the covariance matrix is available at the transmitter. By a careful and

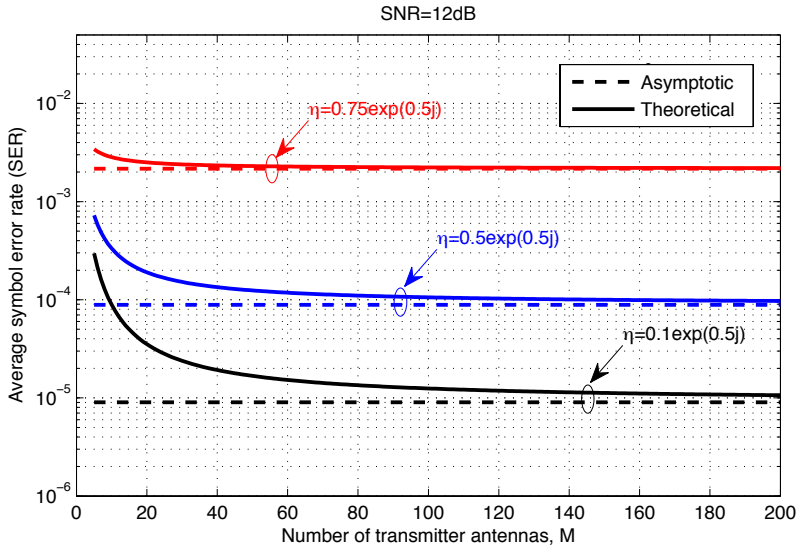


Figure 5.19: Average SEP against number of transmitter antennas, M, with different antenna correlation coefficients η.

thorough investigation of the product majorization relationship among the eigenvalues, singular-values and Cholesky values of the design matrix parameters, we derived a necessary condition for the optimal solution to satisfy and hence, characterized the structure of the optimal solution. With the aid of these results, we converted the original non-convex optimization problem into a convex geometrical programming problem which was efficiently solved using an interior point method. Our computer simulations showed that the error performance of our scheme outperformed those in [40, 46]. In addition, with channel correlation increasing and $N > M$, the SER performance of our optimally precoded system with the V-BLAST detector catches up with and surpasses that of the unprecoded system with the same detector. However, when $N = M$, the ZF-DF receiver for our optimal system has almost the same error performance as the V-BLAST receiver, but achieves substantial SNR gains over the unprecoded system with the V-BALST detector.

In addition, we have investigated the asymptotic behavior for the QRS precoded correlated MIMO communication systems with the ZF-DF detector. By fully making use of the characteristic of the large MIMO channels, the structure of the QRS transmitter as well as of the ZF-DF receiver, the Szegö's theorem [44] on large Hermitian Toeplitz matrices, and the well known limit: $\lim_{x\to\infty}(1+1/x)^x = e$, we have attained a simple expression for the SEP limit with a fast convergence rate, which, therefore, is effective and efficient for error performance evaluation for the large MIMO systems. In addition, an explanation of this approach related to the entropy power of the channel was provided. The effect of channel correlation on error performance has been also studied for an abstract one-parameter exponential correlation model, where the covariance matrix is (non-symmetric) Kac-Murdock-Szegö matrix. Simulation results have verified the effectiveness of our analysis.

Chapter 6

Quadrature Amplitude Modulation Division for Multiuser MISO Broadcast Channels

This chapter considers a discrete-time multiuser multiple-input single-output (MISO) Gaussian broadcast channel (BC), in which channel state information (CSI) is available at both the transmitter and the receivers. The flexible and explicit design of a uniquely decomposable constellation group (UDCG) is provided based on pulse amplitude modulation (PAM) and rectangular quadrature amplitude modulation (QAM) constellations. With this, a modulation division (MD) transmission scheme is developed for the MISO BC. The proposed MD scheme enables each receiver to uniquely and efficiently detect their desired signals from the superposition of mutually interfering cochannel signals in the absence of noise. In our design, the optimal transmitter beamforming problem is solved in a closed-form for two-user MISO BC using max-min fairness as a design criterion. Then, for a general case with more than two

receivers, we develop a user-grouping-based beamforming scheme, where the grouping method, beamforming vector design and power allocation problems are addressed by using weighted max-min fairness. It is shown that our proposed approach has a lower probability of error compared with the zero-forcing (ZF) method when the Hermitian angle between the two channel vectors is small in a two-user case. In addition, simulation results also reveal that for the general channel model with more than two users, our user-grouping-based scheme significantly outperforms the ZF, time division (TD), minimum mean-square error (MMSE) and signal-to-leakage-and-noise ratio (SLNR) based techniques in moderate and high SNR regimes when the number of users approaches to the number of base station (BS) antennas and it degrades into the ZF scheme when the number of users is far less than the number of BS antennas in Rayleigh fading channels.

6.1 Modulation Division for Two-User MISO BC

Our primary purpose in this section is to apply the UDCG based on the QAM constellation to the design of an optimal beamformer for a two-user BC. Toward this goal, let us specifically consider a MISO BC having two single-antenna receivers and a BS equipped with M antennas which transmits independent and identically distributed (i.i.d.) signals s_1 and s_2 simultaneously to the two receivers. The channel is assumed to be flat fading and quasi-static. Let $\mathbf{h}_1 = [h_{1,1}, h_{2,1}, \ldots, h_{M,1}]^H$ and $\mathbf{h}_2 = [h_{1,2}, h_{2,2}, \ldots, h_{M,2}]^H$ denote, respectively, the channel links between BS and user 1 and 2, which are perfectly available at the transmitter. Here, our main idea is that BS treats the two channels to be strongly interfered each other and hence, in order to serve the two receivers at the same time, the BS transmits a sum signal

$s = s_1 + s_2$, with *one* common beamforming vector $\mathbf{w} \in \mathbb{C}^{M\times 1}$ to be designed, where s_1 and s_2 are randomly chosen from an aforementioned UDCG $\mathcal{Q} = \mathcal{X}_1 \uplus \mathcal{X}_2$ such that $s_1 \in \mathcal{X}_1$ and $s_2 \in \mathcal{X}_2$. Then, the signal intended for each receiver can be decoded separately by using our fast detection method described in Algorithm 1 or 2.

6.1.1 Modulation Division for Two-User Case

The equivalent complex-baseband channel model for the received signals at the two receivers is given by

$$y_1 = \mathbf{h}_1^H \mathbf{w} s + \xi_1,$$
$$y_2 = \mathbf{h}_2^H \mathbf{w} s + \xi_2,$$

where s is the information carrying symbol for both users with $\mathbb{E}[|s|^2] = 1$ and hence the total transmitted power is $P = \mathbb{E}[|s|^2]\mathbf{w}^H\mathbf{w}{=}\mathbf{w}^H\mathbf{w}$. Also, $\xi_1, \xi_2 \sim \mathcal{CN}(0, \sigma^2)$ are additive circularly-symmetric complex Gaussian noise arising at each receiver. It is worth noting that the case where different receivers have different noise levels can be incorporated into our model by performing a scaling operation on the channel coefficients. Hence, the noises are assumed to be of equal variance. The SNRs for the sum signal s at each receiver are expressed by

$$\mathrm{SNR}_{\mathrm{md}_1} = \frac{|\mathbf{h}_1^H \mathbf{w}|^2}{\sigma^2}, \qquad \mathrm{SNR}_{\mathrm{md}_2} = \frac{|\mathbf{h}_2^H \mathbf{w}|^2}{\sigma^2}.$$

By using a max-min fairness on the received SNR, we aim to solve the following optimization problem:

Problem 2 *Find the beamforming vector* $\mathbf{w}$ *such that*

$$\max_{\mathbf{w}} \min \ \{\mathbf{w}^H\mathbf{h}_1\mathbf{h}_1^H\mathbf{w}, \mathbf{w}^H\mathbf{h}_2\mathbf{h}_2^H\mathbf{w}\}, \tag{6.114a}$$

$$\text{s.t.} \ \ \mathbf{w}^H\mathbf{w} = P. \tag{6.114b}$$

■

Without loss of generality, we assume that $\|\mathbf{h}_1\|, \|\mathbf{h}_2\| \neq 0$, since otherwise, the solution is trivial and in fact we can not achieve reliable communication to both users simultaneously in this case. Now, let

$$\mathbf{A} = \mathbf{h}_1\mathbf{h}_1^H - \mathbf{h}_2\mathbf{h}_2^H. \tag{6.115}$$

1. If $\mathbf{h}_1$ and $\mathbf{h}_2$ are linearly dependent (or equivalently, $\mathbf{h}_1 = \tau\mathbf{h}_2$ for some $\tau \in \mathbb{C}$), then $\mathbf{A} = (|\tau|^2 - 1)\mathbf{h}_2\mathbf{h}_2^H$. Hence, if $|\tau| = 1$, we have $\mathbf{A} = \mathbf{0}$. Otherwise, $\mathbf{A}$ has rank one.

2. If $\mathbf{h}_1$ and $\mathbf{h}_2$ are linearly independent (i.e., $\mathbf{h}_1 \neq \tau\mathbf{h}_2, \forall\tau \in \mathbb{C}$), then, the rank of $\mathbf{A}$ is 2. Let the eigenvalue decomposition of $\mathbf{A}$ be given by

$$\mathbf{A} = \mathbf{V}\mathbf{\Sigma}\mathbf{V}^H, \tag{6.116}$$

where $\mathbf{V}$ is a unitary matrix and $\mathbf{\Sigma} = \text{diag}(\lambda_1, -\lambda_2, 0, \ldots, 0)$ with $\lambda_1 > 0$ and $\lambda_2 > 0$. From (6.116), we can obtain

$$\mathbf{\Sigma} = \mathbf{V}^H(\mathbf{h}_1\mathbf{h}_1^H - \mathbf{h}_2\mathbf{h}_2^H)\mathbf{V} = \tilde{\mathbf{h}}_1\tilde{\mathbf{h}}_1^H - \tilde{\mathbf{h}}_2\tilde{\mathbf{h}}_2^H, \tag{6.117}$$

where $\tilde{\mathbf{h}}_1 = \mathbf{V}^H\mathbf{h}_1 = [\tilde{h}_{1,1}, \tilde{h}_{1,2}, 0, \ldots, 0]^T$ and $\tilde{\mathbf{h}}_2 = \mathbf{V}^H\mathbf{h}_2 = [\tilde{h}_{2,1}, \tilde{h}_{2,2}, 0, \ldots, 0]^T$. We also denote $\tilde{\mathbf{w}} = \mathbf{V}^H\mathbf{w} = [\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_M]^T$ and

$$\tilde{\mathbf{H}} = \mathbf{V}^H\mathbf{H} = [\tilde{\mathbf{h}}_1\ \tilde{\mathbf{h}}_2]. \tag{6.118}$$

Equation (6.117) is equivalent to

$$|\tilde{h}_{1,1}|^2 - |\tilde{h}_{2,1}|^2 = \lambda_1, \tag{6.119a}$$

$$|\tilde{h}_{1,2}|^2 - |\tilde{h}_{2,2}|^2 = -\lambda_2, \tag{6.119b}$$

$$\tilde{h}_{1,1}\tilde{h}_{1,2}^* = \tilde{h}_{2,1}\tilde{h}_{2,2}^*. \tag{6.119c}$$

The above relationships can be characterized by

$$\begin{bmatrix} \tilde{h}_{1,1} & \tilde{h}_{2,1} \\ \tilde{h}_{1,2} & \tilde{h}_{2,2} \end{bmatrix} = \begin{bmatrix} \sqrt{\lambda_1}\sec\theta e^{j\beta} & \sqrt{\lambda_1}\tan\theta e^{j(\gamma+\alpha)} \\ \sqrt{\lambda_2}\tan\theta e^{j(\beta-\alpha)} & \sqrt{\lambda_2}\sec\theta e^{j\gamma} \end{bmatrix} \tag{6.120}$$

where $\theta = \arccos\frac{\sqrt{\lambda_1}}{|\tilde{h}_{11}|}, 0 \leq \theta < \pi/2$ and $\beta = \arg(\tilde{h}_{1,1})$, $\gamma = \arg(\tilde{h}_{2,2})$ and $\alpha = \arg(\tilde{h}_{2,1}) - \arg(\tilde{h}_{2,2})$.

Now, we are ready to state one of our main results in this chapter, i.e., the optimal solution to the two-user beamforming problem in (6.114).

Theorem 10 (Optimal beamforming for two-user cases) *Let* $f(\mathbf{w}) = \min\{\mathbf{w}^H\mathbf{h}_1\mathbf{h}_1^H\mathbf{w},\ \mathbf{w}^H\mathbf{h}_2\mathbf{h}_2^H\mathbf{w}\}$. *Then, the optimal solution* $\mathbf{w}^{\text{opt}}$ *to Problem 2 is determined as follows:*

Scenario 1: $\mathbf{h}_1 = \tau \mathbf{h}_2, \tau \in \mathbb{C}$.

$$\max_{\mathbf{w}^H\mathbf{w}=P} \ f(\mathbf{w}) = \min\{P\|\mathbf{h}_1\|^2, P\|\mathbf{h}_2\|^2\},$$

where $\mathbf{w}^{\text{opt}} = \frac{\sqrt{P}\mathbf{h}_1}{\|\mathbf{h}_1\|} = \frac{\sqrt{P}\mathbf{h}_2}{\|\mathbf{h}_2\|}$.

Scenario 2: $\mathbf{h}_1 \neq \tau \mathbf{h}_2, \forall \tau \in \mathbb{C}$. *The solution is given below:*

1. $\lambda_1 \leq \lambda_2$. *Then*

   (a) *for* $0 \leq \sin\theta \leq \frac{\lambda_1}{\lambda_2}$, *we have*

   $$\max_{\|\mathbf{w}\|^2=P} \ f(\mathbf{w}) = \frac{P\lambda_1\lambda_2}{\lambda_1+\lambda_2}\frac{(1+\sin\theta)^2}{\cos^2\theta},$$

   where $\mathbf{w}^{\text{opt}} = \mathbf{V}\tilde{\mathbf{w}}^{\text{opt}}$ *with* $\tilde{\mathbf{w}}^{\text{opt}} = \left[\sqrt{\frac{P\lambda_2}{\lambda_1+\lambda_2}}e^{j\beta}, \sqrt{\frac{P\lambda_1}{\lambda_1+\lambda_2}}e^{j(\beta-\alpha)}, 0, \ldots, 0\right]^T$.

   (b) *for* $\frac{\lambda_1}{\lambda_2} < \sin\theta < 1$, *we have*

   $$\max_{\|\mathbf{w}\|^2=P} \ f(\mathbf{w}) = \frac{P(\lambda_1+\lambda_2\sin^2\theta)}{\cos^2\theta},$$

   where $\mathbf{w}^{\text{opt}} = \mathbf{V}\tilde{\mathbf{w}}^{\text{opt}}$ *with* $\tilde{\mathbf{w}}^{\text{opt}} = \left[\sqrt{\frac{P\lambda_1}{\lambda_1+\lambda_2\sin^2\theta}}e^{j\beta}, \sqrt{\frac{P\lambda_2\sin^2\theta}{\lambda_1+\lambda_2\sin^2\theta}}e^{j(\beta-\alpha)}, 0, \ldots, 0\right]^T$.

2. $\lambda_1 > \lambda_2$. *Then,*

   (a) *for* $0 \leq \sin\theta \leq \frac{\lambda_2}{\lambda_1}$, *we have*

   $$\max_{\|\mathbf{w}\|^2=P} \ f(\mathbf{w}) = \frac{P\lambda_1\lambda_2}{\lambda_1+\lambda_2}\frac{(1+\sin\theta)^2}{\cos^2\theta},$$

   where $\mathbf{w}^{\text{opt}} = \mathbf{V}\tilde{\mathbf{w}}^{\text{opt}}$ *with* $\tilde{\mathbf{w}}^{\text{opt}} = \left[\sqrt{\frac{P\lambda_2}{\lambda_1+\lambda_2}}e^{j(\gamma+\alpha)}, \sqrt{\frac{P\lambda_1}{\lambda_1+\lambda_2}}e^{j\gamma}, 0, \ldots, 0\right]^T$.

(b) and for $\frac{\lambda_2}{\lambda_1} < \sin\theta < 1$*, we have*

$$\max_{\|\mathbf{w}\|^2=P} \; f(\mathbf{w}) = \frac{P(\lambda_1 \sin^2\theta + \lambda_2)}{\cos^2\theta},$$

where $\mathbf{w}^{\text{opt}} = \mathbf{V}\tilde{\mathbf{w}}^{\text{opt}}$ *with* $\tilde{\mathbf{w}}^{\text{opt}} = \left[\sqrt{\frac{P\lambda_1}{\lambda_1+\lambda_2\csc^2\theta}}e^{j(\gamma+\alpha)}, \sqrt{\frac{P\lambda_2}{\lambda_1\sin^2\theta+\lambda_2}}e^{j\gamma}, 0, \ldots, 0\right]^T$.

■

The proof of Theorem 10 can be found in Appendix A.2. We would like to make the following comments on Theorem 10:

1. The problem dealt with in Theorem 10 is different from the physical-layer multicasting problem discussed in [161], where a group of users are interested in a common message. However, in our model, the information symbols intended for separate users are different and form a UDCG. In addition, despite the fact that the optimization problem in [161] is more general, its solution is numerical and not necessarily global. Our Theorem 10 gives the global solution in the closed form for the two-user case.

2. Here, it should be mentioned clearly that work [162] deals with the same optimization problem as ours for the optimal design of a multicast beamformer with superposition coding. Unfortunately, the optimal solution given in [162] holds only when the condition $0 \leq \sin\theta \leq \frac{\min\{\lambda_1,\lambda_2\}}{\max\{\lambda_1,\lambda_2\}}$ in Theorem 10 is satisfied. In other words, under that condition, the optimal solution is achieved in boundary $\mathbf{w}^H\mathbf{h}_1\mathbf{h}_1\mathbf{w} = \mathbf{w}^H\mathbf{h}_2\mathbf{h}_2\mathbf{w}$, which, however, is not true in general. In fact, the condition under which the solution to the max-min optimization problem is reached in the boundary is thoroughly studied in [163].

6.1.2 The Comparison between MD and ZF Method

In this section, we compare the error performance of our proposed MD beamforming with that of ZF beamforming [76]. For simplicity, we assume that the information rates of the two receivers are the same and that the channel matrix $\mathbf{H} = [\mathbf{h}_1 \ \mathbf{h}_2]$ has full column rank, whose singular values are $\sqrt{\mu_1}$ and $\sqrt{\mu_2}$ with $\mu_1, \mu_2 > 0$. Then, the received SNR for ZF beamforming with the max-min fairness criterion is determined by

$$\mathrm{SNR}_{\mathrm{zf}} = \frac{P}{\sigma^2 \sum_{i=1}^{2} [(\mathbf{H}^H \mathbf{H})^{-1}]_{i,i}} = \frac{P\mu_1\mu_2}{\sigma^2(\mu_1 + \mu_2)}. \tag{6.121}$$

On the other hand, for the MD method, by Theorem 10, the minimum received SNR between the two users for the sum signal is given by

$$\mathrm{SNR}_{\mathrm{md}} = \frac{\max_{\|\mathbf{w}\|^2 = P} \ f(\mathbf{w})}{\sigma^2}. \tag{6.122}$$

Jointly considering (6.118) and (6.120), we can obtain

$$\mu_1 + \mu_2 = \mathrm{tr}(\tilde{\mathbf{H}}^H \tilde{\mathbf{H}}) = (\lambda_1 + \lambda_2) \frac{1 + \sin^2\theta}{\cos^2\theta}, \tag{6.123a}$$

$$\mu_1\mu_2 = \det(\tilde{\mathbf{H}}^H \tilde{\mathbf{H}}) = \lambda_1\lambda_2. \tag{6.123b}$$

Hence, (6.121) can be further represented in terms of λ_1 and λ_2 as

$$\mathrm{SNR}_{\mathrm{zf}} = \frac{P\lambda_1\lambda_2}{\sigma^2(\lambda_1 + \lambda_2)} \frac{\cos^2\theta}{1 + \sin^2\theta}. \tag{6.124}$$

For discussion convenience, we define $\kappa = \lambda_1/\lambda_2$ and the SNR gain as $\eta(\kappa, \theta) =$

$10\log_{10}\frac{\mathrm{SNR}_{\mathrm{md}}}{\mathrm{SNR}_{\mathrm{zf}}}$. Then, by Theorem 10, the SNR gain as a function of κ and θ is given by

Corollary 4 (SNR Gain in terms of κ, θ) *The following statements are true:*

1. *If* $0 < \kappa \leq 1$ *and* $0 \leq \sin\theta \leq \kappa$, *then* $\eta(\kappa,\theta) = 10\log_{10}\frac{1+\sin^2\theta}{(1-\sin\theta)^2}$;
2. *If* $0 < \kappa \leq 1$ *and* $\kappa < \sin\theta < 1$, *then* $\eta(\kappa,\theta) = 10\log_{10}\frac{(1+\sin^2\theta)(\kappa+\sin^2\theta)(1+1/\kappa)}{\cos^4\theta}$;
3. *If* $1 < \kappa$, $0 \leq \sin\theta \leq 1/\kappa$, *then* $\eta(\kappa,\theta) = 10\log_{10}\frac{1+\sin^2\theta}{(1-\sin\theta)^2}$;
4. *If* $1 < \kappa$ *and* $1/\kappa < \sin\theta < 1$, *then* $\eta(\kappa,\theta) = 10\log_{10}\frac{(1+\sin^2\theta)(1/\kappa+\sin^2\theta)(1+\kappa)}{\cos^4\theta}$.

■

By Corollary 4, the SNR gain can be evaluated once $\mathbf{H}$ is obtained. To further appreciate the physical meaning of the SNR gain, we have the following lemma.

Lemma 5 *Given channel* $\mathbf{H} = [\mathbf{h}_1\ \mathbf{h}_2]$, *let* $\sqrt{\mu_1}$ *and* $\sqrt{\mu_2}$ *denote its two singular values,* $\|\mathbf{h}_1\|^2 = a$ *and* $\|\mathbf{h}_2\|^2 = b$, *and* $|\mathbf{h}_1^H\mathbf{h}_2| = c$. *Also we let* λ_1 *and* λ_2 *be defined in* (6.116). *Then, we have* $\mu_1 = \frac{a+b+\sqrt{(a-b)^2+4c^2}}{2}$, $\mu_2 = \frac{a+b-\sqrt{(a-b)^2+4c^2}}{2}$ *and* $\lambda_1 = \frac{a-b+\sqrt{(a+b)^2-4c^2}}{2}$, $\lambda_2 = \frac{-a+b+\sqrt{(a+b)^2-4c^2}}{2}$. ■

The proof of Lemma 5 can be found in Appendix-A.3. By Lemma 5, we can immediately have the following corollary:

Corollary 5 *Let* θ *be defined in* (6.120), *and* a, b *and* c *be defined in Lemma 5. Then, we have* $\sin\theta = \frac{2c}{a+b+\sqrt{(a+b)^2-4c^2}}$.

■

To gain more physical meaning of the SNR gain, we now define $\rho = a/b$ and $\cos\varphi = \frac{|\mathbf{h}_1^H\mathbf{h}_2|}{\|\mathbf{h}_1\|\|\mathbf{h}_2\|}$, where $\varphi \in [0, \pi/2]$ is called the Hermitian angle [164] between

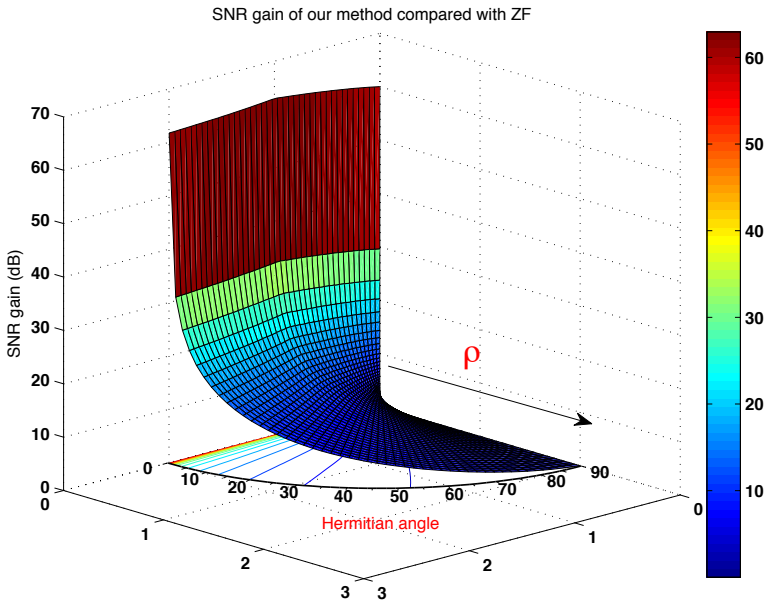


Figure 6.20: SNR Gain in terms of ρ, φ in dB

two channel vectors $\mathbf{h}_1$ and $\mathbf{h}_2$. The SNR gain as a function of ρ and φ is defined by $\nu(\rho, \varphi) = 10 \log_{10} \frac{\mathrm{SNR}_{\mathrm{md}}}{\mathrm{SNR}_{\mathrm{zf}}}$. Inserting θ in Corollary 5 into Corollary 4 and using Lemma 5 and Corollary 5, we can have the following corollary, whose proof is omitted.

Corollary 6 (SNR Gain in terms of ρ, φ) *The following statements are true.*

1. *If* $0 < \rho \leq 1$ *and* $0 \leq \cos\varphi \leq \sqrt{\rho}$ *(i.e.,* $0 < \kappa \leq 1$, $0 \leq \sin\theta \leq \kappa$*), then* $\nu(\rho, \varphi) = 10 \log_{10} \frac{1+\rho}{1+\rho-2\sqrt{\rho}\cos\varphi}$;

2. *If* $0 < \rho \leq 1$ *and* $\sqrt{\rho} < \cos\varphi \leq 1$ *(i.e.,* $0 < \kappa \leq 1$, $\kappa < \sin\theta < 1$*), then* $\nu(\rho, \varphi) = 10 \log_{10} \frac{1+\rho}{1-\cos^2\varphi}$;

3. *If* $1 < \rho$ *and* $0 \leq \cos\varphi \leq 1/\sqrt{\rho}$ *(i.e.,* $1 < \kappa$, $0 \leq \sin\theta \leq 1/\kappa$*), then* $\nu(\rho, \varphi) =$

$10\log_{10}\frac{1+\rho}{1+\rho-2\sqrt{\rho}\cos\varphi}$;

4. *If* $1 < \rho$ *and* $1/\sqrt{\rho} < \cos\varphi \leq 1$ *(i.e.,* $1 < \kappa$, $1/\kappa < \sin\theta < 1$*),* $\nu(\rho,\varphi) = 10\log_{10}\frac{1+1/\rho}{1-\cos^2\varphi}$.

■

From Corrollary 6, it is not hard to obtain $\nu(\rho,\varphi) \geq 0$ for all $\rho > 0, 0 \leq \varphi \leq \pi/2$. Hence, The SNR gain of our proposed MD beamforming is at least as good as that of ZF beamforming. To see it more clearly, the SNR gain in terms of ρ and φ are plotted in Fig. 6.20 for $0 < \rho < 2$ and $0 < \theta < \frac{\pi}{2}$. It can be observed that for given ρ, the SNR gain is determined by the Hermitian angle φ between two channel vectors. When φ approaches zero, i.e., $\mathbf{h}_1$ and $\mathbf{h}_2$ are approximately aligned with each other, the SNR gain is extremely large. For more clarity, the SNR gains for some specific cases are shown in Table 6.2 .

φ (rad)	$\frac{\pi}{180}$	$\frac{5\pi}{180}$	$\frac{15\pi}{180}$	$\frac{30\pi}{180}$	$\frac{45\pi}{180}$	$\frac{90\pi}{180}$
$\rho = 1/16$	35.43	21.46	12.00	6.28	3.27	0
$\rho = 1/8$	35.67	21.71	12.25	6.53	3.52	0
$\rho = 1/4$	36.13	22.16	12.71	6.99	3.98	0
$\rho = 1/2$	36.92	22.96	13.50	7.78	4.77	0
$\rho = 1$	38.17	24.20	14.68	8.73	5.33	0

Table 6.2: SNR gain in term of ρ and φ in dB

Corollary 6 is very convenient for the SNR gain evaluation for the two-user case, since $\|\mathbf{h}_1\|^2$, $\|\mathbf{h}_2\|^2$, and $|\mathbf{h}_1^H\mathbf{h}_2|$ are very easy to compute. As an example, we show how the SNR gain can be evaluated for line-of-sight (LoS) channels.

Example 6 *The channel coefficients for a LoS channel [127] with two users are given*

by

$$\mathbf{h}_1 = \frac{\sqrt{a}e^{j\psi_1}}{\sqrt{M}}[1 \; e^{-j2\pi\Delta\Omega_1} \; e^{-j2\pi 2\Delta\Omega_1} \cdots \; e^{-j2\pi(M-1)\Delta\Omega_1}]^T,$$
$$\mathbf{h}_2 = \frac{\sqrt{b}e^{j\psi_2}}{\sqrt{M}}[1 \; e^{-j2\pi\Delta\Omega_2} \; e^{-j2\pi 2\Delta\Omega_2} \cdots \; e^{-j2\pi(M-1)\Delta\Omega_2}]^T,$$

where a, b *are the channel gain,* Ω_1, Ω_2 *are called the directional cosine with respect to the transmitting antenna array and* Δ *is the normalized transmitting antenna distance, normalized to the unit wavelength of carrier. Then, we have* $\rho_{\text{LoS}} = \frac{a}{b}$ *and the Hermitian angle between two channel vectors* $\varphi, \varphi \in (0, \pi/2)$ *is determined by* $\cos\varphi_{\text{LoS}} = \frac{|\mathbf{h}_1^H \mathbf{h}_2|}{\|\mathbf{h}_1\|\|\mathbf{h}_2\|} = \frac{1}{M}\left|\frac{\sin(\pi M\Delta(\Omega_1-\Omega_2))}{\sin(\pi\Delta(\Omega_1-\Omega_2))}\right|$. *By Corollary 6, the SNR gain can be computed against the directional cosine* Ω_1, Ω_2 *and the normalized antenna length* Δ. ■

6.2 Grouped Modulation Division Transmission for Multiuser MISO BC

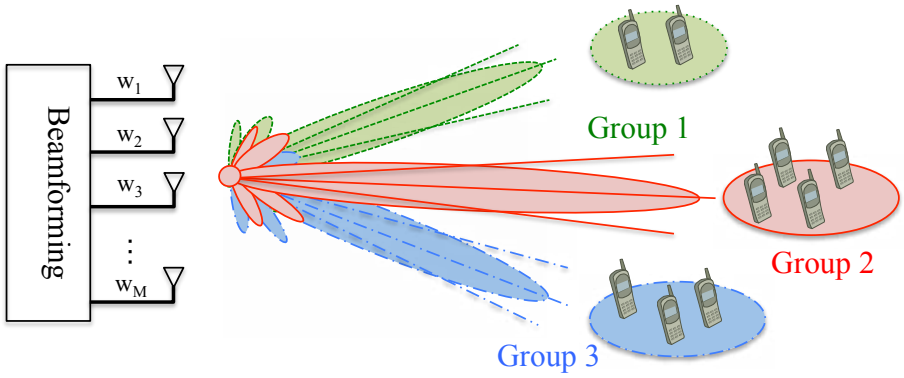


Figure 6.21: Illustration of Precoded MISO BC Model

In this section, a novel grouped modulation division transmission method is proposed for the multiuser MISO BC. The grouping algorithm is developed for cases where each group has at most two users. Then, the optimal beamforming vector and power allocation are all given in a closed-form.

6.2.1 System Model

We consider a communication system with a BS equipped with a set of M transmitting antennas communicating with N single antenna users $\mathcal{U} = \{U_1, U_2, \ldots, U_N\}$ simultaneously in the downlink as illustrated in Fig. 6.21. The channel links from BS to all the receivers can be stacked together into a matrix $\mathbf{H} \in \mathbb{C}^{M\times N}$, the k-th column of which is denoted by $\mathbf{h}_k = [h_{1,k}, h_{2,k}, \cdots, h_{M,k}]^H$, representing the channel link from BS to the k-th receiver. The N different receiver nodes can be further divided into $G \leq N$ groups, with k-th group containing N_k users, say, $U_{k_1}, U_{k_2}, \cdots, U_{k_{N_k}}$, such that $N = \sum_{k=1}^{G} N_k$. For clarity, all the users are relabelled to represent the grouping results. If we let $\mathcal{S}$ denote a set consisting of all the users, then, it can be partitioned into $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \ldots \cup \mathcal{S}_G$, $\mathcal{S}_k \cap \mathcal{S}_\ell = \varnothing, \forall k \neq \ell$, where $\mathcal{S}_k = \{U_{k_1}, U_{k_2}, \ldots, U_{k_{N_k}}\}$. Correspondingly, the channel vector from BS to U_{k_ℓ} is now denoted by $\mathbf{h}_{k_\ell}$ for $k \in \{1, 2, \cdots, G\}$ and $\ell \in \{1, 2, \cdots, N_k\}$ and in turn, the channel matrix between BS and all the users in $\mathcal{S}_k$ is represented by

$$\mathbf{H}_k = [\mathbf{h}_{k_1}, \mathbf{h}_{k_2}, \ldots, \mathbf{h}_{k_{N_k}}]. \tag{6.126}$$

Meanwhile, the matrix containing channel links from BS to all the other users in $\mathcal{S} \setminus \mathcal{S}_k$ is represented by

$$\bar{\mathbf{H}}_k = [\mathbf{H}_1, \ldots, \mathbf{H}_{k-1}, \mathbf{H}_{k+1}, \ldots, \mathbf{H}_G]. \tag{6.127}$$

The grouping strategy of dividing the original user set $\mathcal{U}$ into G mutually disjoint subsets $\mathcal{S}_k$ of $\mathcal{S}$ will be discussed later. For now, let us suppose that the grouping method has been given. Then, the communication process is carried out in the following two steps.

Firstly, we assume that all the users in the group $\mathcal{S}_k$ use one UDCG $\mathcal{Q}_k = \uplus_{\ell=1}^{N_k} \mathcal{X}_{k_\ell}$, with each sub-constellation $\mathcal{X}_{k_\ell}$ adopted by user U_{k_ℓ}. The rate allocation of group $\mathcal{S}_k$ is based on the sum decomposition $K_k = \sum_{\ell=1}^{N_k} K_{k_\ell}, \forall k \in \{1, 2, \ldots, G\}$, where K_k is the sum rate for all the users in group $\mathcal{S}_k$ and $K_{k_\ell} = \log_2(|\mathcal{X}_{k_\ell}|)$ is the rate of user U_{k_ℓ}.

Secondly, a normalized information carrying signal s_k intended for all the users in $\mathcal{S}_k$ is generated by using the UDCG $\mathcal{Q}_k = \uplus_{\ell=1}^{N_k} \mathcal{X}_{k_\ell}$, i.e.,

$$s_k = \frac{1}{\sqrt{\mathbb{E}[|\sum_{\ell=1}^{N_k} s_{k_\ell}|^2]}} \sum_{\ell=1}^{N_k} s_{k_\ell}, \quad \forall k \in \{1, 2, \ldots, G\},$$

where s_{k_ℓ} is assumed to be independently and uniformly drawn from the corresponding sub-constellation $\mathcal{X}_{k_\ell}$. It can be showed that the sum signal s_k is also uniformly distributed over the scaled sum-constellation $\frac{1}{\sqrt{\mathbb{E}[|\sum_{\ell=1}^{N_k} s_{k_\ell}|^2]}} \mathcal{Q}_k$ such that $\mathbb{E}[|s_k|^2] = 1, \forall k$. It is worth pointing out that due to the power normalization, the minimum Euclidean distance of the constellation points of s_k for different user group $\mathcal{S}_k$ might be different. Since the probability of error for the sum signal s_k is dominated by

the minimum Euclidean distance in high SNR regimes, it is anticipated that more transmitting power is required for s_k when the sum-constellation is large with the same target error performance.

Then, all users in the same group $\mathcal{S}_k$ adopt the same precoding vector $\mathbf{w}_k$ and the weighted signals are fed into M transmitter antennas at BS. Hence, the received signal at user U_{k_ℓ} can be expressed by

$$y_{k_\ell} = \underbrace{\mathbf{h}_{k_\ell}^H \mathbf{w}_k s_k}_{\text{intra-group signal}} + \underbrace{\mathbf{h}_{k_\ell}^H \sum_{m=1, m\neq k}^{G} \mathbf{w}_m s_m}_{\text{out of group interference}} + \underbrace{\xi_{k_\ell}}_{\text{noise}}, \tag{6.128}$$

in which $\xi_{k_\ell} \sim \mathcal{CN}(0, \sigma^2)$ is the circularly-symmetric complex Gaussian noise arising at U_{k_ℓ}. Here, the noise variance is assumed to be the same for all the users. The case with different receiver noise level can be incorporated into our model by performing a scaling operation on the channel coefficient $\mathbf{h}_{k_\ell}$.

In the receiver side, all users U_{k_ℓ} can detect the information intended for themselves from the uniquely decomposable signal s_k by using our fast detection method, i.e., Algorithms 1 and 2 while treating the out of group interference as additive noise. However, in high SNR regimes, the out of group interference is the dominant term that limites the error performance of our system. In what follows, a novel transmission scheme is proposed so that the out of group interference is completely cancelled out by using the ZF philosophy while the intra-group interference contained in s_k can be eliminated by taking advantage of the uniquely decomposable property of the sum-constellation.

6.2.2 Weighted Max-Min Fairness Grouped Transmission with ZF and Modulation Division

From (6.128), we know that the cochannel interference for $U_{k\ell}$ consists of two parts: 1) the inter-group interference (i.e., interference originated from users in $\mathcal{S} \setminus \mathcal{S}_k$) and 2) the intra-group interference (i.e., interference due to users in $\mathcal{S}_k \setminus \{U_{k\ell}\}$). In our scheme, we use a ZF method to cancel the inter-group interference, i.e.,

$$\bar{\mathbf{H}}_k^H \mathbf{w}_k = \mathbf{0}, \forall k \in \{1, 2, \ldots, G\}, \tag{6.129}$$

where $\bar{\mathbf{H}}_k$ is defined in (6.127). Now, user $U_{k\ell}$ only suffers from intra-group interference which can be also eliminated later by utilizing the uniquely decomposable property. Under the ZF constraint (6.129), the SNR for the sum signal s_k at user $U_{k\ell}$ can be expressed by $\text{SNR}_{k\ell} = \frac{|\mathbf{h}_{k\ell}^H \mathbf{w}_k|^2}{\sigma^2}, \forall k, \ell$. Let us denote the total transmitted power for all the users in $\mathcal{S}_k$ at BS as

$$P_k = \mathbb{E}[|\mathbf{w}_k s_k|^2] = \mathbf{w}_k^H \mathbf{w}_k, \quad \forall k \in \{1, 2, \ldots, G\}. \tag{6.130}$$

Therefore, we aim to solve the following weighted max-min grouped beamforming optimization problem:

Problem 3 *Find the beamforming vectors* $\mathbf{w}_k$ *such that the worst case weighted received signal power is maximized, i.e.,*

$$\max_{\mathbf{w}_k, \forall k} \min_{\forall k,\ell} \varrho_k |\mathbf{h}_{k\ell}^H \mathbf{w}_k|^2 \tag{6.131a}$$

$$\text{s.t. (6.129) and } \sum_{k=1}^{G} \mathbf{w}_k^H \mathbf{w}_k = P. \tag{6.131b}$$

■

The quantity $\varrho_k |\mathbf{h}_{k\ell}^H \mathbf{w}_k|^2$ in Problem 3 is usually called the weighted received signal power for s_k at $U_{k\ell}$. Using weighted SNR is a common method to balance the QoS among different users [165,166]. The resulting $\text{SNR}_{k\ell}$ is anticipated to increase with ϱ_k decrease. Since the error performance for each user is mainly determined by the SNR of the sum signal s_k and the minimum Euclidean distance of the sum-constellation of user groups $\mathcal{S}_k$, in this chapter we choose ϱ_k to be

$$\varrho_k = \frac{1}{\mathbb{E}[|\sum_{\ell=1}^{N_k} s_{k\ell}|^2]}, \tag{6.132}$$

which reasonably balances the minimum Euclidean distance for the sum signal s_k in different user groups. Other choices of ϱ_k are possible based on different application requirements.

In order to solve Problem 3, we first examine whether or not its feasible domain, $\mathcal{W} = \{\mathbf{w} = (\mathbf{w}_1^T, \mathbf{w}_2^T, \cdots, \mathbf{w}_G^T)^T : \bar{\mathbf{H}}_k^H \mathbf{w}_k = 0 \text{ and } \sum_{k=1}^{G} \mathbf{w}_k^H \mathbf{w}_k = P\}$ is empty, i.e., Problem 3 is feasible. This essentially checks whether constraint $\bar{\mathbf{H}}_k^H \mathbf{w}_k = \mathbf{0}, k \in \{1, 2, \ldots, G\}$, can be satisfied. Since $\bar{\mathbf{H}}_k \in \mathbb{C}^{M \times (N - N_k)}$ has a rank of $N - N_k$, where $N_k \geq 1$, the constraint can be satisfied if $M \geq N - N_k, \forall k$. This

condition is indeed satisfied, since we assume $N \leq M+1$ in this chapter. Therefore, Problem 3 is always feasible. On the other hand, we observe an important fact on the feasible domain. For any fixed P_k, $0 \leq P_k \leq P$ for $k = 1, 2, \cdots, G$, if we let $\mathcal{W}(P_1, P_2, \cdots, P_G) = \{(\mathbf{w}_1^T, \mathbf{w}_2^T, \cdots, \mathbf{w}_G^T)^T : \bar{\mathbf{H}}_k \mathbf{w}_k = \mathbf{0} \,\text{and}\, \mathbf{w}_k^H \mathbf{w}_k = P_k,\ k = 1,\ 2,\ \cdots,\ G\}$. Since $\mathcal{W}$ can be decomposed into a union of all such $\mathcal{W}(P_1, P_2, \cdots, P_G)$, i.e., $\mathcal{W} = \bigcup_{\sum_{k=1} P_k = P} \mathcal{W}(P_1, P_2, \cdots, P_G)$. Therefore, the original optimization Problem 3 can be equivalently split into the following two kinds of sub-optimization problems:

Sub-problem 2.1: For any fixed P_k, $0 < P_k < P$, find the beamforming vectors $\mathbf{w}_k, \forall k \in \{1, 2, \ldots, G\}$ such that

$$\zeta(P_k) = \max_{\mathbf{w}_k} \min_{\forall \ell} \ |\mathbf{h}_{k\ell}^H \mathbf{w}_k|^2 \tag{6.133a}$$

$$\text{s.t. } \bar{\mathbf{H}}_k^H \mathbf{w}_k = \mathbf{0} \text{ and } \mathbf{w}_k^H \mathbf{w}_k = P_k, \tag{6.133b}$$

■

Sub-problem 2.2: Once sub-problem 2.1 has been solved, find an optimal power allocation strategy for all user groups $\mathcal{S}$ such that

$$\max_{P_k, \forall k} \ \min_{\forall k} \ \varrho_k \zeta(P_k) \qquad \text{s.t. } \sum_{k=1}^{G} P_k = P. \tag{6.134}$$

■

In general, the optimization problem (6.133) for arbitrary $N_k \geq 3, \forall k$ is hard to solve. However, since the power required for using a large sum-constellation $\mathcal{Q}_k = \uplus_{\ell=1}^{N_k} \mathcal{X}_{k_\ell}$ with certain error target is huge if N_k is too large, in this chapter we primarily restrict ourself in the case with $N_k \leq 2$. In this case, we assume that $N \leq M+1$.

Let us consider (6.133) first, where P_k is temporarily regarded as a fixed number. For group $\mathcal{S}_k$ with $N_k = 1$, by the Cauchy-Swarz inequality we have $\mathbf{w}_k = \sqrt{P_k}\frac{\mathbf{h}_{k_1}}{\|\mathbf{h}_{k_1}\|}$. For user group $\mathcal{S}_k$ with $N_k = 2$, the sub-optimization problem 2.1 can be reformulated as

$$\begin{aligned}\zeta(P_k) = \max_{\|\mathbf{w}_k\|^2=P_k} \min\ & \{\mathbf{w}_k^H \mathbf{h}_{k_1}\mathbf{h}_{k_1}^H \mathbf{w}_k, \mathbf{w}_k^H \mathbf{h}_{k_2}\mathbf{h}_{k_2}^H \mathbf{w}_k\} \\ \text{s.t. } & \bar{\mathbf{H}}_k^H \mathbf{w}_k = \mathbf{0}, \quad \forall k \in \{1, 2, \dots G\}. \end{aligned} \tag{6.135}$$

Let us consider the constraint of (6.135) first. For $N_k = 2$, we have $\bar{\mathbf{H}}_k \in \mathbb{C}^{M\times(N-2)}$, which is a tall matrix of full column rank, since $N \leq M+1$. This constraint essentially requires that $\mathbf{w}_k$ lies in the orthogonal complement subspace of span($\bar{\mathbf{H}}_k$). Since $\bar{\mathbf{H}}_k$ has full column rank, the orthogonal complement projector for span($\bar{\mathbf{H}}_k$) is determined by $\mathbf{P}_k = \mathbf{I} - \bar{\mathbf{H}}_k(\bar{\mathbf{H}}_k^H\bar{\mathbf{H}}_k)^{-1}\bar{\mathbf{H}}_k^H \in \mathbb{C}^{M\times M}$, where $\bar{\mathbf{H}}_k^H\mathbf{P}_k = \mathbf{0}$. We know that the rank of $\mathbf{P}_k$ is $(M - N + 2)$. Now we want to find an orthonormal basis for $\mathbf{w}_k$. To do that, let the QR-decomposition of $\mathbf{P}_k$ be $\mathbf{P}_k = \mathbf{Q}_k\mathbf{R}_k$, where $\mathbf{Q}_k \in \mathbb{C}^{M\times(M-N+2)}$ is a column-wise unitary matrix. If we let $\mathbf{w}_k = \mathbf{Q}_k\check{\mathbf{w}}_k$, then, problem (6.135) is equivalent to

$$\zeta(P_k) = \max_{\|\check{\mathbf{w}}_k\|^2=P_k} \min\ \{\check{\mathbf{w}}_k^H \check{\mathbf{h}}_{k_1}\check{\mathbf{h}}_{k_1}^H \check{\mathbf{w}}_k, \check{\mathbf{w}}_k^H \check{\mathbf{h}}_{k_2}\check{\mathbf{h}}_{k_2}^H \check{\mathbf{w}}_k\},$$

where $\check{\mathbf{h}}_{k_\ell} = \mathbf{Q}_k^H\mathbf{h}_{k_\ell}, \forall k, \ell$. The above optimization problem can be solved by using Theorem 10, with the optimal value, i.e., $\zeta(P_k)$, being linear in terms of P_k, i.e.,

$$\zeta(P_k) = P_k\varsigma_k, \tag{6.136}$$

in which ς_k is determined by channel coefficients and is independent of P_k. Our next goal is to solve sub-problem 2.2 in this case. Substituting (6.136) into (6.134) yields

$$\max_{P_k,\forall k} \min_{\forall k} P_k \varrho_k \varsigma_k \qquad \text{s.t.} \sum_{k=1}^{G} P_k = P. \tag{6.137}$$

Its optimal value is attained when all $P_k \varrho_k \varsigma_k$ for $k = 1, 2, \cdots, G$ are equal to each other, hence, leading to

$$P_k^{\text{opt}} = \frac{P \prod_{\ell \neq k}^{G} \varrho_\ell \varsigma_\ell}{\sum_{m=1}^{G} \prod_{n \neq m}^{G} \varrho_n \varsigma_n}, \quad \forall k \in \{1, 2, \ldots, G\}.$$

Thus far, we have solved the problem (6.131) for given grouping method $\mathcal{S}$ with $N_k \leq 2, \forall k \in \{1, 2, \ldots, G\}$.

6.2.3 User Grouping for $N_k \leq 2, \forall k \in \{1, 2, \ldots, G\}$

As we have mentioned before, the performance of our proposed transmission method is closely related to the user grouping strategy. In this subsection, we consider the user grouping method for cases with $N_k \leq 2, \forall k \in \{1, 2, \ldots, G\}$. In these cases, we require that $G \leq N \leq 2G$ and as a consequence, there are $N - G$ groups with each having 2 users and $2G - N$ groups with each having one user. For example, if $G = N$, each group has only one user. If $G = N/2$, each group has exactly two users. Since $\mathcal{S}$ and $\mathcal{S}_k$ are all unordered sets, for the given number of groups G, we have $\frac{\prod_{k=0}^{N-G-1} \binom{N-2k}{2}}{(N-G)!}$ different grouping methods if $G \leq N - 1$ and only one method if $G = N$. Since for $N_k \leq 2$, $\lceil N/2 \rceil \leq G \leq N$, we have in total $1 + \sum_{m=\lceil N/2 \rceil}^{N-1} \frac{\prod_{k=0}^{N-m-1} \binom{N-2k}{2}}{(N-m)!}$ different grouping ways. For small N, the optimal grouping method can be found by brute-force search. However, it would be prohibitively complicated for large N, which makes our

design hard to implement in practice. Therefore, we now propose a suboptimal user-grouping method to make trade-off between performance and complexity by setting a threshold γ_T, which is a predefined level to balance the error performance among different groups.

Example 7 *Consider a two-user MISO BC with ZF beamforming, where each user employs a square K-ary QAM constellation and hence, the sum-rate of this network is $2\log_2 K$. In contrast, for the modulation division method, the sum-constellation is set to be a K^2-ary QAM constellation such that the sum-rate is $\log_2 K^2 = 2\log_2 K$, which is the same as ZF method. Assume that the average power of the transmitted symbol x_k for the ZF method and that of s_k for the sum-constellation are all normalized to 1. Then the minimum Euclidean distance of the constellation points of x_k is $d_{\text{zf}}(K) = \sqrt{\frac{6}{K-1}}$ and that of s_k is $d_{\text{md}}(K^2) = \sqrt{\frac{6}{K^2-1}}$. As a consequence we can set $\gamma_T = 10\log_{10}\frac{d^2_{\text{zf}}(K)}{d^2_{\text{md}}(K)} = 10\log_{10}(K+1)$ to compensate the SNR loss due to using a larger constellation. For example, every user is using a 4-QAM, we would expect $\gamma_T = 6.99dB$, $\gamma_T = 12.30dB$ for 16-QAM, $\gamma_T = 18.13dB$ for 64-QAM and $\gamma_T = 24.10dB$ for 256-QAM.* ■

Algorithm 3 (Grouping method) *The grouping method for $N_k \leq 2, \forall k \in \{1, 2, \ldots, G\}$ is given as follows for N receivers such that $N = \sum_{k=1}^{G} N_k$. There are $\binom{N}{2} = \frac{N(N-1)}{2}$ possible grouping methods for one group with exactly two users.*

1. *Enumeration: Find all the possible groups with two users and generate the grouping index.*

2. *Grouping gain calculation: For all the grouping indexes, calculate the coding gain. For example, suppose that users m and n ($m < n$) are grouped*

together with channel matrix $\breve{\mathbf{H}}_{m,n} = [\mathbf{h}_m, \mathbf{h}_n]$ *and another matrix containing all the channel links of users in* $\mathcal{U} \setminus \{U_m, U_n\}$*, which is denoted by* $\breve{\mathbf{H}}_{m,n} = [\mathbf{h}_1, \ldots, \mathbf{h}_{m-1}, \mathbf{h}_{m+1}, \ldots, \mathbf{h}_{n-1}, \mathbf{h}_{n+1}, \ldots, \mathbf{h}_N]$*. Let* $\breve{\mathbf{P}}_{m,n} = \mathbf{I} - \breve{\mathbf{H}}_{m,n}(\breve{\mathbf{H}}_{m,n}^H \breve{\mathbf{H}}_{m,n})^{-1} \breve{\mathbf{H}}_{m,n}^H$ *and its QR-decomposition be* $\breve{\mathbf{P}}_{m,n} = \breve{\mathbf{Q}}_{m,n} \breve{\mathbf{R}}_{m,n}$*. Compute* $\rho_{m,n} = \frac{\|\breve{\mathbf{h}}_m\|^2}{\|\breve{\mathbf{h}}_n\|^2}$ *and* $\cos\varphi_{m,n} = \frac{|\breve{\mathbf{h}}_m^H \breve{\mathbf{h}}_n|}{\|\breve{\mathbf{h}}_m\| \|\breve{\mathbf{h}}_n\|}$*, where* $\breve{\mathbf{h}}_m = \breve{\mathbf{Q}}_{m,n}^H \mathbf{h}_m$ *and* $\breve{\mathbf{h}}_n = \breve{\mathbf{Q}}_{m,n}^H \mathbf{h}_n$*. Then, by Corollary 6, calculate the grouping gain* $\nu_{m,n} = \nu(\rho_{m,n}, \cos\varphi_{m,n})$ *as a function of* $\rho_{m,n}$, $\cos\varphi_{m,n}$ *if users* m *and* n *are grouped together.*

3. *Sorting: Now the* $\frac{N(N-1)}{2}$ *grouping gains* $\nu_{m,n}$ *are sorted in descending order, forming a vector* $[\nu_{m_1,n_1}, \nu_{m_2,n_2}, \ldots, \nu_{m_{\frac{N(N-1)}{2}},n_{\frac{N(N-1)}{2}}}]^T$*.*

4. *Grouping: If* $\nu_{m_1,n_1} > \gamma_T$*, then, users* m_1 *and* n_1 *are grouped together. Otherwise, go to the next step and no users are grouped together* [1]*. Then, consider* ν_{m_2,n_2}*. Again, if* $\nu_{m_2,n_2} \leq \gamma_T$*, go to the next step and terminate the grouping procedure. Otherwise, if* $\nu_{m_2,n_2} > \gamma_T$ *and either* m_2 *or* n_2 *has not been grouped yet, then,* m_2 *and* n_2 *are grouped together. Repeat this process until all the* $\nu_{m,n}$ *have already been considered. The remaining users is left ungrouped.*

5. *Stop and output the grouping index.* ■

In our model, the grouping operation is carried out at the BS and then, the grouping indexes are informed to all the receivers. In fact, each receiver $\mathcal{U}_{k_\ell}$ only needs to know the grouping index k_ℓ to obtain the sum-constellation used by the group and the corresponding sub-constellation of itself.

[1] Then our method essentially degrade into the ZF method.

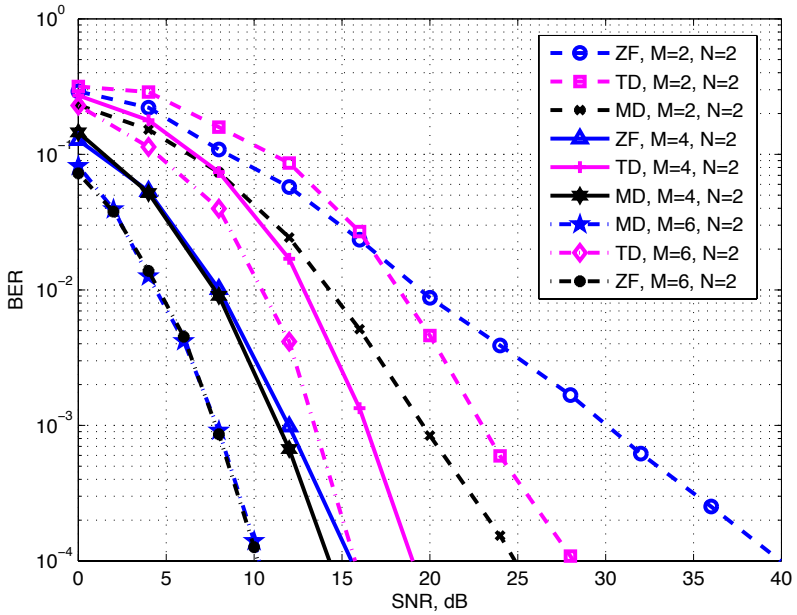


Figure 6.22: BER performance against SNR with $M = 2, 4, 6$, $N = 2$ for i.i.d. Rayleigh channel, i.e., no transmitter correlation ($\rho = 0$); Each user uses a 4-QAM.

6.3 Computer Simulations and Discussions

In this section, computer simulations are carried out to verify our theoretical analysis and to assess the effectiveness of our proposed modulation division transmission method in a multiuser MISO BC. Throughout our simulations, we assume that the channel links from the BS are potentially correlated, but are uncorrelated between different users, i.e.,

$$\mathbb{E}[\mathbf{h}_k \mathbf{h}_\ell^H] = \begin{cases} \mathbf{0} & \forall k \neq \ell, \\ \mathbf{\Sigma} & \text{otherwise} \end{cases} \tag{6.138}$$

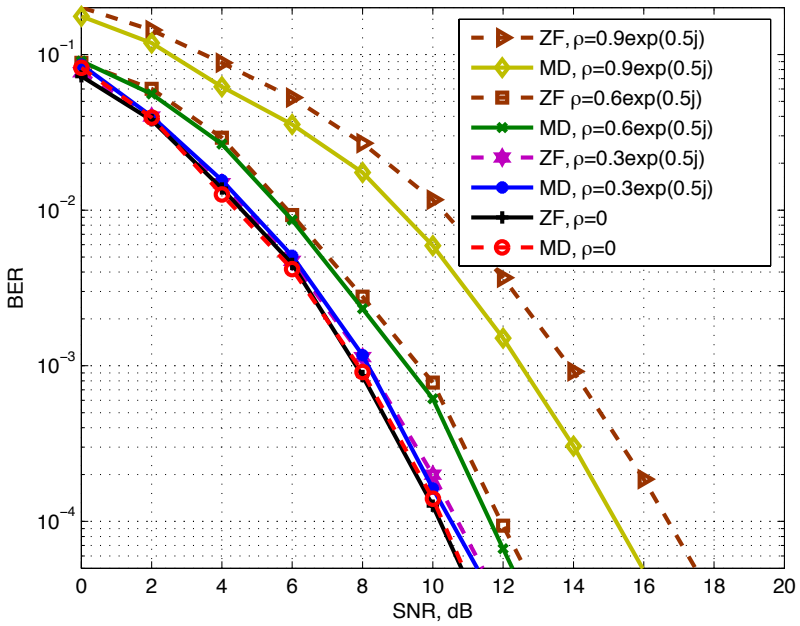


Figure 6.23: Average BER performance against SNR with $M = 6$, $N = 2$ with different ρ. Each user uses a 4-QAM.

In addition, to help with our simulation, we assume that the channel covariance matrix $\mathbf{\Sigma}$ is the commonly-used non-symmetric Kac-Murdock-Szegö (KMS) matrix [38], the (m, n)-th entry of which is denoted by $\sigma(m-n)$, i.e.,

$$\sigma(m-n) = [\mathbf{\Sigma}]_{mn} = \begin{cases} \rho^{n-m} & m \leq n \\ [\mathbf{\Sigma}]^*_{nm} & m > n, \end{cases} \tag{6.139}$$

where $0 \leq |\rho| < 1$ indicates the degree of correlation. In particular, if $\rho = 0$, then $\mathbf{\Sigma} = \mathbf{I}$, i.e., all the entries of $\mathbf{H}$ are i.i.d. Gaussian. Under all these assumptions, we perform five kinds of simulations to test our proposed MD transmission scheme in terms of uncoded BER.

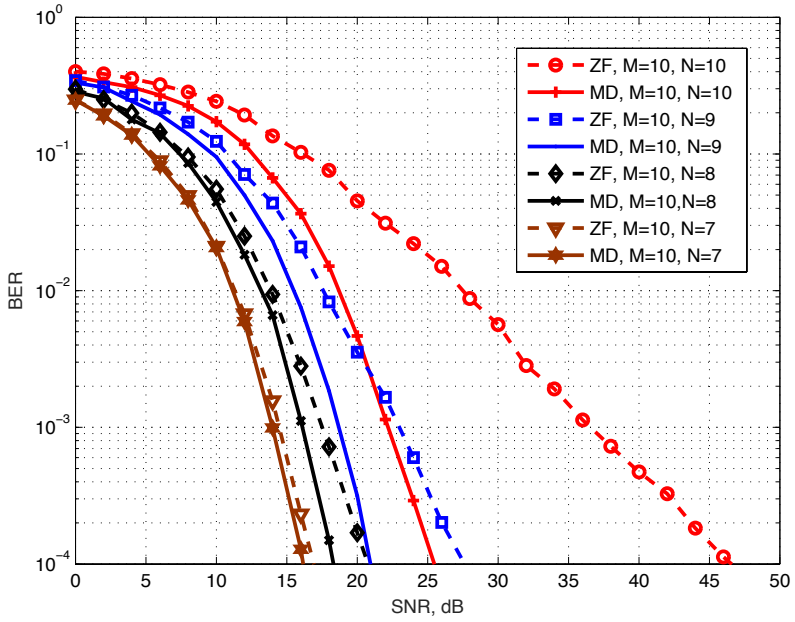


Figure 6.24: Average BER among all users against SNR, $M = 10$, $N = 7, 8, 9, 10$ with $\rho = 0$.

The first kind of simulations is to test our MD method for the two user case. Its error performance comparison with the ZF beamforming method and the time division (TD) method is plotted in Fig. 6.22, where our grouping method for two-user case is examined, with $\rho = 0$ and $M = 2, 4, 6$ transmitting antennas, $N = 2$ receivers. For the ZF method, each user uses 4-QAM and for the MD and TD methods, the sum-constellation is 16-QAM. It can be observed that the BER performance of the proposed MD method is always better than those of the ZF beamforming and the TD methods. Specifically, the SNR gain at BER 10^{-4} is approximately 15dB for $M = 2$ and $N = 2$. However, as the number of transmitter antennas is increased to $M = 4$, the BER performance of MD is still better than that of ZF, but the gap between the two methods decreases. Particularly when $M = 6$, the performance of our proposed

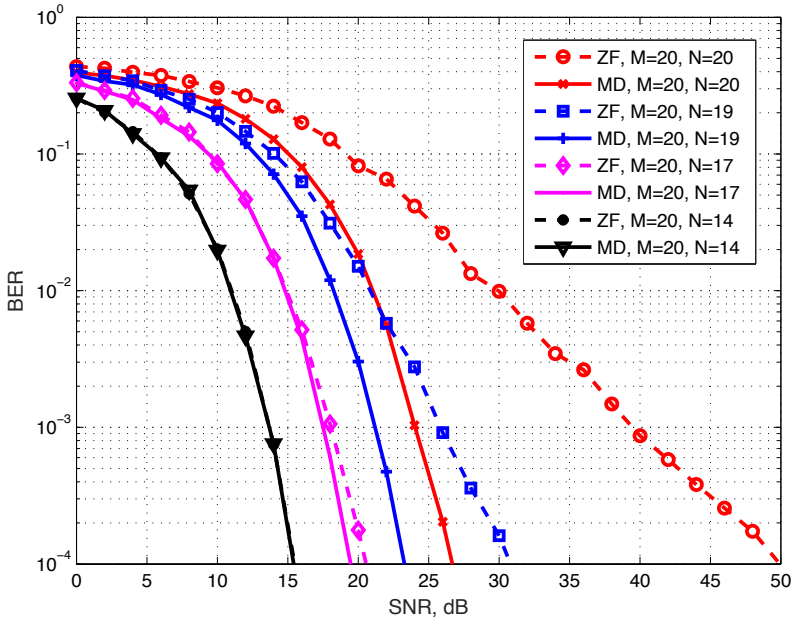


Figure 6.25: Average BER among all users against SNR, $M = 20$, $N = 14, 17, 19, 20$ with $\rho = 0$.

method is almost the same as that of ZF.

The second kind of simulations is to examine that how the correlations among the transmitter antennas affect the error performance of our MD method in the two user situation. The simulation results are shown in Fig. 6.23. It can be seen that the error performance gap between the MD and the ZF method becomes large with $|\rho|$ increasing. In this case with mild correlation, e.g., $\rho = 0.3\exp(0.5j)$, the performance gain of our method is not observable. However, when the transmitter antennas are severely correlated, e.g., $\rho = 0.9\exp(0.5j)$, our method attains at least 1.5dB gain at BER 10^{-4} over the ZF scheme.

The third kind of simulations is to test our proposed suboptimal grouping method for the multi-user ($N \geq 3$) MISO BC, as shown in Fig. 6.24, where $M = 10, N =$

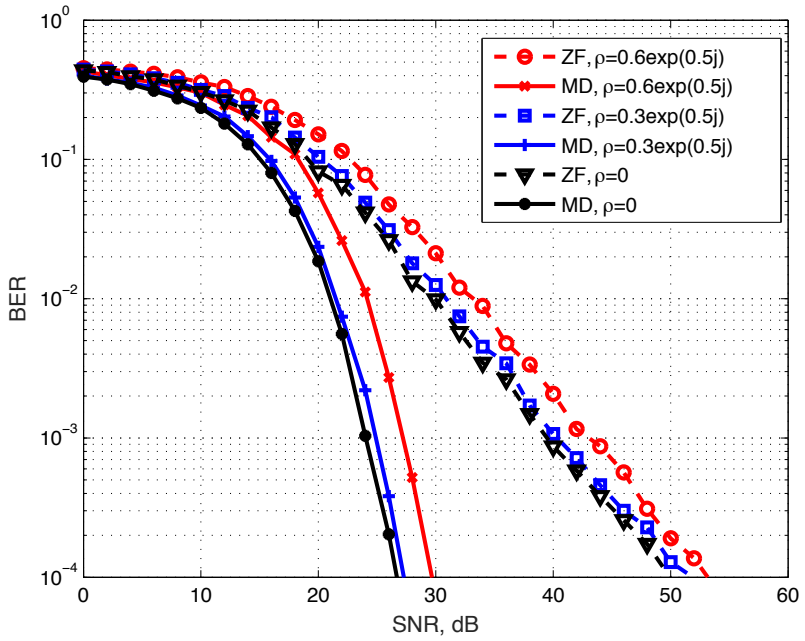


Figure 6.26: Average BER among all users against SNR, $M = 20$, $N = 20$ with $\rho = 0, 0.3\exp(0.5j)$ and $0.6\exp(0.5j)$.

$7, 8, 9, 10$ and the channels are i.i.d. Rayleigh fading. It can be observed that the MD method with user grouping always outperforms the ZF method. Specifically for the case of $M = N = 10$, the SNR gain is approximately 20dB at BER 10^{-4}. We find that the closer the number of users N is to the number of transmitters M, the larger the performance gap between the proposed MD strategy and ZF method becomes, since when M is close to N, there is a higher probability that the Hermitian angle between the channel vector of two users is small and as a consequence, the grouping gain becomes large. A similar conclusion can be drawn from the case with $M = 20$, $N = 14, 17, 19, 20$, as shown in Fig. 6.25.

Similar to the second kind of simulations, the fourth kind of simulations is to investigate how the channel correlations among the transmitter antennas affect the

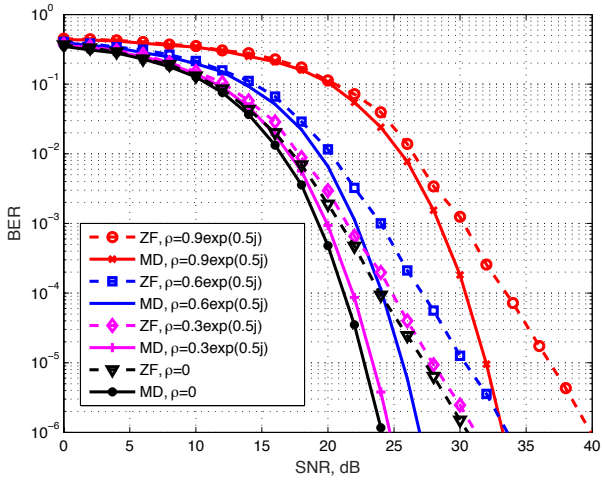


Figure 6.27: Average BER among all users against SNR, $M = 20$, $N = 18$ with $\rho = 0, 0.3\exp(0.5j)$, $0.6\exp(0.5j)$ and $0.9\exp(0.5j)$.

error performance of our proposed MD-based grouping scheme. The simulation results are shown in Figs. 6.26 and 6.27. In Fig. 6.26, we consider the case with $M = N = 20$ and different correlation coefficients. It can be expected that the BER performance of our proposed method is much better than that of ZF method. In addition, it is not surprising that the BER performance becomes worse when the channel links from the BS becomes more correlated. Similar observations are also verified for the case with $M = 20$ and $N = 18$, as shown in Fig. 6.27.

The last kind of simulation is to compare our proposed MD method with other existing precoding methods in Figs. 6.28 and 6.29. Toward this end, in Fig. 6.28, we compare the average BER of the MD approach with SLNR based scheme in [75] and MMSE method with equal power allocation as well as TD and ZF methods when $M = 4, N = 4$. For the MD method, we consider both the proposed suboptimal grouping method described in Algorithm 3 and an exhaustive search grouping scheme

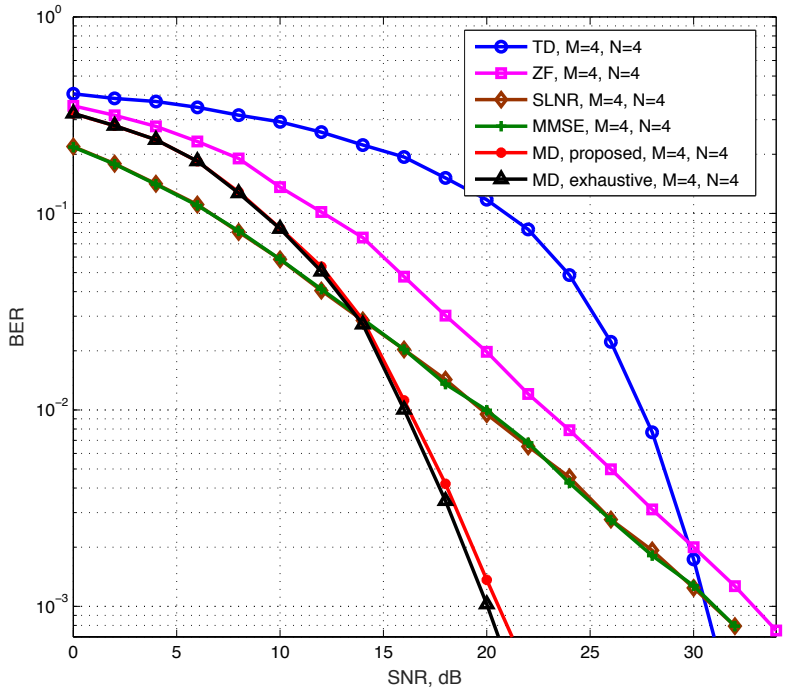


Figure 6.28: Comparison of the proposed MD method, the exhaustive MD method, ZF, TD, MMSE and SLNR methods with $M = 4$ and $N = 4$.

that enumerates all the possible grouping methods for seeking the best possible max-min weighted SNR in Problem 3. It can be observed that the TD method has the worst BER in low and moderate SNR regimes. It can also be noticed that the SLNR and MMSE methods with equal power allocation have the same BER performance as proved by [167] and both of them outperform the ZF method. In addition, we can see that in a low SNR regime, the MMSE and SLNR methods have a lower BER than the MD method. However, in a moderate and high SNR regime, the MD method outperforms all the other methods in terms of BER. Despite the fact that it has better error performance than the proposed MD approach with the suboptimal grouping

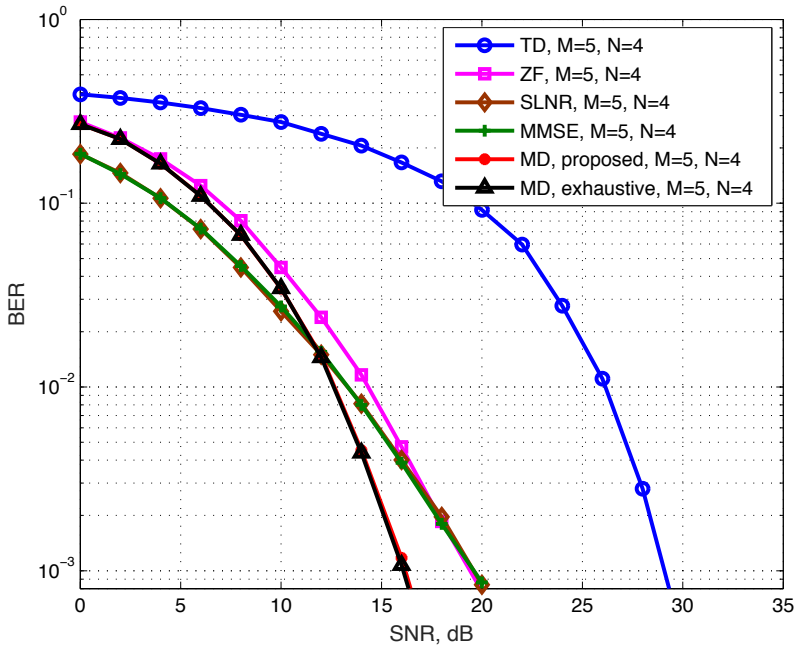


Figure 6.29: Comparison of the proposed MD method, the exhaustive MD method, ZF, TD, MMSE and SLNR methods with $M = 5$ and $N = 4$.

method, the exhaustive grouping MD method obtains the marginal BER gain and costs much higher computational complexity. Therefore, the proposed suboptimal grouping method is greatly desirable in practice. To further demonstrate the error performance comparison of our proposed MD method with other precoding schemes, the scenario with $M = 5, N = 4$ is given in Fig. 6.29, where similar conclusion can be drawn.

Here, it should be pointed out clearly that our optimal beamformer design, the optimal grouping scheme and all the resulting simulations are based on the prior assumption that the BS has the knowledge of the perfect CSI. However, in practice, it is difficult to obtain the perfect knowledge of the CSI at BS. Therefore, it would be

very necessary to analyze the error performance of our proposed algorithm with the imperfect CSI, especially for the case of multiuser massive MIMO BC. Unfortunately, this problem is too big and too important to have space for any investigation in this chapter, but will be further studied in our future research.

In addition, it also should be mentioned explicitly that in terms of computational complexity, channel state information required and other overheads, there are some drawbacks in our proposed MD scheme when compared with the ZF scheme, as listed in Table 6.3.

Table 6.3: Comparison of MD and ZF Method

Aspects	MD	ZF
Complexity	$\mathcal{O}(M^5)$	$\mathcal{O}(M^3)$
CSI	CSIT and CSIR	CSIT
Overhead	Grouping Index needed	No additional overhead

6.4 Conclusion

In this chapter, we have revealed an important property on PAM and QAM constellations for multiuser communications that any PAM constellation or QAM constellation of large size can be uniquely decomposed into the sum of a group of the scaled version of the PAM or QAM constellations of varieties of small flexible sizes. In addition, we have developed two detection algorithms, showing that one of significant advantages of such unique decomposition is that once the sum signal is detected, each individual user signal can be efficiently decoded. Then, we consider a MISO BC with two users. For the special case with two receivers, the optimal beamforming vector is given in

a closed-form based on a max-min criterion on the received SNR for the sum-signal. In addition, the SNR gain compared with ZF method is also given explicitly, which can be used to evaluate the effectiveness of our MD method in different channel coefficients. In the case with more than two receivers, a novel low-complexity grouping transmission scheme based on MD, which aims at improving the condition number of the channel matrix, is proposed, with each group having one or two users.

Finally, the simulation results have demonstrated that for the Rayleigh channel, if the number of the receivers is far less than the number of BS antennas, our method has the same error performance as the ZF method. However, when the number of users is very close to that of the BS antennas, the error performance of our proposed MD scheme is substantially better than ZF. Moreover, our computer simulations have also shown that when the transmitter antennas are correlated, our presented method is still better than ZF, even if the number of transmitter antennas is larger than that of the receivers.

In conclusion, our QAM MD transmission scheme can be considered as a feasible and concrete approach to the general NOMA method that introduces interference with proper power level superimposed on the desired signal and is possible to be applied to multiuser networks, which would enable a new promising multiple access method.

Chapter 7

QAM Division for Multiuser Uplink Massive SIMO Communications

In this chapter, we consider the design of multiuser space-time modulation (MUSTM) for an uplink massive single-input and multiple-output (SIMO) system, where the base station (BS) and all users know the large scale channel coefficients. For such system, a novel concept called uniquely factorable (UF)-MUSTM is invented. In order to assure that each transmitted signal matrix is able to be uniquely determined in a noise-free case when the number of the BS antennas goes to infinity, an important constraint of full row rank on each transmitted matrix is found to assure that the channel is able to be uniquely identified in the noise-free case as well as reliably estimated in the noisy case with the least square error criterion if the transmitted signal matrix is perfectly estimated. A new definition of full receiver diversity gain and coding gain exponent in terms of the number of the BS antennas is introduced. Then, using our

recently developed framework on uniquely decomposable constellation group (UDCG) with quadrature amplitude modulation (QAM) constellations and properly and timely assigning each sub-constellation to each user at each time slot, we develop a machinery method for systematically designing a family of invertible UF-MUSTM with flexible data rates in order to assure the reliable estimation of the transmitted signal as well as of the channel for the multiuser massive SIMO system. In addition, a simple training correlation receiver (TCR) is proposed to efficiently and effectively detect such UF-MUSTM and its pair-wise error probability (PEP) is derived, showing that our proposed UF-MUSTM enables full receiver diversity. Furthermore, the optimal closed-form power allocation and user constellation assignment are found to maximize the worst-case coding gain exponent under a peak power constraint on each user.

7.1 System Model with MUSTM for Multiuser Massive SIMO Communications

Consider an uplink massive SIMO multiple access channel (MAC), where the BS with M antennas serves N single-antenna users on the same time-frequency band ($M \gg N$). In this chapter, the channel links between the BS and all the users are assumed to be in block fading, i.e., the channel coefficients are quasi-static in the current block and change independently to other values in next consecutive block. For such a network, at the t-th time slot, a relationship between transmitted signals from all the users and a received signal in a discrete-time complex baseband-equivalent

model can be expressed by

$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{n}_t,$$

where $\mathbf{y}_t = [y_{1,t}, y_{2,t}, \ldots, y_{M,t}]^T \in \mathbb{C}^{M\times 1}$ is the $M \times 1$ received signal vector at the BS, $\mathbf{x}_t = [x_{1,t}, x_{2,t}, \ldots, x_{N,t}]^T \in \mathbb{C}^{N\times 1}$ denotes an $N \times 1$ signal vector from all the N different users and $\mathbf{n}_t \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ is an $M \times 1$ circularly-symmetric complex Gaussian (CSCG) distributed noise vector, and $\mathbf{H}$ denotes an $M \times N$ channel matrix consisting of small and large scale fading coefficients, i.e., $\mathbf{H} = \mathbf{G}\mathbf{D}^{1/2} \in \mathbb{C}^{M\times N}$, with the n-th column vector $\mathbf{h}_n = [h_{1,n}, h_{2,n}, \cdots, h_{M,n}]^T$ representing the channel links connecting User-n to all the M antennas of BS. We assume that all the entries of $\mathbf{G}$ are i.i.d. CSCG random variables with each having zero mean and unit variance (Rayleigh fading), characterizing the local scattering fading, and that $\mathbf{D} = \text{diag}\{\beta_1, \beta_2, \cdots, \beta_N\}$ is a diagonal matrix capturing the propagation loss due to near-far distances and shadowing. Throughout this chapter, we assume that $\mathbf{D}$ is available to the BS and all the N users, since it changes much slower than the first-order channel statistics and can be estimated by using training sequences such as in [168].

Now, we consider a transmission block with T time slots and thus, all T received signal vectors can be stacked together into a matrix written by

$$\mathbf{Y}_T = \mathbf{H}\mathbf{X}_T + \mathbf{N}_T, \tag{7.140}$$

where $\mathbf{Y}_T = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T]$, $\mathbf{X}_T = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T]$ and $\mathbf{N}_T = [\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_T]$.

Let us now consider what is the best way for the design of space-time signals in a noise-free case for the multiuser massive SIMO communication system.

Unique identification of transmitted signals: For the purpose on the estimation of space-time signals, let us reveal what is a necessary ability for the reliable massive communication system must have. By central limit theorem, we have $\lim_{M\to\infty} \frac{\mathbf{G}^H\mathbf{G}}{M} = \mathbf{I}$ and $\lim_{M\to\infty} \frac{\mathbf{N}_T^H\mathbf{N}_T}{M} = \sigma^2\mathbf{I}$. If we let $\mathbf{R}_M = \frac{\mathbf{Y}_T^H\mathbf{Y}_T}{M}$, we must have $\lim_{M\to\infty} \mathbf{R}_M = \mathbf{R}$, where $\mathbf{R}$ is an $N \times N$ positive semidefinite matrix satisfying $\mathbf{R} = \mathbf{X}_T^H\mathbf{D}\mathbf{X}_T$. Now, a necessary condition becomes more clear: any reliable communication for the massive SIMO system must have the ability to uniquely determine each transmitted signal matrix $\mathbf{X}_T$ once $\mathbf{R}$ has been received. In other words, any reliable communication for the massive SIMO system requires that the transmitter should carefully design such finite transmitted matrix signal set, $\mathcal{X}_T \subseteq \mathbb{C}^{N\times T}$, that if there exist any two matrices $\mathbf{X}_T, \widetilde{\mathbf{X}}_T \in \mathcal{X}_T$ satisfying $\mathbf{X}_T^H\mathbf{D}\mathbf{X}_T = \widetilde{\mathbf{X}}_T^H\mathbf{D}\widetilde{\mathbf{X}}_T$, then, we must have $\mathbf{X}_T = \widetilde{\mathbf{X}}_T$. This leads us to formally introducing the following concept:

Definition 6 (UF-MUSTM) *A MUSTM $\mathcal{S} \subseteq \mathbb{C}^{N\times T}$ is said to form a uniquely factorable (UF)-MUSTM if there exists a pair of codewords $\mathbf{S}, \tilde{\mathbf{S}} \in \mathcal{S}$ such that $\mathbf{S}^H\mathbf{S} = \tilde{\mathbf{S}}^H\tilde{\mathbf{S}}$, then, we have $\mathbf{S} = \tilde{\mathbf{S}}$.* ■

Unique identification of channels: For the purpose on the estimation of the channel, let us reveal another necessary condition for the channel to be uniquely determined if the transmitted signal matrix is perfectly estimated. In other words, can we uniquely solve the equation $\mathbf{Y}_T = \mathbf{H}\mathbf{X}_T$ for the channel matrix $\mathbf{H}$ if both $\mathbf{Y}_T$ and $\mathbf{X}_T$ are known? The answer comes up to us immediately: the unique solution for $\mathbf{H}$ requires that $\mathbf{X}_T$ must have full row rank, implying coherence time $T_\mathrm{c} \geq T \geq N$. Under this condition and even in a noisy case, we can still use the least square error criterion to reliably obtain the estimate of the channel matrix as $\hat{\mathbf{H}} = \mathbf{Y}_T\mathbf{X}_T^H(\mathbf{X}_T\mathbf{X}_T^H)^{-1}$.

Therefore, the reliable estimation of the transmitted signal as well as of the channel

for the massive MIMO communication system require that the transmitter should design UF-MUSTM with each codeword matrix having full row rank, i.e., full row rank UF-MUSTM. The following property, which can be directly verified by Definition 6, provides us with another strong evidence to further support our argument.

Proposition 6 *Let* $T_{\rm c} \geq T \geq N$ *and* $\mathbf{X}_T = [\mathbf{X}_N, \overline{\mathbf{X}}_N]$ *satisfying*

$$\mathbf{X}_T^H \mathbf{D} \mathbf{X}_T = \mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix},$$

where $\mathbf{R}$ *is known. If* $\mathbf{X}_N$ *is invertible and can be uniquely determined from* $\mathbf{R}_{11} = \mathbf{X}_N^H \mathbf{D} \mathbf{X}_N$*, then,* $\overline{\mathbf{X}}_N$ *can be uniquely determined from* $\mathbf{R}_{12} = \mathbf{X}_N^H \mathbf{D} \overline{\mathbf{X}}_N$*,* ■

Proposition 6 reduces the design of UF-MUSTM $\mathcal{X}_T$ to the design of invertible UF-MUSTM $\mathcal{X}_N$. Therefore, our primary task in the rest of this chapter is to propose a new method for the systematic design of such invertible UF-MUSTM with $T = N$.

7.2 Design of UF-MUSTM using QAM Division

In order to fulfill our task, in this section we will develop a novel machinery method for systematically designing a family of invertible UF-MUSTM by making use of our recently developed framework on UDCG with QAM constellations [169].

7.2.1 QAM Division-based Multiuser Space-Time Modulation

Our goal in this subsection is to propose a novel QAMD transmission method for the multiuser uplink massive SIMO systems by taking advantage of UDCG with the

commonly used energy efficient QAM signalling. To this end, each transmitted signal matrix $\mathbf{X}$ has the following structure:

$$\mathbf{X} = \begin{bmatrix} \sqrt{q_1} & x_{1,2} & x_{1,3} & \dots & x_{1,N} \\ \sqrt{q_2} & x_{2,2} & x_{2,3} & \dots & x_{2,N} \\ \sqrt{q_3} & x_{3,2} & x_{3,3} & \dots & x_{3,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sqrt{q_N} & x_{N,2} & x_{N,3} & \dots & x_{N,N} \end{bmatrix} = \mathbf{D}^{-1/2} \underbrace{\begin{bmatrix} 1 & s_{1,2} & s_{1,3} & \dots & s_{1,N} \\ 1 & s_{2,2} & s_{2,3} & \dots & s_{2,N} \\ 1 & s_{3,2} & s_{3,3} & \dots & s_{3,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & s_{N,2} & s_{N,3} & \dots & s_{N,N} \end{bmatrix}}_{\mathbf{S}} \mathbf{P}^{1/2}$$

$$= \mathbf{D}^{-1/2}\mathbf{S}\mathbf{P}^{1/2}, \tag{7.141}$$

where the (k,t)-th entry of $\mathbf{X}$ is transmitted from the k-th user at the t-th time slot and $\mathbf{P} = \text{diag}\{p_1, p_2, \cdots, p_N\}, p_n > 0$ for $n = 1, 2, \cdots, N$ is a diagonal power loading matrix to be optimized. We assume that the constellations used in each time slot can be permuted, i.e., $s_{k,t} \in \mathcal{A}_{\pi_t(k)}$, where $\pi_t = \big(\pi_t(1), \pi_t(2), \cdots, \pi_t(N)\big), t = 2, 3, \cdots, N$ is a permutation on N-tuple $(1, 2, \ldots, N)$ and $\mathcal{A}_k, \forall k$ constitute a UDCG with sum-QAM constellation $\mathcal{A}$, i.e., $\mathcal{A} = \uplus_{k=1}^{N} \mathcal{A}_k$. The rate allocation between the N users are based on the sum-decomposition such that $\sum_{i=1}^{N} K_i = K$, with $K_i = \log_2 |\mathcal{A}_i|$ being the rate of the user constellation $\mathcal{A}_i$. All such transmitted matrices form a set $\mathcal{X}$, which enjoys the following very interesting and important property.

Proposition 7 *The following statements are true:*

1. *If there exists a pair of* $\mathbf{X}$ *and* $\widetilde{\mathbf{X}}$ *belonging to* $\mathcal{X}$ *such that* $\mathbf{X}^H\mathbf{D}\mathbf{X} = \widetilde{\mathbf{X}}^H\mathbf{D}\widetilde{\mathbf{X}}$*, then, we have* $\mathbf{X} = \widetilde{\mathbf{X}}$*.*

2. *For any* $\mathbf{X} \in \mathcal{X}$*, if* $\pi_k(k-1) = 1$ *for* $k = 2, 3, \cdots, N$*, then,* $\mathbf{X}$ *is invertible.*

3. *For the channel model described in* (7.140)*, if the transmitted signal matrix* $\mathbf{X} \in \mathcal{X}$ *is perfectly estimated, then, the channel matrix* $\mathbf{H}$ *can be uniquely determined in the absence of noise.* ■

Proof of Statement 1: Let $\mathbf{X} = \mathbf{D}^{-1/2}\mathbf{S}\mathbf{P}^{1/2}$ and $\widetilde{\mathbf{X}} = \mathbf{D}^{-1/2}\widetilde{\mathbf{S}}\mathbf{P}^{1/2}$. Then, if $\mathbf{X}^H\mathbf{D}\mathbf{X} = \widetilde{\mathbf{X}}^H\mathbf{D}\widetilde{\mathbf{X}}$, we have $\mathbf{S}^H\mathbf{S} = \widetilde{\mathbf{S}}^H\widetilde{\mathbf{S}}$. As a consequence, $\sum_{k=1}^{N} s_{k,t} = \sum_{k=1}^{N} \tilde{s}_{k,t}$, where $s_{k,t}, \tilde{s}_{k,t} \in \mathcal{A}_{\pi_t(k)}$ for $t = 2, 3, \cdots, N$. Since all $\mathcal{A}_k$ for $k = 1, 2, \cdots, N$ form a UDCG, we attain $s_{k,t} = \tilde{s}_{k,t}, \forall k, t$, or equivalently, $\mathbf{S} = \tilde{\mathbf{S}}$. This completes the proof of Statement 1.

Proof of Statement 2: Recall that $\mathbf{X} = \mathbf{D}^{-1/2}\mathbf{S}\mathbf{P}^{1/2}$ in (7.141). Hence, we have $\det(\mathbf{X}) = \det(\mathbf{D}^{-1/2})\det(\mathbf{S})\det(\mathbf{P}^{1/2})$. Since $\det(\mathbf{D}^{-1/2}) > 0$ and $\det(\mathbf{P}^{1/2}) > 0$, proving that $\mathbf{X}$ is invertible is equivalent to proving that $\mathbf{S}$ is invertible. In what follows, we will prove that $\det(\mathbf{S}) \neq 0$. We know from the construction of all the sub-constellations that the real and imaginary part of $s_{k-1,k} \in \mathcal{A}_1, k = 2, 3, \cdots N$ are odd numbers over two and all the other information carrying signals are Gaussian integers. By using the Laplace expansion, the determinant of $\mathbf{S}$ can be represented by

$$\det(\mathbf{S}) = \alpha_0 + 2^{-1}\alpha_1 + \cdots + 2^{-(N-1)}\alpha_{N-1}, \tag{7.142}$$

where α_k are the product of $(N-1)$ Gaussian integers whose real part and imaginary part are both odd numbers or odd numbers time a power of two. As a result, we have that α_k is dividable by $(1+j)^{N-1}$ for $k = 0, 1, \cdots, N-1$. Multiplying both sides

of (7.142) by 2^{N-1}, we have

$$2^{N-1}\det(\mathbf{S}) = 2^{N-1}\alpha_0 + 2^{N-2}\alpha_1 + \cdots + 2\alpha_{N-2} + \alpha_{N-1}. \tag{7.143}$$

Since α_k is dividable by $(1+j)^{(N-1)}$ and 2 is dividable by $(1+j)$, then, $2^{N-1}\alpha_0 + 2^{N-2}\alpha_1 + \ldots + 2\alpha_{N-2}$ must be dividable by $(1+j)^N$. Now, we can finish the proof by contradiction. Suppose that $\det(\mathbf{S}) = 0$. Then, we would have $(1+j)^N | \alpha_{N-1}$. Since $\alpha_{N-1} = (-2)^{N-1}\prod_{k=2}^{N} s_{k-1,k}$, there exists some $s_{k-1,k}$ such that $2s_{k-1,k}$ is dividable by $(1+j)^2 = 2j$, which is impossible, since $2s_{k-1,k}$ are Gaussian integers whose real and imaginary parts are odd numbers. This contradicts the assumption that $\det(\mathbf{S}) = 0$. Therefore, we must have $\det(\mathbf{S}) \neq 0$. This completes the proof of Statement 2.

Proof of Statement 3: In the noise-free case, $\mathbf{Y} = \mathbf{HX}$. If $\mathbf{X}^{-1}$ exists, then we have $\hat{\mathbf{H}} = \mathbf{YX}^{-1} = \mathbf{H}$. Hence, $\mathbf{H}$ can be uniquely identified and this completes the proof of Statement 3 and hence, Proposition 2. □

7.3 Training Correlation Receiver, Error Performance Analysis and Optimal Signalling

Our primary purpose in this section is to first propose a training correlation receiver (TCR) for the QAMD MUSTM designed in Section 7.2 and then, analyze its pair-wise error probability (PEP). Particularly when each user constellation is either BPSK or QPSK, a symbol error probability formula will be derived. With all these error performance results, an new power loading scheme is finally developed that either maximizes the worst-case coding gain or minimizes the average symbol error probability subject

to average power and peak power constraints.

7.3.1 Training Correlation Receiver for UF-MUSTM

Here, by taking advantage of each coding matrix structure in the design of QAMD MUSTM, we consider a simple receiver. Since $\mathbf{x}_1 = [\sqrt{q_1}, \sqrt{q_2}, \ldots, \sqrt{q_N}]^T$, the received noisy training signal at BS in the first time slot can be written as

$$\mathbf{y}_1 = \sum_{k=1}^{N} \mathbf{h}_k \sqrt{q_k} + \mathbf{n}_1. \tag{7.144}$$

Then, the information-carrying signals from all the users are transmitted concurrently. Hence, the received signal at the t-th time slot can be represented by

$$\mathbf{y}_t = \sum_{k=1}^{N} \mathbf{h}_k x_{k,t} + \mathbf{n}_t \tag{7.145}$$

for $t = 2, 3, \cdots, N$. Now, we attempt to extract the transmitted signals for all the users from $\mathbf{R}_M$ using a so called training correlation receiver (TCR). That being said, we first calculate the correlation between the received training signal $\mathbf{y}_1$ and information carrying signal $\mathbf{y}_t$, i.e., the off-diagonal values of matrix $\mathbf{R}_M$ in the first row, $r_t = \frac{1}{M}\mathbf{y}_1^H \mathbf{y}_t,\ t = 2, 3, \cdots, N$. In order to make a reasonable decision on r_t for estimating all the transmitted signals, we need to establish the following lemma.

Lemma 6 *Let $S_t = \sum_{k=1}^{N} s_{k,t}$. If all the entries of $\mathbf{G}$ and $\mathbf{N}$ are i.i.d. CSCG random variables and independent to each other, then, we have $\mathbb{E}[r_t] = \sqrt{p_1 p_t} S_t, \mathrm{var}(r_t) = \frac{\delta_t^2}{M}$,*

where

$$\delta_t = \sqrt{(Np_1 + \sigma^2)(p_t \sum_{k=1}^{N} |s_{k,t}|^2 + \sigma^2)}. \tag{7.146}$$

Moreover, as M goes to infinity[1]*, r_t converges in distribution to a complex Gaussian random variable, i.e., $r_t \xrightarrow{d} \mathcal{CN}(\sqrt{p_1 p_t} S_t, \frac{\delta_t^2}{M})$.* ■

The proof of Lemma 6 is given in Appendix A.4. Therefore, by Lemma 6, the probability density function (PDF) of r_t conditioned on S_t can be approximated by

$$f(r_t|S_t) \doteq \frac{M}{\pi \delta_t^2} \exp\Big(- \frac{M|r_t - \sqrt{p_1 p_t} S_t|^2}{\delta_t^2} \Big). \tag{7.147}$$

Therefore, TCR is to solve the following optimization problem: $\{\hat{s}_{k,t}\}_{k=1}^{N} = \arg\min_{s_{k,t}} \frac{|r_t - \sqrt{p_1 p_t} S_t|^2}{\sigma_t^2} + \frac{\ln \sigma_t^2}{M}$ for any fixed t. Here, it should be pointed out that TCR has to be implemented by performing an exhaustive search over all the possible values of $s_{k,t}$, or, equivalently, S_t. However, when each point $s_{k,t}$ is either binary or QPSK constellation point, TCR is significantly reduced to $\hat{S}_t = \arg\min_{S_t} |r_t - \sqrt{p_1 p_t} S_t|$ for $t = 2, 3, \cdots, N$. Then, using Algorithm 2, we can quickly obtain all the estimates of user signals $s_{k,t}$.

7.3.2 Error Performance Analysis of TCR

Let us now consider the error performance analysis of TCR. From (7.146), it can be observed that the equivalent noise variance term δ_t^2 depends on the transmitted sum signal S_t and as a result, the exact symbol error probability is hard to be derived.

[1]We are interested in the massive MIMO systems, in which there are typically several hundreds or even thousands of available antennas and hence, the above assumption can be justified.

Instead, we aim to derive PEP for the received sum signal. We know that S_t is taken from a 2^K-QAM constellation and the 2^K possible values of S_t are denoted by $S_t^{(k)}, k \in \{1, 2, \ldots, 2^K\}$, respectively. Then, the PEP of sending $S_t^{(k)}$ and detecting it as $S_t^{(\ell)}$ is given by

$$P_r(S_t^{(k)} \to S_t^{(\ell)}) = P_r\bigg(\frac{|r_t - \sqrt{p_1 p_t} S_t^{(\ell)}|^2}{(\delta_t^{(\ell)})^2} + \frac{\ln(\delta_t^{(\ell)})^2}{M} < \frac{|r_t - \sqrt{p_1 p_t} S_t^{(k)}|^2}{(\delta_t^{(k)})^2} + \frac{\ln(\delta_t^{(k)})^2}{M}\bigg). \tag{7.148}$$

To further evaluate this probability, we need the following lemma.

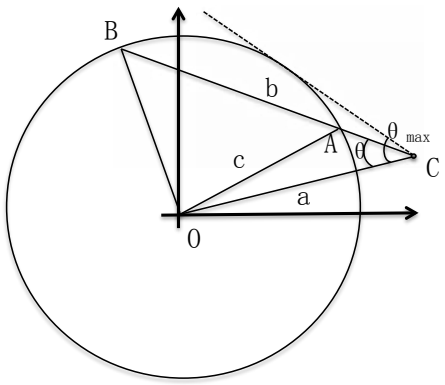


Figure 7.30: Case 1 with $a > c$.

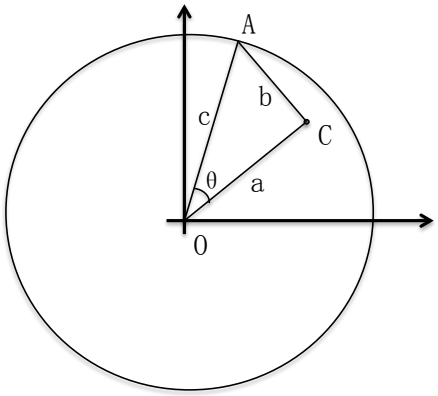


Figure 7.31: Case 2 with $a < c$.

Lemma 7 *Let a random variable $r \sim \mathcal{CN}(\mu_c, \nu^2)$, where μ_c is the coordinate of point C in the complex plane. If we let P_e denote the probability of r falling into the circle $(a > c)$ as illustrated in Fig. 7.30 or that of r falling outside the circle $(a < c)$ as depicted in Fig. 7.31, Then, we have $P_e < \exp\big(-\frac{(a-c)^2}{\nu^2}\big)$, where a and c are constant as given in Fig. 7.30 and 7.31, respectively.* ■

The proof of Lemma 7 is given in Appendix A.5. With the aid of Lemma 7, the PEP

expression (7.148) can be significantly simplified as the following theorem:

Theorem 11 *Let* $S_t^{(k)} = \sum_{m=1}^{N} s_{m,t}^{(k)}, S_t^{(\ell)} = \sum_{m=1}^{N} s_{m,t}^{(\ell)}, (\delta_t^{(k)})^2 = (Np_1 + \sigma^2)(p_t \sum_{m=1}^{N} |s_{m,t}^{(k)}|^2 + \sigma^2)$ *and* $(\delta_t^{(\ell)})^2 = (Np_1 + \sigma^2)(p_t \sum_{m=1}^{N} |s_{m,t}^{(\ell)}|^2 + \sigma^2)$. *If* $M \gg \max_{\forall S_t^{(k)} \neq S_t^{(\ell)}} \frac{2(\delta_t^{(\ell)})^2 \ln(\delta_t^{(k)}/\ln \delta_t^{(\ell)})}{p_1 p_t |\Delta S_t^{(k,\ell)}|^2}$, *then, PEP can be upper bounded by* $P_r(S_t^{(k)} \to S_t^{(\ell)}) < \exp\Big(- \frac{M p_1 p_t |\Delta S_t^{(k,\ell)}|^2}{(\delta_t^{(k)} + \delta_t^{(\ell)})^2} \Big)$, *where* $\Delta S_t^{(k,\ell)} = S_t^{(k)} - S_t^{(\ell)}$. ■

Theorem 11, whose proof is postponed into Appendix A.6, reveals that TCR achieves full receiver diversity M.[2] In addition, it also tells us that the upper bound for PEP at the t-th time slot is dominated by the following important quantity:

Definition 7 *The coding gain index at the t-th time slot is defined as*

$$G(p_1, p_t) = \min_{\forall S_t^{(k)} \neq S_t^{(\ell)}} \frac{p_1 p_t \, |\Delta S_t^{(k,\ell)}|^2}{(\delta_t^{(k)} + \delta_t^{(\ell)})^2}. \tag{7.149}$$

■

On one hand, we notice that the minimum value of $|\Delta S_t^{(k,\ell)}|^2$ in the numerator of the objective coding gain index function is one from the construction of our QAMD MUSTM. On the other hand, we note that $(\delta_t^{(k)} + \delta_t^{(\ell)})^2$ in the denominator of the objective coding gain index function is maximized when one of $S_t^{(k)}$ and $S_t^{(\ell)}$ is the corner point and the other is the nearest edge point. Specifically, we denote these two sum signal points as $\dot{S}_t = \sum_{m=1}^{N} \dot{s}_{m,t}$ and $\ddot{S}_t = \sum_{m=1}^{N} \ddot{s}_{m,t}$, respectively. Correspondingly, we let $\dot{\delta}_t^2 = (Np_1 + \sigma^2)(p_t \dot{E}_{s,t} + \sigma^2)$ and $\ddot{\delta}_t^2 = (Np_1 + \sigma^2)(p_t \ddot{E}_{s,t} + \sigma^2)$, where $\dot{E}_{s,t} = \sum_{m=1}^{N} |\dot{s}_{m,t}|^2$ and $\ddot{E}_{s,t} = \sum_{m=1}^{N} |\ddot{s}_{m,t}|^2$. Hence, the coding gain index (7.149)

[2] It is said that a detector Det achieves full receiver diversity M if there exist two positive constants C and $\rho < 1$ independent of M such that $P_{\mathrm{Det}}(\mathbf{X}_N \to \widetilde{\mathbf{X}}_N) \leq C\rho^M$ for large M.

can be significantly simplified into

$$G(p_1, p_t) = \frac{1}{\left(\sqrt{(N+\frac{\sigma^2}{p_1})\left(\dot{E}_{s,t}+\frac{\sigma^2}{p_t}\right)}+\sqrt{(N+\frac{\sigma^2}{p_1})\left(\ddot{E}_{s,t}+\frac{\sigma^2}{p_t}\right)}\right)^2}. \tag{7.150}$$

7.3.3 Power Loading under Average Power Constraint

In this subsection, we consider an optimal power loading scheme under an average power constraint where the user constellation permutation is given. As all the N users are geographically separated, an average power constraint can be imposed on each user over N time slots in each block, i.e.,

$$\frac{1}{N}\Big(q_k + \sum_{t=2}^{N} \mathbb{E}[|x_{k,t}|^2]\Big) \leq \bar{P}_k, \quad k = 1, 2, \ldots, N, \tag{7.151}$$

where $\bar{P}_k$ is a predefined average power constraint for User k. From (7.141), we know that $x_{k,t} = \frac{\sqrt{p_t}}{\sqrt{\beta_k}} s_{k,t}$. Hence, if we let $\mathbb{E}[|s_{k,t}|^2] = E_{\pi_t(k)}$, then $\mathbb{E}[|x_{k,t}|^2] = \frac{p_t}{\beta_k} E_{\pi_t(k)}$. Consequently, the average power constraint can be reformulated as

$$\frac{1}{N}\Big(\frac{p_1}{\beta_k} + \sum_{t=2}^{N} \frac{E_{\pi_t(k)}}{\beta_k}\, p_t\Big) \leq \bar{P}_k, \quad k = 1, 2, \ldots, N. \tag{7.152}$$

Inequality (7.152) can be rewritten in a more compact matrix form as

$$\underbrace{\begin{bmatrix} 1 & E_{\pi_2(1)} & E_{\pi_3(1)} & \dots & E_{\pi_N(1)} \\ 1 & E_{\pi_2(2)} & E_{\pi_3(2)} & \dots & E_{\pi_N(2)} \\ 1 & E_{\pi_2(3)} & E_{\pi_3(3)} & \dots & E_{\pi_N(3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & E_{\pi_2(N)} & E_{\pi_3(N)} & \dots & E_{\pi_N(N)} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_N \end{bmatrix}}_{\mathbf{p}} \leq \underbrace{\begin{bmatrix} N\bar{P}_1\beta_1 \\ N\bar{P}_2\beta_2 \\ N\bar{P}_3\beta_3 \\ \vdots \\ N\bar{P}_N\beta_N \end{bmatrix}}_{\mathbf{b}}. \tag{7.153}$$

That is $\mathbf{Ap} \leq \mathbf{b}$, where $\mathbf{A}$, $\mathbf{p}$ and $\mathbf{b}$ are given in (7.153). It can be observed that the feasible region of $\mathbf{p}$ is a polyhedron, which is a convex set.

Power allocation maximizing the worst-case coding gain

As we have seen from Theorem 11 that PEP is controlled by the coding gain index. Therefore, we now aim to maximize the worst-case coding gain index subject to the average power constraint on each user, i.e.,

Problem 4 *Find a power loading that maximizes the minimum coding gain index over N time slots subject to the average power constraint, i.e.,*

$$\max_{\mathbf{p}\in\mathbb{C}^{N\times 1}} \min_{\forall t} \{G(p_1, p_t)\} \tag{7.154a}$$

$$\text{s.t. } \mathbf{p} \geq \mathbf{0} \text{ and } \mathbf{Ap} \leq \mathbf{b}. \tag{7.154b}$$

■

In order to efficiently solve this problem, we first establish the following lemma.

Lemma 8 *Let* $p_u = \min\{N\bar{P}_k\beta_k\}_{k=1}^N$. *Then, for any given* p_1 *with* $0 \leq p_1 \leq p_u$, *the solution to the following optimization problem:*

$$\max_{\bar{\mathbf{p}}_1 \in \mathbb{C}^{(N-1)\times 1}} \min_{\forall t \geq 2} \{p_t\} \tag{7.155a}$$

$$\text{s.t. } \bar{\mathbf{p}}_1 \geq \mathbf{0} \text{ and } \bar{\mathbf{A}}_1\bar{\mathbf{p}}_1 \leq \bar{\mathbf{b}}_1, \tag{7.155b}$$

is given by $\bar{\mathbf{p}}_1^\star = p^\star \mathbf{1}_{N-1}$ *with* $p^\star = f(p_1) = \min\left\{\frac{N\bar{P}_k\beta_k - p_1}{\sum_{t=2}^N E_{\pi_t(k)}}\right\}_{k=1}^N$, *where* $\bar{\mathbf{p}}_1 = (p_2, p_3, \cdots, p_N)^T, \bar{\mathbf{b}}_1 = \mathbf{b} - p_1\mathbf{1}_N$ *and*

$$\bar{\mathbf{A}}_1 = \begin{bmatrix} E_{\pi_2(1)} & E_{\pi_3(1)} & \dots & E_{\pi_N(1)} \\ E_{\pi_2(2)} & E_{\pi_3(2)} & \dots & E_{\pi_N(2)} \\ E_{\pi_2(3)} & E_{\pi_3(3)} & \dots & E_{\pi_N(3)} \\ \vdots & \vdots & \ddots & \vdots \\ E_{\pi_2(N)} & E_{\pi_3(N)} & \dots & E_{\pi_N(N)} \end{bmatrix}$$

■

Proof: First, we prove that $\bar{\mathbf{p}}_1^\star$ is achievable, i.e., it is in the feasible domain. Since $0 \leq p_1 \leq p_u$, we have $\tilde{\mathbf{p}}_1^\star \geq \mathbf{0}$. In addition, substituting $\bar{\mathbf{p}}_1^\star$ into (7.155b) and using the fact that all the entries of $\bar{\mathbf{A}}_1$ are positive yield $\bar{\mathbf{A}}_1\bar{\mathbf{p}}_1^\star = p^\star\bar{\mathbf{A}}_1\mathbf{1}_{N-1} \leq \bar{\mathbf{b}}_1$. Hence, $\bar{\mathbf{p}}_1^\star$ is a feasible solution. In what follows, we will show that any point $\bar{\mathbf{p}}_1$ for $\max_{\bar{\mathbf{p}}_1} \min\{p_t\}_{t=2}^N > p^\star$ is not achievable. Suppose that there exists some $\bar{\mathbf{p}}_1^* = [p_2^*, p_3^*, \ldots, p_K^*]^T$ in the feasible domain such that $\min\{p_t^*\}_{t=2}^N > p^\star$. Then, there must exist some $\epsilon > 0$ such that $\min\{p_t^*\}_{t=2}^N = p^\star + \epsilon$. Therefore, we have $p_t \geq p^\star + \epsilon$ for $t = 2, 3, \cdots, N$, i.e., $(p^\star + \epsilon)\mathbf{I} \leq \tilde{\mathbf{p}}^*$. Combining this with the fact that all the entries of $\bar{\mathbf{A}}_1$ are positive results in $p^\star\bar{\mathbf{A}}_1\mathbf{1}_{N-1} < (p^\star + \epsilon)\bar{\mathbf{A}}_1\mathbf{1}_{N-1} \leq \bar{\mathbf{A}}_1\bar{\mathbf{p}}_1^* \leq \bar{\mathbf{b}}_1$, which

contradicts with the definition of $p^\star$. This completes the proof of Lemma 8. □

Theorem 12 *Let p_u and function $f(p_1)$ be defined in Lemma 8. Then, the optimal solution to Problem 4 is determined by $\tilde{\mathbf{p}}$, where $\tilde{p}_1 = \arg\max_{0\le p_1\le p_u} G(p_1, f(p_1))$ and $\tilde{p}_t = f(\tilde{p}_1)$ for $t = 2, 3, \cdots, N$.* ■

Proof: First, from the constraint we know $p_1 \le N\bar{P}_k\beta_k$ for $k = 1, 2, \cdots, N$. Hence. we have $p_1 \le \min\{N\bar{P}_k\beta_k\}_{k=1}^N = p_u$. On the other hand, we notice that for any given p_1 with $0 \le p_1 \le p_u$, the coding gain index function in terms of p_t is nondecreasing. Hence, $\min_{2\le t\le N} G(p_1, p_t) = G(p_1, \min_{2\le t\le N} p_t)$ and as a consequence, for any given p_1 with $0 \le p_1 \le p_u$, we have $\max_{\bar{\mathbf{p}}_1} \min_{2\le t\le N} G(p_1, p_t) = G(p_1, \max_{\bar{\mathbf{p}}_1} \min_{2\le t\le N} p_t)$. Now, by Lemma 8, we obtain $\max_{\bar{\mathbf{p}}_1} \min_{2\le t\le N} G(p_1, p_t) = G(p_1, f(p_1))$. Therefore, we attain $\max_{\mathbf{p}} \min_{2\le t\le N} G(p_1, p_t) = \max_{p_1} \max_{\bar{\mathbf{p}}_1} \min_{2\le t\le N} G(p_1, p_t) = \max_{p_1} G(p_1, f(p_1))$. This completes the proof of Theorem 12.

The Special Case where $\mathcal{A}_k$ are BPSK or 4-QAM with Different Scales

In this subsection, we will restrict our sub-constellations for each user to be BPSK or 4-QAM with different scales. In this case, $|s_{k,t}|^2 = \mathbb{E}[|s_{k,t}|^2] = E_{\pi_t(k)}$ and (7.146) becomes

$$\tilde{\delta}_t^2 = (Np_1 + \sigma^2)(p_t E_s + \sigma^2), \tag{7.156}$$

where $E_s = \sum_{k=1}^N E_k$. It can be observed that $\tilde{\delta}_t^2$ keeps constant for different transmitted signal S_t. For the considered large SIMO system, if S_t is transmitted,

$r_t \dot{\sim} \mathcal{CN}(\sqrt{p_1 p_t} S_t, \frac{\tilde{\delta}_t^2}{M})$ and the PDF conditioned on S_t can be approximated by

$$f(r_t|S_t) \dot{=} \frac{M}{\pi \tilde{\delta}_t^2} \exp\bigg(- \frac{M|r_t - \sqrt{p_1 p_t} S_t|^2}{\tilde{\delta}_t^2} \bigg). \tag{7.157}$$

As a result, the sum signal S_t can be estimated based on the approximated $f(r_t|S_t)$ above in massive MIMO systems,

$$\hat{S}_t = \arg\max_{S_t} \ln\big(f(r_t|S_t)\big) = \arg\min_{S_t} |r_t - \sqrt{p_1 p_t} S_t|^2.$$

Once $\hat{S}_t$ has been detected, $\hat{s}_{k,t}$ can be determined for each user. Due to the decision region is the same as QAM constellations, the performance measure can be based on the average symbol error rate (SER) for the sum signal. In particular, for the N users each using $\mathcal{A}_k$ that are all 4-QAM constellations with different scales, the received sum signal S_t is taken from a 4^N-square QAM constellation. The SER over the t-th time slot for the sum signal S_t can be approximated by

$$P_e(p_1, p_t) \doteq 4\Big(1 - \frac{1}{2^N}\Big) Q\bigg(\frac{\sqrt{M}}{2\sqrt{(N + \frac{\sigma^2}{p_1})(E_s + \frac{\sigma^2}{p_t})}} \bigg) - 4\Big(1 - \frac{1}{2^N}\Big)^2 Q^2\bigg(\frac{\sqrt{M}}{2\sqrt{(N + \frac{\sigma^2}{p_1})(E_s + \frac{\sigma^2}{p_t})}} \bigg).$$

The case when all the users are using BPSK is similar and hence omitted. As a consequence, the approximated average probability of error

$$P_e(\mathbf{p}) \doteq \frac{1}{N-1} \sum_{t=2}^{N} P_e(p_1, p_t), \tag{7.158}$$

where $\mathbf{p} = [p_1, p_2, \ldots, p_N]^T$.

Problem 5 (Minimize the average SER) *We aim to minimize the average SER of S_t in one frame by performing optimization on $\mathbf{p}$, i.e., we aim to solve the following*

optimization problem

$$\min_{\mathbf{p}} P_e(\mathbf{p}) \tag{7.159a}$$

$$\text{s.t. } \mathbf{p} \geq \mathbf{0} \text{ and } \mathbf{Ap} \leq \mathbf{b}. \tag{7.159b}$$

■

The above problem is a convex optimization problem, and the proof is given in Appendix A.7. As a result, the above problem can be solved efficiently by using an interior-point method [47].

7.3.4 Optimal Signalling under Peak Power Constraints

In this subsection, we consider an optimal power loading scheme that maximizes the worst-case coding gain index when each user has a respective peak power constraint [170] over each time slot (e.g., due to the limited linear region of the radio frequency power amplifier), i.e., $\max_{s_{k,t}\in\mathcal{A}_{\pi_t(k)}} |x_{k,t}|^2 \leq \hat{P}_k, k, t = 1, 2, \cdots, N$, where $\hat{P}_k, \forall k$ are predefined peak power upper bound. If we denote the maximum instantaneous power (i.e., the power of the corner point) of sub-constellation $\mathcal{A}_i$ by D_i, then, the peak power constraint is equivalent to $\frac{p_1}{\beta_k} \leq \hat{P}_k$, and $\frac{D_{\pi_t(k)}p_t}{\beta_k} \leq \hat{P}_k$ for $k, t = 1, 2, \cdots, N$. From Theorem 1 and 2, we have $D_i = \left(2^{K_i} - \frac{1}{2}\right)^2 \times 2^{\sum_{n=1}^{i-1} 2K_n}$ for PAM-UDCG and $D_i = \left(2^{K_i^{(c)}} - \frac{1}{2}\right)^2 \times 2^{\sum_{n=1}^{i-1} 2K_n^{(c)}} + \left(2^{K_i^{(s)}} - \frac{1}{2}\right)^2 \times 2^{\sum_{n=1}^{i-1} 2K_n^{(s)}}$ for QAM-UDCG for $i = 1, 2, \cdots, N$. Without loss of generality and for discussion simplicity, we assume that all the users are labelled such that $\hat{P}_1\beta_1 \leq \hat{P}_2\beta_2 \leq \cdots \leq \hat{P}_N\beta_N$.

Power allocation maximizing the worst-case coding gain index

Now, our optimal power allocation and sub-constellation assignment method under the peak power constraint is formally stated as follows:

Problem 6 *Find a power and sub-constellation assignment with* $\pi_t(t-1)=1$ *for* $t=2,3,\cdots,N$ *that maximizes the worst-case coding gain index under the peak power constraint, i.e.,* $\max_{\mathbf{p}\in\mathbb{C}^{N\times 1}} \min_{\forall t} \ \{G(p_1,p_t)\}$ s.t. $p_1 \leq \hat{P}_k\beta_k,\ p_t \leq \frac{\hat{P}_k\beta_k}{D_{\pi_t(k)}}$ *for* $k=1,2,\cdots,N$ *and* $\pi_t(t-1)=1$ *for* $t=2,3,\cdots,N$. ■

In order to obtain the solution to Problem 6, we need the following lemma.

Lemma 9 *Suppose that two positive sequences* $\{a_n\}_{n=1}^N$ *and* $\{b_n\}_{n=1}^N$ *are arranged both in a nondecreasing order. If we let* $\mathcal{U}$ *denote the set containing all the possible permutations of* $1,2,\cdots,N$*, then, the solution to the optimization problem,* $\max_{\pi\in\mathcal{U}} \min\left\{\frac{a_k}{b_{\pi(k)}}\right\}_{k=1}^N$*, is given by* $\pi^*(k)=k$ *for* $k=1,2,\cdots,N$. ■

We are now in a position to formally state our main result in this chapter.

Theorem 13 *The optimal solution to Problem 6 is given as follows:* $p_1^\star = \min\{\hat{P}_k\beta_k\}_{k=1}^N = \hat{P}_1\beta_1, p_t^\star = \min\left\{\frac{\hat{P}_k\beta_k}{D_{\pi_t^*(k)}}\right\}_{k=1}^N$ *for* $t=2,3,\cdots,N$*, and is* $\pi_t^* = (2,3,\ldots,t-1,1,t,\cdots,N)$ *for* $t=2,3,\cdots,N$. ■

Proof: For any fixed sub-constellation assignment $\pi_t = \big(\pi_t(1),\pi_t(2),\ldots,\pi_t(N)\big)$, since $G(p_1,p_t)$ is a monotonically increasing function of p_1 and p_t for $t=2,\cdots,N$, we have $p_1^\star = \min\{\hat{P}_k\beta_k\}_{k=1}^N = \hat{P}_1\beta_1$ and $p_t^\star = \min\left\{\frac{\hat{P}_k\beta_k}{D_{\pi_t(k)}}\right\}_{k=1}^N$. Now, we need to find an optimal permutation with $\pi_t(t-1)=1$ to further maximize $p_t^\star$. Again, since $G(p_1,p_t)$ is monotonically increasing with respect to p_t, such optimal sub-constellation permutation can be attained by solving such max-min problem:

$\max_{\pi_t} \min_k \left\{ \frac{\hat{P}_k \beta_k}{D_{\pi_t(k)}} \right\}$ s.t. $\pi_t(t-1) = 1$ for $t = 2, 3, \cdots, N$. As $\frac{\hat{P}_{t-1}\beta_{t-1}}{E_1}$ is fixed for $\pi_t(t-1) = 1$, the overall optimal permutation can be found by just solving $\max_{\pi_t} \min \left\{ \frac{\hat{P}_1\beta_1}{D_{\pi_t(1)}}, \ldots, \frac{\hat{P}_{t-2}\beta_{t-2}}{D_{\pi_t(t-2)}}, \frac{\hat{P}_t\beta_t}{D_{\pi_t(t)}}, \ldots, \frac{\hat{P}_N\beta_N}{D_{\pi_t(N)}} \right\}$, Notice $\hat{P}_1\beta_1 \leq \ldots \leq \hat{P}_{t-2}\beta_{t-2} \leq \hat{P}_t\beta_t \leq \ldots \leq \hat{P}_N\beta_N$. Hence, by Lemma 9, the optimal solution to the above max-min optimization problem is achieved by seeking for π_t such that $D_{\pi_t(1)} \leq \ldots \leq D_{\pi_t(t-2)} \leq D_{\pi_t(t)} \leq \ldots \leq D_{\pi_t(N)}$. From the definition of D_i, we know $D_1 \leq D_2 \leq \ldots \leq D_N$, and hence, the optimal sub-constellation assigment is $\pi_t^* = (2, 3, \ldots, t-1, 1, t, \ldots, N)$. This completes the proof. □

The Special Case where $\mathcal{A}_k$ are BPSK or 4-QAM with Different Scales

In this case, the average SER for the sum signal in the $N-1$ time slots can be expressed by

$$P_e(\mathbf{p}) = \frac{1}{N-1} \sum_{t=2}^{N} P_e(p_1, p_t). \tag{7.160}$$

Now, we aim to solve the following optimization problem:

Problem 7 (Minimize the Average SER) *We aim to minimize the average SER subject to the peak power constraint, i.e.,*

$$\min_{\mathbf{p} \geq \mathbf{0}} P_e(\mathbf{p}) \tag{7.161a}$$

$$\text{s.t. } p_1 \leq \hat{P}_k \beta_k \text{ and } p_t \leq \frac{\hat{P}_k \beta_k}{E_{\pi_t(k)}}, \quad k, t = 2, 3, \ldots, N. \tag{7.161b}$$

■

Again, we find that $P_e(\mathbf{p})$ is a monotonically decreasing function of p_1 and p_t, hence

$p_1^\star = \min_k\{\hat{P}_k\beta_k\} = \hat{P}_1\beta_1$ and $p_t^\star = \min_k\left\{\frac{\hat{P}_k\beta_k}{E_{\pi_t(k)}}\right\}, t = 2, 3, \ldots, N$.

In conclusion, for both the general case and the special case, the optimal solution to the power allocation problem takes the same form. For the peak power constraint, since the solution has the above special form, we can have the above optimal sub-constellation assignment method.

7.4 Comparison with Other Receivers

In this section, we briefly discuss Riemannian distance receiver, non-coherent maximum likelihood receiver and orthogonal pilot training receiver with zero-forcing equalization for our UF-MUST modulation scheme. Then, we compare their error performances with that of TCR by computer simulations.

7.4.1 The Minimum Riemannian Distance Detector

As we have discussed in Sections 7.2 and 7.3, each transmitted signal matrix in the UF-MUST modulation scheme can be uniquely identified in the absence of noise when $M \to \infty$. Furthermore, the channel matrix can be also uniquely identified in the noise-free case if the transmitted signal matrix is perfectly estimated.

Now, we aim to estimate the transmitted signal matrix and the channel matrix jointly in a noisy environment with a Riemannian distance based receiver when the noise statistic is unknown. In this scenario, it is known from noncoherent communication theory that a simple receiver is the least square error receiver, which is to solve the following optimization problem: $\{\hat{\mathbf{X}}, \hat{\mathbf{H}}\} = \arg\min_{\mathbf{X},\mathbf{H}} \|\mathbf{Y} - \mathbf{H}\mathbf{X}\|_F^2$. Its solution, in general, is given by $\hat{\mathbf{H}} = \mathbf{Y}\hat{\mathbf{X}}^H(\hat{\mathbf{X}}\hat{\mathbf{X}}^H)^{-1}$ if $\mathbf{X}$ has full row rank,

where $\hat{\mathbf{X}} = \arg\max \mathrm{tr}\big(\mathbf{Y}\mathbf{X}^H(\mathbf{X}\mathbf{X}^H)^{-1}\mathbf{X}\mathbf{Y}^H\big)$. It, however, fails if $\mathbf{X}$ is square, since $\mathbf{X}^H(\mathbf{X}\mathbf{X}^H)^{-1}\mathbf{X} = \mathbf{I}$. This is a motivation for us to study the Riemannian distance receiver. Note that

$$\begin{aligned}\|\mathbf{Y} - \mathbf{H}\mathbf{X}\|_F^2 &= \|\mathbf{Y} - \mathbf{G}\mathbf{T}\|_F^2 \\ &= \mathrm{tr}(\mathbf{Y}^H\mathbf{Y}) - 2\,\Re\,\mathrm{tr}(\mathbf{T}\mathbf{Y}^H\mathbf{G}) + \mathrm{tr}(\mathbf{T}^H\mathbf{G}^H\mathbf{G}\mathbf{T}),\end{aligned} \tag{7.162}$$

where $\mathbf{T} = \mathbf{S}\mathbf{P}^{1/2}$. On the other hand, we observe that when M is very large, $\frac{1}{\sqrt{M}}\mathbf{G}$ approaches to a unitary matrix, i.e., $\lim_{M\to\infty}\frac{1}{M}\mathbf{G}^H\mathbf{G} = \mathbf{I}$. Hence, for a fixed $\mathbf{T}$, we attempt to find $\hat{\mathbf{G}}$ by solving the following optimization problem, $\hat{\mathbf{G}} = \arg\max_{\mathbf{G}:\frac{1}{M}\mathbf{G}^H\mathbf{G}=\mathbf{I}} \Re\big(\mathrm{tr}(\mathbf{T}\mathbf{Y}^H\mathbf{G})\big)$. Its solution can be obtained by using the following property.

Property 4 (7.4.9, [141]) *Let $\mathbf{A} \in \mathbb{C}^{N\times M}(N \le M)$ be a given matrix of rank-N, and also, let the singular value decomposition of $\mathbf{A}$ be $\mathbf{A} = \mathbf{V}_A\mathbf{\Sigma}_A\mathbf{W}_A^H$, where $\mathbf{V}_A$ is an $N\times N$ unitary matrix, $\mathbf{W}_A$ is an $N\times M$ row-wise unitary matrix and $\mathbf{\Sigma}_A = \mathrm{diag}(\sigma_1(\mathbf{A}), \sigma_2(\mathbf{A}), \cdots, \sigma_N(\mathbf{A}))$ with $\sigma_n(\mathbf{A})$ for $n = 1, 2, \cdots, N$ be the singular values of $\mathbf{A}$. Then, the problem $\max_{\mathbf{U}^H\mathbf{U}=\mathbf{I}} \Re\big(\mathrm{tr}\{\mathbf{A}\mathbf{U}\}\big)$ has the solution $\mathbf{U} = \mathbf{W}_A\mathbf{V}_A^H$ and the maximum is $\sigma_1(\mathbf{A}) + \cdots + \sigma_n(\mathbf{A})$.* ■

Now using Property 4 yields $\hat{\mathbf{G}} = \sqrt{M}\mathbf{W}\mathbf{V}^H$, where the singular value decomposition of $\mathbf{T}\mathbf{Y}^H$ is $\mathbf{T}\mathbf{Y}^H = \mathbf{V}\mathbf{\Sigma}\mathbf{W}^H$. Therefore, minimizing (7.162) can be simplified as a so called the Remannian distance receiver into $\min_{\mathbf{T}}\mathrm{tr}(\mathbf{Y}^H\mathbf{Y}/M) - 2\mathrm{tr}\Big(\mathbf{T}\mathbf{Y}^H\mathbf{Y}\mathbf{T}^H/M\Big)^{1/2} + \mathrm{tr}(\mathbf{T}^H\mathbf{T}) = \min_{\mathbf{T}} d_{R_1}^2(\mathbf{T}^H\mathbf{T}, \mathbf{Y}^H\mathbf{Y}/M)$, where d_{R_1} is defined as the first kind of the Riemannian distances [171] and thus, we have $\hat{\mathbf{X}} = \mathbf{D}^{-1/2}\hat{\mathbf{T}}$.

7.4.2 The Non-coherent ML Detector

Recall that the system model is given by $\mathbf{Y} = \mathbf{GT} + \mathbf{N}$, The conditional PDF of the received signal at the BS for any given $\mathbf{T}$ is given by $p(\mathbf{y}|\mathbf{T}) = \frac{1}{\pi^{NM}\det(\mathbf{R}_{\mathbf{y}|\mathbf{T}})}\exp(-\mathbf{y}^H\mathbf{R}_{\mathbf{y}|\mathbf{T}}^{-1}\mathbf{y})$, where $\mathbf{R}_{\mathbf{y}|\mathbf{T}} = \mathbf{I}\otimes(\mathbf{T}^H\mathbf{T}+\sigma^2\mathbf{I})$. Note $\det(\mathbf{R}_{\mathbf{y}|\mathbf{T}}) = \big(\det(\mathbf{T}^H\mathbf{T}+\sigma^2\mathbf{I})\big)^M$ and $\mathbf{y}^H\mathbf{R}_{\mathbf{y}|\mathbf{T}}^{-1}\mathbf{y} = \mathrm{tr}\big(\mathbf{Y}(\mathbf{T}^H\mathbf{T}+\sigma^2\mathbf{I})^{-1}\mathbf{Y}^H\big)$. Therefore, the non-coherent maximum likelihood detector aims to estimate the transmitted information carrying matrix by solving the following optimization problem, $\min_{\mathbf{T}}\ \mathrm{tr}\big(\mathbf{Y}(\mathbf{T}^H\mathbf{T}+\sigma^2\mathbf{I})^{-1}\mathbf{Y}^H\big) + M\log\det(\mathbf{T}^H\mathbf{T}+\sigma^2\mathbf{I})$.

7.4.3 Orthogonal Pilot Training Receiver with Zero-Forcing Equalization

This method basically uses an orthogonal pilot signal for training. The system model is given by $\mathbf{Y} = \mathbf{HX} + \mathbf{N}$. In the training phase, the k-th user uses its peak power $\hat{P}_k$, i.e., $\mathbf{X}_{tp} = \mathrm{diag}\{\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_N\}$. The estimated channel is given by $\hat{\mathbf{H}} = \mathbf{Y}\mathbf{X}_{tp}^{-1}$. In the information communication phase, we have $\mathbf{y}_{N+1} = \mathbf{H}\mathbf{x}_{N+1} + \mathbf{n}_{N+1}$, Then a zero-forcing (ZF) receiver can be used so that $\hat{\mathbf{x}}_{N+1} = (\hat{\mathbf{H}}^H\hat{\mathbf{H}})^{-1}\hat{\mathbf{H}}^H\mathbf{y}_{N+1} = (\mathbf{X}_{tp}^{-1}\mathbf{Y}^H\mathbf{Y}\mathbf{X}_{tp}^{-1})^{-1}\mathbf{X}_{tp}^{-1}\mathbf{Y}^H\mathbf{y}_{N+1} = \mathbf{X}_{tp}(\mathbf{Y}^H\mathbf{Y})^{-1}\mathbf{Y}^H\mathbf{y}_{N+1}$.

7.5 Simulation Results and Discussions

In this section, computer simulations are performed to verify the theoretical analysis in this chapter. We first consider the channel model and then compare the error performance of the TCR, the modified Riemannian distance receiver and the

non-coherent ML receiver under both average and peak power constraints. The performance of the TCR against the conventional training based method is studied as well.

7.5.1 The Combined Path-loss and Shadowing Model

To achieve a tradeoff between accuracy and simplicity, we consider a combined path-loss and shadowing model [135] in which the power fall-off due to distance and random attenuation are both captured. We assume that the transmitted and the received power at each antenna element are denoted by P_t and P_r, respectively. Then, the pathloss as a function of transmission distance d at antenna far-field can be approximated by

$$\frac{P_r}{P_t} = \kappa\Big(\frac{d_0}{d}\Big)^{\gamma}\psi, \quad d \geq d_0,$$

where κ is a unit-less constant that depends on the antenna characteristics and free-space pathloss up to far-filed close-in reference distance d_0, γ is the path-loss exponent and ψ is the random shadowing attenuation. The detailed explanation of these parameters are given as follows.

- The value of κ is sometimes set to the free-space path gain at distance d_0 assuming omni-directional antenna:

$$\kappa = \Big(\frac{\lambda}{4\pi d_0}\Big)^2,$$

  where $\lambda = \frac{3\times10^8 \text{ m/s}}{f_c}$ is the wavelength of the carrier while f_c is the carrier frequency.

- The pathloss exponent γ depends on the propagation environment and it typically takes values in the range of $[2, 6]$ for most wireless environment. In this simulation, we set $\gamma = 3.71$.

- In this model, ψ captures the shadowing fading resulting from blockage of objects in the signal path which gives rise to random variations of the received power at given distance. Here, ψ is assumed to be random with a log-normal distribution given by

$$p(\psi) = \frac{\xi}{\sqrt{2\pi}\sigma_{\psi_{dB}}\psi} \exp\Big(- \frac{(10\log_{10}\psi - \mu_{\psi_{dB}})^2}{2\sigma^2_{\psi_{dB}}}\Big), \quad \psi > 0,$$

  where $\xi = 10/\ln 10$, $\mu_{\psi_{dB}} = \mathbb{E}[10\log_{10}\psi]$ and σ_{dB} is the standard deviation of ψ_{dB} in decibels. Assuming that $\psi_{dB} = 10\log_{10}\psi$, it is Gaussian distributed with mean $\mu_{\psi_{dB}}$ and standard deviation $\sigma_{\psi_{dB}}$ given by

$$p(\psi_{dB}) = \frac{1}{\sqrt{2\pi}\sigma_{\psi_{dB}}} \exp\left(- \frac{(\psi_{dB} - \mu_{\psi_{dB}})^2}{2\sigma^2_{\psi_{dB}}}\right).$$

  In our simulation, we assume that $\mu_{\psi_{dB}} = 0$ and $\sigma_{\psi_{dB}} = 3.16\,\text{dB}$.

From the above discussion, the dB attenuation is given by

$$10\log_{10}\frac{P_r}{P_t} = 10\log_{10}\kappa - 10\gamma\log_{10}\frac{d}{d_0} + \psi_{dB}.$$

For the receiver, the power of noise is $P_n = N_0 B_w$, where N_0 is the power spectral density of noise and B_w is the channel bandwidth. For the thermal noise, it is assumed that $N_0 = k_0 T_0 10^{F_0/10}$, where $k_0 = 1.38 \times 10^{-23}$ J/K is the Boltzman's constant, T_0 is a reference temperature and F_0 is the noise figure.

7.5.2 System Setup

Table 7.4: Simulation Parameters

Cell radius $d_{\max}$	1000 m
Reference distance d_0	100 m
Carrier frequency f_c	3 GHz
Channel bandwidth B_w	20 MHz
Pathloss exponent γ	3.71
Reference temperature / Noise figure	290 K / 6 dB
Standard deviation of shadow fading $\sigma_{\psi_{dB}}$	3.16 dB

In our simulation, we set set T_0 to 290 K ("room temperature") and noise figure F_0=6 dB. The channel bandwidth is assumed to be B_w=20 MHz and hence, the noise power is $P_n = 3.2 \times 10^{-13}$ W, or equivalently,

$$10 \log_{10} P_n = 10 \log_{10} 3.2 \times 10^{-10} w = -124.95 \text{ dBW}.$$

The distance dependent pathloss, we assume that $\gamma = 3.71$, $d_0 = 100$ m, the carrier frequency is assumed to be f_c =3 GHz. Hence, the pathloss is

$$10 \log_{10} P_L = -20 \log_{10} \frac{\lambda}{4\pi d_0} + 10\gamma \log_{10} \frac{d}{d_0} = 81.98 \text{ dB} + 37.1 \log_{10} \frac{d}{d_0}.$$

The small-scale fading is assumed to be the normalized Rayleigh fading. As a result, the end-to-end SNR is

$$10 \log_{10} \frac{P_r}{P_n} = 10 \log_{10} P_t + 42.97 - 37.1 \log_{10} \frac{d}{d_0}.$$

For example, if d=100 m, then $10 \log_{10} \frac{P_r}{P_n} = 10 \log_{10} P_t + 42.97$ dB and if d=1000 m,

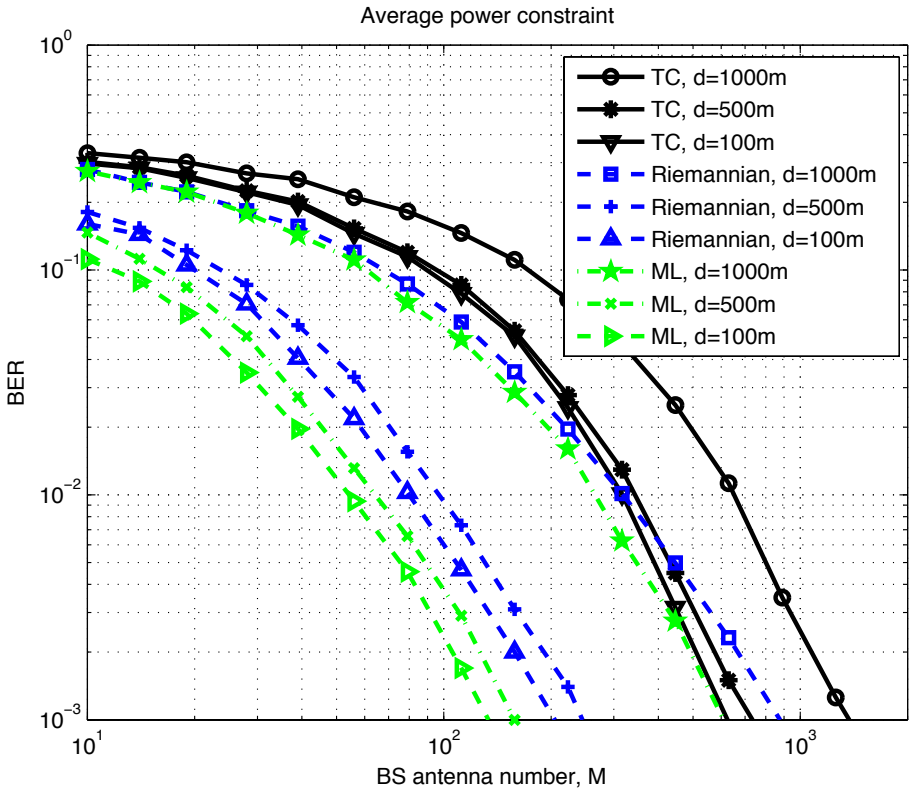


Figure 7.32: Average BER of all users versus M, for different d, $N = 3$ and 4-QAM are used by all the users with average power constraint.

we have $10\log_{10}\frac{P_r}{P_n} = 10\log_{10} P_t + 5.87$. For clarity, all the simulation parameters are summarized in Table 7.4.

7.5.3 Simulation Results

We first examine the error performance of all the different receivers under the average power constraint as illustrated in Fig. 7.32. It is assumed that the average power upper bound is $\bar{P}_k$=316 mW (25 dBm), $\forall k$. All the N users are assumed to be uniformly distributed on the circle with radius d to the cell centre. It can be observed that the

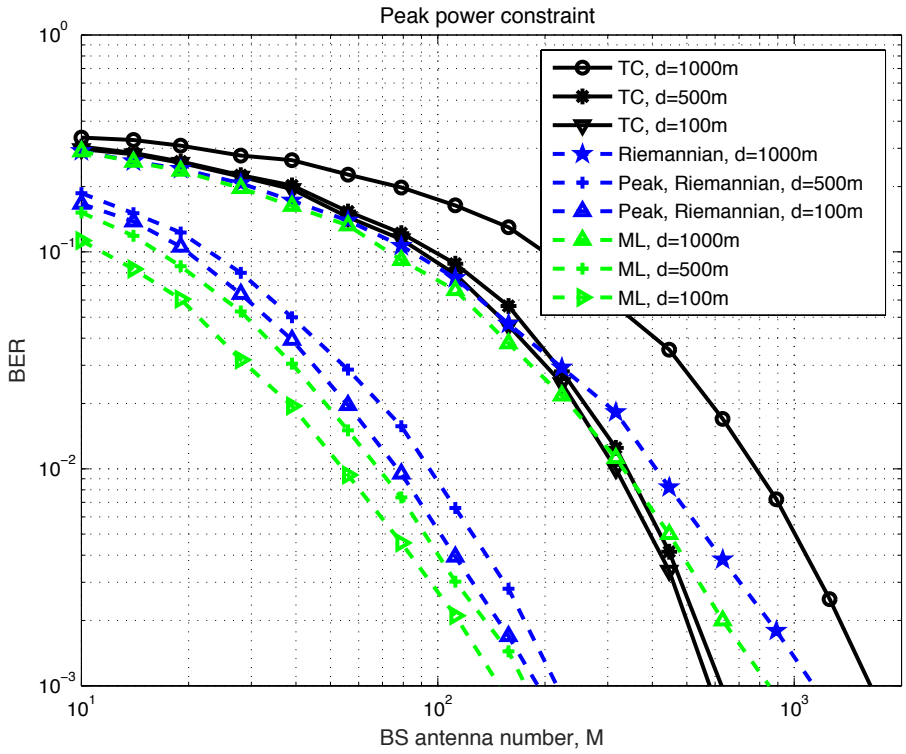


Figure 7.33: Average BER among all users against M, with different d, $N = 3$ and all users use 4-QAM with peak power constraint.

non-coherent ML receiver (denoted by ML) have the best error performance while the BER of the modified Riemannian distance receiver (labelled by Riemannian) is higher than the ML receiver but it has a better error performance than the TCR (represented by TC). For all the three different receivers, with the increased distance, the simulated BER increases as expected. It can also be observed that, to have an average BER 10^{-3}, the TCR needs roughly four times more antennas compared with the ML receiver and three times more antennas than the Riemannian distance receiver.

The error performance of all the three receivers against the number of BS antenna

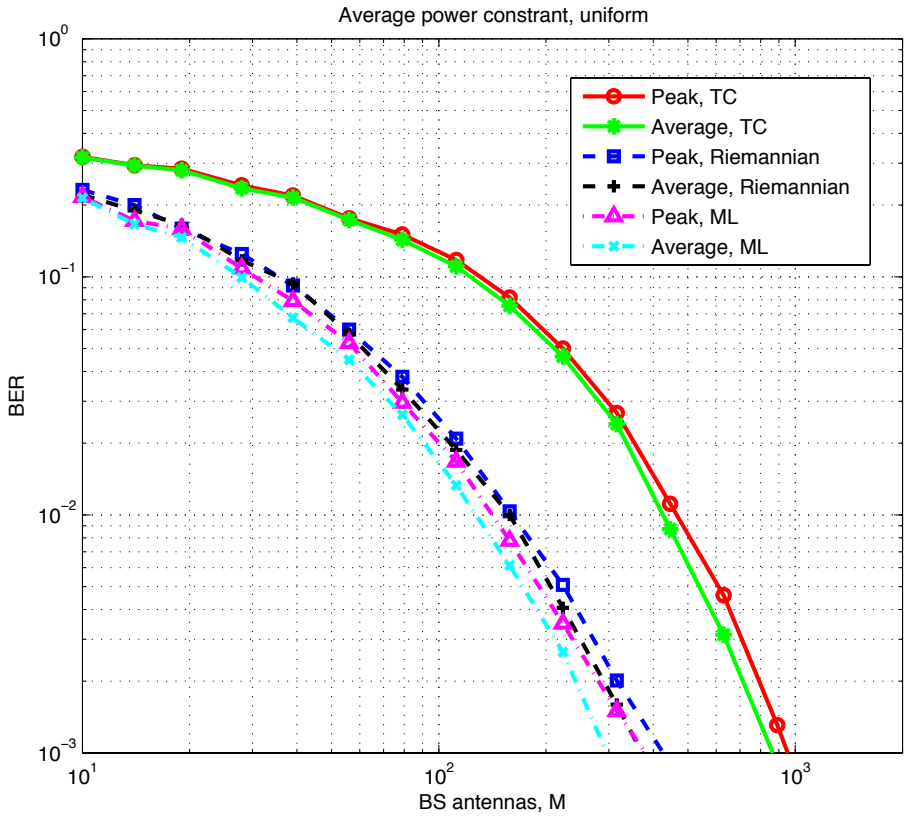


Figure 7.34: Average BER of all users against M, with different d, $N = 3$ and all users use 4-QAM with peak power constraint.

number M under the peak power constraint is plotted in Fig. 7.33. Likewise, the ML receiver has the best error performance while the average BER of the TCR is highest. From both Fig. 7.32 and 7.33, we argue that the proposed receiver works in an efficient symbol-by-symbol detection mode for the sum-signal in each time slot while the computational complexity of the ML receiver is prohibitively high when there are a lot of users. As a result, when the antenna number is large (e.g., in large MIMO system with hundreds or even thousands of available antennas), the proposed receiver is a good candidate compared with the ML and the Riemannian distance

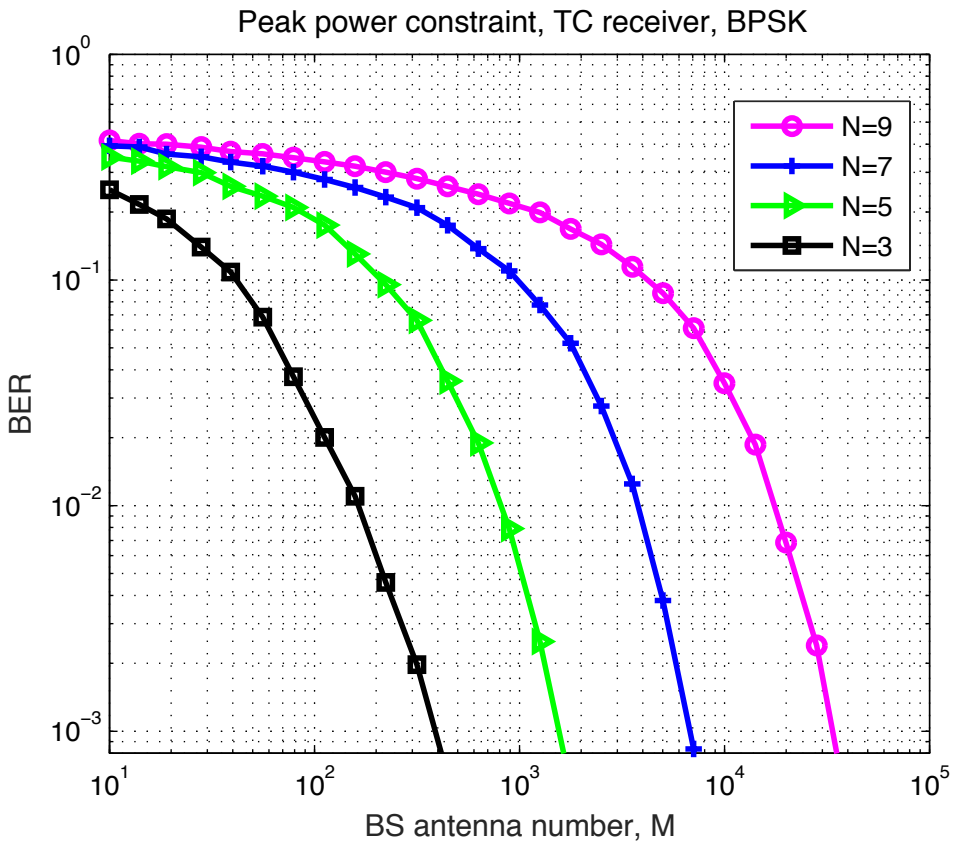


Figure 7.35: Average BER of all users against M for different N, BPSK

receiver.

Next, we study the error performance of all the three receivers under different power constraints in Fig. 7.34. In our simulation, all the users are assumed to be uniformly distributed over the cell disk with radius from 100m to 1000m and $\bar{P}_k$ and $\hat{P}_k$ are all set to 316mW. It can also be observed that, given the same total transmitted power, the average power constraint case always have a better error performance since it can allocate transmitting power more flexibly among all the users which results in larger feasible regions. On the other hand, the gain of the average power constraint

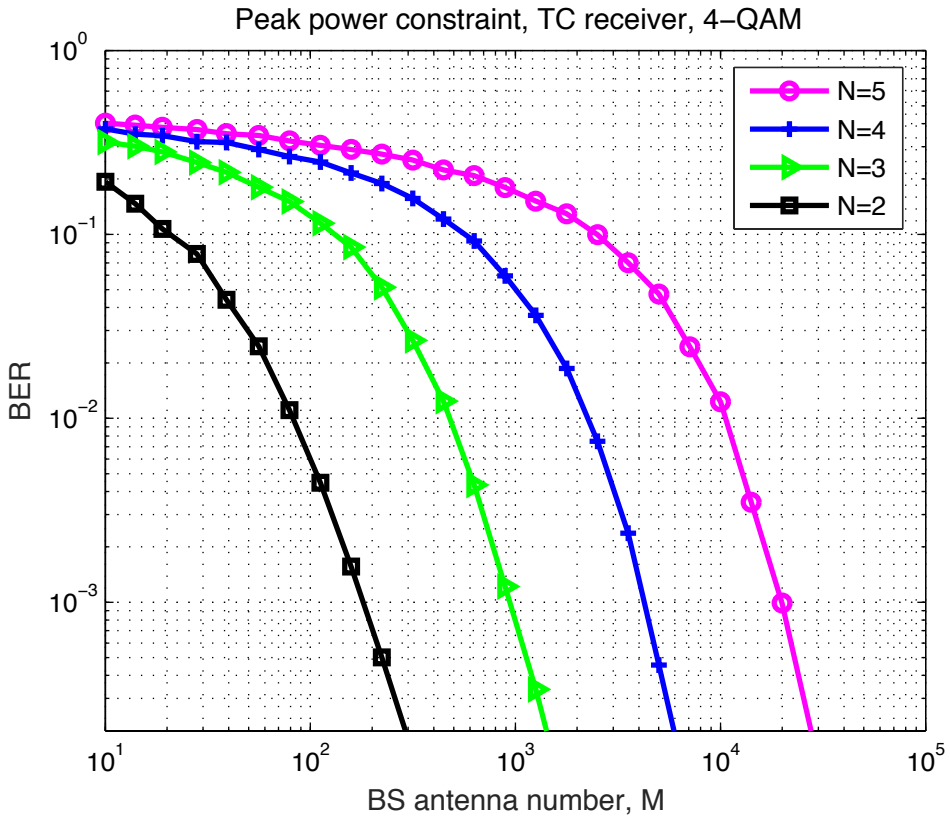


Figure 7.36: Average BER of all users against M, for different N, 4-QAM.

case against the peak power constraint case is marginal. Thus, in what follows, we will mainly concentrate on the peak power constraint case due to its simplicity.

The effectiveness of our proposed TCR with different number of users N is studied in Fig. 7.35. To guarantee that $\mathbf{X}$ is invertible and the sum-constellation is a rectangular QAM constellation, we assume that $\mathcal{X}_1$ is 4-QAM and all the other constellations are BPSK. Again, all the users are assumed to be uniformly distributed over the cell disk with radius from 100m to 1000m. It can be observed that as the number of users increase, the required number of antennas M to achieves the same error performance increases. The case where each user is using a 4-QAM constellation

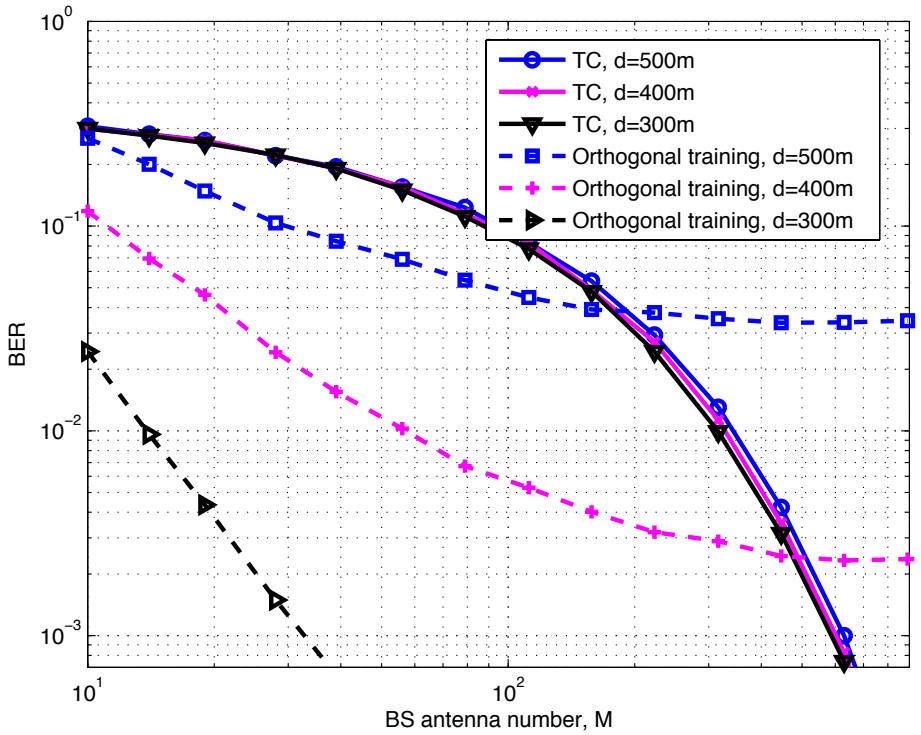


Figure 7.37: The comparison between the TC receiver and the orthogonal training method with $N = 3$ users and 4 time slot.

is given in Fig. 7.36.

Finally, we compare the performance of the TCR with the conventional training sequence based receiver in Fig. 7.37. It can be observed that, when the antenna number M is small, the training based outperforms the TCR in term of BER. However, when the antenna number is extremely large, the TCR has a better error performance compared with the training based receiver, especially at the cell edge.

Here, it should be mentioned that a related non-coherent multiuser massive SIMO system is considered by using DPSK constellation in [172] with correlation based receiver. The transmitted information of all the users is modulated on the phase offset

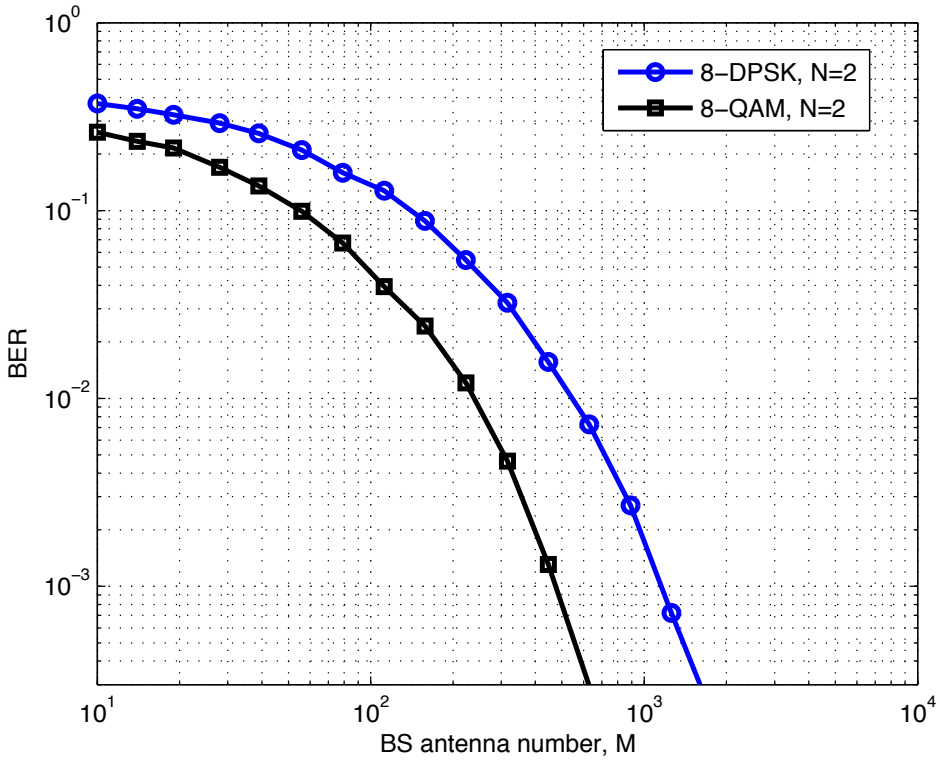
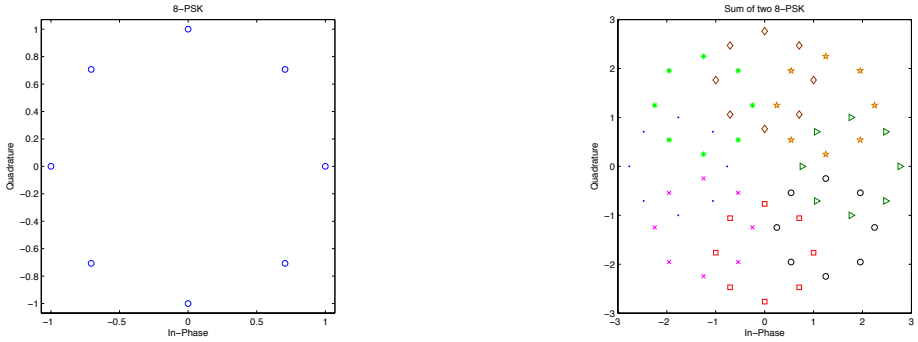


Figure 7.38: The comparison between the TC receiver and the non-coherent receiver with 8-QAM and 8-DPSK, respectively.

between successive symbols. In fact, the DBPSK and the DQPSK constellations with optimal scale between every sub-constellation are the special case of our QAMD. However, for larger constellations such as 8-DPSK, our QAMD has greater normalized minimal Euclidean distance. The resulting sum-constellation of two 8-DQPSK is not a regular constellation anymore, just as studied in [111]. Also, in [172], the actual transmitted power of each user is not given explicitly, and hence, the optimal power allocation under both the average and the peak power constraint case is hard to evaluate. To make a comparison, especially when the constellation size is large, we compare the 8-DPSK constellation suggested in [172] with the optimal scale 1.765

between the two sub-constellations with the rectangular 8-QAM constellation in our case. The constellation of the 8-DPSK with the optimal scale is plotted in Fig. 7.39. The error performance of [172] and our TCR with two users, each using 8-DPSK and 8-QAM, is studied in Fig. 7.38. It can be observed that our scheme with 8-QAM sub-constellation has a better error performance than [172] with 8-DPSK constellation, since the normalized minimal distance for our constellation is larger. Also, it should be pointed out that the resulting sum-constellation in [172] is not a regular constellation and it must be either computed or stored in advance. The detection of the sum-constellation typically requires a exhaustive search over the whole constellation. In addition, the optimal power scale for general DPSK needs to be optimized by numerical methods.



(a) 8-PSK (b) Sum-constellation

Figure 7.39: The sum constellation of two 8-PSK with scale 1.765 between them.

7.6 Conclusion

In this chapter, a QAMD multiple access framework is proposed for the uplink multiuser massive SIMO systems based on the MUSTM scheme. For the MUSTM code

design, a simple and systematic construction method based on the concept of QAM-UDCG is devised. Assuming that the large scale fading coefficients are known to all the terminals, the detailed transmission scheme is proposed followed by an efficient TCR. To enable that the channel coefficients of all the users can be estimated, a sub-constellation allocation method is also proposed. The optimal power allocation problem under the average and the peak power constraint cases are both considered. In particular, for the peak power constraint scenario, the optimal sub-constellation allocation method is given in closed-form. Then, a Riemannian distance receiver, a non-coherent ML receiver are also proposed for our transmission scheme. Computer simulations reveal that, in general, the ML receiver has the best error performance. Although the TCR requires more antennas than the ML and the Riemannian distance receiver for the same error performance, it is much simpler and easier to implement which is especially suitable for the massive MIMO systems with arguably unlimited number of antennas.

Most importantly, in our design, the minimum number of required training sequences is exactly one and it is considerably smaller than the conventional training sequence based method. Therefore, the training overhead can be reduced significantly. Also, more terminals can be supported in fast fading environment where the number of users is limited by the channel coherent time and delay spread for the conventional training based method.

In conclusion, by fully taking advantage of the finite-alphabet property of the digital communication signals to let them cooperate with each other to form a sum-constellation with good geometric structures as well as unlimited number of available antennas at the BS, the design of massive SIMO systems can be greatly simplified

and much more terminals can be supported simultaneously even in the fast fading environment.

Chapter 8

Conclusion and Future Work

In this thesis, we concentrate on the modulation division for multiuser single-hop and multihop wireless communication networks.

In Chapter 1, we first introduce the research motivations and major contributions of this thesis. Then, in Chapter 2, the definition of uniquely factorable constellation pair (UFCP) based on PSK and square-QAM constellations and uniquely decodable constellation group (UDCG) generated from square-QAM constellation are given explicitly.

In Chapter 3, we consider a two-hop relaying network consisting of two single-antenna terminal nodes and a relay node equipped with two antennas. By jointly processing the signals from the two antennas and taking advantage of the Alamouti coding structure twice at the relay node, the proposed code design made the equivalent channel between source and destination be a product of the two Alamouti channels, which is called distributed concatenated Alamouti STBC. The SEP analysis shows that the optimal diversity function is achieved for the maximum likelihood (ML) detector.

In Chapter 4, a single-hop point-to-point correlated MIMO channel is considered in which fully CSI is available at the receiver, but only the zero-mean and the covariance matrix is available at the transmitter. The optimal precoder is first developed for the ZF receiver which minimizes the average SEP over the Rayleigh fading channels. Then, by scaling up the antenna array size of both sides while keeping their ratio constant, simple asymptotic SEP formulas with the PAM, PSK and square QAM constellations are derived when the the correlation matrix is Hermitian Toeplitz with the aid of Szegö's theorem. This new results has a simple expression and can be used to evaluate the SEP performance conveniently and efficiently. It should be emphasized that the Szegö's theorem [44] is power mathematical tool for the systematic study of asymptotic behaviors on the large MIMO systems from both information-theoretic and detection viewpoints.

In Chapter 5, we also consider a point-to-point correlated MIMO channel where the optimal precoder is developed for the ZF-DF detector. By a thorough investigation on the product majorization relationship among the eigenvalues, singular-values and Cholesky values of the design matrix parameters, the original non-convex precoder design problem is converted into a convex geometrical programming problem which can be efficiently solved by using interior-point method. In addition, the asymptotic SEP analysis is also proposed for a QRS based precoder when the number of both transmitter and receiver antenna elements go unbounded. By making use of the characteristic of the large MIMO channels, the structure of the QRS transmitter as well as of the ZF-DF receiver, the Szegö's theorem [44] on large Hermitian Toeplitz matrices, we have attained a simple expression for the SEP limit with a fast convergence rate. In addition, an explanation of this approach related to the entropy power

of the channel is provided.

In Chapter 6, we first show the important fact that PAM or QAM constellations of large size can be uniquely decomposed into the sum of a group of the scaled version of the PAM or QAM constellations of variety of small flexible sizes. Then the application of such a design in multi-antenna MISO BCs is provided. For the special case with two receivers, the optimal beamforming vector is given in a closed-form by adopting a max-min criterion on the received SNR for the sum-signal. In addition, the SNR gain compared with ZF method is also given explicitly, which can be used to evaluate the effectiveness of our MD method in different channel coefficients. For the more general case with more than two receivers, a novel low-complexity grouping transmission scheme based on MD, which aims at improving the condition number of the channel matrix, is proposed, with each group having one or two users. Computer simulation results have revealed that for the Rayleigh channel, if the number of the receivers is far less than the number of BS antennas, our method degrade into the ZF method. However, when the number of users is very close to that of the BS antennas, the error performance of our proposed MD scheme is substantially better than ZF. Moreover, when the transmitter antennas are correlated, our presented method is still better than ZF, even if the number of transmitter antennas is larger than that of the receivers. In conclusion, our QAM MD transmission scheme can be considered as a feasible and concrete approach to the general NOMA method that introduces interference with proper power level superimposed on the desired signal and is potential and applicable to multiuser networks, which would enable a new promising multiple access method.

In Chapter 7, we concentrate on the uplink SIMO channel where a QAMD multiple

access framework is proposed on the MUSTM scheme. First, a simple and systematic construction of the MUSTM scheme based on the concept of QAM-UDCG is devised. Then, the detailed transmission scheme for the SIMO channel proposed followed by an efficient TCR by assuming that the large scale fading coefficients are known to all the terminals. We have developed a sub-constellation allocation method which enables the estimation of the channel coefficients of all the users. The optimal power allocation problem under the average and the peak power constraint cases are both considered where the optimal sub-constellation allocation method is given in closed-form. In addition, a Riemannian distance receiver and a non-coherent ML receiver are also proposed for our transmission scheme. Computer simulations reveal that in general, the ML receiver has the best error performance. It is shown that although the TCR requires more antennas than the ML and the Riemannian distance receiver for the same error performance, it is much simpler and easier to implement and is suitable for the massive MIMO systems with arguably unlimited number of antennas. It should be pointed out that in our design, the minimum number of required training sequences is exactly one and it is considerably smaller than the conventional training sequence based method. Therefore, the training overhead can be reduced significantly. Also, more terminals can be supported in fast fading environment where the number of users is limited by the channel coherent time and delay spread for the conventional training based method.

We will generalize our design to more complicated network topologies such as multi-hop networks with at least three hops or multiuser MIMO systems where each user has multiple antennas. In particular, in large MIMO systems, the transmission scheme when the large scale fading coefficients are unkown to all the terminals should

be addressed. Moreover, the application of modulation divison method in interference channels such as the Z-channel and the X-channel will be very promising. Also, the performace analysis for the considered networks with modualtion division should be performed in the future.

Appendix A

A.1 Proof of Algorithm 1

For any given $y = g + \xi$, the ML detection of g is exactly equivalent to that of $p = g + \frac{2^K-1}{2} \in \{k\}_{k=0}^{2^K-1}$ from $p + \xi$, which is given by

$$\hat{p} = \begin{cases} 0, & y + \frac{2^K-1}{2} \leq 0; \\ \lfloor y + \frac{2^K-1}{2} + \frac{1}{2} \rfloor, & 0 < y + \frac{2^K-1}{2} \leq 2^K - 1; \\ 2^K - 1, & y + \frac{2^K-1}{2} > 2^K - 1. \end{cases}$$

leading us to (2.14).

For $K_i = 0$, (2.15) and (2.16) indeed hold. Therefore, we only consider the case $K_i \neq 0$. For presentation simplicity, we define p_i as follows: 1) for any $x_1 \in \mathcal{X}_1$, $p_1 = x_1 + \frac{2^{K_1}-1}{2}$; and 2) for any $x_i \in \mathcal{X}_i$, $p_i = x_i \times 2^{-\sum_{\ell=1}^{i-1} K_\ell} + \frac{2^{K_i}-1}{2}$. From the definitions of $\mathcal{X}_i$, we can attain that $p_i \in \{k\}_{k=0}^{k=2^{K_i}-1}$. In addition, Theorem 1 tells us that given $\hat{g} \in \mathcal{G}$ defined by (2.14), $\hat{g}$ can be uniquely decomposed into $\hat{g} = \sum_{i=1}^{N} \hat{x}_i$,

where $\hat{x}_i \in \mathcal{X}_i$ and thus, be equivalently rewritten into

$$\hat{g} = \sum_{i=1}^{N} \hat{x}_i = \hat{p}_1 + \sum_{i=2}^{N} \hat{p}_i \times 2^{\sum_{\ell=1}^{i-1} K_\ell} - \frac{2^K - 1}{2} \tag{A.1}$$

where $\hat{p}_i \in \{k\}_{k=0}^{k=2^{K_i}-1}$. It is noticed that for any $\hat{g} \in \mathcal{G} = \{\pm(k - \frac{1}{2})\}_{k=1}^{k=2^{K-1}}$, we have $\{\hat{g} + \frac{2^K-1}{2} : \hat{g} \in \mathcal{G}\} = \{k\}_{k=0}^{2^K-1}$. Then, letting $\hat{p} = \hat{g} + \frac{2^K-1}{2}$ produces an equivalent form of (A.1) as $\hat{p} = \hat{p}_1 + \sum_{i=2}^{N} \hat{p}_i \times 2^{\sum_{\ell=1}^{i-1} K_\ell}$, where $\hat{p}_i \in \{k\}_{k=0}^{k=2^{K_i}-1}$. Now, we notice that $\hat{p}_1 = \hat{p} \mod 2^{K_1}$ because of the fact that $\sum_{i=2}^{N} \hat{p}_i \times 2^{\sum_{\ell=1}^{i-1} K_\ell} \mod 2^{K_1} = 0$. This result gives us that $\hat{x}_1 = \left(\hat{g} + \frac{2^K-1}{2}\right) \mod 2^{K_1} - \frac{2^{K_1}-1}{2}$, proving (2.15). Also, we observe that $\frac{\hat{p}-\hat{p}_1}{2^{K_1}} = \hat{p}_2 + \sum_{i=3}^{N} \hat{p}_i \times 2^{\sum_{\ell=2}^{i-1} K_\ell}$ and thus, attain that $\hat{p}_2 = \frac{\hat{p}-\hat{p}_1}{2^{K_1}} \mod 2^{K_2}$, leading to $\hat{x}_2 = \left(\hat{p}_2 - \frac{2^{K_2}-1}{2}\right) \times 2^{K_1}$. Following this process, we can obtain that for $2 \leq i \leq N$, $\hat{p}_i = \left(\frac{\hat{p}-\hat{p} \mod 2^{\sum_{\ell=1}^{i-1} K_\ell}}{2^{\sum_{\ell=1}^{i-1} K_\ell}}\right) \mod 2^{K_i}$. In addition, our notation $\hat{p}_i = \hat{x}_i \times 2^{-\sum_{\ell=1}^{i-1} K_\ell} + \frac{2^{K_i}-1}{2}$ allows us to arrive at the desired (2.16).

Therefore, this verifies Algorithm 1. ■

A.2 Proof of Theorem 10

Scenario 1: $\mathbf{h}_1 = \tau \mathbf{h}_2, \tau \in \mathbb{C}$. If $|\tau| \geq 1$, then, we have $\mathbf{w}^H \mathbf{h}_1 \mathbf{h}_1^H \mathbf{w} - \mathbf{w}^H \mathbf{h}_2 \mathbf{h}_2^H \mathbf{w} = (|\tau|^2 - 1|) \mathbf{w}^H \mathbf{h}_2 \mathbf{h}_2^H \mathbf{w} \geq 0$ for any $\mathbf{w}^H \mathbf{w} = P$ and as a result, the original optimization problem degrades into $\max_{\|\mathbf{w}\|^2 = P} \mathbf{w}^H \mathbf{h}_2 \mathbf{h}_2^H \mathbf{w}$. Using Cauchy-Schwarz inequality, $\mathbf{w}^{\text{opt}} = \frac{\sqrt{P}\mathbf{h}_2}{\|\mathbf{h}_2\|}$ (or $\mathbf{w}^{\text{opt}} = \frac{\sqrt{P}\mathbf{h}_1}{\|\mathbf{h}_1\|}$). The case when $|\tau| < 1$ is similar and we omit it here.

Scenario 2: $\mathbf{h}_1 \neq \tau \mathbf{h}_2, \forall \tau \in \mathbb{C}$. In this case, the overall optimization problem can be divided into the following two problems by introducing restrictions on the original feasible region and the global optimum is the maximum of the two.

- *Case 1: SNR of user 1 is not worse than that of user 2*

$$\max_{\mathbf{w}} \ \mathbf{w}^H \mathbf{h}_2 \mathbf{h}_2^H \mathbf{w} \tag{A.2a}$$

$$\text{s.t.} \ \ \mathbf{w}^H \mathbf{A} \mathbf{w} \geq 0 \text{ and } \mathbf{w}^H \mathbf{w} = P. \tag{A.2b}$$

- *Case 2: SNR of user 2 is better than that of user 1*

$$\max_{\mathbf{w}} \ \mathbf{w}^H \mathbf{h}_1 \mathbf{h}_1^H \mathbf{w} \tag{A.3a}$$

$$\text{s.t.} \ \ \mathbf{w}^H \mathbf{A} \mathbf{w} < 0 \text{ and } \mathbf{w}^H \mathbf{w} = P. \tag{A.3b}$$

Let us consider *Case 1* first. Using the notation in (6.117), the optimization problem can be reformulated as

$$\max_{\tilde{\mathbf{w}}} \ \tilde{\mathbf{w}}^H \tilde{\mathbf{h}}_2 \tilde{\mathbf{h}}_2^H \tilde{\mathbf{w}} \tag{A.4a}$$

$$\text{s.t.} \ \ \lambda_1 |\tilde{w}_1|^2 \geq \lambda_2 |\tilde{w}_2|^2 \text{ and } \sum_{\ell=1}^{M} |\tilde{w}_\ell|^2 = P. \tag{A.4b}$$

We have the following two observations on the above optimization problem:

1. From the objective function $\tilde{\mathbf{w}}^H \tilde{\mathbf{h}}_2 \tilde{\mathbf{h}}_2^H \tilde{\mathbf{w}} = |\sum_{\ell=1}^{M} \tilde{w}_\ell^* \tilde{h}_{2,\ell}|^2$ and the constraint, one optimal choice of the angle of $\tilde{w}_\ell$ is $\arg(\tilde{w}_\ell) = \arg(\tilde{h}_{2,\ell}), \forall \ell$.

2. Since $\tilde{\mathbf{h}}_2 = [\tilde{h}_{21}, \tilde{h}_{22}, 0, \ldots, 0]^T$, we should let $|\tilde{w}_\ell| = 0, \forall \ell \geq 3$. Otherwise we could let $|\tilde{w}_\ell| = 0, \forall \ell \geq 3$ and increase $|\tilde{w}_1|$ and $|\tilde{w}_2|$ slightly without violating the constraint, resulting in an increased objective function.

Hence, the optimization problem can be simplified into

$$\max_{\tilde{w}_1} \ |\tilde{w}_1||\tilde{h}_{2,1}| + \sqrt{P - |\tilde{w}_1|^2}|\tilde{h}_{2,2}| \tag{A.5a}$$

$$\text{s.t.} \ \sqrt{\frac{P\lambda_2}{\lambda_1 + \lambda_2}} \leq |\tilde{w}_1| \leq \sqrt{P}. \tag{A.5b}$$

With the help of (6.120), we can denote the objective function as

$$g_1(|\tilde{w}_1|) = |\tilde{w}_1|\sqrt{\lambda_1}\tan\theta + \sqrt{P - |\tilde{w}_1|^2}\sqrt{\lambda_2}\sec\theta$$

whose derivative is $g_1'(|\tilde{w}_1|) = \sqrt{\lambda_1}\tan\theta - \frac{|\tilde{w}_1|}{\sqrt{P-|\tilde{w}_1|^2}}|\sqrt{\lambda_2}\sec\theta$

$$g_1'(|\tilde{w}_1|) = \begin{cases} = 0 & |\tilde{w}_1| = \sqrt{\frac{P\lambda_1}{\lambda_1+\lambda_2\csc^2\theta}}, \\ > 0 & 0 \leq |\tilde{w}_1| < \sqrt{\frac{P\lambda_1}{\lambda_1+\lambda_2\csc^2\theta}}, \\ < 0 & \sqrt{\frac{P\lambda_1}{\lambda_1+\lambda_2\csc^2\theta}} < |\tilde{w}_1| \leq \sqrt{P}. \end{cases}$$

Therefore, the solution to problem (A.5) is given below:

1. $0 \leq \sin\theta \leq \frac{\lambda_2}{\lambda_1}$ (i.e., $0 \leq \sqrt{\frac{P\lambda_1}{\lambda_1+\lambda_2\csc^2\theta}} \leq \sqrt{\frac{P\lambda_2}{\lambda_1+\lambda_2}}$), then $\max_{\sqrt{\frac{P\lambda_2}{\lambda_1+\lambda_2}} \leq |\tilde{w}_1| \leq \sqrt{P}} \ g_1(|\tilde{w}_1|) = g_1\left(\sqrt{\frac{P\lambda_2}{\lambda_1+\lambda_2}}\right) = \sqrt{\frac{P\lambda_1\lambda_2}{\lambda_1+\lambda_2}}\frac{1+\sin\theta}{\cos\theta}$.
2. $\frac{\lambda_2}{\lambda_1} < \sin\theta \leq 1$ (i.e., $\sqrt{\frac{P\lambda_2}{\lambda_1+\lambda_2}} < \sqrt{\frac{P\lambda_1}{\lambda_1+\lambda_2\csc^2\theta}} \leq \sqrt{\frac{P\lambda_1}{\lambda_1+\lambda_2}}$), then $\max_{\sqrt{\frac{P\lambda_2}{\lambda_1+\lambda_2}} \leq |\tilde{w}_1| \leq \sqrt{P}} \ g_1\left(|\tilde{w}_1|\right) = g_1\left(\sqrt{\frac{P\lambda_1}{\lambda_1+\lambda_2\csc^2\theta}}\right) = \frac{\sqrt{P(\lambda_1\sin^2\theta+\lambda_2)}}{\cos\theta}$.

For *Case 2*, by a similar argument on the objective function $\tilde{\mathbf{w}}^H\tilde{\mathbf{h}}_1\tilde{\mathbf{h}}_1^H\tilde{\mathbf{w}} =$

$|\sum_{\ell=1}^{M} \tilde{w}_\ell^* \tilde{h}_{1,\ell}|^2$, we have $\arg(\tilde{w}_\ell) = \arg(\tilde{h}_{1,\ell}), \forall \ell$, and $|\tilde{w}_\ell| = 0, \forall \ell \geq 3$. The resulting optimization problem can be reformulated as

$$\max_{\tilde{w}_1} \ |\tilde{w}_1||\tilde{h}_{1,1}| + \sqrt{P - |\tilde{w}_1|^2}|\tilde{h}_{1,2}| \tag{A.6a}$$

$$\text{s.t.} \ \ 0 \leq |\tilde{w}_1| \leq \sqrt{\frac{P\lambda_2}{\lambda_1 + \lambda_2}}. \tag{A.6b}$$

With the aid of (6.120), we can represent the objective function as $g_2(|\tilde{w}_1|) = |\tilde{w}_1|\sqrt{\lambda_1}\sec\theta + \sqrt{P - |\tilde{w}_1|^2}\sqrt{\lambda_2}\tan\theta$. Since its first order derivative is $g_2'(|\tilde{w}_1|) = \sqrt{\lambda_1}\sec\theta - \frac{|\tilde{w}_1|}{\sqrt{P-|\tilde{w}_1|^2}}\sqrt{\lambda_2}\tan\theta$, we have

$$g_2'(|\tilde{w}_1|) = \begin{cases} = 0 & |\tilde{w}_1| = \sqrt{\frac{P\lambda_1}{\lambda_1 + \lambda_2 \sin^2\theta}}, \\ > 0 & 0 \leq |\tilde{w}_1| < \sqrt{\frac{P\lambda_1}{\lambda_1 + \lambda_2 \sin^2\theta}}, \\ < 0 & \sqrt{\frac{P\lambda_1}{\lambda_1 + \lambda_2 \sin^2\theta}} < |\tilde{w}_1| \leq 1. \end{cases}$$

Therefore, the solution to (A.6) can be determined as follows:

1. $0 \leq \sin\theta < \frac{\lambda_1}{\lambda_2}$ (i.e., $\sqrt{\frac{P\lambda_2}{\lambda_1+\lambda_2}} < \sqrt{\frac{P\lambda_1}{\lambda_1+\lambda_2\sin^2\theta}} \leq \sqrt{P}$), then $\max_{0 \leq |\tilde{w}_1| \leq \sqrt{\frac{P\lambda_2}{\lambda_1+\lambda_2}}} \ g_2\Big(|\tilde{w}_1|\Big) = g_2\Big(\sqrt{\frac{P\lambda_2}{\lambda_1+\lambda_2}}\Big) = \sqrt{\frac{P\lambda_1\lambda_2}{\lambda_1+\lambda_2}}\frac{1+\sin\theta}{\cos\theta}$.
2. $\frac{\lambda_1}{\lambda_2} \leq \sin\theta \leq 1$ (i.e., $0 < \sqrt{\frac{P\lambda_1}{\lambda_1+\lambda_2\sin^2\theta}} \leq \sqrt{\frac{P\lambda_2}{\lambda_1+\lambda_2}}$), then $\max_{0 \leq |\tilde{w}_1| \leq \sqrt{\frac{P\lambda_2}{\lambda_1+\lambda_2}}} \ g_2(|\tilde{w}_1|) = g_2\Big(\sqrt{\frac{P\lambda_1}{\lambda_1+\lambda_2\sin^2\theta}}\Big) = \frac{\sqrt{P(\lambda_1+\lambda_2\sin^2\theta)}}{\cos\theta}$.

The overall maximum value of the original problem is the maximum of the two cases.

A.3 Proof of Lemma 5

Let the singular value decomposition (SVD) of $\mathbf{H}$ be $\mathbf{H} = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^H$, where $\mathbf{U}_1 \in \mathbb{C}^{M\times 2}$ is a tall column-wise unitary matrix, $\mathbf{V}_1 = [\mathbf{v}_1\ \mathbf{v}_2] = \begin{bmatrix} v_{1,1} & v_{1,2} \\ v_{2,1} & v_{2,2} \end{bmatrix} \in \mathbb{C}^{2\times 2}$ is a unitary matrix, and $\mathbf{\Sigma}_1 = \text{diag}(\sqrt{\mu}_1, \sqrt{\mu}_2)$. Then, we have $[\mathbf{h}_1\ -\mathbf{h}_2] = \mathbf{U}_1\mathbf{\Sigma}_1 \begin{bmatrix} v_{1,1}^* & -v_{2,1}^* \\ v_{1,2}^* & -v_{2,2}^* \end{bmatrix}$, and thus, equation (6.115) can be represented by $\mathbf{A} = [\mathbf{h}_1\ -\mathbf{h}_2][\mathbf{h}_1\ \mathbf{h}_2]^H = \mathbf{U}_1\mathbf{B}\mathbf{U}_1^H$, where

$$\mathbf{B} = \mathbf{\Sigma}_1 \begin{bmatrix} |v_{1,1}|^2 - |v_{2,1}|^2 & v_{1,1}^* v_{1,2} - v_{2,1}^* v_{2,2} \\ v_{1,2}^* v_{1,1} - v_{2,2}^* v_{2,1} & |v_{1,2}|^2 - |v_{2,2}|^2 \end{bmatrix} \mathbf{\Sigma}_1. \tag{A.7}$$

Notice that matrix $\mathbf{A}$ and $\mathbf{B}$ have the same non-zero eigenvalues, i.e., $\text{diag}(\lambda_1, -\lambda_2) = \text{eig}(\mathbf{B})$. Hence, in order to use a, b and c to represent $\lambda_1, \lambda_2, \mu_1$ and μ_2, we need to calculate $\mathbf{V}_1$. Since $\|\mathbf{h}_1\|^2 = a$, $\|\mathbf{h}_2\|^2 = b$ and $|\mathbf{h}_1^H\mathbf{h}_2| = c$, if we let $\arg(\mathbf{h}_1^H\mathbf{h}_2) = \phi_c$, then, we have

$$\mathbf{H}^H\mathbf{H} = \mathbf{V}_1\mathbf{\Sigma}_1^2\mathbf{V}_1^H = \begin{bmatrix} a & ce^{j\phi_c} \\ ce^{-j\phi_c} & b \end{bmatrix}. \tag{A.8}$$

Hence, $\mathbf{V}_1$ is the eigenvector matrix of $\mathbf{H}^H\mathbf{H}$ and μ_1 and μ_2 are its eigenvalues, which must satisfy the following characteristic equation, $\begin{vmatrix} a-\mu_i & ce^{j\phi_c} \\ ce^{-j\phi_c} & b-\mu_i \end{vmatrix} = 0$, for $i = 1, 2$. Therefore, we have $\mu_1 = \frac{a+b+\sqrt{(a-b)^2+4c^2}}{2}$, $\mu_2 = \frac{a+b-\sqrt{(a-b)^2+4c^2}}{2}$.

Correspondingly, the two column vectors of $\mathbf{V}_1 = [\mathbf{v}_1\ \mathbf{v}_2]$ must satisfy the following

equations:

$$\begin{bmatrix} a-\mu_1 & ce^{j\phi_c} \\ ce^{-j\phi_c} & b-\mu_1 \end{bmatrix} \mathbf{v}_1 = \mathbf{0}, \quad \begin{bmatrix} a-\mu_2 & ce^{j\phi_c} \\ ce^{-j\phi_c} & b-\mu_2 \end{bmatrix} \mathbf{v}_2 = \mathbf{0}.$$

Let $d = \sqrt{(a-b)^2 + 4c^2}$. Then, we have

$$\mathbf{v}_1 = \left[\frac{-2ce^{j\phi_c}}{\sqrt{2d^2 - 2(a-b)d}} \ \frac{a-b-d}{\sqrt{2d^2 - 2(a-b)d}} \right]^T,$$
$$\mathbf{v}_2 = \left[\frac{-2ce^{j\phi_c}}{\sqrt{2d^2 + 2(a-b)d}} \ \frac{a-b+d}{\sqrt{2d^2 + 2(a-b)d}} \right]^T,$$

and $\mu_1 = \frac{a+b+d}{2}$, $\mu_2 = \frac{a+b-d}{2}$ and $\sqrt{\mu_1\mu_2} = \sqrt{ab-c^2}$. Hence, (A.7) can be expressed by

$$\begin{aligned} \mathbf{B} &= \frac{1}{d} \begin{bmatrix} (a-b)\mu_1 & 2c\sqrt{\mu_1\mu_2} \\ 2c\sqrt{\mu_1\mu_2} & -(a-b)\mu_2 \end{bmatrix} \\ &= \frac{1}{2d} \begin{bmatrix} (a-b)(a+b+d) & 4c\sqrt{ab-c^2} \\ 4c\sqrt{ab-c^2} & -(a-b)(a+b-d) \end{bmatrix}. \end{aligned}$$

On the other hand, the eigenvalues of $\mathbf{A}$, i.e., $\lambda_1, -\lambda_2$ must satisfy the following characteristic equation:

$$\begin{vmatrix} \frac{(a-b)(a+b+d)}{2d} - \lambda & \frac{4c\sqrt{ab-c^2}}{2d} \\ \frac{4c\sqrt{ab-c^2}}{2d} & \frac{-(a-b)(a+b-d)}{2d} - \lambda \end{vmatrix} = 0. \tag{A.9}$$

Solving this equation and after some algebraic manipulations, we attain $\lambda_1 =$

$\frac{a-b+\sqrt{(a+b)^2-4c^2}}{2}$, $\lambda_2 = \frac{-a+b+\sqrt{(a+b)^2-4c^2}}{2}$.

This completes the proof of Lemma 5. □

A.4 Proof of Lemma 6

We first calculate the expectation $\mathbb{E}[r_t]$ and variance $\text{var}(r_t)$ of r_t. Recall that $\mathbf{y}_t = [y_{1,1}, y_{2,1}, \ldots, y_{M,t}]^T$, where $y_{m,1} = \sum_{k=1}^{N} g_{m,k}\sqrt{\beta_k q_k} + n_{m,1} = \mathbf{c}^H\mathbf{f}_m + n_{m,1}$, and $y_{m,t} = \sum_{k=1}^{N} g_{m,k}\sqrt{\beta_k}x_{k,t} + n_{m,t} = \mathbf{d}_t^H\mathbf{f}_m + n_{m,t}, m = 1, 2, \ldots, M, t = 2, 3, \ldots, N$, in which $\mathbf{f}_m = [g_{m,1}, g_{m,2}, \ldots, g_{m,N}]^T$. Let us denote $z_{m,t} = y_{m,1}^* y_{m,t}$. Then, we have

$$z_{m,t} = (\mathbf{c}^H\mathbf{f}_m + n_{m,1})^H(\mathbf{d}_t^H\mathbf{f}_m + n_{m,t}) = \mathbf{f}_m^H\mathbf{c}\mathbf{d}_t^H\mathbf{f}_m + n_{m,1}^*\mathbf{d}_t^H\mathbf{f}_m + n_{m,t}\mathbf{f}_m^H\mathbf{c} + n_{m,1}^* n_{m,t}.$$

Hence, the mean value of $z_{m,t}$ is given by

$$\mathbb{E}[z_{m,t}] = \text{tr}(\mathbf{c}\mathbf{d}_t^H\mathbb{E}[\mathbf{f}_m\mathbf{f}_m^H]) + n_{m,1}^*\mathbf{d}_t^H\mathbb{E}[\mathbf{f}_m] + n_{m,t}\mathbb{E}[\mathbf{f}_m^H]\mathbf{c} + \mathbb{E}[n_{m,1}^*]\mathbb{E}[n_{m,t}] = \mathbf{d}_t^H\mathbf{c},$$

Meanwhile, the variance of $z_{m,t}$ is calculated by taking the expectation over $\mathbf{f}_m$ and $n_{m,t}$ again,

$$\text{var}(z_{m,t}) = \mathbb{E}[z_{m,t}^* z_{m,t}] - \mathbb{E}[z_{m,t}^*]\mathbb{E}[z_{m,t}]. \tag{A.12}$$

We first calculate the first term on the right hand side of (A.12) as follows:

$$
\begin{aligned}
&\mathbb{E}[z_{m,t}^* z_{m,t}] = \mathbb{E}[(\mathbf{f}_m^H \mathbf{c}\mathbf{d}_t^H \mathbf{f}_m + n_{m,1}^* \mathbf{d}_t^H \mathbf{f}_m + n_{m,t}\mathbf{f}_m^H \mathbf{c} + n_{m,1}^* n_{m,t})^H \\
&\quad \times (\mathbf{f}_m^H \mathbf{c}\mathbf{d}_t^H \mathbf{f}_m + n_{m,1}^* \mathbf{d}_t^H \mathbf{f}_m + n_{m,t}\mathbf{f}_m^H \mathbf{c} + n_{m,1}^* n_{m,t})] \\
&= \mathbb{E}[|\mathbf{f}_m^H \mathbf{c}\mathbf{d}_t^H \mathbf{f}_m|^2] + \mathbb{E}[n_{m,1}^* \mathbf{f}_m^H \mathbf{d}_t \mathbf{c}^H \mathbf{f}_m \mathbf{d}_t^H \mathbf{f}_m] + \mathbb{E}[n_{m,t}\mathbf{f}_m^H \mathbf{d}_t \mathbf{c}^H \mathbf{f}_m \mathbf{f}_m^H \mathbf{c}] + \mathbb{E}[n_{m,1}^* n_{m,t}\mathbf{f}_m^H \mathbf{d}_t \mathbf{c}^H \mathbf{f}_m] \\
&+ \mathbb{E}[n_{m,1}\mathbf{f}_m^H \mathbf{d}_t \mathbf{f}_m^H \mathbf{c}\mathbf{d}_t^H \mathbf{f}_m] + \mathbb{E}[n_{m,1}^* n_{m,1}\mathbf{f}_m^H \mathbf{d}_t \mathbf{d}_t^H \mathbf{f}_m] + \mathbb{E}[n_{m,1} n_{m,t}\mathbf{f}_m^H \mathbf{d}_t \mathbf{f}_m^H \mathbf{c}] + \mathbb{E}[n_{m,1} n_{m,1}^* n_{m,t}\mathbf{f}_m^H \mathbf{d}_t] \\
&+ \mathbb{E}[n_{m,t}^* \mathbf{c}^H \mathbf{f}_m \mathbf{f}_m^H \mathbf{c}\mathbf{d}_t^H \mathbf{f}_m] + \mathbb{E}[n_{m,t}^* n_{m,1}^* \mathbf{c}^H \mathbf{f}_m \mathbf{d}_t^H \mathbf{f}_m] + \mathbb{E}[n_{m,t}^* n_{m,t} \mathbf{c}^H \mathbf{f}_m \mathbf{f}_m^H \mathbf{c}] + \mathbb{E}[n_{m,t}^* n_{m,1}^* n_{m,t} \mathbf{c}^H \mathbf{f}_m] \\
&+ \mathbb{E}[n_{m,1} n_{m,t}^* \mathbf{f}_m^H \mathbf{c}\mathbf{d}_t^H \mathbf{f}_m] + \mathbb{E}[n_{m,1} n_{m,t}^* n_{m,1}^* \mathbf{d}_t^H \mathbf{f}_m] + \mathbb{E}[n_{m,1} n_{m,t}^* n_{m,t} \mathbf{f}_m^H \mathbf{c}] + \mathbb{E}[n_{m,1} n_{m,t}^* n_{m,1}^* n_{m,t}] \\
&\overset{(a)}{=} \mathbb{E}[|\mathbf{f}_m^H \mathbf{c}\mathbf{d}_t^H \mathbf{f}_m|^2] + \mathbb{E}[n_{m,1}^* n_{m,1} \mathbf{f}_m^H \mathbf{d}_t \mathbf{d}_t^H \mathbf{f}_m] + \mathbb{E}[n_{m,t}^* n_{m,t} \mathbf{c}^H \mathbf{f}_m \mathbf{f}_m^H \mathbf{c}] + \mathbb{E}[n_{m,1} n_{m,t}^* n_{m,1}^* n_{m,t}] \\
&= \mathbb{E}[\mathrm{tr}(\mathbf{d}_t \mathbf{c}^H \mathbf{f}_m \mathbf{f}_m^H \mathbf{c}\mathbf{d}_t^H \mathbf{f}_m \mathbf{f}_m^H)] + \mathbb{E}[n_{m,1}^* n_{m,1}]\mathrm{tr}(\mathbf{d}_t \mathbf{d}_t^H \mathbb{E}[\mathbf{f}_m \mathbf{f}_m^H]) \\
&+ \mathbb{E}[n_{m,t}^* n_{m,t}]\mathrm{tr}(\mathbf{c}\mathbf{c}^H \mathbb{E}[\mathbf{f}_m \mathbf{f}_m^H]) + \mathbb{E}[n_{m,1}^* n_{m,1}]\mathbb{E}[n_{m,t}^* n_{m,t}] \\
&\overset{(b)}{=} \mathrm{tr}(\mathbf{d}_t \mathbf{c}^H (\mathbf{c}\mathbf{d}_t^H + \mathrm{tr}(\mathbf{c}\mathbf{d}_t^H)\mathbf{I})) + \sigma^2 \mathbf{d}_t^H \mathbf{d}_t + \sigma^2 \mathbf{c}^H \mathbf{c} + \sigma^4 \\
&= \|\mathbf{c}\|^2 \|\mathbf{d}_t\|^2 + |\mathbf{c}^H \mathbf{d}_t|^2 + \sigma^2 \|\mathbf{d}_t\|^2 + \sigma^2 \|\mathbf{c}\|^2 + \sigma^4,
\end{aligned}
$$

where (a) holds since the channel coefficients and noise are assumed to be with zero mean and independent to each other, and (b) results from the following lemma.

Lemma 10 *[173] In general for $\mathbf{W} \sim \mathcal{CW}_N(\mathbf{\Sigma}, m) \in \mathbb{C}^{N \times N}$ which follows the complex Wishart distribution of m degree of freedom, we have*

$$\mathbb{E}[\mathbf{W}] = m\mathbf{\Sigma}, \quad \mathbb{E}[\mathbf{WRW}] = m^2 \mathbf{\Sigma R \Sigma} + m\mathrm{tr}(\mathbf{R\Sigma})\mathbf{\Sigma}.$$

■

Let $\mathbf{W}_m = \mathbf{f}_m \mathbf{f}_m^H$. Then, $\mathbf{W}_m$ follows the complex Wishart distribution with 1 degree of freedom and covariance $\mathbf{I}$, i.e., $\mathbf{W}_m \sim \mathcal{CW}_N(\mathbf{I}, 1)$ and (b) holds. Now, the variance of $z_{m,t}$ can be calculated by

$$\begin{aligned}\delta_t^2 \triangleq \mathrm{var}(z_{m,t}) &= \mathbb{E}[z_{m,t}^* z_{m,t}] - \mathbb{E}[z_{m,t}^*]\mathbb{E}[z_{m,t}] \\ &= \|\mathbf{c}\|^2\|\mathbf{d}_t\|^2 + \sigma^2(\|\mathbf{d}_t\|^2 + \|\mathbf{c}\|^2) + \sigma^4.\end{aligned}$$

In what follows, we will consider the distribution of r_t. Recall that the random variable $z_{m,t}$ is defined by $z_{m,t} = \mathbf{f}_m^H \mathbf{c}\mathbf{d}_t^H \mathbf{f}_m + n_{m,1}^* \mathbf{d}_t^H \mathbf{f}_m + n_{m,t}\mathbf{f}_m^H \mathbf{c} + n_{m,1}^* n_{m,t}$ and also

$$r_t = \frac{1}{M}\mathbf{y}_1^H \mathbf{y}_t = \frac{1}{M}\sum_{m=1}^{M} y_{m,1}^* y_{m,t} = \frac{1}{M}\sum_{m=1}^{M} z_{m,t}.$$

From the assumption above, we know that $\mathbf{f}_m$ and $n_{m,1}$ and $n_{m,t}$ are i.i.d. for different m. As a consequence, $z_{m,t}$ are also i.i.d. for different m and hence,

$$\begin{aligned}\mathbb{E}[r_t] &= \frac{1}{M}\sum_{m=1}^{M}\mathbb{E}[z_{m,t}] = \mathbf{d}_t^H \mathbf{c}, \\ \mathrm{var}(r_t) &= \frac{1}{M^2}\mathrm{var}\Big(\sum_{m=1}^{M} z_{m,t}\Big) = \frac{\delta_t^2}{M}.\end{aligned}$$

By using the CLT and noticing that $z_{m,t}$ are i.i.d. for different m, when M is large enough, we have $r_t \dot{\sim} \mathcal{CN}\Big(\mathbf{d}_t^H \mathbf{c}, \frac{\delta_t^2}{M}\Big)$. This completes the proof. □

A.5 Proof of Lemma 7

In Fig. 7.30, $a = |\overline{OC}|$ and $c = |\overline{OA}|$ is the radius of the circle where $a > c$. Also, $b(\theta)$ is the length of the line segment in the secant line with $b_L(\theta) = |\overline{AC}|$ and

$b_U(\theta) = |\overline{BC}|$. According to the law of cosines and note that $|\overline{OB}| = |\overline{OA}| = c$, we have $a^2 + b^2(\theta) - 2ab(\theta)\cos\theta = c^2$. As a result, we have

$$b_L(\theta) = a\cos\theta - \sqrt{a^2\cos^2\theta - (a^2 - c^2)},$$
$$b_U(\theta) = a\cos\theta + \sqrt{a^2\cos^2\theta - (a^2 - c^2)}.$$

We know that $\theta_{\max} = \arcsin(\frac{c}{a})$, and note that $c < a$. Thus, $\theta_{\max} < \frac{\pi}{2}$. For a random variable $r \sim \mathcal{CN}(\mu_c, \nu^2)$, where μ_c is the coordinate of point C, the probability of r falling into the circle is given by

$$\begin{aligned} P_{e_1} &= 2\int_0^{\theta_{\max}} \int_{b_L(\theta)}^{b_U(\theta)} \frac{1}{\pi\nu^2} e^{-\frac{b^2(\theta)}{\nu^2}} b(\theta)\, db(\theta)\, d\theta \\ &= \frac{1}{\pi}\int_0^{\theta_{\max}} \left(e^{-\frac{b_L^2(\theta)}{\nu^2}} - e^{-\frac{b_U^2(\theta)}{\nu^2}}\right) d\theta \\ &< \frac{1}{\pi}\int_0^{\theta_{\max}} e^{-\frac{b_L^2(\theta)}{\nu^2}} d\theta < e^{-\frac{(a-c)^2}{\nu^2}}. \end{aligned}$$

Similarly for Fig. 7.31, $a = |\overline{OC}|$ and $c = |\overline{OA}|$ is the radius of the circle and $b = |\overline{AC}|$. By using the law of cosines again, $a^2 + c^2 - 2ac\cos\theta = b^2(\theta)$. As a result, $b(\theta) = \sqrt{a^2 + c^2 - 2ac\cos\theta}$, where $\theta \in [0, \pi)$. Note that $c > a$ and $b_{\min}(\theta) = c - a$. For a random variable $r \sim \mathcal{CN}(\mu_c, \nu^2)$, the probability of r falling outside the circle is given by

$$\begin{aligned} P_{e_2} &= 2\int_0^{\pi} \int_{b(\theta)}^{\infty} \frac{1}{\pi\nu^2} e^{-\frac{b^2(\theta)}{\nu^2}} b(\theta)\, d\, b(\theta)\, d\theta \\ &= \frac{1}{\pi}\int_0^{\pi} e^{-\frac{b^2(\theta)}{\nu^2}} d\theta < e^{-\frac{(c-a)^2}{\nu^2}}. \end{aligned}$$

Hence, in both cases, we can attain $P_e < e^{-\frac{(c-a)^2}{\nu^2}}$. This completes the proof. □

A.6 Proof of Theorem 11

We know that r_t, $S_t^{(k)}$ and $S_t^{(\ell)}$ are complex numbers and $\sqrt{p_1 p_t}$, $\delta_t^{(k)}$, $\delta_t^{(\ell)}$ are positive real numbers. Then, (7.148) is equivalent to $P_r(S_t^{(k)} \to S_t^{(\ell)}) = P_r\Big((\delta_t^{(k)})^2|r_t - \sqrt{p_1 p_t}S_t^{(\ell)}|^2 - (\delta_t^{(\ell)})^2|r_t - \sqrt{p_1 p_t}S_t^{(k)}|^2 < \frac{2(\delta_t^{(k)}\delta_t^{(\ell)})^2 \ln(\delta_t^{(k)}/\delta_t^{(\ell)})}{M}\Big)$. Now let $\tilde{r}_t = r_t - \sqrt{p_1 p_t}S_t^{(k)}$ and note that $r_t \dot{\sim} \mathcal{CN}(\sqrt{p_1 p_t}S_t^{(k)}, \frac{(\delta_t^{(k)})^2}{M})$. Hence, we have $\tilde{r}_t \dot{\sim} \mathcal{CN}(0, \frac{(\delta_t^{(k)})^2}{M})$ and as a result,

$$\begin{aligned} P_r(S_t^{(k)} \to S_t^{(\ell)}) &= P_r\Big((\delta_t^{(k)})^2|\tilde{r}_t + \sqrt{p_1 p_t}\Delta S_t^{(k,\ell)}|^2 - (\delta_t^{(\ell)})^2|\tilde{r}_t|^2 < \frac{2(\delta_t^{(k)}\delta_t^{(\ell)})^2 \ln(\delta_t^{(k)}/\delta_t^{(\ell)})}{M}\Big) \\ &= P_r\Big(\big((\delta_t^{(k)})^2 - (\delta_t^{(\ell)})^2\big)|\tilde{r}_t|^2 + 2\sqrt{p_1 p_t}(\delta_t^{(k)})^2\, \Re(\tilde{r}_t \Delta S_t^{*(k,\ell)}) \\ &\quad + p_1 p_t (\delta_t^{(k)})^2 |\Delta S_t^{(k,\ell)}|^2 - \frac{2(\delta_t^{(k)}\delta_t^{(\ell)})^2 \ln(\delta_t^{(k)}/\delta_t^{(\ell)})}{M}\Big) < 0\Big). \end{aligned}$$

Now, we consider the following cases:

1. If $\delta_t^{(k)} = \delta_t^{(\ell)}$, then

$$\begin{aligned} P_r(S_t^{(k)} \to S_t^{(\ell)}) =& P_r\Big(2\sqrt{p_1 p_t}(\delta_t^{(k)})^2 \Re(\tilde{r}_t \Delta S_t^{*(k,\ell)}) + p_1 p_t (\delta_t^{(k)})^2 |\Delta S_t^{(k,\ell)}|^2 < 0\Big) \\ =& P_r\Big(2|\tilde{r}_t| \cos\vartheta + \sqrt{p_1 p_t}|\Delta S_t^{(k,\ell)}| < 0\Big), \end{aligned}$$

where $\vartheta = \arg(\tilde{r}_t) - \arg(\Delta S_t^{(k,\ell)})$.

- If $\vartheta \in [0, \frac{\pi}{2}] \cup [\frac{3\pi}{2}, 2\pi)$, then $\cos\vartheta \geq 0$, and we have

$$P_r(S_t^{(k)} \to S_t^{(\ell)}) \leq P_r(\sqrt{p_1 p_t}|\Delta S_t^{(k,\ell)}| < 0) = 0.$$

- Else if $\vartheta \in (\frac{\pi}{2}, \frac{3\pi}{2})$, then $\cos\vartheta < 0$, and we have

$$\begin{aligned} P_r(S_t^{(k)} \to S_t^{(\ell)}) &= P_r\Big(|\tilde{r}_t| > \frac{\sqrt{p_1 p_t}|\Delta S_t^{(k,\ell)}|}{-2\cos\vartheta}\Big) \\ &= \int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} \int_{\frac{\sqrt{p_1 p_t}|\Delta S_t^{(k,\ell)}|}{-2\cos\vartheta}}^{\infty} \frac{M}{\pi(\delta_t^{(k)})^2} \exp\left(-\frac{M|\tilde{r}_t|^2}{(\delta_t^{(k)})^2}\right) |\tilde{r}_t| d|\tilde{r}_t|\, d\vartheta \\ &= \frac{1}{2\pi} \int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} \exp\left(-\frac{M p_1 p_t |\Delta S_t^{(k,\ell)}|^2}{4(\delta_t^{(k)})^2 \cos^2\vartheta}\right) d\vartheta \\ &< \exp\Big(-\frac{M p_1 p_t |\Delta S_t^{(k,\ell)}|^2}{4(\delta_t^{(k)})^2}\Big). \end{aligned}$$

Hence, regardless of the angle between $\tilde{r}_t$ and $\Delta S_t^{(k,\ell)}$, we can always attain that $P_r(S_t^{(k)} \to S_t^{(\ell)}) < \exp\left(-\frac{M p_1 p_t |\Delta S_t^{(k,\ell)}|^2}{4(\delta_t^{(k)})^2}\right) = \exp\left(-\frac{M p_1 p_t |\Delta S_t^{(k,\ell)}|^2}{(\delta_t^{(k)}+\delta_t^{(\ell)})^2}\right)$.

2. If $\delta_t^{(k)} > \delta_t^{(\ell)}$, then

$$\begin{aligned} P_r(S_t^{(k)} \to S_t^{(\ell)}) &= P_r\Big(\big((\delta_t^{(k)})^2 - (\delta_t^{(\ell)})^2\big)|\tilde{r}_t|^2 + 2\sqrt{p_1 p_t}(\delta_t^{(k)})^2\, \Re(\tilde{r}_t \Delta S_t^{*(k,\ell)}) \\ &\quad + p_1 p_t (\delta_t^{(k)})^2 |\Delta S_t^{(k,\ell)}|^2 - \frac{2(\delta_t^{(k)}\delta_t^{(\ell)})^2 \ln(\delta_t^{(k)}/\delta_t^{(\ell)})}{M}\Big) < 0\Big). \\ &= P_r\Big(\Big|\tilde{r}_t + \frac{\sqrt{p_1 p_t}(\delta_t^{(k)})^2 \Delta S_t^{(k,\ell)}}{(\delta_t^{(k)})^2 - (\delta_t^{(\ell)})^2}\Big|^2 - \frac{p_1 p_t (\delta_t^{(k)})^4 |\Delta S_t^{(k,\ell)}|^2}{\Big((\delta_t^{(k)})^2 - (\delta_t^{(\ell)})^2\Big)^2} \\ &\quad + \frac{p_1 p_t (\delta_t^{(k)})^2 |\Delta S_t^{(k,\ell)}|^2}{(\delta_t^{(k)})^2 - (\delta_t^{(\ell)})^2} - \frac{2(\delta_t^{(k)}\delta_t^{(\ell)})^2 \ln(\delta_t^{(k)}/\delta_t^{(\ell)})}{M((\delta_t^{(k)})^2 - (\delta_t^{(\ell)})^2)}\Big) < 0\Big) \\ &= P_r\Big(\Big|\tilde{r}_t + \frac{\sqrt{p_1 p_t}(\delta_t^{(k)})^2 \Delta S_t^{(k,\ell)}}{(\delta_t^{(k)})^2 - (\delta_t^{(\ell)})^2}\Big|^2 < \frac{p_1 p_t (\delta_t^{(k)})^2 (\delta_t^{(\ell)})^2 |\Delta S_t^{(k,\ell)}|^2}{\Big((\delta_t^{(k)})^2 - (\delta_t^{(\ell)})^2\Big)^2} \\ &\quad + \frac{2(\delta_t^{(k)}\delta_t^{(\ell)})^2 \ln(\delta_t^{(k)}/\delta_t^{(\ell)})}{M((\delta_t^{(k)})^2 - (\delta_t^{(\ell)})^2)}\Big). \end{aligned}$$

Note that $M \gg \max_{\forall S_t^{(k)}, S_t^{(\ell)}} \frac{2(\delta_t^{(\ell)})^2 \ln(\delta_t^{(k)}/\ln\delta_t^{(\ell)})}{p_1 p_t |\Delta S_t^{(k,\ell)}|^2}$ and by our assumption, we have

$$\lim_{M\to\infty} \frac{\sqrt{p_1p_t}(\delta_t^{(k)})^2|\Delta S_t^{(k,\ell)}|}{(\delta_t^{(k)})^2-(\delta_t^{(\ell)})^2} - \sqrt{\frac{p_1p_t(\delta_t^{(k)})^2(\delta_t^{(\ell)})^2|\Delta S_t^{(k,\ell)}|^2}{\Big((\delta_t^{(k)})^2-(\delta_t^{(\ell)})^2\Big)^2} + \frac{2(\delta_t^{(k)}\delta_t^{(\ell)})^2(\ln(\delta_t^{(k)}/\delta_t^{(\ell)})}{M((\delta_t^{(k)})^2-(\delta_t^{(\ell)})^2)}}$$
$$= \frac{\sqrt{p_1p_t}\delta_t^{(k)}|\Delta S_t^{(k,\ell)}|(\delta_t^{(k)}-\delta_t^{(\ell)})}{(\delta_t^{(k)})^2-(\delta_t^{(\ell)})^2} = \frac{\sqrt{p_1p_t}\delta_t^{(k)}|\Delta S_t^{(k,\ell)}|}{\delta_t^{(k)}+\delta_t^{(\ell)}} > 0.$$

Now, using lemma 7 for case 1 produces

$$P_r(S_t^{(k)}\to S_t^{(\ell)}) < \exp\left(-\frac{p_1p_t(\delta_t^{(k)})^2|\Delta S_t^{(k,\ell)}|^2}{\Big(\delta_t^{(k)}+\delta_t^{(\ell)}\Big)^2\frac{(\delta_t^{(k)})^2}{M}}\right) = \exp\left(-\frac{Mp_1p_t|\Delta S_t^{(k,\ell)}|^2}{\Big(\delta_t^{(k)}+\delta_t^{(\ell)}\Big)^2}\right).$$

3. $\delta_t^{(k)} < \delta_t^{(\ell)}$. In this case, we have

$$P_r(S_t^{(k)}\to S_t^{(\ell)}) = P_r\Big(((\delta_t^{(\ell)})^2-(\delta_t^{(k)})^2)|\tilde{r}_t|^2 - 2\sqrt{p_1p_t}(\delta_t^{(k)})^2\Re(\tilde{r}_t\Delta S_t^{*(k,\ell)})$$
$$- p_1p_t(\delta_t^{(k)})^2|\Delta S_t^{(k,\ell)}|^2 + \frac{2(\delta_t^{(k)}\delta_t^{(\ell)})^2\ln(\delta_t^{(k)}/\delta_t^{(\ell)})}{M} > 0\Big)$$
$$= P_r\Big(\Big|\tilde{r}_t - \frac{\sqrt{p_1p_t}(\delta_t^{(k)})^2\Delta S_t^{(k,\ell)}}{(\delta_t^{(\ell)})^2-(\delta_t^{(k)})^2}\Big|^2 > \frac{p_1p_t(\delta_t^{(k)})^2(\delta_t^{(\ell)})^2|\Delta S_t^{(k,\ell)}|^2}{\Big((\delta_t^{(\ell)})^2-(\delta_t^{(k)})^2\Big)^2}$$
$$- \frac{2(\delta_t^{(k)}\delta_t^{(\ell)})^2\ln(\delta_t^{(k)}/\delta_t^{(\ell)})}{M\Big((\delta_t^{(\ell)})^2-(\delta_t^{(k)})^2\Big)}\Big).$$

Again for $M \gg \max_{\forall S_t^{(k)},S_t^{(\ell)}} \frac{2(\delta_t^{(\ell)})^2\ln(\delta_t^{(k)}/\ln\delta_t^{(\ell)})}{p_1p_t|\Delta S_t^{(k,\ell)}|^2}$, we have

$$\lim_{M\to\infty} \sqrt{\frac{p_1p_t(\delta_t^{(k)})^2(\delta_t^{(\ell)})^2|\Delta S_t^{(k,\ell)}|^2}{\Big((\delta_t^{(\ell)})^2-(\delta_t^{(k)})^2\Big)^2} - \frac{2(\delta_t^{(k)}\delta_t^{(\ell)})^2\ln(\delta_t^{(k)}/\delta_t^{(\ell)})}{M\Big((\delta_t^{(\ell)})^2-(\delta_t^{(k)})^2\Big)}\Big)} - \frac{\sqrt{p_1p_t}(\delta_t^{(k)})^2|\Delta S_t^{(k,\ell)}|}{(\delta_t^{(\ell)})^2-(\delta_t^{(k)})^2}$$
$$= \frac{\sqrt{p_1p_t}\delta_t^{(k)}|\Delta S_t^{(k,\ell)}|(\delta_t^{(\ell)}-\delta_t^{(k)})}{(\delta_t^{(\ell)})^2-(\delta_t^{(k)})^2} = \frac{\sqrt{p_1p_t}\delta_t^{(k)}|\Delta S_t^{(k,\ell)}|}{\delta_t^{(k)}+\delta_t^{(\ell)}} > 0.$$

Now, using lemma 7 for case 2 results in

$$P_r(S_t^{(k)} \to S_t^{(\ell)}) < \exp\left(-\frac{p_1 p_t (\delta_t^{(k)})^2 |\Delta S_t^{(k,\ell)}|^2}{(\delta_t^{(k)} + \delta_t^{(\ell)})^2 \frac{(\delta_t^{(k)})^2}{M}}\right) = \exp\left(-\frac{M p_1 p_t |\Delta S_t^{(k,\ell)}|^2}{\left(\delta_t^{(k)} + \delta_t^{(\ell)}\right)^2}\right).$$

Therefore, summing up the above all cases, we obtain $P_r(S_t^{(k)} \to S_t^{(\ell)}) < \exp\left(-\frac{M p_1 p_t |\Delta S_t^{(k,\ell)}|^2}{\left(\delta_t^{(k)} + \delta_t^{(\ell)}\right)^2}\right)$. This completes the proof of theorem 11. □

A.7 Proof of the Convexity of Problem 5

The Hessian matrix of $P_e(\mathbf{p})$ is given as

$$\bigtriangledown^2 P_e(\mathbf{p}) = \begin{bmatrix} \frac{\partial^2 P_e(\mathbf{p})}{\partial p_1^2} & \frac{\partial^2 P_e(\mathbf{p})}{\partial p_1 \partial p_2} & \frac{\partial^2 P_e(\mathbf{p})}{\partial p_1 \partial p_3} & \dots & \frac{\partial^2 P_e(\mathbf{p})}{\partial p_1 \partial p_N} \\ \frac{\partial^2 P_e(\mathbf{p})}{\partial p_2 \partial p_1} & \frac{\partial P_e(\mathbf{p})}{\partial p_2^2} & 0 & \dots & 0 \\ \frac{\partial^2 P_e(\mathbf{p})}{\partial p_3 \partial p_1} & 0 & \frac{\partial^2 P_e(\mathbf{p})}{\partial p_3^2} & \dots & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ \frac{\partial^2 P_e(\mathbf{p})}{\partial p_N \partial p_1} & 0 & 0 & \dots & \frac{\partial P_e(\mathbf{p})}{\partial p_N^2} \end{bmatrix}$$

We know that $P_e(\mathbf{p})$ is a convex function of $\mathbf{p}$ iff $\bigtriangledown^2 P_e(\mathbf{p}) \succeq \mathbf{0}$ for $\mathbf{p} > \mathbf{0}$ by the second-order conditions [47].

For our purpose, we now prefer to use another expression for Gaussian Q-function [139, 140] i.e.,

$$Q(x) = \frac{1}{\pi}\int_0^{\pi/2} \exp\left(-\frac{x^2}{2\sin^2\theta}\right) d\theta, \quad Q^2(x) = \frac{1}{\pi}\int_0^{\pi/4} \exp\left(-\frac{x^2}{2\sin^2\theta}\right) d\theta, \quad x \geq 0. \tag{A.17}$$

As a result, for $t = 2, \ldots, N$

$$\begin{aligned}
P_e(p_1, p_t) &\doteq \frac{1}{N-1}\sum_{t=2}^{N} 4\Big(1-\frac{1}{2^N}\Big)Q\left(\frac{\sqrt{M}}{2\sqrt{(N+\frac{\sigma^2}{p_1})(E_s+\frac{\sigma^2}{p_t})}}\right) \\
&\quad - 4\Big(1-\frac{1}{2^N}\Big)^2 Q^2\Big(\frac{\sqrt{M}}{2\sqrt{(N+\frac{\sigma^2}{p_1})(E_s+\frac{\sigma^2}{p_t})}}\Big) \\
&= \frac{1}{N-1}\sum_{t=2}^{N}\frac{4\Big(1-1/2^N\Big)}{\pi}\int_{\pi/4}^{\pi/2}\exp\left(-\frac{M}{8\sin^2\theta(N+\frac{\sigma^2}{p_1})(E_s+\frac{\sigma^2}{p_t})}\right)d\theta \\
&\quad + \frac{4\Big(1-1/2^N\Big)}{2^N\pi}\int_{0}^{\pi/4}\exp\left(-\frac{M}{8\sin^2\theta(N+\frac{\sigma^2}{p_1})(E_s+\frac{\sigma^2}{p_t})}\right)d\theta.
\end{aligned}$$

Then, we have

$$\begin{aligned}
\frac{\partial P_e(\mathbf{p})}{\partial p_1^2} &= \frac{4\Big(1-1/2^N\Big)}{(N-1)\pi}\int_{\pi/4}^{\pi/2}\exp\Big(-\frac{M}{8\sin^2\theta(N+\frac{\sigma^2}{p_1})(E_s+\frac{\sigma^2}{p_t})}\Big) \\
&\quad \times\left(\frac{M\sigma^2}{8\sin^2\theta(p_1N+\sigma^2)^2(E_s+\frac{\sigma^2}{p_t})}\right)\left(\frac{M\sigma^2}{8\sin^2\theta(p_1N+\sigma^2)^2(E_s+\frac{\sigma^2}{p_t})}+\frac{2N}{p_1N+\sigma^2}\right)d\theta \\
&\quad + \frac{4\Big(1-1/2^N\Big)}{2^N(N-1)\pi}\int_{0}^{\pi/4}\exp\Big(-\frac{M}{8\sin^2\theta(N+\frac{\sigma^2}{p_1})(E_s+\frac{\sigma^2}{p_t})}\Big) \\
&\quad \left(\frac{M\sigma^2}{8\sin^2\theta(p_1N+\sigma^2)^2(E_s+\frac{\sigma^2}{p_t})}\right)\left(\frac{M\sigma^2}{8\sin^2\theta(p_1N+\sigma^2)^2(E_s+\frac{\sigma^2}{p_t})}+\frac{2N}{p_1N+\sigma^2}\right)d\theta > 0.
\end{aligned}$$

For $t = 2, \ldots, N$, we also have

$$\frac{\partial P_e(\mathbf{p})}{\partial p_t^2} = \frac{4\Big(1 - 1/2^N\Big)}{(N-1)\pi} \int_{\pi/4}^{\pi/2} \exp\bigg(- \frac{M}{8\sin^2\theta(N + \frac{\sigma^2}{p_1})(E_s + \frac{\sigma^2}{p_t})}\bigg)$$
$$\bigg(\frac{M\sigma^2}{8\sin^2\theta(N + \frac{\sigma^2}{p_1})(p_t E_s + \sigma^2)^2}\bigg)\bigg(\frac{M\sigma^2}{8\sin^2\theta(N + \frac{\sigma^2}{p_1})(p_t E_s + \sigma^2)^2} + \frac{2E_s}{p_t E_s + \sigma^2}\bigg)\, d\theta$$
$$+ \frac{4\Big(1 - 1/2^N\Big)}{2^N(N-1)\pi} \int_{0}^{\pi/4} \exp\bigg(- \frac{M}{8\sin^2\theta(N + \frac{\sigma^2}{p_1})(E_s + \frac{\sigma^2}{p_t})}\bigg)$$
$$\bigg(\frac{M\sigma^2}{8\sin^2\theta(N + \frac{\sigma^2}{p_1})(p_t E_s + \sigma^2)^2}\bigg)\bigg(\frac{M\sigma^2}{8\sin^2\theta(N + \frac{\sigma^2}{p_1})(p_t E_s + \sigma^2)^2} + \frac{2E_s}{p_t E_s + \sigma^2}\bigg)\, d\theta > 0.$$

Also, we have

$$\frac{\partial P_e(\mathbf{p})}{\partial p_1 \partial p_t} = \frac{4\Big(1 - 1/2^N\Big)}{(N-1)\pi} \int_{\pi/4}^{\pi/2} \exp\bigg(- \frac{M}{8\sin^2\theta(N + \frac{\sigma^2}{p_1})(E_s + \frac{\sigma^2}{p_t})}\bigg)$$
$$\bigg(\frac{M\sigma^2}{8\sin^2\theta(Np_1 + \sigma^2)^2(E_s + \frac{\sigma^2}{p_t})}\bigg)\bigg(\frac{M\sigma^2}{8\sin^2\theta(N + \frac{\sigma^2}{p_1})(E_s p_t + \sigma^2)^2} + \frac{\sigma^2}{E_s p_t^2 + \sigma^2 p_t}\bigg)\, d\theta$$
$$+ \frac{4\Big(1 - 1/2^N\Big)}{2^N(N-1)\pi} \int_{0}^{\pi/4} \exp\bigg(- \frac{M}{8\sin^2\theta(N + \frac{\sigma^2}{p_1})(E_s + \frac{\sigma^2}{p_t})}\bigg)$$
$$\bigg(\frac{M\sigma^2}{8\sin^2\theta(Np_1 + \sigma^2)^2(E_s + \frac{\sigma^2}{p_t})}\bigg)\bigg(\frac{M\sigma^2}{8\sin^2\theta(N + \frac{\sigma^2}{p_1})(E_s p_t + \sigma^2)^2} + \frac{\sigma^2}{E_s p_t^2 + \sigma^2 p_t}\bigg)\, d\theta > 0.$$

Clearly, $\bigtriangledown^2 P_e(\mathbf{p}) \succeq \mathbf{0}$ for $\forall \mathbf{p} > \mathbf{0}$. This proves that $P_e(\mathbf{p})$ is a convex function of $\mathbf{p} > \mathbf{0}$. The constraint $\mathbf{p} \geq \mathbf{0}$ is a cone and $\mathbf{Ap} \leq \mathbf{b}$ defines a polyhedron which are both convex feasible regions. Hence the overall optimization problem is convex. This completes the proof. □

A.8 Lemma on the Quotient of Ordered Sequences

Let $m = \arg\min_{k=1,2,\ldots,N} \left\{\frac{a_k}{b_{\pi^*(k)}}\right\} = \arg\min_k \left\{\frac{a_k}{b_k}\right\}$. In other words, m is the index such that $q_m = \frac{a_m}{b_m} = \min_{k=1,2,\ldots,N} \left\{\frac{a_k}{b_k}\right\}$. Now, we want to show that $q_m = \max_{(\pi(1),\pi(2),\ldots,\pi(N))\in\mathcal{U}} \min_{k=1,2,\cdots,N} \left\{\frac{a_k}{b_{\pi(k)}}\right\}$. To that end, we divide $\mathcal{U}$ into two mutually exclusive subsets, i.e., $\mathcal{P} = \{(\pi(1),\pi(2),\ldots,\pi(N))|\pi(m) \neq m\}$ and $\mathcal{U}\setminus\mathcal{P} = \{(\pi(1),\pi(2),\ldots,\pi(N))|\pi(m) = m\}$. Consider the following cases:

- $(\pi'(1),\pi'(2),\ldots,\pi'(N)) \in \mathcal{P}$. In this case, there exists an $\ell \neq m$ such that $\pi'(\ell) = m$ and hence $b_{\pi'(\ell)} = b_m$. If $\ell < m$, then, we have $\frac{a_\ell}{b_{\pi'(\ell)}} = \frac{a_\ell}{b_m} \leq \frac{a_m}{b_m} = q_m$. If $\ell > m$, there exits an $n \leq m$ such that $\pi'(n) > m$ by the property of permutation. Then, we have $\frac{a_n}{b_{\pi'(n)}} \leq \frac{a_m}{b_{\pi'(n)}} \leq \frac{a_m}{b_m} = q_m$. Therefore, we conclude $\min_{k=1,2,\ldots,N} \left\{\frac{a_k}{b_{\pi'(k)}}\right\} \leq q_m$ for any $(\pi'(1),\pi'(2),\ldots,\pi'(N)) \in \mathcal{P}$. Or equivalently, $\max_{(\pi(1),\pi(2),\ldots,\pi(N))\in\mathcal{P}} \min_{k=1,2,\cdots,N} \left\{\frac{a_k}{b_{\pi(k)}}\right\} \leq q_m$.

- $(\pi'(1),\pi'(2),\ldots,\pi'(N)) \in \mathcal{U}\setminus\mathcal{P}$. In this case, $\pi'(m) = m$ and hence, we have $\min_{k=1,2,\cdots,N} \left\{\frac{a_k}{b_{\pi'(k)}}\right\} \leq \frac{a_m}{b_{\pi'(m)}} = \frac{a_m}{b_m} = q_m$. Therefore, $\max_{(\pi(1),\pi(2),\cdots,\pi(N))\in\mathcal{U}\setminus\mathcal{P}} \min_{k=1,2,\cdots,N} \left\{\frac{a_k}{b_{\pi(k)}}\right\} \leq q_m$.

In conclusion, we have $\max_{\pi\in\mathcal{U}} \min_{k=1,2,\cdots,N} \left\{\frac{a_k}{b_{\pi(k)}}\right\} \leq q_m$. In the following, we aim to prove that the equality is achievable for certain $(\pi(1),\pi(2),\ldots,\pi(K))$. By setting $(\pi(1),\pi(2),\cdots,\pi(N)) = (\pi^*(1),\pi^*(2),\ldots,\pi^*(N))$ and then, from the construction process above, we can find that for the given sequences $a_1 \leq a_2 \leq \cdots \leq a_N$ and $b_1 \leq b_2 \leq \cdots \leq b_N$, $\min_{k=1,2,\cdots,N} \left\{\frac{a_k}{b_{\pi^*(k)}}\right\} = \frac{a_m}{b_m} = q_m$. Hence, the equality is achievable for $(\pi^*(1),\pi^*(2),\cdots,\pi^*(N))$. This completes the proof. □

Bibliography

[1] C. E. Shannon, "Two-way communication channels," *Proc. 4th Berkeley Symp. Math. Statist. and Prob.*, vol. 1, pp. 611–644, 1961. 3

[2] E. van der Meulen, "A survey of muti-way channels in information theory," *IEEE Trans. Inf. Theory*, vol. 23, pp. 1–37, Jan. 1977. 3

[3] T. M. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, pp. 572–584, Sep. 1979. 3

[4] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity — Part I: System description," *IEEE Trans. Commun.*, vol. 51, pp. 1927–1938, Nov. 2003. 3

[5] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity — Part II: Implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, pp. 1939–1948, Nov. 2003. 3

[6] K. Azarian, H. E. Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4152–4157, Dec. 2005. 3

[7] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2415–2425, Oct. 2003. 3

[8] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2062–3080, Dec. 2004. 3

[9] R. U. Nabar, H. Bölcskei, and F. W. Kneubühler, "Fading relay channels: Performance limits and space-time signal design," *IEEE J. Sel. Areas Commun.*, vol. 22, pp. 1099–1109, Aug. 2004. 3

[10] W. Zhang and K. B. Y. Letaief, "Bandwidth efficient cooperative diversity for wireless networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM'07)*, pp. 2942–2946, 2007. 3

[11] B. Rankov and A. Wittneben, "Spectral efficient protocols for half-duplex fading relay channels," *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 379–389, Feb. 2007. 3

[12] Y. Chang and Y. Hua, "Diversity analysis of orthogonal space-time modulation for distributed wireless relays," in *Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, pp. 561–564, Apr. 2004. 3

[13] Y. Jing and B. Hassibi, "Distributed space-time coding in wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 3524–3536, Dec. 2006. 3

[14] S. Yang and J.-C. Belfiore, "Optimal space-time codes for the MIMO amplify-and-forward cooperative channel," *IEEE Trans. Inf. Theory*, vol. 53, pp. 647–663, Feb. 2007. 3

[15] G. Wang, J. K. Zhang, M. Amin, and K. M. Wong, "Nested distributed space-time encoding protocol for wireless networks with high energy efficiency," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 521–531, Feb. 2008. 3

[16] T. Cui, F. Gao, T. Ho, and A. Nallanathan, "Distributed space-time coding for two-way wireless relay networks," *IEEE Trans. Signal Process.*, vol. 57, pp. 658–671, Feb. 2009. 3

[17] L. Xiong and J.-K. Zhang, "Energy-efficient uniquely-factorable constellation designs for noncoherent SIMO channels," *IEEE Trans. Veh. Technol.*, vol. 61, pp. 2130–2144, Jun. 2012. 4, 19

[18] D. Xia, J. K. Zhang, and S. Dumitrescu, "Energy-efficient full diversity collaborative unitary space-time block code designs via unique factorization of signals," *IEEE Trans. Inf. Theory*, vol. 59, pp. 1678–1703, Mar. 2013. 4, 19, 20, 21, 24

[19] L. Zhou, J.-K. Zhang, and K. M. Wong, "A novel signaling scheme for blind unique identification of Alamouti space-time block coded channel," *IEEE Trans. Signal Processing*, vol. 55, pp. 2570–2582, June 2007. 4

[20] J.-K. Zhang and W.-K. Ma, "Full diversity blind Alamouti space-time block codes for unique identification of flat fading channels," *IEEE Trans. Signal Processing*, vol. 57, pp. 635–644, Feb. 2008. 4

[21] J.-K. Zhang, F. Huang, and S. Ma, "Full diversity blind space-time block codes," *IEEE Trans. Inform. Theory*, vol. 57, Sept. 2011. 4

[22] F.-K. Gong, J.-K. Zhang, Y.-J. Zhu, and J.-H. Ge, "Energy-efficient collaborative alamouti codes," *IEEE Wireless Comm. Lett.*, vol. 1, pp. 512–515, Oct. 2012. 4

[23] F.-K. Gong, J.-K. Zhang, and J.-H. Ge, "Distributed concatenated alamouti codes for two-way relaying networks," *Wireless Communications Letters, IEEE*, vol. 1, pp. 197 –200, june 2012. 4, 40

[24] I. Telatar, "Capacity of multiple antenna Gaussian channels," *Europ. Trans. Telecommu.*, vol. 10, pp. 585–595, Nov.-Dec. 1999. 5, 10, 62

[25] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1073–1096, May 2003. 5

[26] S. Jin, X. Gao, and X. You, "On the ergodic capacity of rank-1 Ricean-fading MIMO channels," *IEEE Trans. Inf. Theory*, vol. 53, pp. 502–517, Feb. 2007. 5

[27] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, pp. 40–60, Jan. 2013. 5, 14

[28] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, pp. 186–195, Feb. 2014. 5, 14

[29] C. Shepard, H. Yu, N. Anand, L. E. Li, T. L. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *Proc. ACM Int. Conf. Mobile Computing, Networking, MobiCom'12*, (Istanbul, Turkey), Aug. 2012. 5

[30] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," *IEEE J. Sel. Areas Commun.*, vol. 31, pp. 160–171, Feb. 2013. 5

[31] C.-N. Chuah, D. Tse, J. Kahn, and R. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, pp. 637–650, Mar. 2002. 5

[32] V. A. Marcenko and L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices," *Math USSR-Sbornik*, vol. 1, pp. 457–483, 1967. 5

[33] X. Mestre, J. Fonollosa, and A. Pages-Zamora, "Capacity of MIMO channels: asymptotic evaluation under correlated fading," *IEEE J. Sel. Areas Commun.*, vol. 21, pp. 829–838, Jun. 2003. 5, 6

[34] A. Moustakas, S. Simon, and A. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: a (not so) large N analysis," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2545–2561, Oct. 2003. 5

[35] M. McKay, I. Collings, and A. Tulino, "Achievable sum rate of MIMO MMSE receivers: A general analytic framework," *IEEE Trans. Inf. Theory*, vol. 56, pp. 396–410, Jan. 2010. 5, 6

[36] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, "A new approach for mutual information analysis of large dimensional multi-antenna channels," *IEEE Trans. Inf. Theory*, vol. 54, pp. 3987–4004, Sep. 2008. 5

[37] D. shan Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 502–513, Mar. 2000. 5, 62

[38] S. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, pp. 369–371, Sep. 2001. 5, 62, 76, 103, 112, 138

[39] A. Moustakas and P. Kazakopoulos, "SINR statistics of correlated MIMO linear receivers," *IEEE Trans. Inf. Theory*, vol. 59, pp. 6490–6500, Oct. 2013. 6

[40] M. Kiessling and J. Speidel, "Statistical prefilter design for MIMO ZF and MMSE receivers based on majorization theory," in *Int. Conf. Acoust., Speech, Signal Process. (ICASSP'04)*, (Montreal, Canada), May 2004. 6, 7, 58, 61, 106, 108, 109, 113

[41] S. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1165–1175, Jul. 2004. 6

[42] A. L. Moustakas, "Communication through a diffusive medium: Coherence and capacity," *Science*, vol. 287, pp. 287–290, Jan. 2000. 6

[43] T.-T. Liu, J.-K. Zhang, and K.-M. Wong, "Optimal precoder design for correlated MIMO communication systems using zero-forcing decision feedback equalization," *IEEE Trans. Signal Process.*, vol. 57, pp. 3600–3612, Sep. 2009. 6

[44] R. M. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006. 6, 63, 79, 114, 184

[45] J. M. Cioffi and G. D. Forney, *Generalized decision–feedback equalization packet transmission with ISI and Gaussain noise*. Kluwer, Boston: Communications, Computation, Control and Signal Processing, 1997. 7

[46] T.-T. Liu, J.-K. Zhang, and K.-M. Wong, "Optimal precoder design for correlated MIMO communication systems using zero-forcing decision feedback equalization," *IEEE Trans. Signal Processing*, vol. 57, pp. 3600–3612, Sept. 2009. 7, 99, 106, 109, 113

[47] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge Univ. Press, 2004. 8, 98, 164, 202

[48] J.-K. Zhang, A. Kavčic, and K. M. Wong, "Equal-diagonal QR decomposition and its application to precoder design for successive cancellation detection," *IEEE Trans. Inform. Theory*, pp. 154–171, Jan. 2005. 8, 83, 100, 101

[49] T. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. 18, pp. 2–14, Jan. 1972. 9

[50] P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inf. Theory*, vol. 19, pp. 197–207, Mar. 1973. 9

[51] R. G. Gallager, "Coding and capacity for degraded broadcast channels," *Probl. Peredachi Inf.*, vol. 10, pp. 185–193, Oct. 1974. 9

[52] E. van der Meulen, "A survey of multi-way channels in information theory: 1961-1976," *IEEE Trans. Inf. Theory*, vol. 23, pp. 1–37, Jan. 1977. 9

[53] T. Cover, "Comments on broadcast channels," *IRE Trans. Inf. Theory*, vol. 44, pp. 2524–2530, Oct. 1998. 9

[54] A. E. Gamal and Y.-H. Kim, *Network Information Theory.* Cambridge Univ. Press, 2012. 9, 16

[55] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antenna," *Wireless Personal Commun.*, vol. 6, pp. 311–335, Mar. 1998. 10

[56] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1691–1706, Jul. 2003. 10

[57] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, pp. 439–441, May 1983. 10

[58] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2658–2668, Oct. 2003. 10

[59] P. Viswanath and D. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1912–1921, Aug. 2003. 10

[60] W. Yu and J. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, pp. 1875–1892, Sep. 2004. 10

[61] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3936–3964, Sep. 2006. 10

[62] N. Deshpande and B. S. Rajan, "Constellation constrained capacity of two-user broadcast channels," in *Proc. IEEE Global Commun. Conf. (GLOBECOM'09)*, pp. 1–6, Nov. 2009. 10

[63] R. Ghaffar and R. Knopp, "Near optimal linear precoder for multiuser MIMO for discrete alphabets," in *Proc. IEEE Int. Conf. Commun. (ICC'10)*, pp. 1–5, May 2010. 10

[64] C. Xiao, Y. R. Zheng, and Z. Ding, "Globally optimal linear precoders for finite alphabet signals over complex vector gaussian channels," *IEEE Trans. Signal Process.*, vol. 59, pp. 3301–3314, Jul. 2011. 10

[65] Y. Wu, M. Wang, C. Xiao, Z. Ding, and X. Gao, "Linear precoding for MIMO broadcast channels with finite-alphabet constraints," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 2906–2920, Aug. 2012. 10

[66] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electron. Lett.*, vol. 7, pp. 138–139, Mar. 1971. 10

[67] C. Windpassinger, R. Fischer, T. Vencel, and J. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1305–1316, Jul. 2004. 10

[68] A. Garcia-Rodriguez and C. Masouros, "Power-efficient Tomlinson-Harashima precoding for the downlink of multi-user MISO systems," *IEEE Trans. Commun.*, vol. 62, pp. 1884–1896, Jun. 2014. 10

[69] H. Viswanathan, S. Venkatesan, and H. Huang, "Downlink capacity evaluation of cellular networks with known-interference cancellation," *IEEE J. Sel. Areas Commun.*, vol. 21, pp. 802–811, Jun. 2003. 10

[70] W. Yu, D. Varodayan, and J. Cioffi, "Trellis and convolutional precoding for transmitter-based interference presubtraction," *IEEE Trans. Commun.*, vol. 53, pp. 1220–1230, Jul. 2005. 10

[71] F. Rashid-Farrokhi, K. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 1437–1450, Oct. 1998. 11

[72] M. Schubert and H. Boche, "Joint 'dirty paper' pre-coding and downlink beamforming," in *Proc. IEEE 7th International Symposium on Spread Spectrum Techniques and Applications,*, vol. 2, pp. 536–540 vol.2, 2002. 11

[73] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, pp. 18–28, Jan. 2004. 11

[74] C. Peel, B. Hochwald, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, pp. 195–202, Jan. 2005. 11

[75] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1711–1721, May 2007. 11, 142

[76] A. Wiesel, Y. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 56, pp. 4409–4418, Sep. 2008. 11, 122

[77] Q. Spencer, A. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, pp. 461–471, Feb. 2004. 11

[78] D. Samardzija and N. Mandayam, "Multiple antenna transmitter optimization schemes for multiuser systems," in *Proc. IEEE 58th Vehi. Tech. Conf. (VTC Fall'03)*, vol. 1, pp. 399–403 Vol.1, Oct. 2003. 11

[79] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, pp. 49–60, Jan. 1981. 11

[80] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, pp. 3425–3441, Aug. 2008. 11, 16

[81] G. Bresler, A. Parekh, and D. N. C. Tse, "The approximate capacity of the many-to-one and one-to-many Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 56, pp. 4566–4592, Sep. 2010. 11, 12, 17

[82] Y. Wu, S. S. Shitz, and S. Verdú, "Information dimension and the degrees of freedom of the interference channel," *IEEE Trans. Inf. Theory*, vol. 61, pp. 256–279, Jan. 2015. 11

[83] Y. Liu, Y. Zhang, R. Yu, and S. Xie, "Integrated energy and spectrum harvesting for 5G wireless communications," *IEEE Network*, vol. 29, pp. 75–81, May 2015. 11

[84] G. Zheng, I. Krikidis, C. Masouros, S. Timotheou, D. A. Toumpakaris, and Z. Ding, "Rethinking the role of interference in wireless networks," *IEEE Commun. Mag.*, vol. 52, pp. 152–158, Nov. 2014. 11

[85] C. Masouros, T. Ratnarajah, M. Sellathurai, C. B. Papadias, and A. K. Shukla, "Known interference in the cellular downlink: a performance limiting factor or a source of green signal power?," *IEEE Commun. Mag.*, vol. 51, pp. 162–171, Oct. 2013. 11

[86] C. Masouros and E. Alsusa, "Dynamic linear precoding for the exploitation of known interference in MIMO broadcast systems," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 1396–1404, Mar. 2009. 11

[87] C. Masouros and G. Zheng, "Exploiting known interference as green signal power for downlink beamforming optimization," *IEEE Trans. Signal Process.*, vol. 63, pp. 3628–3640, Jul. 2015. 11

[88] C. Masouros and T. Ratnarajah, "Interference as a source of green signal power in cognitive relay assisted co-existing MIMO wireless transmissions," *IEEE Trans. Commun.*, vol. 60, pp. 525–536, Feb. 2012. 11

[89] Y. Zeng and R. Zhang, "Optimized training design for wireless energy transfer," *IEEE Trans. Commun.*, vol. 63, pp. 536–550, Feb. 2015. 11

[90] X. Lu, P. Wang, D. Niyato, D. I. Kim, and Z. Han, "Wireless networks with RF energy harvesting: A contemporary survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 757–789, 2015. 11

[91] I. Krikidis, S. Timotheou, S. Nikolaou, G. Zheng, D. W. K. Ng, and R. Schober, "Simultaneous wireless information and power transfer in modern communication systems," *IEEE Commun. Mag.*, vol. 52, pp. 104–110, Nov. 2014. 11

[92] D. W. K. Ng, E. S. Lo, and R. Schober, "Wireless information and power transfer: Energy efficiency optimization in OFDMA systems," *IEEE Trans. Wireless Commun.*, vol. 12, pp. 6352–6370, Dec. 2013. 11

[93] R. Zhang and C. K. Ho, "MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 12, pp. 1989–2001, May 2013. 11

[94] K. Huang and V. K. N. Lau, "Enabling wireless power transfer in cellular networks: Architecture, modeling and deployment," *IEEE Trans. Wireless Commun.*, vol. 13, pp. 902–912, Feb. 2014. 11

[95] Z. Ding, I. Krikidis, B. Sharif, and H. V. Poor, "Wireless information and power transfer in cooperative networks with spatially random relays," *IEEE Trans. Wireless Commun.*, vol. 13, pp. 4440–4453, Aug. 2014. 11

[96] E. Hossain, M. Rasti, H. Tabassum, and A. Abdelnasser, "Evolution toward 5G multi-tier cellular wireless networks: An interference management perspective," *IEEE Wireless Commun.*, vol. 21, pp. 118–127, Jun. 2014. 11

[97] G. Zheng, I. Krikidis, J. Li, A. P. Petropulu, and B. Ottersten, "Improving physical layer secrecy using full-duplex jamming receivers," *IEEE Trans. Signal Process.*, vol. 61, pp. 4962–4974, Oct. 2013. 12

[98] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, pp. 1872–1905, Apr. 2011. 12, 17

[99] T. Kasami and S. Lin, "Coding for a multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 22, pp. 129–137, Mar. 1976. 12, 27

[100] T. Kasami and S. Lin, "Bounds on the achievable rates of block coding for a memoryless multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 24, pp. 187–197, Mar. 1978. 12, 27

[101] P. van den Braak and H. van Tilborg, "A family of good uniquely decodable code pairs for the two-access binary adder channel," *IEEE Trans. Inf. Theory*, vol. 31, pp. 3–9, Jan. 1985. 12, 27

[102] R. Ahlswede and V. Balakirsky, "Construction of uniquely decodable codes for the two-user binary adder channel," *IEEE Trans. Inf. Theory*, vol. 45, pp. 326–330, Jan. 1999. 12, 27

[103] P. Chevillat, "N-user trellis coding for a class of multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 27, pp. 114–120, Jan. 1981. 12, 27

[104] H. Murata and S. Yoshida, "Trellis-coded cochannel interference canceller for microcellular radio," *IEEE Trans. Commun.*, vol. 45, pp. 1088–1094, Sep. 1997. 12, 27

[105] Y. Li, H. Murata, and S. Yoshida, "Coding for multi-user detection in interference channel," in *Proc. IEEE Global Commun. Conf. (GLOBECOM'98)*, vol. 6, pp. 3596–3601 vol.6, 1998. 12, 27

[106] W. Zhang, C. D'Amours, and A. Yongacoglu, "Trellis coded modulation design for multi-user systems on AWGN channels," in *Proc. IEEE 59th Vehi. Tech. Conf. (VTC Spring'04)*, vol. 3, pp. 1722–1726 Vol.3, May 2004. 12, 27

[107] K. Ramchandran, A. Ortega, K. Uz, and M. Vetterli, "Multiresolution broadcast for digital HDTV using joint source/channel coding," *IEEE J. Sel. Areas Commun.*, vol. 11, pp. 6–23, Jan. 1993. 12, 27

[108] L.-F. Wei, "Coded modulation with unequal error protection," *IEEE Trans. Commun.*, vol. 41, pp. 1439–1449, Oct. 1993. 12, 27

[109] J. Hossain, M.-S. Alouini, and V. Bhargava, "Multi-user opportunistic scheduling using power controlled hierarchical constellations," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1581–1586, May 2007. 12, 27

[110] D. Malladi, "Hierarchical modulation for communication channels in single-carrier frequency division multiple access," Sep. 2012. US Patent 8,259,848. 12, 27

[111] J. Harshan and B. Rajan, "On two-user Gaussian multiple access channels with finite input constellations," *IEEE Trans. Inf. Theory*, vol. 57, pp. 1299–1327, Mar. 2011. 12, 27, 179

[112] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (noma) for cellular future radio access," in *Proc. IEEE 77th Vehi. Tech. Conf. (VTC Spring'13)*, pp. 1–5, June 2013. 14

[113] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C. I, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *CoRR*, vol. abs/1511.08610, 2015. 14

[114] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 3590–3600, Nov. 2010. 14, 15

[115] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?," *IEEE J. Sel. Areas Commun.*, vol. 32, pp. 1065–1082, Jun. 2014. 14

[116] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, pp. 74–80, Feb. 2014. 14

[117] L. Lu, G. Li, A. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, pp. 742–758, Oct. 2014. 14, 15

[118] O. Elijah, C. Leow, T. Rahman, S. Nunoo, and S. Iliya, "A comprehensive survey of pilot contamination in massive MIMO-5G system," *IEEE Commun. Surveys Tuts.*, vol. PP, no. 99, pp. 1–1, 2015. 14

[119] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang, "Large-scale measurement and characterization of cellular machine-to-machine traffic," *IEEE/ACM Trans. Netw.*, vol. 21, pp. 1960–1973, Dec. 2013. 15

[120] M. Hasan, E. Hossain, and D. Niyato, "Random access for machine-to-machine communication in LTE-advanced networks: issues and approaches," *IEEE Commun. Mag.*, vol. 51, pp. 86–93, Jun. 2013. 15

[121] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, 2015. 15

[122] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, pp. 951–963, Apr. 2003. 15

[123] 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, pp. 884–893, Mar. 2006. 15

[124] T. L. Marzetta, "How much training is required for multiuser MIMO?," in *Proc. Asilomar Conf. Signals Systems Comput.*, pp. 359–363, Oct. 2006. 15, 17

[125] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *Proc. IEEE Int. Symp. Inf. Theory*, pp. 1137–1141, Jul. 2012. 15

[126] W. C. Y. Lee, *Mobile Communications Engineering: Theory and Applications*. McGraw-Hill Inc.,US, 2nd ed., 1998. 15

[127] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005. 16, 125

[128] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. Poor, *MIMO Wireless Communications*. Cambridge Univ. Press, 2007. 16

[129] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 2nd ed., 2006. 16

[130] The 5G-Infrastructure-PPP, "5G Vision," *https://5g-ppp.eu/wp-content/uploads/2015/02/5G-Vision-Brochure-v1.pdf*, 2015. 17

[131] Z. Dong, Y.-Y. Zhang, and J.-K. Zhang, "Quadrature amplitude modulation division for multiuser MISO broadcast channels," *accepted for publication by IEEE J. Sel. Topics Signal Process.*, 2016. 19

[132] L.-K. Hua, *Introduction to Number Theory*. Berlin ; New York: Springer-Verlag, 1982. 23

[133] E. Schimmerling, *A Course on Set Theory*. Cambridge Univ. Press, 2011. 28

[134] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. San Diego: CA Academic Press, 7th ed., 2007. 41, 67

[135] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005. 44, 170

[136] 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. Sym. on Signal, Systems, and Electron. (ISSSE'98)*, pp. 586–603, Sep. 1998. 48

[137] T. W. Anderson, *An introduction to multivariate statstical analysis*. New York: John Wiley & Sons, INC, 1971. 50, 52

[138] R. J. Muirhead, *Aspects of multivariate statistical theory*. New York: John Wiley & Sons, INC, 1982. 50, 52, 83, 88

[139] J. W. Craig, "A new, simple, and exact result for calculating the probability of error for two-dimensional signal constellations," in *Proc. IEEE Milit. Commun. Conf. (MILCOM'91)*, pp. 571–575, Oct. 1991. 52, 202

[140] M. K. Simon and M.-S. Alouini, "A unified approach to the perforance analysis of digital communication over generalized fading channels," *Proc. IEEE*, vol. 86, pp. 1860–1877, Sep. 1998. 52, 54, 202

[141] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, MA: Cambridge University Press, 1985. 57, 168

[142] Y. Ding, T. Davidson, Z.-Q. Luo, and K. M. Wong, "Minimum BER block precoders for zero-forcing equalization," *IEEE Trans. Signal Process.*, vol. 51, pp. 2410–2423, Sep. 2003. 58, 61

[143] D. Palomar, J. Cioffi, and M.-A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, pp. 2381–2401, Sep. 2003. 58, 90

[144] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, INC, 1991. 59, 103, 104

[145] A. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*. New York: Academic Press, 1979. 59, 90, 91

[146] R. G. Bartle, *The Elements of Integration and Lebesgue Measure*. New York: John Wiley & Sons, INC, 1995. 65

[147] M. Kac, W. Murdock, and G. Szegö, "On the eigenvalues of certain Hermitian forms," *J. Rat. Mech. and Anal. 2*, pp. 787–800, 1953. 66

[148] M. Dow, "Explicit inverses of toeplitz and associated matrices," *ANZIAM J.*, vol. 44, pp. E185–E215, Jan. 2003. 66, 68

[149] W.-C. Yueh, "Eigenvalues of several tridiagonal matrices," in *Applied Mathematics E-notes*, pp. 5–66, 2005. 70

[150] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P.-W. Wolniansky, "Detection algorithm and initial laboratory results uisng V-BLAST space-time communication architecture," *Electron. Lett.*, vol. 35, pp. 14–15, Jan. 1999. 82, 106

[151] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore: The Johns Hopkins University Press, 1983. 83, 87

[152] T. Guess, "Optimal sequences for CDMA with decision-feedback receivers," *IEEE Trans. Inform. Theory*, vol. 49, pp. 886–900, Apr. 2003. 90, 91, 99

[153] Y. Jiang, J. Li, and W. W. Hager, "Uniform channel decomposition for MIMO communications," *IEEE Trans. Signal Processing*, vol. 53, pp. 4283–4294, Nov. 2005. 90

[154] M. B. Shenouda and T. N. Davidson, "A framework for designing MIMO systems with decision feedback equalization or Tomlison Harashima precoding," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, (Honolulu, USA), pp. 209–212, Apr. 2007. 90

[155] A. A. Damico, "Tomlinson-Harashima precoding in MIMO systems: A unified approach to transceiver optimization based on multiplicative Schur-convexity," *IEEE Trans. Signal Processing*, vol. 56, pp. 3662–3677, Aug. 2008. 90

[156] E. Jorswieck and H. Boche, "Majorization and matrix monotone wireless communications," *Foundations and trends in communications and information theory*, vol. 3, pp. 553–701, June 2007. 90

[157] Y. Jiang, W. W. Hager, and J. Li, "The generalized triangular decomposition," *Math. Comp.*, Oct. 2007. 99

[158] J.-K. Zhang and K. Wong, "Fast QRS decomposition of matrices and its applications to numerical optimization," in *http://www.ece.mcmaster.ca/jkzhang/*, 2005. 99

[159] F. Xu, T. N. Davidson, J.-K. Zhang, and K. M. Wong, "Design of block transceivers with decision feedback detection," *IEEE Trans. Signal Process.*, vol. 54, pp. 964–978, Mar. 2006. 100

[160] Y.-H. Kim, "Feedback capacity of stationary gaussian channels," *IEEE Trans. Inform. Theory*, vol. 56, pp. 57–85, Jan 2010. 103

[161] N. Sidiropoulos, T. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, pp. 2239–2251, Jun. 2006. 121

[162] J. Choi, "Minimum power multicast beamforming with superposition coding for multiresolution broadcast and application to NOMA systems," *IEEE Trans. Commun.*, vol. 63, pp. 791–800, Mar. 2015. 121

[163] Y. Liang, V. Veeravalli, and H. Poor, "Resource allocation for wireless fading relay channels: Max-Min solution," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3432–3453, Oct. 2007. 121

[164] K. Scharnhorst, "Angles in complex vector spaces," *Acta Applicandae Mathematica*, vol. 69, pp. 95–103, Oct. 2001. 123

[165] E. Karipidis, N. Sidiropoulos, and Z.-Q. Luo, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," *IEEE Trans. Signal Process.*, vol. 56, pp. 1268–1279, Mar. 2008. 131

[166] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Weighted fair multicast multigroup beamforming under per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 62, pp. 5132–5142, Oct. 2014. 131

[167] P. Patcharamaneepakorn, S. Armour, and A. Doufexi, "On the equivalence between SLNR and MMSE precoding schemes with single-antenna receivers," *IEEE Commun. Lett.*, vol. 16, pp. 1034–1037, Jul. 2012. 143

[168] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 2640–2651, Aug. 2011. 149

[169] Z. Dong, Y. Y. Zhang, J. K. Zhang, and X. C. Gao, "Quadrature amplitude modulation division for multiuser MISO broadcast channels," *IEEE J. Sel. Topics Signal Process.*, no. 99, DOI:10.1109/JSTSP.2016.2607684, 2016. 151

[170] S. Shamai and I. Bar-David, "The capacity of average and peak-power-limited quadrature Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 41, pp. 1060–1071, Jul. 1995. 164

[171] Y. Li and K. M. Wong, "Riemannian distances for signal classification by power spectral density," *IEEE J. Sel. Topics Signal Process.*, vol. 7, pp. 655–669, Aug. 2013. 168

[172] A. G. Armada and L. Hanzo, "A non-coherent multi-user large scale SIMO system relaying on M-ary DPSK," in *Proc. IEEE ICC' 15*, pp. 2517–2522, Jun. 2015. 178, 179, 180

[173] J. A. Tague and C. I. Caldwell, "Expectations of useful complex wishart forms," *Multidimensional Systems Signal Processing*, vol. 5, pp. 263–279, Jul. 1994. 196