SIGNAL DESIGN FOR MULTIPLE-ANTENNA SYSTEMS AND WIRELESS NETWORKS

by

Xiaofei Song

B.E., Central South University, 2000

M.E., Shandong University, 2003

M.S., University of Pittsburgh, 2005

Submitted to the Graduate Faculty of

the School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2007

UNIVERSITY OF PITTSBURGH

SCHOOL OF ENGINEERING

This dissertation was presented

by

Xiaofei Song

It was defended on

July 23, 2007

and approved by

Heung-No Lee, Assistant Professor, Electrical and Computer Engineering Department

J. Robert Boston, Professor, Electrical and Computer Engineering Department

Zhi-Hong Mao, Assistant Professor, Electrical and Computer Engineering Department

Mickle H. Marlin, Professor, Electrical and Computer Engineering Department

Xinfu Chen, Professor, Department of Mathematics

Dissertation Director: Heung-No Lee, Assistant Professor, Electrical and Computer Engineering Department

SIGNAL DESIGN FOR MULTIPLE-ANTENNA SYSTEMS AND WIRELESS NETWORKS

Xiaofei Song, PhD

University of Pittsburgh, 2007

This dissertation is concerned with the signal design problems for Multiple Input and Multiple Output (MIMO) antenna systems and wireless networks. Three related but distinct problems are considered.

The first problem considered is the design of space time codes for MIMO systems in the case when neither the transmitter nor the receiver knows the channel. We present the theoretical concept of communicating over block fading channel using Layered Unitary Space Time Codes (LUSTC), where the input signal is formed as a product of a series of unitary matrices with corresponding dimensionality. We show the channel capacity using isotropically distributed (i.d.) input signaling and optimal decoding can be achieved by layered i.d. signaling scheme along with a low complexity successive decoding. The closed form layered channel capacity is obtained, which serves as a design guideline for practical LUSTC. In the design of LUSTC, a successive design method is applied to leverage the problem of optimizing over lots of parameters.

The feedback of channel state information (CSI) to the transmitter in MIMO systems is known to increase the forward channel capacity. A suboptimal power allocation scheme for MIMO systems is then proposed for limited rate feedback of CSI. We find that the capacity loss of this simple scheme is rather small compared to the optimal water-filling solution. This knowledge is applied for the design of the feedback codebook. In the codebook design, a generalized Lloyd algorithm is employed, in which the computation of the centroid is formulated as an optimization problem

and solved optimally. Numerical results show that the proposed codebook design outperforms the existing algorithms in the literature.

While it is not feasible to deploy multiple antennas in a wireless node due to the space limitation, user cooperation is an alternative to increase performance of the wireless networks. To this end, a coded user cooperation scheme is considered in the dissertation, which is shown to be equivalent to a coding scheme with the encoding done in a distributive manner. Utilizing the coding theoretic bound and simulation results, we show that the coded user cooperation scheme has great advantage over the non-cooperative scheme.

TABLE OF CONTENTS

LIST OF FIGURES

PREFACE

I would like to express my gratitude to my advisor, Professor Heung-No Lee, for his tremendous support and encouragement. His expertise and experience proved invaluable in the development of this dissertation.

I am honored by having in my committee scientists of highest caliber. I want to thank Professor Mickle Marlin, Professor Zhi-Hong Mao, Professor Robert Boston and Professor Xinfu Chen for serving in my committee and offering important help to my research work.

I am indebted to many people who have helped me along the way of my study. Special thanks go to Dr. Michael McCloud who worked closely with me for 8 months. I am also grateful to Professor Marwan Simaan and Professor Mahmoud El Nokali for their help and support. I would also like to thank Sandy for all her assistance on my research works.

Working in Communications Research Lab was a very enjoyable process, I want to thank my labmates for providing a stimulating and fun environment where I can learn and grow.

My friends and family, wherever they are, have always been the most important part of my life. I want to thank them all, particularly Jialan, for their love and support.

I dedicate this dissertation to my parents.

1.0 INTRODUCTION

1.1 WIRELESS COMMUNICATIONS CHALLENGES

The past decade has witnessed a tremendous surge in wireless communications industry. The second generation of cellular network has achieved huge success in providing digital voice service, and wireless network has also become indispensable for millions of people's requirements in data access. With dazzling new products and services emerging based upon foundations of wireless communications, wireless has and will continue to impact people's daily life in many ways.

The widespread acceptance of wireless technologies has in return triggered a huge demand for increased data rate and quality of service. The quality of service and data rate provided by wireless systems though have increased rapidly over the past years, they are still not comparable to their wireline counterparts. However, people expect to "see" the same quality of services anywhere anytime even in a wireless networks. The customers' desire for seamless and high quality connectivity has driven wireless engineers to seek all possible means to increase the capacity and quality of wireless network.

Two possible ways of increasing the throughput are to use more bandwidth and more transmit power. A celebrated result in Shannon's 1948 paper states that the capacity of an Additive White Gaussian Noise (AWGN) channel is given by:

$$C = B_w \log(1 + \frac{P}{B_w N_0}),$$

where N_0 is the power spectral density, B_w is the total available bandwidth and P is the transmit power. This is the fundamental limit in throughput of reliable communications over the AWGN

channel, and which can be used to understand other types of channels, in particular, wireless environment.

While both approaches are possible to improve the throughput in a point to point wireless channel, they are not desired in a wireless network. The spectrum is a scarce and expensive resource and the available bandwidth for most applications, e.g. cellular system, is fixed and limited. Increasing the bandwidth for one user reduces other users' data rate within the network. On the other hand, as the capacity scales only logarithmatically with the transmit power, increasing the transmit power is inefficient to improve the throughput and reduces the battery life of the mobile devices. Moreover, the overall throughput of some wireless systems, e.g. CDMA cellular networks, is interference limited and boosting up the transmit power for one user will pose strong interference to the other users. From the perspective of the entire system, increasing the transmit power can not efficiently improve the overall throughput of the network. Therefore, it is of interests to develop approaches to enhance the throughput without increasing the bandwidth or the transmit power.

1.2 MULTIPATH FADING AND MULTIPLE ANTENNAS SYSTEM

The limitation in the throughput for wireless communication is largely attributed to the physical channel. Due to the effect of physical environment on radio signal propagation such as reflection, refraction, and scattering, a transmitted signal will travel along different paths to arrive at the intended receiver. The signals from different paths have different phase shifts and signal strengthes. They can add up constructively or destructively, which causes the overall signal strength to go up and down. This random fluctuation in signal level, known as fading, severely affects the quality and reliability of wireless communication.

Over the years, multipath induced fading has been widely considered to be an impairment to wireless communications and wireless engineers were seeking various approaches to compensate or eliminate the fading effect. One of the means to mitigate multipath fading and to improve the link reliability is to deploy multiple receive antennas. If these antennas are put sufficiently apart,

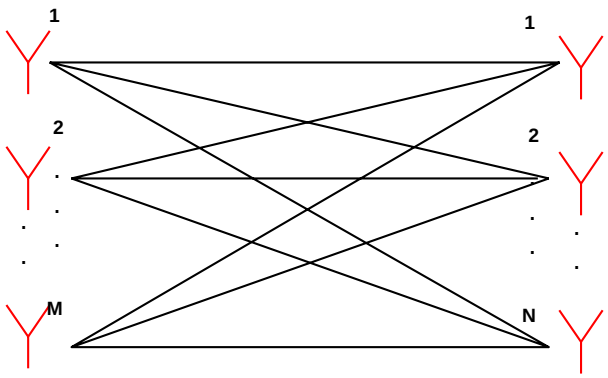

Figure 1: Multiple input and multiple output antennas system

the signal from the transmit antenna will undergo independent fading to arrive at each receive antenna. As the number of receive antenna grows, the probability that all of the received signals are in deep fading decays by a power law. This power law decay is called the diversity gain and this use of multiple antennas is to mitigate the fading effect to achieve the diversity gain.

Another view of multipath fading was offered by the landmark works [3,4]. The main result in these works states that the capacity of a wireless channel equipped with M transmit and N receive antennas (shown in Fig.1) in a rich scattering environment is given by:

$$C \approx \min(M, N) \log(\mathrm{SNR}),$$

in the high Signal to Noise Ratio (SNR) region. That is the capacity of a MIMO system scales linearly with the minimum number of transmit and receive antennas. The scaling factor $\min(M, N)$ is also called the multiplexing gain. The remarkable capacity improvement of MIMO systems owes largely to the effect of multipath fading. Fading can now also be seen as providing an opportunity

to significantly improve the capacity of such systems. This use of multiple antennas is to exploit the fading effect and achieve the multiplexing gain.

Zheng and Tse [5] show that one can exploit the diversity and multiplexing gains promised by the MIMO channel simultaneously with an existence of a fundamental tradeoff between the two. Motivated by the two fold gains offered by MIMO system, lots of researches are currently being done to increase the throughput and reliability of wireless systems.

1.3 MOTIVATIONS OF THE DISSERTATION

The tremendous capacity and diversity gains in MIMO wireless systems rely heavily on the availability of the channel state information (CSI). With regard to the availability of CSI, MIMO systems can be categorized as follows: 1) No CSI is available at both the transmitter and receiver, 2) CSI is available at the receiver only and 3) CSI is available at both the transmitter and receiver. While most works in the literature consider the second scenario, the main focus in this dissertation is on the other two.

1.3.1 Space Time Code Design without CSI

It is well known that the capacity achieving input signal of a MIMO system is Gaussian distributed when the channel fading is known to the receiver (CSI at the receiver only or CSI at both the transmitter and the receiver). In contrast, the optimal input signal distribution in the case of no CSI, is still not known except for some special cases. In the high SNR region, the throughput of an unknown MIMO fading channel is shown to be maximized by isotropically distributed (i.d.) unitary matrix input signals [6]. Motivated by this finding, a practical coding scheme called Unitary Space Time Coding (USTC) is also designed in [7] to realize the diversity and capacity gains promised by MIMO system in such a case. Nonetheless, the decoding complexity of USTC signals is exponential in the transmission rate and thus making USTC impractical for MIMO systems.

In the context of receiver CSI, an efficient decoding - sphere decoding can be performed to implement the maximum likelihood detector [8, 9]. The prerequisite to the sphere decoder is to decompose the channel matrix first via a QR factorization. This raises the following question yet to be answered:

Can we still implement the maximum likelihood decoding efficiently without CSI at the receiver?

The answer is "Yes". Rather than the decomposition of the channel, we seek a somewhat analogous approach - decomposition of the transmit signal. We show any i.d. unitary matrix can be written as a product of a number of i.d. unitary matrices. Via this decomposition, we transfer the original channel with USTC into a number of equivalent layered sub-channels with the input signal being the corresponding i.d. layered unitary matrix. According to the chain rule in information theory, the channel capacity with USTC can be actually achieved by using the layered i.d. unitary signalling scheme in the transmitter and successive decoding in the receiver. We give three different detection algorithms for LUSTC, namely successive cancelation, Generalized Likelihood Ratio Test (GLRT) decoding and sphere decoding. These three practical decoding algorithms all have low complexity as compared to the exponentially complex decoding of USTC. Roughly speaking, the LUSTC scheme can be generally viewed as the non-coherent version of the Bell Lab Layered Space Time coding scheme (BLAST) [9].

1.3.2 Limited Feedback Codebook Design

When available, CSI at the transmitter can be utilized to adapt resources and the transmission strategy to improve the throughput of a MIMO system. While full knowledge of CSI at the receiver can be assumed, CSI at the transmitter (CSIT) assumption is not reasonable in general. In practice, the channel estimates are obtained at the receiver through training with pilot signals, and the estimated CSI somehow needs to be fed back in order for the transmitter to know the channel.

Unlike for a scalar channel case, for a MIMO system, the number of channel coefficients which need to be fed back to the transmitter is large and a naive feedback method (feedback the scaler quantization of each and every channel coefficients) would require a large capacity in the feedback channel. Thus, the design of limited feedback codebook for MIMO system is an important problem.

Our goal of feedback codebook design is to maximize the forward channel capacity under the constraint on the feedback rate. Thus, each codeword in the codebook shall contain the most essential information to enhance the forward channel capacity. To this end, we address the following question first :

Are all the channel state information equally important to the forward channel capacity?

The answer to this question is "No". We first investigate the information theoretic loss of a Sub-optimal Water-Filling (WF) scheme compared to the optimal WF scheme with full CSIT. Using the simple Talyor series expansion, a family of bounds are derived, which offer great insights into the capacity difference between the two schemes. Making use of the analysis, we determine that the most important factor for the SWF scheme is the identification of those subchannels to which nonzero transmit power should be allocated.

This knowledge is then applied for the design of the limited rate feedback mechanism and codebook. In the codebook design, a similar generalized Lloyd algorithm as in [1] is employed, in which different from [1] the computation of the centroid is formulated as an optimization problem and solved optimally rather than approximately. The proposed algorithm is also adaptive to the channel statistics and the SNR variation. Numerical results show that the proposed codebook design outperforms the existing algorithms reported in the literature.

1.3.3 Cooperation in Wireless Networks

Inspired by the great capacity and diversity improvement promised by MIMO system, lots of researchers are now seeking to improve the reliability of wireless networks by employing multiple antennas. Unfortunately, it is usually not feasible to deploy multiple antennas in a single wireless

node due to the limitation in node size. One example is the uplink of cellular systems, where the size of the mobile device is the limiting factor. In such a case, cooperation among wireless nodes can be exploited to increase the communication capability and information transmission reliability.

Various user cooperation schemes have appeared in the literature. They can be roughly divided into two categories, namely diversity cooperation and coded cooperation schemes. In the diversity cooperation scheme, the transmit node utilizes neighbors as "virtual" antennas to achieve transmit diversity and thus only diversity gain can be obtained. On the other hand, coded user cooperation scheme can realize another gain - coding gain on top of the diversity gain that can be obtained.

We focus on a specific coded user cooperation protocol in the dissertation. The two phase transmission protocol is defined and by the protocol the coded user cooperation scheme is equivalent to a channel coding with the encoding done in a distributed manner. A number of analytical tools have been developed to understand the coded user cooperation scheme. Simulation results show a huge gain of the coded user cooperation over non-cooperation scheme can be obtained.

1.4 OUTLINE OF THE DISSERTATION

The rest of the dissertation is composed of three major parts. In Chapter 2, we consider the design of space time code for non-coherent MIMO system. We review the block-faded channel model and summarize some known results about USTC in Section 2.2. In Section 2.3, we introduce LUSTC and get the coding theorem of achieving the channel capacity via layered unitary signaling scheme at the transmitter and successive decoding at the receiver. The layered capacity of the equivalent channel is analyzed in Section 2.4. In Section 2.5, three different practical decoding algorithms for LUSTC are introduced. The practical layered signal constellation design was considered in Section 2.6. In Section 2.7, the simulation results are given. In Section 2.8, the conclusions are drawn.

When the CSI is perfectly known to the receiver, it needs to be fed back to the transmitter to improve the forward channel capacity. In Chapter 3, we study the design of feedback codebook

when the feedback channel has a rate limitation. In Section 3.2, the MIMO channel model considered in this chapter is introduced and two different transmission schemes with transmitter CSI, the optimal Water-filling and a sub-optimal water-filling schemes are defined. In Section 3.3.1, bounds on the capacity difference between WF and SWF are developed. Numerical results on the capacity difference between the two schemes are also obtained in this section. In Section 3.3.2, the possible transmission scheme based on SWF scheme with limited rate feedback is discussed. The limited rate feedback codebook design algorithm is shown in Section 3.4 and the numerical results are presented in Section 3.5. Conclusions are drawn in Section 3.6.

We study coded user cooperation scheme in Chapter 4. The basic scenario considered is discussed in Section 4.2 and the coded user cooperation protocol is also defined there. In Section 4.3, we identify the similarity and difference between coded user cooperation scheme and conventional channel coding. A number of analytical tools are given in Section 4.4 and Section 4.5. These analytical prediction of the performance along with simulation results are shown in Section 4.6. Conclusions are drawn in Section 4.7.

Chapter 5 contains some future research directions.

2.0 LAYERED UNITARY SPACE TIME CODING

2.1 INTRODUCTION

It has been shown in [3] [4] that Multiple-antenna wireless communication links promise very high date rates in Rayleigh flat fading environment, provided that the complex-valued propagation coefficients between all pairs of transmitter and receiver antennas are statistically independent. Motivated by the huge potential throughput increase, practical schemes [10, 11, 9, 12] have also been developed, which indeed exhibit much better performance (in terms of throughput or error probability) over single transmit and receive antenna system.

However, these results are derived under the assumption that the instantaneous fading coefficients are perfectly known to the receiver (but not to the transmitter). This assumption is reasonable in slow fading environment, where the transmitter can send training sequence that allows the receiver to track the channel. In mobile or other "fast" fading environment, the fading coefficients can change quite rapidly and the estimation of channel parameters becomes difficult particularly as the number of transmit and receive antennas grows. In such a situation, an alterative approach is non-coherent space time communication, for which it is not required to explicitly estimate the fading coefficients.

A line of work [6, 7, 13, 14] was initiated by a group of researchers (Hochwald, Marzetta, Hassibi, et al.) to study both the theoretical capacity limit and practical design of non-coherent Space Time Codes. The capacity-achieving signal is shown [6] [15] to be the product of an isotropically distributed (i.d.) unitary matrix (orthonomal column matrix) and a diagonal matrix with

non-negative real elements on it's diagonal. In high SNR region, the diagonal matrix becomes deterministic and all the information is carried in the i.d. unitary matrix. Therefore, the problem of designing signal constellations becomes finding a finite set of good (in terms of certain performance measure) matrices with orthonomal columns in such case. Related work on such signal constellation design – so called Unitary Space Time Codes (USTC) can be found in [13] by systematic design approach and in [14] by numerical methods.

Nonetheless, the existing problem of these signal constellations is their detection algorithm suffers from exponential complexity, i.e., the receiver has to test all possible candidates to get the optimal estimate of the transmitted signal. To alleviate this problem, unitary space time codes via Cayley transformation was proposed in [16], where polynomial time algorithms – successive cancellation or sphere decoding [17] can be used for decoding the transmitted signal. Albeit efficient, the sphere decoding algorithm is only an approximate rather than the exact Maximum Likelihood (ML) decoding of the Cayley codes and the exploitation of it for decoding consequently subjects to certain performance penalty. In [18, 19, 20] the author proposed a layered unitary space time coding (LUSTC) scheme, wherein the overall transmitted signal is built as the product of a number of layered signals with each of them having different dimensionality. In contrast to the Caley codes, this layered unitary signal structure allows the receiver to perform an exact ML decoding by implementing the sphere decoding algorithm.

Unfortunately, no design guideline (e.g. how many signals should we put in each and every layer) on the layered signals is provided in these papers, which consequently makes the LUSTC scheme not achieving the potential benefits as it shall. In this chapter, we attempt to offer a theoretical design methodology on designing the non-coherent layered space time signal constellations. To this end, however, the layered signal structure in this chapter differs from [18, 19, 20] in the layered signal dimensionality setting. By imposing this change of signal structure, we are able to give the theoretical concept of layered unitary space time coding and provide the design guideline for practical signal constellation design. The main contribution of this chapter is summarized as follows:

- We first show any $D \times M$ $(D > M)$ i.d. unitary matrix can be decomposed into product of a number of i.d. unitary matrices, where each of the i.d. unitary matrix of dimensionality $d \times (d-1)$ $(M+1 \leq d \leq D)$ is called a layered signal. Via this decomposition, we transfer the original channel with USTC into a number of equivalent layered sub-channels with the input signal being the corresponding i.d. layered unitary matrix.

- We give a coding theorem of this LUSTC. According to the chain rule in information theory, the channel capacity with USTC and ML decoding can be actually achieved by using the layered i.d. unitary signalling scheme in the transmitter and successive decoding in the receiver. In essence, the LUSTC scheme can be generally viewed as the non-coherent version of the Bell Lab Layered Space Time coding scheme (BLAST) [9].

- We derive the closed form layered channel capacity of applying i.d. layered unitary signalling scheme in the transmitter and successive decoding scheme in the receiver. The numerical values of the layered capacity are also obtained and then used as a design guideline in building the practical layered signal constellations.

- We give three different detection algorithms for LUSTC, namely successive cancelation, generalized likelihood ratio test decoding (GLRT) and sphere decoding. These three decoding algorithms have low complexity as compared to the exponentially complex decoding of USTC.

- In designing of the layered signal constellation set, we employ a successive design approach to search for the optimal layered signal constellations layer by layer. The signal constellation design starts from the innermost layer. Once the signals of the inner layers are designed, they are taken into account for designing the outer layer signals. Iterating this process layer by layer, we can then eventually design all the layered signal set. The successive design approach can decrease the overall design complexity.

2.2 OVERVIEW OF UNITARY SPACE TIME CODING

2.2.1 Channel Model

We adopt the usual block-fading channel model with multiple transmit and multiple receive antennas, where the channel is constant for some time interval D, after which it changes to an independent value and holds for another interval D, and so on. Within one block of D symbols, the received signal is of the form

$$\mathbf{X} = \sqrt{\frac{\rho D}{M}}\mathbf{S}\mathbf{H} + \mathbf{W}, \tag{2.1}$$

where $\mathbf{X}$ is a $D \times N$ received complex signal matrix, the dimension N is the number of receive antennas. The transmitted signal is $\mathbf{S}$, a $D \times M$ complex matrix where M is the number of transmit antennas. $\mathbf{H}$ is $M \times N$ matrix, with entries independent identically complex Gaussian distributed $\mathcal{CN}(0,1)$. $\mathbf{W}$ is $D \times N$ noise with complex Gaussian $\mathcal{CN}(0,1)$ i.i.d entries. We impose an extra power constraint on the transmitted signal $\mathbf{S}$, i.e.,

$$\frac{1}{M}\sum_{m=1}^{M}\mathbb{E}|\mathbf{S}_{tm}|^2 = \frac{1}{D}, t = 1, 2, ...D.$$

or $\mathrm{tr}(\mathbf{S}\mathbf{S}^*) = M$, which means the average expected power over the M transmitted antennas is kept constant for each channel use. Hence, here ρ represents the expected SNR at each receive antenna.

2.2.2 Unitary Space Time Coding

We cite theorem 2 in [6] (which is also theorem 2 in [7]) as the following lemma:

Lemma 1 *(Structure of signal that achieves capacity) The signal matrix that achieves capacity can be written as* $\mathbf{S} = \mathbf{\Phi}\mathbf{V}$*, where* $\mathbf{\Phi}$ *is an* $D \times M$ *isotropically distributed unitary matrix, and* $\mathbf{V}$ *is an independent* $M \times M$ *real, nonnegative, diagonal matrix. Furthermore, we can choose the joint density of the diagonal elements of* $\mathbf{V}$ *to be unchanged.*

Definition: An ***isotropically distributed (i.d.)*** $D \times M$ matrix $\mathbf{\Phi}$ has a probability density that is unchanged when pre-multiplied by any deterministic $D \times D$ unitary matrix or post-multiplied by any deterministic $M \times M$ unitary matrix.

This Lemma characterizes the structure of the input signal structure, but the optimal distribution of $\mathbf{V}$ that achieves capacity is generally unknown. However, it has been shown that it is optimal to choose a deterministic $\mathbf{V} = \mathbf{I}_M$ in high SNR region when $M \leq \min(D/2, N)$ [21] (or $D \gg M$ [6]), which means that the diagonal matrix $\mathbf{V}$ doesn't bear any information and all the information is carried in the i.d. unitary matrix $\mathbf{\Phi}$ in these cases. In the sequel, we only consider the input signal to be unitary, which is called Unitary Space Time Coding (USTC).

2.2.3 Unitary Space Time Coding Signal Design

Even though the input signal matrix distribution of USTC is fully characterized to be i.d. unitary in high SNR region, in practice, the designed signal codebook needs to form a finite set to be decodable. Assume we have a set of unitary signals $\{\mathbf{\Phi}\}_{k=1}^{L}$ where L is the cardinality of this signal set. Then from the channel model (2.1), conditioned on transmitted signal $\mathbf{\Phi}_k$, the received signal $\mathbf{X}$ is zero-mean circularly symmetric complex Gaussian distributed. Thus, the likelihood function is

$$p(\mathbf{X}|\mathbf{\Phi}_k) = \frac{1}{\pi^{DN} \det^N \left(\mathbf{I}_D + \frac{\rho D}{M}\mathbf{\Phi}_k\mathbf{\Phi}_k^*\right)} \cdot \exp\left(-\mathrm{tr}\left\{\mathbf{X}^*\left[\mathbf{I}_D + \frac{\rho D}{M}\mathbf{\Phi}_k\mathbf{\Phi}_k^*\right]^{-1}\mathbf{X}\right\}\right),$$

where tr denotes "trace" of a matrix and $\det$ denotes the "determinant" of a matrix. The optimal ML detector for the transmitted signal is

$$\hat{k} = \arg\max p(\mathbf{X}|\mathbf{\Phi}_k).$$

Given the unitary structure of the signals $\{\mathbf{\Phi}_k\}_{k=1}^{L}$, the optimal ML detector reduces to

$$\hat{k} = \arg\max \mathrm{tr}(\mathbf{X}^*\mathbf{\Phi}_k\mathbf{\Phi}_k^*\mathbf{X}).$$

Thus, the problem of designing signal constellations for USTC becomes searching for a set of unitary matrices such that the probability of making errors using this detector is minimized. As the exact probability of making errors is difficult to derive, the researchers in [13] turned to use the Chernoff bound of the pairwise error probability. The pairwise error probability (the probability of receiving $\mathbf{\Phi}_k$ given $\mathbf{\Phi}_{k'}$ is transmitted) can be upper-bounded as

$$P_{k,k'} \leq \frac{1}{2} \prod_{m=1}^{M} \left[\frac{1}{1 + \frac{(\rho D/M)^2(1-d_m^2)}{4(1+\rho D/M)}} \right]^N, \tag{2.2}$$

where $\{d_m\}_{m=1}^M$ are the singular values of the $M \times M$ correlation matrix $\mathbf{\Phi}_k^* \mathbf{\Phi}_{k'}$.

This chernoff bound can be further upper-bounded as

$$P_{k,k'} \leq \frac{1}{2} \left[\frac{1}{1 + \frac{(\rho D/M)^2}{4(1+\rho D/M)}} \right]^{N(M - \lceil M \|\mathbf{\Phi}_k^* \mathbf{\Phi}_{k'}\|_F \rceil^2)}. \tag{2.3}$$

Accordingly, a much simpler measure

$$\delta = \max_{\substack{1 \leq k,k' \leq L \\ k \neq k'}} \|\mathbf{\Phi}_k^* \mathbf{\Phi}_{k'}\|_F = \max_{\substack{1 \leq k,k' \leq L \\ k \neq k'}} \mathrm{tr}(\mathbf{\Phi}_{k'}^* \mathbf{\Phi}_k \mathbf{\Phi}_k^* \mathbf{\Phi}_{k'})$$

is applied in [13] that needs to be minimized to design the optimal signal constellations. This performance measure is the so called chordal distance in grassmannian space and it has been widely studied on packing spheres in Grassman manifold [22].

In the past, the unitary space time signal constellations design falls largely into two different approaches. One is systematic construction of signals [13] [23], where by constraining the signal constellations to form a finite group, the signal constellations can either be constructed or searched with relatively less effort. (In some references, similar ideas are also employed in designing signal constellations for differential unitary space time coding, see e.g. [24] [25] [26]). The other approach is to design signal matrices using numerical methods as in [14] [16]. The numerical design is in general a cumbersome and difficult optimization problem and requires certain parameterizations the unitary matrix. In [14], the optimization is based on Givens angles parameterizations of unitary matrix, whereas in [16] the search is based on the Cayley transform.

2.3 LAYERED UNITARY SPACE TIME CODING

One drawback of the USTC is it's high complexity decoding algorithm which makes it not feasible in practical communications system. In order to build signals that can be efficiently detected at the receiver, layered unitary signal design was first proposed in [18, 19, 20]. The layered structure of signal constellations entails the use of a low complexity sphere decoding algorithm [17] for detection at the receiver.

2.3.1 Layered Signal Structure

The overall input signals in [18, 19, 20] are formed from the product

$$\mathbf{S} = \mathbf{S}_D \mathbf{S}_{D-1} \cdots \mathbf{S}_{2M} \mathbf{E},$$

where

$$\mathbf{E} = \begin{bmatrix} \mathbf{I}_M \\ \mathbf{0}_{(D-M)\times M} \end{bmatrix},$$

and

$$\mathbf{S}_d = \left[\begin{array}{c|c} \mathbf{K}_d & \mathbf{0}_{d\times(D-d)} \\ \hline \mathbf{0}_{(D-d)\times d} & \mathbf{I}_{(D-d)} \end{array}\right]. \tag{2.4}$$

Notice that for each layered $D \times D$ signal $\mathbf{S}_d$, the information is carried only in the $d \times d$ sub-matrix $\mathbf{K}_d$. Thus, the signal constellations are totally specified by $\mathbf{K}_d$'s for $2M \leq d \leq D$[1].

By imposing this layered structure, the signals can be efficiently detected via the low complexity sphere decoding algorithm. Two practical layered signal design examples are also given to illustrate the performance in these papers. However, the number of signals put in each layer is chosen in an ad hoc way as no design rules are given. This degrades the potential performance of the layered signalling scheme. To solve this problem, we attempt to analyze the layered capacity in this chapter and use the numerical values of layered capacity as a guideline to build the layered

[1] $2M \leq d$ is necessary for designing signal constellations that achieve full diversity. If design of full diversity signal constellations is not an issue, we can in general have $d \geq M + 1$.

signals. To this end, however, the layered signals structure imposed in this chapter is somewhat different, and which is shown in the following lemma.

Lemma 2 *(Decomposition of i.d. unitary matrix) Any i.d.* $D \times M$ $(D > M)$ *unitary matrix* $\mathbf{S}$ *can be written as the product of a series of i.d. unitary matrices, i.e.,*

$$\begin{aligned} \mathbf{S} &= \mathbf{S}_D \mathbf{S}_{D-1} \cdots \mathbf{S}_{M+1} \\ &= \prod_{d=M+1}^{D} \mathbf{S}_d, \end{aligned} \tag{2.5}$$

where $\mathbf{S}_d$ *is i.d.* $d \times (d-1)$ *unitary matrix.*

Proof See Appendix.

Via this lemma, our transmitted unitary signals can be formed by the product of a series of i.d. unitary matrices. Each of the i.d. unitary matrix with corresponding dimensionality is also called a layered signal. The layered structure differs from the above mentioned one in the following:

- Our layered signals have different dimensionality as compared to the previously introduced one. By careful examination, one can see that the square unitary structure (2.4) is not necessary as the last column of $\mathbf{S}_d$ does not contribute to the overall resultant signal $\mathbf{S}$ when it is multiplied with the matrix $\mathbf{E}$.
- Our layered signal structure ensures the overall input signal to be i.d. unitary. Therefore, in information theoretic sense this layered structure does not incur any capacity loss as compared to USTC.

2.3.2 Equivalent Layered Channel Model

By lemma 2, we can transform the channel model (2.1) as

$$\begin{aligned} \mathbf{X} &= \sqrt{\frac{\rho D}{M}} \mathbf{S}\mathbf{H} + \mathbf{W} \\ &= \sqrt{\frac{\rho D}{M}} \mathbf{S}_D \mathbf{S}_{D-1} ... \mathbf{S}_{M+1} \mathbf{H} + \mathbf{W}. \end{aligned} \tag{2.6}$$

To facilitate the analysis, we define $\mathbf{Q}_{d-1} := \mathbf{S}_{d-1}\mathbf{S}_{d-2}\cdots\mathbf{S}_{M+1}$ and $\mathbf{P}_{d+1} := \mathbf{S}_D\mathbf{S}_{D-1}\cdots\mathbf{S}_{d+1}$ for convenience purpose. According to Lemma 2, $\mathbf{Q}_{d-1}$ and $\mathbf{P}_{d+1}$ are i.d. unitary matrix with dimensionality $(d-1)\times M$ and $D\times d$ respectively. Substitute $\mathbf{Q}_{d-1}$ and $\mathbf{P}_{d+1}$ into (2.6), we then have an equivalent channel model for each layer d ($d > M+1$)

$$\mathbf{X} \;=\; \sqrt{\frac{\rho D}{M}}\mathbf{P}_{d+1}\mathbf{S}_d\mathbf{Q}_{d-1}\mathbf{H} + \mathbf{W}. \tag{2.7}$$

Assume the outer layer signals have been decoded successfully, i.e., $\mathbf{P}_{d+1}$ is known, we can then rewrite the equivalent channel model (2.7) as the following by multiplying both sides by $\mathbf{P}^*_{d+1}$,

$$\mathbf{P}^*_{d+1}\mathbf{X} = \sqrt{\frac{\rho D}{M}}\mathbf{S}_d\mathbf{Q}_{d-1}\mathbf{H} + \mathbf{P}^*_{d+1}\mathbf{W}. \tag{2.8}$$

Thus, the following equivalent channel model is obtained,

$$\mathbf{X}_d = \sqrt{\frac{\rho D}{M}}\mathbf{S}_d\mathbf{Q}_{d-1}\mathbf{H} + \mathbf{W}_d, \quad d > M+1, \tag{2.9}$$

where $\mathbf{X}_d = \mathbf{P}^*_{d+1}\mathbf{X}$ and $\mathbf{W}_d = \mathbf{P}^*_{d+1}\mathbf{W}$. Notice that statistically each of the elements in $\mathbf{W}_d$ is still $\mathcal{CN}(0,1)$ as the columns of $\mathbf{P}_{d+1}$ are orthonormal.

For the innermost layer ($d = M+1$), as there is no inner layer signal, the equivalent channel model is simply

$$\mathbf{X}_{M+1} = \sqrt{\frac{\rho D}{M}}\mathbf{S}_{M+1}\mathbf{H} + \mathbf{W}_{M+1}. \tag{2.10}$$

2.3.3 Layered Unitary Space Time Coding Theorem

Theorem 1 *(Layered unitary space time coding theorem) The non-coherent isotropically distributed (i.d.) Unitary Space Time Coding capacity C is equal to the sum of the layered i.d. unitary space time coding capacities C^d, i.e.,*

$$C = \sum_{d=M+1}^{D} C^d,$$

where $\{C^d\}_{d=M+1}^{D}$ are the corresponding layered capacities for layer d .

Proof : Apply the chain rule of mutual information [27], we get

$$\begin{aligned} I(\mathbf{X};\mathbf{S}) &= I\left(\mathbf{X};\mathbf{S}_{M+1},\mathbf{S}_{M+2},...,\mathbf{S}_{D}\right) \\ &= I(\mathbf{X};\mathbf{S}_{D}) + I(\mathbf{X};\mathbf{S}_{D-1}|\mathbf{S}_{D}) + ... + I(\mathbf{X};\mathbf{S}_{M+1}|\mathbf{S}_{D},\mathbf{S}_{D-1},...,\mathbf{S}_{M+2}) \\ &= I(\mathbf{X};\mathbf{S}_{D}) + I(\mathbf{X}_{D-1};\mathbf{S}_{D-1}) + ... + I(\mathbf{X}_{M+1};\mathbf{S}_{M+1}), \end{aligned} \tag{2.11}$$

where (2.11) follows directly from the equivalent channel model (2.9). Therefore, the capacity of non-coherent communications via USTC is equal to the LUSTC capacities, that is

$$C = \sum_{d=M+1}^{D} C^{d}.$$

To achieve the capacity, we can use i.d. layered signaling scheme in the transmitter and successive decoding scheme in the receiver. The layered transmitted signals can be decoded layer by layer starting from the outmost layer. Once the transmitted signals of the outer layers are known, they can be taken into account to decode the inner layer signals. According to Shannon's Channel Coding theorem, as long as the individual rate of each layer is chosen to be less than the corresponding capacity of that layer, the transmitted signal can be decoded without error.

Conversely, if any of the individual rate R^{d} is greater than the layered capacity C^{d}, no error free transmission is possible.

Remarks:

1. When the fading channel is known to the receiver, a well known BLAST system (it has various forms such as V-BLAST, H-BLAST) was proposed in [9] [12], wherein layered signalling scheme and the successive cancellation are also employed to perform a low complexity detection. The low complexity detection of BLAST is achieved by performing certain matrix decomposition (e.g. QR decomposition) of the known channel, whereas in our channel knowledge unknown case it is achieved by the decomposition of the transmitted signal. However different in nature, the LUSTC resembles the BLAST system and it can be generally viewed as a non-coherent version of it.

2. The layering setup is not necessarily $d \times (d-1)$. For example, when $D = 8$, $M = 1$, in principal the signals can have up to 7 layers ($d = 2, 3, \cdots, 8$), with each layered signal having dimensionality $d \times (d-1)$. One the other hand, if we only want 3 layers, we then can group $d = 2, 3, 4$ as the innermost, $d = 5, 6$ as the middle one and the rest as the outmost layer respectively. However, it's worthwhile to note that the more the layers, the more efficient the detection will be under the same condition.

2.4 LAYERED CAPACITY ANALYSIS

In designing layered signals for use in LUSTC, one fundamental question to be asked is: What the rate should be chosen in each and every layer? To answer this question, in this section, we calculate the capacity of each layer, and use the numerical values of the layered capacity as the design rule.

As the equivalent channel model for the innermost layer (2.10) differs from that for all other layers (2.9), the innermost layer capacity calculation also varies from that for all other layers. For the innermost layer, the channel model (2.10) is almost the same as USTC case except that the received SNR at each receiver antenna in this case is $\frac{\rho D}{M+1}$ in stead of ρ. The channel capacity of USTC is already derived in [28] and hence is omitted here (Please see Theorem 2 in [28] and references therein). The results in [28] can be directly used to calculate the innermost layer capacity with the modification of SNR.

For all other layers but the innermost one, to compute the channel capacity of the equivalent channel model (2.9), the following quantity need to be evaluated.

$$I(\mathbf{X}_d; \mathbf{S}_d) = \frac{1}{D}\mathbb{E}\left\{\log \frac{p(\mathbf{X}_d|\mathbf{S}_d)}{p(\mathbf{X}_d)}\right\}, \tag{2.12}$$

where $\mathbb{E}\{\cdot\}$ is the expectation over all random matrices inside the brackets and $\log$ is base 2.

One can see from this equation that in order to obtain the layered capacity the two density functions $p(\mathbf{X}_d)$ and $p(\mathbf{X}_d|\mathbf{S}_d)$ need to be computed beforehand.

2.4.1 Computation of $p(\mathbf{X}_d)$

As a direct consequence of lemma 2, the input signal $\mathbf{S}_d\mathbf{Q}_{d-1}$ is an i.d. $d \times M$ unitary matrix as well. To obtain a closed form density of the received signal $p(\mathbf{X}_d)$, we follow our previous notation and rewrite $\mathbf{S}_d\mathbf{Q}_{d-1} = \mathbf{Q}_d$, thus

$$\begin{aligned} p(\mathbf{X}_d) &= \mathbb{E}_{|\mathbf{Q}_d} p(\mathbf{X}_d|\mathbf{Q}_d) \\ &= \mathbb{E}_{|\mathbf{Q}_d} \left\{ \frac{\exp\left\{ -\operatorname{tr}\left[\mathbf{X}_d^*(\mathbf{I}_d + \rho\beta\mathbf{Q}_d\mathbf{Q}_d^*)^{-1}\mathbf{X}_d\right]\right\}}{\pi^{dN}\det{}^N\left(\mathbf{I}_d + \rho\beta\mathbf{Q}_d\mathbf{Q}_d^*\right)} \right\} \\ &= \frac{\mathbb{E}_{|\mathbf{Q}_d} \exp\left\{\operatorname{tr}[-\mathbf{X}_d^*\mathbf{Q}_d(\mathbf{I}_M + \frac{1}{\rho\beta}\mathbf{I}_M)^{-1}\mathbf{Q}_d^*\mathbf{X}_d]\right\}}{\pi^{dN}\det{}^N(\mathbf{I}_M + \rho\beta\mathbf{I}_M)} \\ &= \frac{1}{\pi^{dN}} \cdot \frac{\exp\left[-\operatorname{tr}(\mathbf{X}_d^*\mathbf{X}_d)\right]}{(1+\rho\beta)^{MN}} \cdot \mathbb{E}_{|\mathbf{Q}_d} \exp\left[\operatorname{tr}(\alpha\mathbf{X}_d^*\mathbf{Q}_d\mathbf{Q}_d^*\mathbf{X}_d)\right], \end{aligned} \tag{2.13}$$

where we define $\beta := \frac{D}{M}$ and $\alpha := \frac{\rho\beta}{1+\rho\beta}$, $\mathbb{E}_{|\mathbf{Q}_d}\{\cdot\}$ is the expectation taken over $\mathbf{Q}_d$ only.

Therefore, to calculate the layered capacity, the only thing left is to compute the density function $p(\mathbf{X}_d|\mathbf{S}_d)$ for the layers other than the innermost one[2].

2.4.2 Computation of $p(\mathbf{X}_d|\mathbf{S}_d)$

One way of computing the conditional density $p(\mathbf{X}_d|\mathbf{S}_d)$ is to treat $\mathbf{Q}_{d-1}\mathbf{H}$ in (2.9) as the new channel $\tilde{\mathbf{H}}$ and obtain the density $p(\tilde{\mathbf{H}})$ first. Once the distribution of $\tilde{\mathbf{H}}$ is known, the conditional density $p(\mathbf{X}_d|\mathbf{S}_d)$ can then be readily derived as the entries of the additive noise are assumed to be complex Gaussian. At first glance, one may think that the distribution of new channel matrix $\tilde{\mathbf{H}}$ is also gaussian as $\mathbf{Q}_{d-1}$ is unitary and $\mathbf{H}$ is gaussian distributed. However, $\mathbf{Q}_{d-1}$ has orthonormal columns in stead of orhonormal rows, the equivalent channel $\tilde{\mathbf{H}}$ is thus not gaussian any more.

To circumvent the tedious approach of first obtaining $p(\tilde{\mathbf{H}})$ and then $p(\mathbf{X}_d|\mathbf{S}_d)$, we pursue an alternating approach to calculate the conditional density $p(\mathbf{X}_d|\mathbf{S}_d)$ directly. For all the layers but the innermost one, as all random matrices appeared in both sides of (2.9) are i.d., then statistically

[2]For the innermost layer, $p(\mathbf{X}_d|\mathbf{S}_d)$ is nothing but Gaussian distributed.

the input output relationship stay unchanged when pre-multiply both sides of (2.9) by any $d \times d$ unitary matrix $\mathbf{\Psi}$(or it's conjugate transpose $\mathbf{\Psi}^*$), that is

$$p\big(\mathbf{\Psi}^*\mathbf{X}_d|\mathbf{\Psi}^*\mathbf{S}_d\big) = p(\mathbf{X}_d|\mathbf{S}_d).$$

We choose $\mathbf{\Psi}$ particularly as

$$\mathbf{\Psi} = \left[\begin{array}{c|c} \mathbf{S}_d & \mathbf{S}_d^{\perp} \end{array}\right],$$

where the $d \times 1$ unitary matrix $\mathbf{S}_d^{\perp}$ denotes the orthogonal complement of $\mathbf{S}_d$, i.e.,

$$\mathbf{S}_d\mathbf{S}_d^* + \mathbf{S}_d^{\perp}\mathbf{S}_d^{\perp *} = \mathbf{I}_d \;\text{ and }\; \mathbf{S}_d^{\perp *}\mathbf{S}_d^{\perp} = \mathbf{I}_{d-M}.$$

This yields

$$\mathbf{\Psi}^*\mathbf{X}_d = \left[\begin{array}{c} \sqrt{\rho\beta}\mathbf{Q}_{d-1}\mathbf{H} \\ \mathbf{0}_{1\times N} \end{array}\right] + \mathbf{\Psi}^*\mathbf{W}_d \equiv \left[\begin{array}{c} \sqrt{\rho\beta}\mathbf{Q}_{d-1}\mathbf{H} \\ \mathbf{0}_{1\times N} \end{array}\right] + \mathbf{W}_d,$$

where $\equiv$ means statical equivalence.

This leads to

$$\begin{aligned} p(\mathbf{X}_d|\mathbf{S}_d) &= p\left\{\left[\begin{array}{c} \sqrt{\rho\beta}\mathbf{Q}_{d-1}\mathbf{H} \\ \mathbf{0}_{1\times N} \end{array}\right] + \mathbf{W}_d\Big|\left[\begin{array}{c} \mathbf{I}_{d-1} \\ \mathbf{0}_{1\times(d-1)} \end{array}\right]\right\} \\ &= p\left\{\left[\begin{array}{c} \sqrt{\rho\beta}\mathbf{Q}_{d-1}\mathbf{H} + \mathbf{W}_d(1:d-1,:) \\ \mathbf{W}_d(d,:) \end{array}\right]\right\} \\ &= p\Big\{\mathbf{X}_{d-1}, \mathbf{W}_d(d,:)\Big\}, \end{aligned}$$

where $\mathbf{W}_d(d,:)$ denotes the d-th row of $d \times N$ Gaussian distributed matrix $\mathbf{W}_d$.

Let $\bar{\mathbf{W}} = \mathbf{W}_d(d,:)$, thus

$$\begin{aligned}
& p(\mathbf{X}_d|\mathbf{S}_d) \\
= \; & p(\bar{\mathbf{W}})p(\mathbf{X}_{d-1}) \\
= \; & p(\bar{\mathbf{W}}) \cdot \frac{\exp\left[-\mathrm{tr}(\mathbf{X}_{d-1}^*\mathbf{X}_{d-1})\right]}{\pi^{N(d-1)}\det^N(\mathbf{I}_M+\rho\beta\mathbf{I}_M)} \cdot \mathbb{E}_{|\mathbf{Q}_{d-1}} \exp\left[\mathrm{tr}(\alpha\mathbf{X}_{d-1}^*\mathbf{Q}_{d-1}\mathbf{Q}_{d-1}^*\mathbf{X}_{d-1})\right] \qquad (2.14) \\
= \; & \frac{\exp\left[-\mathrm{tr}(\bar{\mathbf{W}}^*\bar{\mathbf{W}})\right]}{\pi^{Nd}} \cdot \frac{\exp\left[-\mathrm{tr}(\mathbf{X}_{d-1}^*\mathbf{X}_{d-1})\right]}{\det^N(\mathbf{I}_M+\rho\beta\mathbf{I}_M)} \cdot \mathbb{E}_{|\mathbf{Q}_{d-1}} \exp\left[\mathrm{tr}(\alpha\mathbf{X}_{d-1}^*\mathbf{Q}_{d-1}\mathbf{Q}_{d-1}^*\mathbf{X}_{d-1})\right] \\
= \; & \frac{1}{\pi^{Nd}} \cdot \frac{\exp\left[-\mathrm{tr}(\mathbf{X}_d^*\mathbf{X}_d)\right]}{(1+\rho\beta)^{MN}} \cdot \mathbb{E}_{|\mathbf{Q}_{d-1}} \exp\left[\mathrm{tr}(\alpha\mathbf{X}_{d-1}^*\mathbf{Q}_{d-1}\mathbf{Q}_{d-1}^*\mathbf{X}_{d-1})\right]. \qquad (2.15)
\end{aligned}$$

(2.14) follows from (2.13) by replacing d with $d-1$ and (2.15) follows from the fact that

$$\mathrm{tr}(\mathbf{X}_d^*\mathbf{X}_d) = \mathrm{tr}(\mathbf{\Psi}^*\mathbf{X}_d^*\mathbf{X}_d\mathbf{\Psi}) = \mathrm{tr}(\bar{\mathbf{W}}^*\bar{\mathbf{W}}) + \mathrm{tr}(\mathbf{X}_{d-1}^*\mathbf{X}_{d-1}).$$

2.4.3 An Alternative Derivation of $p(\mathbf{X}_d|\mathbf{S}_d)$

From (2.9), $\mathbf{X}_d$ is Gaussian distributed given the transmitted signal $\mathbf{S}_d\mathbf{Q}_{d-1}$,

$$p(\mathbf{X}_d|\mathbf{S}_d,\mathbf{Q}_{d-1}) = \frac{1}{\pi^{dN}} \cdot \frac{\exp\left\{-\mathrm{tr}\left[\mathbf{X}_d^*(\mathbf{I}_d+\rho\beta\mathbf{S}_d\mathbf{Q}_{d-1}\mathbf{Q}_{d-1}^*\mathbf{S}_d^*)^{-1}\mathbf{X}_d\right]\right\}}{\det^N(\mathbf{I}_d+\rho\beta\mathbf{S}_d\mathbf{Q}_{d-1}\mathbf{Q}_{d-1}^*\mathbf{S}_d^*)}. \qquad (2.16)$$

Rewrite (2.16), we have the following

$$\begin{aligned}
p(\mathbf{X}_d|\mathbf{S}_d,\mathbf{Q}_{d-1}) &= \frac{1}{\pi^{dN}} \cdot \frac{\exp\left\{-\mathrm{tr}\left[\mathbf{X}_d^*\mathbf{X}_d - \mathbf{X}_d^*\mathbf{S}_d\mathbf{Q}_{d-1}(\mathbf{I}_M+\frac{1}{\rho\beta}\mathbf{I}_M)^{-1}\mathbf{Q}_{d-1}^*\mathbf{S}_d^*\mathbf{X}_d\right]\right\}}{\det^N(\mathbf{I}_M+\rho\beta\mathbf{I}_M)} \\
&= \frac{1}{\pi^{dN}} \cdot \frac{\exp\left[-\mathrm{tr}(\mathbf{X}_d^*\mathbf{X}_d)\right]}{\det^N(\mathbf{I}_M+\rho\beta\mathbf{I}_M)} \cdot \exp\left[\mathrm{tr}(\alpha\mathbf{X}_d^*\mathbf{S}_d\mathbf{Q}_{d-1}\mathbf{Q}_{d-1}^*\mathbf{S}_d^*\mathbf{X}_d)\right] \\
&= \frac{1}{\pi^{dN}} \cdot \frac{\exp\left[-\mathrm{tr}(\mathbf{X}_d^*\mathbf{X}_d)\right]}{(1+\rho\beta)^{MN}} \cdot \exp\left[\mathrm{tr}(\alpha\mathbf{X}_{d-1}^*\mathbf{Q}_{d-1}\mathbf{Q}_{d-1}^*\mathbf{X}_{d-1})\right],
\end{aligned}$$

where the last step is due to $\mathbf{S}_d^*\mathbf{X}_d = \mathbf{X}_{d-1}$.

Thus, averaging out $\mathbf{Q}_{d-1}$, we get exactly the same expression for $p(\mathbf{X}_d|\mathbf{S}_d)$ as shown in (2.15).

2.4.4 Layered Capacity Calculation

According to the previous derivation of $p(\mathbf{X}_d)$ (2.13) and $p(\mathbf{X}_d|\mathbf{S}_d)$(2.15), the equivalent channel capacity for all layers but the innermost one is:

$$\begin{aligned} I(\mathbf{X}_d;\mathbf{S}_d) &= \mathbb{E}\left\{\log\left(\frac{p(\mathbf{X}_d|\mathbf{S}_d)}{p(\mathbf{X}_d)}\right)\right\} \\ &= \mathbb{E}\left\{\log\left(\frac{\mathbb{E}_{|\mathbf{Q}_{d-1}}\exp\left[\mathrm{tr}(\alpha\mathbf{X}_{d-1}^*\mathbf{Q}_{d-1}\mathbf{Q}_{d-1}^*\mathbf{X}_{d-1})\right]}{\mathbb{E}_{|\mathbf{Q}_d}\exp\left[\mathrm{tr}(\alpha\mathbf{X}_d^*\mathbf{Q}_d\mathbf{Q}_d^*\mathbf{X}_d)\right]}\right)\right\}. \end{aligned} \tag{2.17}$$

In order to calculate the layered capacity numerically, we include the results in [28] as the following lemma:

Lemma 3 *: For $T\times N$ matrix $\mathbf{X}=\sqrt{\frac{\rho D}{M}}\mathbf{\Phi H}+\mathbf{W}$, where $\mathbf{\Phi}$ $T\times M$ i.d. unitary matrix and $\mathbf{W}$ is $T\times N$ matrix with complex normal distribution $\mathcal{CN}(0,1)$, we have*

$$\mathbb{E}_{|\mathbf{\Phi}}\exp\mathrm{tr}(\alpha\mathbf{X}^*\mathbf{\Phi}\mathbf{\Phi}^*\mathbf{X}) = (-1)^{M(M-1)/2}\cdot\frac{\Gamma(T)...\Gamma(T+1-M)}{\Gamma(M)...\Gamma(1)}\det\mathbf{F},$$

where $\mathbf{F}$ is a $M\times M$ Hankel matrix whose entries are given by

$$\mathbf{F}_{mn}=\sum_{k=1}^{K}\frac{1}{\prod_{l\neq k}(\alpha\sigma_k-\alpha\sigma_l)}\sum_{q=0}^{\infty}\frac{(\alpha\sigma_k)^q}{(q+T-(m+n+K)+2)!}, m,n=1,...,M$$

where $K=\min(T,N)$, and σ_k are the nonzero eigenvalues of $\mathbf{X}\mathbf{X}^*$.

or, equivalently,

$$\mathbb{E}_{|\mathbf{\Phi}}\exp\mathrm{tr}(\alpha\mathbf{X}^*\mathbf{\Phi}\mathbf{\Phi}^*\mathbf{X}) = (-1)^{(T-M)(T-M-1)/2}\cdot\frac{\Gamma(T)...\Gamma(M+1)}{\Gamma(T-M)...\Gamma(1)}\det\mathbf{G},$$

where $\mathbf{G}$ is a $(T-M)\times(T-M)$ Hankel matrix whose entries are given by

$$\mathbf{G}_{mn}=\sum_{k=1}^{K}\frac{1}{\prod_{l\neq k}(\alpha\sigma_k-\alpha\sigma_l)}\sum_{q=0}^{\infty}\frac{(-\alpha\sigma_k)^q}{(q+T-(m+n+K)+2)!}, m,n=1,...,T-M$$

Using Lemma 3, we can rewrite (2.17) as follows,

$$I(\mathbf{X}_d;\mathbf{S}_d)=\mathbb{E}\left\{\log\left(\frac{\Gamma(d-M)}{\Gamma(d)}\cdot\frac{\det\mathbf{F}_{d-1}}{\det\mathbf{F}_d}\right)\right\},$$

where $\{\mathbf{F}_d\}_{d=M+1}^{D}$ are the corresponding Hankel matrices w.r.t. $\{\mathbf{X}_d\}_{d=M+1}^{D}$.

Or, equivalently

$$I(\mathbf{X}_d;\mathbf{S}_d)=\mathbb{E}\left\{\log\left((-1)^{d-M-1}\frac{\exp(\alpha\mathbf{X}_{d-1}^{*}\mathbf{X}_{d-1})}{\exp(\alpha\mathbf{X}_{d}^{*}\mathbf{X}_{d})}\cdot\frac{\Gamma(d-M)}{\Gamma(d)}\cdot\frac{\det\mathbf{G}_{d-1}}{\det\mathbf{G}_d}\right)\right\}.$$

In summary, we have the following theorem:

Theorem 2 *(Mutual Information for Layered Isotropically Distributed Unitary Input) Consider the channel model*

$$\begin{aligned}\mathbf{X} &= \sqrt{\frac{\rho D}{M}}\mathbf{S}\mathbf{H}+\mathbf{W}\\ &= \sqrt{\frac{\rho D}{M}}\left(\prod_{d=M+1}^{D}\mathbf{S}_d\right)\mathbf{H}+\mathbf{W},\end{aligned}$$

where $\mathbf{H}\in\mathcal{CN}^{D\times M}(0,1)$ *and* $\mathbf{W}\in\mathcal{CN}^{D\times N}(0,1)$ *and the channel is constant for* D *channel use, the layered channel capacity at each layer by using i.d. unitary signalling scheme is*

$$C^d=\begin{cases}\rho DN\log e-\log\frac{\Gamma(M+1)\ldots\Gamma(2)}{\Gamma(M)\ldots\Gamma(1)}-\mathbb{E}\left\{\log\left((-1)^{M(M-1)/2}\det\mathbf{F}_d\right)\right\} & \text{for } d=M+1\\ \mathbb{E}\left\{\log\left(\frac{\Gamma(d-M)}{\Gamma(d)}\frac{\det\mathbf{F}_{d-1}}{\det\mathbf{F}_d}\right)\right\} & \text{for } d\geq M+1.\end{cases}\tag{2.18}$$

or equivalently,

$$C^d=\begin{cases}-\frac{1}{1+\rho\beta}\rho\beta N\log e-\log\frac{\Gamma(M+1)}{\Gamma(1)}-\mathbb{E}\left\{\log\left(\det\mathbf{G}_d\right)\right\} & \text{for } d=M+1\\ \mathbb{E}\left\{\log\left((-1)^{d-M-1}\frac{\exp(\alpha\mathbf{X}_{d-1}^{*}\mathbf{X}_{d-1})}{\exp(\alpha\mathbf{X}_{d}^{*}\mathbf{X}_{d})}\cdot\frac{\Gamma(d-M)}{\Gamma(d)}\cdot\frac{\det\mathbf{G}_{d-1}}{\det\mathbf{G}_d}\right)\right\} & \text{for } d\geq M+1.\end{cases}\tag{2.19}$$

Remarks:

1. The capacity for the innermost layer are derived using exactly the same approach as in [28].

2. In principal, both (2.18) and (2.19) can be used to compute the mutual information via Monte Carlo simulation. As explained in [28], however, (2.18) may be numerically instable when SNR is high or the number of transmit and receive antennas is large. Due to this reason, the numerical values of the layered capacity shown in Section 2.7 are all obtained by using (2.19).

3. We follow the same layered setting convention as the previous section for no particular reason other than convenience. The capacities of any other valid layering setting can be obtained in a similar way or by simply adding some layered capacities up.

2.5 LAYERED UNITARY SIGNAL DETECTION

In this section, we present three different detection rules for LUSTC when the input unitary signals are constrained to form a finite set. These rules are respectively the multistage or successive cancellation, generalized likelihood ratio test detection (GLRT) and sphere decoding. The performance of GLRT detector and sphere decoding detector is explored in Section 2.7.

2.5.1 MultiStage Decoding or Successive Cancellation

The chain rule of mutual information (2.11) suggests one natural low complexity detection algorithm – multistage decoding (MSD) as in [29] or the successive cancellation as in [9]. The detection process starts from the outmost layer signal set. Once the outer layer signals are decoded successfully, they can be taken into account for decoding the second outmost layer signals. Iterating this process layer by layer, one can eventually decode the signals at all layers successfully.

Assume the d-th layered signal constellations form a finite set with cardinality L_d and the signals for layers less than d are all i.d., the detection rule for the d-th layer signal can be rewritten

as:

$$\begin{aligned}\hat{k} &= \arg \max_{1\le k\le L_d} p(\mathbf{X}_d|\mathbf{S}_d^k) \\ &= \arg \max_{1\le k\le L_d} \mathbb{E}_{|\mathbf{Q}_{d-1}} \exp\left[\mathrm{tr}(\alpha \mathbf{X}_d^* \mathbf{S}_d^k \mathbf{Q}_{d-1} \mathbf{Q}_{d-1}^* \mathbf{S}_d^{k*} \mathbf{X}_d)\right] \qquad (2.20) \\ &= \arg \max_{1\le k\le L_d} (-1)^{M(M-1)/2} \det \mathbf{F}_{d-1}^k, \qquad (2.21)\end{aligned}$$

where $\mathbf{F}_{d-1}^k$ is the Hankel matrix w.r.t. $\mathbf{S}_d^{k*}\mathbf{X}_d$.

The prerequisite of this detector is that the outer layer signals have been decoded successfully, otherwise error propagation from outer layers becomes an issue especially when the code rate is not properly chosen. In practice, we can apply capacity achieving channel codes (e.g. LDPC code or Turbo code) in concatenation with LUSTC scheme to improve the reliability of transmission in each and every layer. As long as the individual rate in each layer is chosen to be the equivalent channel capacity, asymptotically in the code length, this detection rule subjects to no penalty as compared to the overall ML decoding.

Nevertheless, in stead of choosing the inner layer signals being i.d., in practice, each of the inner layer signal codebook needs to form a finite set as well. In such case, assuming all the previous layered signals are transmitted equally probable, the decoding metric for the d-th layer signal can be derived as

$$\hat{k} = \arg \max_{1\le k\le L_d} \sum_j \exp\left\{\mathrm{tr}\left[\alpha \mathbf{X}_d^* \mathbf{S}_d^k \mathbf{Q}_{d-1}^j \mathbf{Q}_{d-1}^{j*} \mathbf{S}_d^{k*} \mathbf{X}_d\right]\right\} \qquad (2.22)$$

Unfortunately, as the presence of the summation over the previous layered signals in this detection rule, it is actually more complex rather than simpler as compared to the ML decoding rule of the unitary signal. Therefore, it is not desirable to use this detection rule especially when the spectral efficiency is high.

In order to circumvent this drawback, it is convenient to resort to the Generalized Likelihood Ration Test (GLRT) approach as it does not require knowing any previous layer signal information.

2.5.2 GLRT Detection

Notice that from the equivalent channel model (2.9), we can treat $\mathbf{Q}_{d-1}\mathbf{H} := \tilde{\mathbf{H}}$ as the new channel. Assume the channel matrix $\tilde{\mathbf{H}}$ is fully known to the receiver, then the optimum coherent ML detector can be shown as

$$\hat{k} = \arg \min_{1 \le k \le L_d} \mathrm{tr}\Big\{ (\mathbf{X}_d - \mathbf{S}_d^k \tilde{\mathbf{H}})^* (\mathbf{X}_d - \mathbf{S}_d^k \tilde{\mathbf{H}}) \Big\}. \tag{2.23}$$

When no knowledge is assumed for the new channel propagation $\tilde{\mathbf{H}}$, the receiver needs to first estimate $\tilde{\mathbf{H}}$. The optimal least square estimate of $\tilde{\mathbf{H}}$ can be shown as

$$\tilde{\mathbf{H}} = (\mathbf{S}_d^{k*} \mathbf{S}_d^k)^{-1} \mathbf{S}_d^{k*} \mathbf{X}_d$$

Treat the estimate of $\tilde{\mathbf{H}}$ as true and take it into the coherent detector (2.23), the detection rule can be written as

$$\hat{k} = \arg \max_{1 \le k \le L_d} \mathrm{tr}\Big\{ \mathbf{X}_d^* \mathbf{S}_d^k (\mathbf{S}_d^{k*} \mathbf{S}_d^k)^{-1} \mathbf{S}_d^{k*} \mathbf{X}_d \Big\}.$$

When the input signal $\mathbf{S}_d^k$ is further constrained to be unitary, the detector reduces to

$$\hat{k} = \arg \max_{1 \le k \le L_d} \mathrm{tr}\Big\{ \mathbf{X}_d^* \mathbf{S}_d^k \mathbf{S}_d^{k*} \mathbf{X}_d \Big\}.$$

This is the GLRT detector for the d-th layered signals. Similarly to the successive cancellation, GLRT detector can also decode the layered signals successively starting from the outmost layer. The advantages of the GLRT lies in it's low complexity as it does not require any operation with the inner layer signals while decoding the current layer signals. Nonetheless, from the perspective of symbol error probability, the GLRT detector is suboptimal relative to the ML decoding of the unitary signals. Moreover, it also has error propagation problem as successful decoding of the inner layer signals depends upon the outer layer signals being successfully decoded.

2.5.3 Sphere Decoding

The sphere decoding algorithm was illustrated in [18, 19, 20] to detect the layered signals. Even though the layered signal structure in this chapter is somewhat different, the sphere decoding algorithm can still be used as an efficient detection rule for our layered signals.

As the signals in each layer are all constrained to be unitary, the resultant overall input signal $\mathbf{Q}_D$ is then consequently unitary. Thus, the optimal detector is:

$$\hat{k} = \arg \max_{1 \le k \le L} \mathrm{tr}(\mathbf{X}_D^* \mathbf{Q}_D^k \mathbf{Q}_D^{k*} \mathbf{X}_D),$$

where $L = \prod_{d=M+1}^{D} L_d$. Let $\mathbf{Q}_D^k = \mathbf{S}_D^m \mathbf{Q}_{D-1}^n$, with $1 \le m \le L_D$ and $1 \le n \le \prod_{d=M+1}^{D-1} L_d$ then we have

$$\begin{aligned}
\hat{k} &= \arg \max_{\substack{1 \le m \le L_D \\ 1 \le n \le \prod_{d=M+1}^{D-1} L_d}} \mathrm{tr}[\mathbf{X}_D^* \mathbf{S}_D^m \mathbf{Q}_{D-1}^n \mathbf{Q}_{D-1}^{n*} \mathbf{S}_D^{k*} \mathbf{X}_D] \\
&= \arg \max_{\substack{1 \le m \le L_D \\ 1 \le n \le \prod_{d=M+1}^{D-1} L_d}} \Big\{ \mathrm{tr}[\mathbf{X}_D^* \mathbf{S}_D^m \mathbf{S}_D^{m*} \mathbf{X}_D] - \mathrm{tr}[\mathbf{X}_D^* \mathbf{S}_D^m \mathbf{Q}_{D-1}^{n\perp} \mathbf{Q}_{D-1}^{n\perp *} \mathbf{S}_D^{m*} \mathbf{X}_D] \Big\} \\
&= \arg \min_{\substack{1 \le m \le L_D \\ 1 \le n \le \prod_{d=M+1}^{D-1} L_d}} \Big\{ \mathrm{tr}\left[\mathbf{X}_D^* \mathbf{S}_D^{m\perp} \mathbf{S}_D^{m\perp *} \mathbf{X}_D\right] + \mathrm{tr}\left[\mathbf{X}_D^* \mathbf{S}_D^m \mathbf{Q}_{D-1}^{n\perp} \mathbf{Q}_{D-1}^{n\perp *} \mathbf{S}_D^{m*} \mathbf{X}_d\right] \Big\} \quad (2.24)
\end{aligned}$$

This minimization search over all possible signal candidates can be done efficiently via the tree search sphere decoding algorithm. We can first find the set of signals from the outmost layer D such that

$$\mathrm{tr}\Big\{ \mathbf{X}_D^* \mathbf{S}_D^{m\perp} \mathbf{S}_D^{m\perp *} \mathbf{X}_D \Big\} < r^2,$$

for some pre-chosen sphere radius r. This can prune a portion of signal candidates at the outmost layer D.

In a similar manner, the second term inside the minimization (2.24) can also be expanded as two summations with one term has only the second outmost layer signals involved. Thus, we can employ a similar search to eliminate some signal candidates in the second outmost layer, with a recomputed radius. Iterate this process layer by layer, essentially a large portion of possible

candidates can be pruned by a carefully chosen radius r. This approach greatly reduce the detection complexity via pruning the possible signal candidates in every layer.

Unlike the GLRT decoding algorithm, sphere decoder does not make decision at each step but only gives admissible signal candidates. Whenever the final admissible signal set is not empty, it must include the optimal ML estimate of the transmitted signal. The GLRT detector, however, can be viewed as first minimize the first term in (2.24), and then treat the estimated signal for the outmost layer as true to minimize the second term. This is in general not equivalent to minimize the sum of the two as in (2.24) and therefore no optimality of this approach is guaranteed. In practice, we expect this detector will outperform the GLRT detector, albeit more complex, which will be verified later in the simulation.

2.6 DESIGN OF LAYERED UNITARY SPACE TIME CODING

Although the input signal in each layer is characterized as i.d. unitary matrix, the practical LUSTC signal constellations have to be constrained in a finite set to be decodable as the USTC case. Nonetheless, different from the USTC codebook design, for LUSTC we need to build $D - 2M + 1$ [3] layers signal constellations. Each of the layered signal constellations has orthonormal columns, i.e, for any codeword $\mathbf{S}_d^k \in \mathcal{S}_d$ with $k \leq L_d$ (L_d is the cardinality of $\mathcal{S}_d$ and is determined by the layered capacity), we have

$$\mathbf{S}_d^k \in \begin{cases} \mathcal{U}(d \times (d-1)), & \text{if } d > 2M \\ \mathcal{U}(2M \times M), & \text{if } d = 2M \end{cases}$$

, where $\mathcal{U}$ is the set of matrices with orthonormal columns.

The layered signal constellation design closely relates to the specific decoding algorithm. For the different detection rules introduced in Section 2.5, the design criteria also varies. In this section,

[3] We set the $d = 2M$ as the first layer to design the signal constellations that achieve full diversity. If this requirement is relaxed, the first layer can start as $d = M + 1$.

we design the signal constellations based on ML or the sphere decoding of the layered signals. The design approach based on the GLRT receiver, though interesting, is not investigated in this chapter as 1) the design criteria is not directly available for the non-gaussian new channel propagation matrix $\tilde{\mathbf{H}}$ 2) the GLRT detector in principal under performs the sphere decoding algorithm.

We present a successive design approach to construct the layered unitary signal constellations. This design approach is based on a numerical optimization through the Given's angle parametrization of unitary matrices, which we will introduce next.

2.6.1 Parametrization of Unitary Matrix

We only briefly introduce the parametrization below, please refer to [14] and references therein for details. As shown in [14], any unitary matrix can be written as the product of a set of basic rotation matrices and one diagonal matrix. The basic $d \times d$ rotation matrix $\mathbf{U}^{pq}(\phi, \sigma)$ specified by $p < q$ and $\phi, \sigma \in [-\pi, \pi)$ is

$$\mathbf{U}^{pq}_{jk}(\phi,\sigma) = \begin{cases} 1 & \text{if } j = k \text{ and } j \neq p, q \\ cos(\phi) & \text{if } j = k \text{ and } j = p, q \\ -sin(\phi)\exp(-i\sigma) & \text{if } j = p \text{ and } k = q \\ sin(\phi)\exp(i\sigma) & \text{if } j = q \text{ and } k = p \\ 0 & \text{otherwise.} \end{cases}$$

Thus, any $d \times d$ unitary matrix $\mathbf{U}$ can be rewritten as

$$\mathbf{U} = \mathbf{V}\prod_{p=1}^{d-1}\prod_{q=p+1}^{d}\mathbf{U}^{p,q}(\phi_{p,q}, \sigma_{p,q}),$$

where $\mathbf{V} = \mathrm{diag}(\exp(-i\delta_1)\exp(-i\delta_2)\cdots\exp(-i\delta_d))$.

To parameterize a $d \times d$ unitary matrix, totally d^2 parameters are needed. In a similar manner, we can parameterize the set of $d \times M$ (where $d > M$) unitary matrices by $2dM - M^2$ parameters. However, to simplify the parametrization, we choose to over parameterize any $d \times M$ unitary

matrix by d^2 parameters as well. Practically, we need only to choose the first M columns of $d \times d$ unitary matrix as our desired matrix.

One good property of this Given's angle parametrization is it turned the constraint optimization problem into an unconstraint one.

2.6.2 Performance Measure

Different from the design criteria used in the previous literature, we adopt an alternative performance measure that is independent of the SNR and the number of receive antennas in this chapter to design the layered signals. On the high SNR region ($\rho \to \infty$), minimizing the chernoff bound of the pairwise error probability (2.2) is equivalent to minimize the following

$$\prod_{m=1}^{M} (1 - d_m^2)^{-1} = \frac{1}{\det(\mathbf{I} - \mathbf{\Phi}_k^* \mathbf{\Phi}_k \mathbf{\Phi}_{k'}^* \mathbf{\Phi}_{k'})}. \tag{2.25}$$

We utilize the union of the asymptotic pairwise result as the performance measure

$$\sum_{k=1}^{L} \sum_{k'>k}^{L} \frac{1}{\det(\mathbf{I} - \mathbf{\Phi}_k^* \mathbf{\Phi}_k \mathbf{\Phi}_{k'}^* \mathbf{\Phi}_{k'})}, \tag{2.26}$$

where L is the cardinality of the codebook.

In [30], it was proved that the codebook achieves full diversity if for any pair of codewords $\mathbf{\Phi}_k$ and $\mathbf{\Phi}_{k'}$, the matrix

$$\begin{bmatrix} \mathbf{\Phi}_k^* \\ \mathbf{\Phi}_{k'}^* \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_k & \mathbf{\Phi}_{k'} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{kk} & \mathbf{R}_{kk'} \\ \mathbf{R}_{k'k} & \mathbf{R}_{k'k'} \end{bmatrix}$$

has full rank.

It's not hard to show that the determinant of this matrix is equal to the denominator of (2.25) given the unitary structure of the signals. Thus, (2.26) ensures the designed codebook to achieve full transmit diversity as long as it's finite.

As pointed out in [31], one necessary condition to have full diversity signal constellations is the coherence time of the channel is greater than $2M$. Therefore, we restrict our innermost layer signal to have the dimensionality $d = 2M$ to construct signal constellations that achieve full diversity.

2.6.3 Layered Unitary Signal Design

Via Given's angle parametrization, in principal an unconstrained optimization problem can be set up to design the optimal layered unitary signal constellations for all d ($2M \leq d \leq D$) at once. However, as there are generally too many parameters to optimize over, this method is not feasible even for small D. As a matter of fact, in [20], the authors admit that they are not able to find satisfactory designs based on Given's angle parametrization for more than two layers signal constellations.

To alleviate the problem of optimizing over too many parameters, we choose to design our signal constellations successively. We start to build the signals from the innermost layer. Once the innermost layer signals are designed they can be taken into account to further design the second innermost layer signals. Iterate this process layer by layer, we are able to design the optimal signal constellations with relative low complexity.

For the innermost layer $d = 2M$, the signal constellation design problem is the same as usual USTC case. We use (2.26) as the objective function and minimize it over the Given's angle parametrization parameters. Detailed approach on designing USTC signal constellations via Given's angle parametrization can be found in [32] and is omitted here.

The signal constellation design for all other layers but the innermost one varies with the detection rule. If the ML or sphere decoding algorithm is applied for detection, then essentially we can view $\mathbf{S}_d\mathbf{Q}_{d-1}$ in (2.9) as the input signal for USTC. Motivated by this, we need also minimize the performance measure (2.26) w.r.t. $\{\mathbf{\Phi}\} = \mathcal{S}_d \bigotimes \mathcal{Q}_{d-1}$. Here $\bigotimes$ is the Kronecker product. As the inner layer signal constellations $\mathcal{Q}_{d-1}$ are already obtained in the previous step, we then need only to design L_d signal constellations for layer d, where L_d is determined by the layered capacity[4].

In summary, the design algorithm of layered signal constellations can be stated as follows:

step 1: Set $d = 2M$

step 2. Design L_{2M} $2M \times M$ *unitary matrices* $\{\mathbf{\Phi}_k\}_{k=1}^{L_{2M}}$ *according to the design metric (2.26)*

[4]For designing signal constellations that work below the capacity, other design rule (e.g. error exponent) is of practical interests and needs further investigation

and set $\mathcal{Q}_{2M} = \{\mathbf{\Phi}_k\}_{k=1}^{L_{2M}}$

step 3. Repeat the following until $d = D$

- *Set* $d = d + 1$
- *Design* L_d $d \times (d-1)$ *unitary matrices* $\mathcal{S}_d$ *according to (2.26), where* $\{\mathbf{\Phi}_k\}_{k=1}^{\prod_{d'=2M}^{d} L_{d'}} = \mathcal{S}_d \bigotimes \mathcal{Q}_{d-1}$
- *Set* $\mathcal{Q}_d = \{\mathbf{\Phi}_k\}_{k=1}^{\prod_{d'=2M}^{d} L_{d'}}$

2.6.4 A Heuristic Design for the Layered Signals

Applying the successive design algorithm at each layer d, we have totally $L_d \cdot d^2$ parameters to optimize over, which may make the optimization problem inefficient while designing higher spectral efficiency signal constellations especially as d increases. To solve this problem, we present a heuristic design approach to design the layered signals as the layer d or L_d is large.

Rather than pursue the time consuming approach to minimize (2.26) numerically, we can simply generate the Given's angle parameters σ's, ϕ's and δ's randomly. Using (2.26) of the optimized signal constellations as the baseline, we get that we can generate fairly good signal constellations by choosing all the parameters according to normal distribution $\mathcal{N}(0,1)$. Practically, we can randomly generate the set of parameters a hundred times and choose the best set of parameters (in terms of (2.26)).

2.7 SIMULATION RESULTS

2.7.1 Numerical Examples of the Layered Channel Capacity

The equivalent channel capacity of the LUSTC is evaluated numerically using Monte-Carlo simulation. In the simulation, we randomly generate i.d. $(d-1) \times M$ unitary matrix $\mathbf{Q}_{d-1}$, $d \times (d-1)$ i.d. unitary matrix $\mathbf{S}_d$ and complex normal channel matrix $\mathbf{H}$ and $\mathbf{W}$. The i.d. unitary

matrix, say $\mathbf{Q}_{d-1}$, can be generated by first performing singular value decomposition of a complex matrix with it's entries being i.i.d normal distributed and then choose the first M columns of any of the two unitary matrices.

In Fig.2, the equivalent channel capacity is evaluated for $M = N = 2, D = 20$ with different SNR values. In Fig.3, the simulation parameters chosen are $M = N = 4, D = 20$. The layered channel capacities are normalized by the channel coherence time D. As can be seen from the simulation results, the capacity of equivalent layered channel is decreasing with layer increasing.

In Fig.4, the normalized layered capacity is plotted as a function of SNR for $M = 2$, $N = 1$, $D = 6$. In Fig.5, the normalized layered capacity iss plotted for $M = 2$, $N = 2$ and $D = 6$ case. Similarly as the previous obtained results, the layered capacity decreases as layer increases. These numerical values of the layered capacities are used next to design practical layered unitary signal constellations. As suggested by these results, in designing the practical LUSTC, we shall put more signals in the inner layers and less in the outer layers.

2.7.2 Simulation Results on the Performance of Designed LUSTC with Practical Decoding Algorithms

In Fig.6, we compare the performance of our successively designed layered signal constellations with the optimized unitary space time constellations. The simulation parameters are $D = 6$, $M = 2$, $N = 1$. According to the previous layered capacity analysis (Fig.4), we put 17, 3, 2 signals respectively in layer $4, 5, 6$. The spectral efficiency is then $\log(102)/6 \approx 1.1$ Bits/Sec/Hz. The USTC signal constellations (totally 102 6×2 unitary matrices) are obtained by numerically optimizing (2.26). It can be seen from the figure, the successively designed LUSTC has less than 2 dB performance degradation as compared to USTC signal constellations. However, as pointed out earlier, the LUSTC enjoys lower decoding complexity. We further plot the performance of the same LUSTC signal constellations when GLRT detector is used for decoding. Seen from the figure, the GLRT decoding has around 8 dB performance loss compared to sphere decoding at the symbol error rate around 10^{-3}. It's worthwhile to note that the signal constellation designed

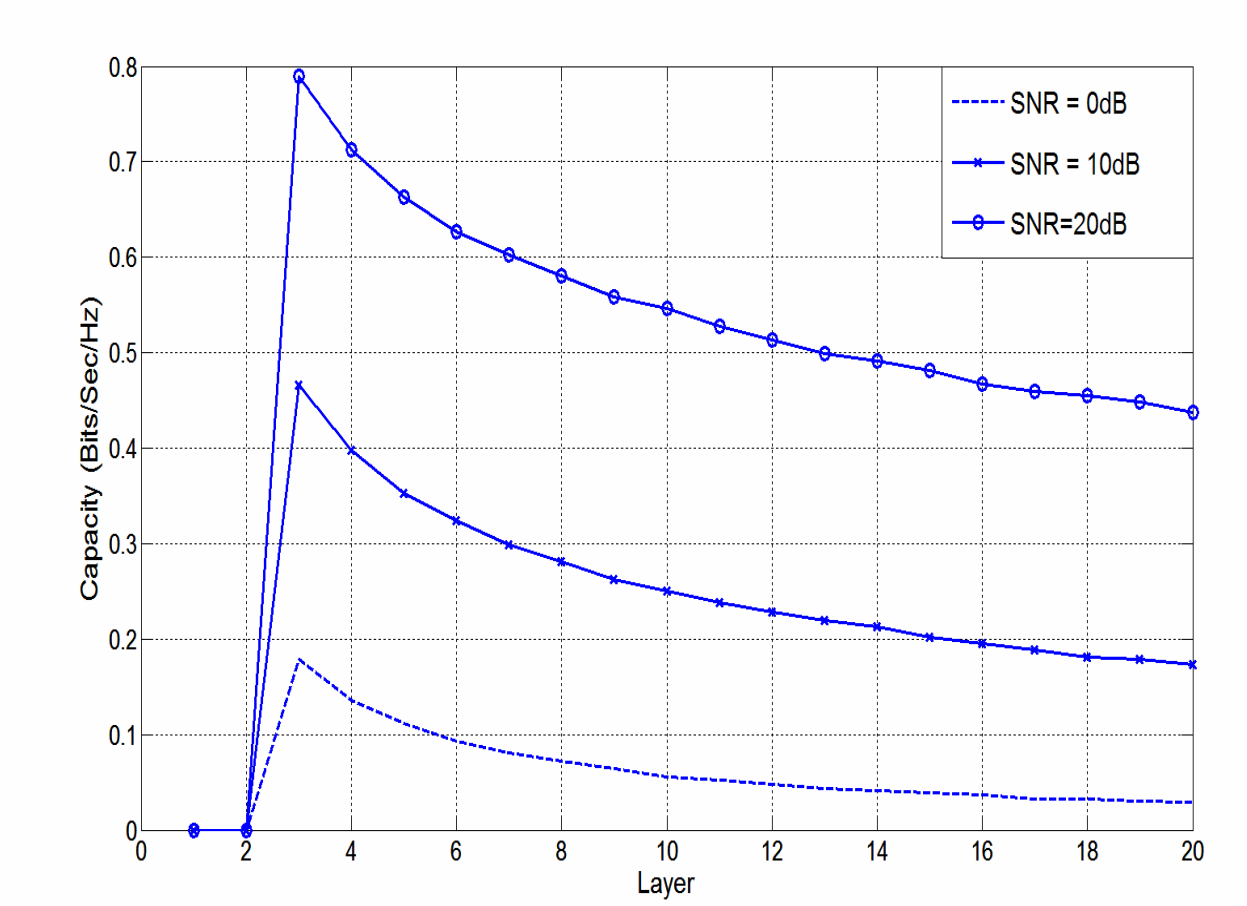

Figure 2: Layered capacity at different SNR, $M = N = 2, D = 20$

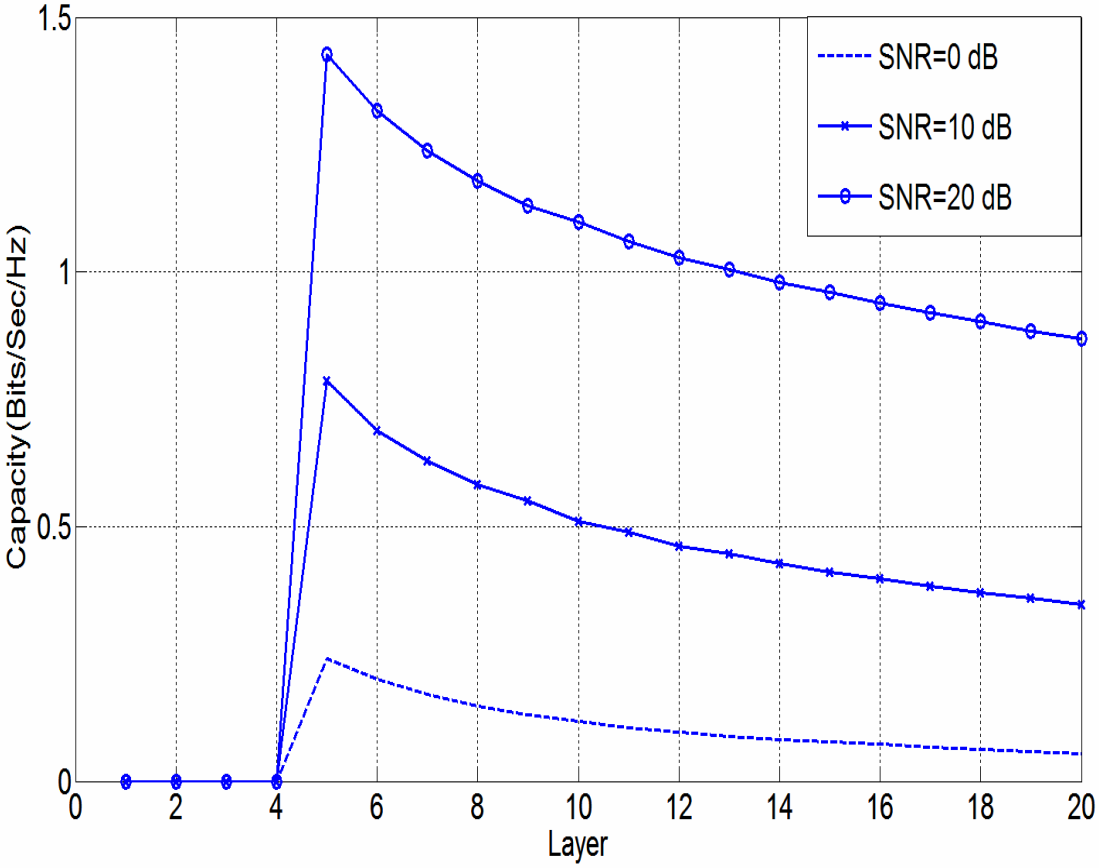

Figure 3: Layered capacity at different SNR, $M = N = 4, D = 20$

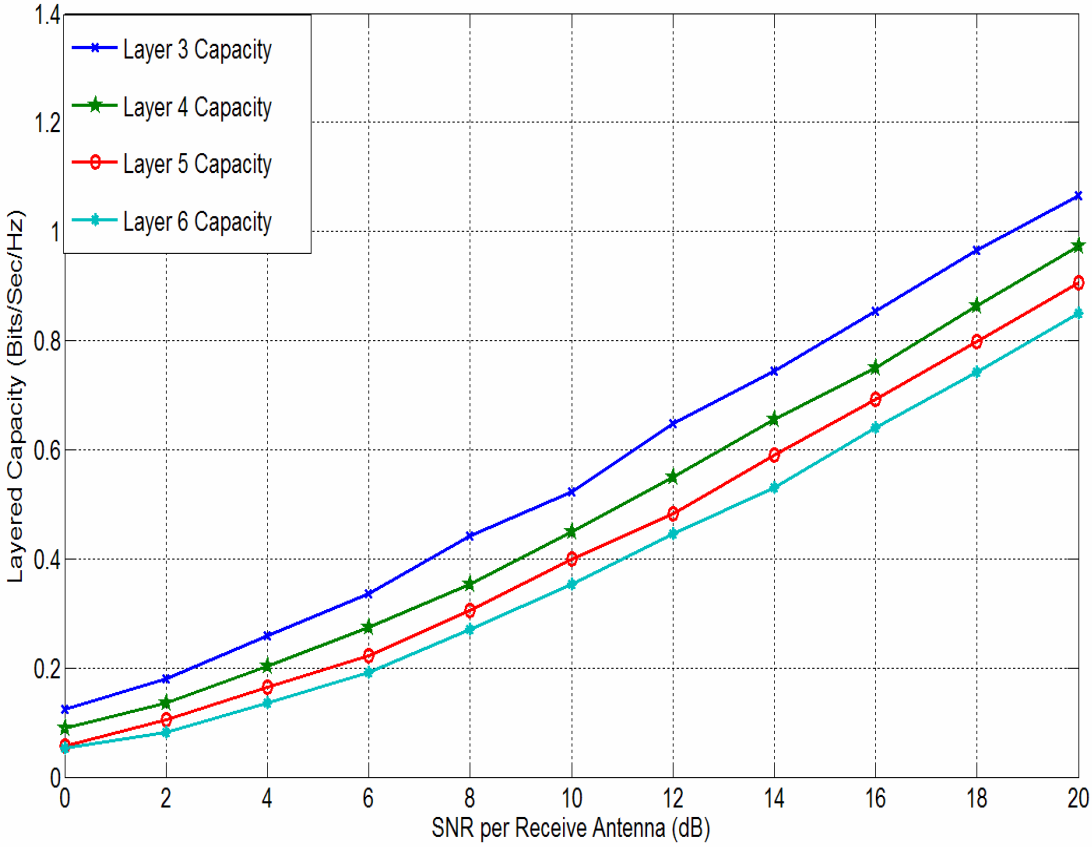

Figure 4: Layered capacity vs SNR, $M = 2, N = 1, D = 6$

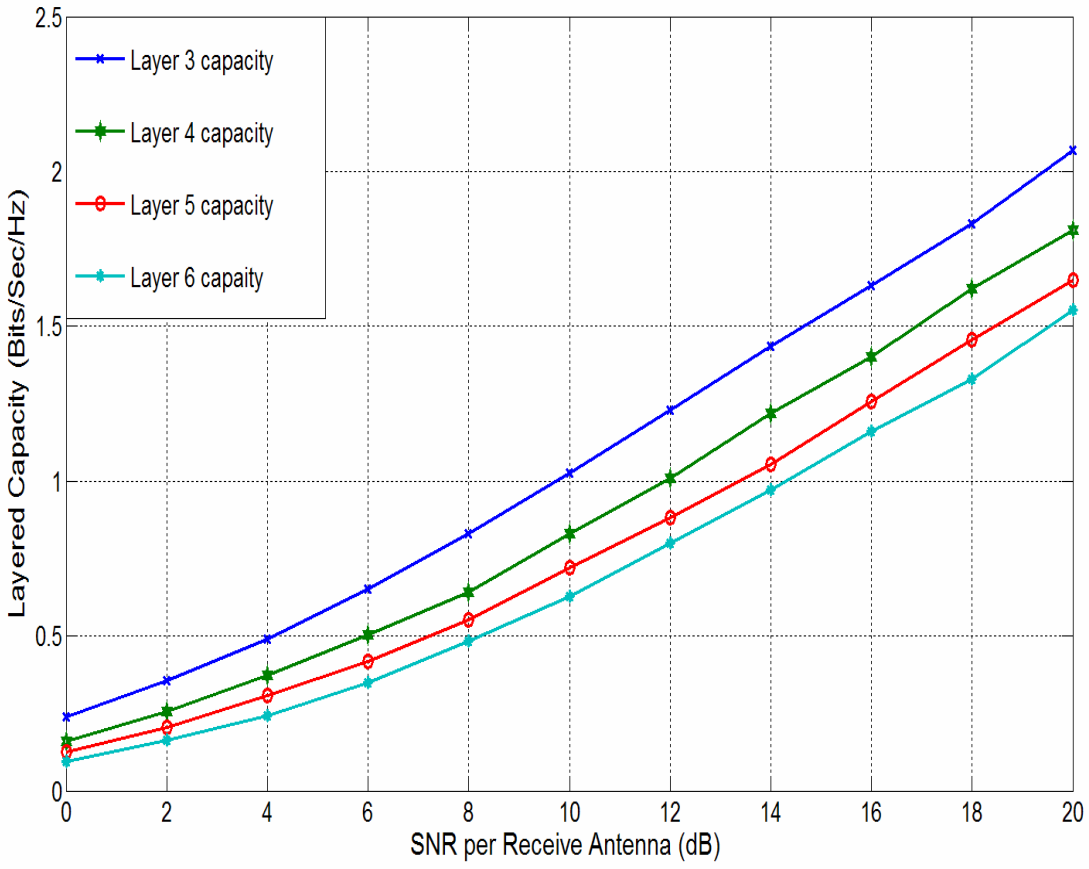

Figure 5: Layered capacity vs SNR, $M = N = 2$, $D = 6$

according to our successive design approach is not optimal in terms of the GLRT detector. We conjecture that the optimal design based on the GLRT detector can make the gap smaller.

Using the obtained successively designed LUSTC signal constellations, we plot the performance of GLRT detection rule at every layer in Fig.7. We observe that the performance of the inner layer signals is always worse than that of the outer layers and the overall GLRT detector performance is almost identical to that of the innermost layer. This is due to error propagation in decoding the layered signals thus the overall performance is bounded below by the innermost layer signals. In practice, the reliability of the outer layer decoding can be strengthened by combining the LUSTC with channel codes (e.g. LDPC or Turbo Codes).

In Fig.8, the performance of the successive designed signal constellations with two different practical decoding algorithms is shown. The parameters in the simulation are $D = 6$, $M = 2$, $N = 2$. The number of signals put in layer $4, 5, 6$ is $171, 6, 4$ respectively according to the previously obtained layered capacity and the spectral efficiency is 2.005 Bits/Sec/Hz. The layered signals are designed according to our heuristic algorithm. The same spectral efficiency optimized unitary space time codes (totally 4104 6×2 unitary matrices) is not available for comparison as it becomes very inefficient to perform the optimization. Seen from the figure, the sphere decoding outperforms the GLRT algorithm for roughly 4 dB .

2.8 CONCLUSION

We considered the problem of communicating over MIMO fading channels using layered unitary space time codes (LUSTC), where no channel information is assumed, but the fading coefficients remain constant for a coherence interval of length D symbol times. A coding theorem of Layered Unitary Space Time Coding scheme was given in the chapter. We showed the channel capacity of applying Unitary Space Time Coding (asymptotically optimal in SNR or channel coherence time) can be achieved by layered signaling in the transmitter and successive decoding in the receiver, with no loss in optimality. The advantage of LUSTC over USTC scheme lied in it's

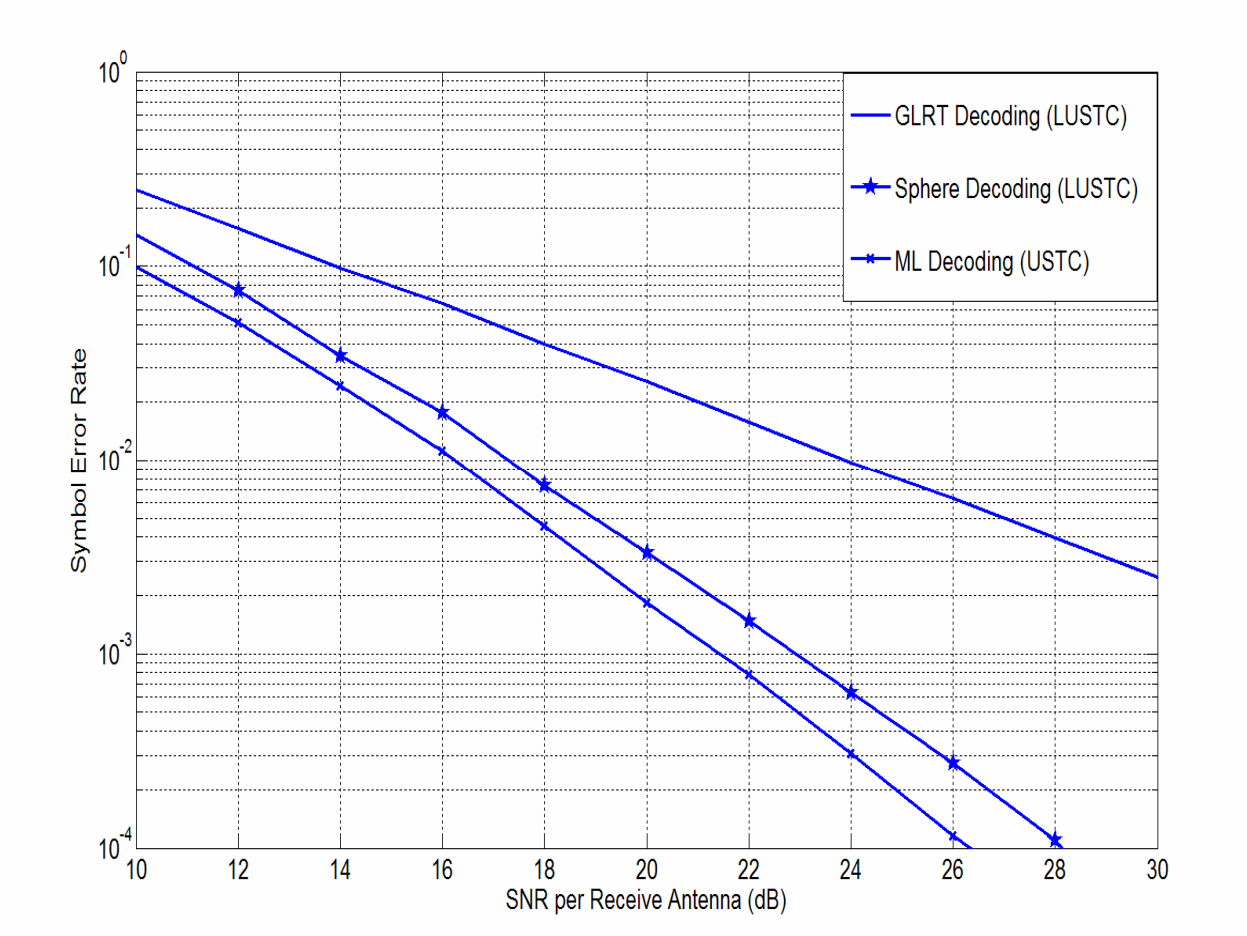

Figure 6: Comparison of optimized USTC and Successive deisgned LUSTC $D = 6$, $M = 2$, $N = 1$

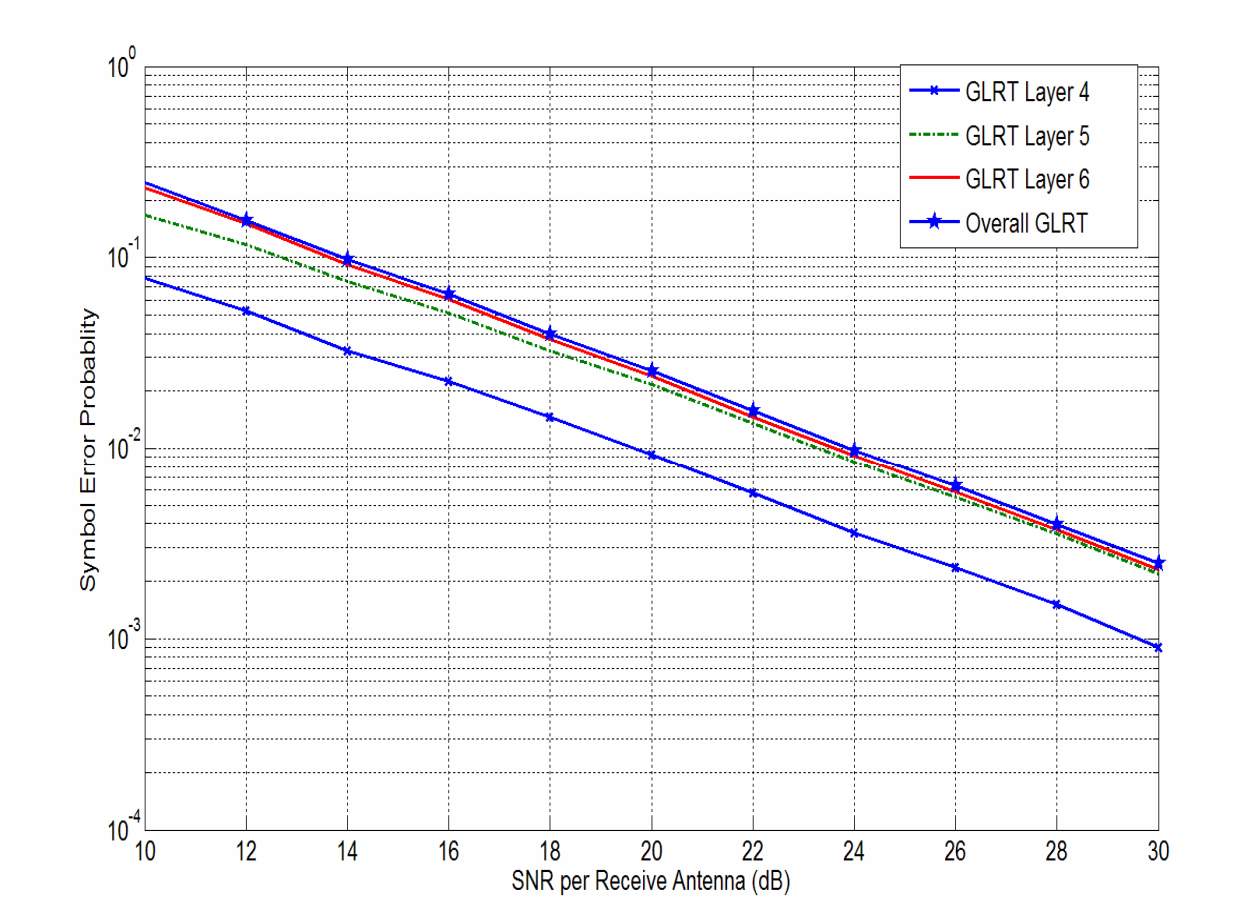

Figure 7: Comparison of GLRT at different layer $D = 6$, $M = 2$, $N = 1$

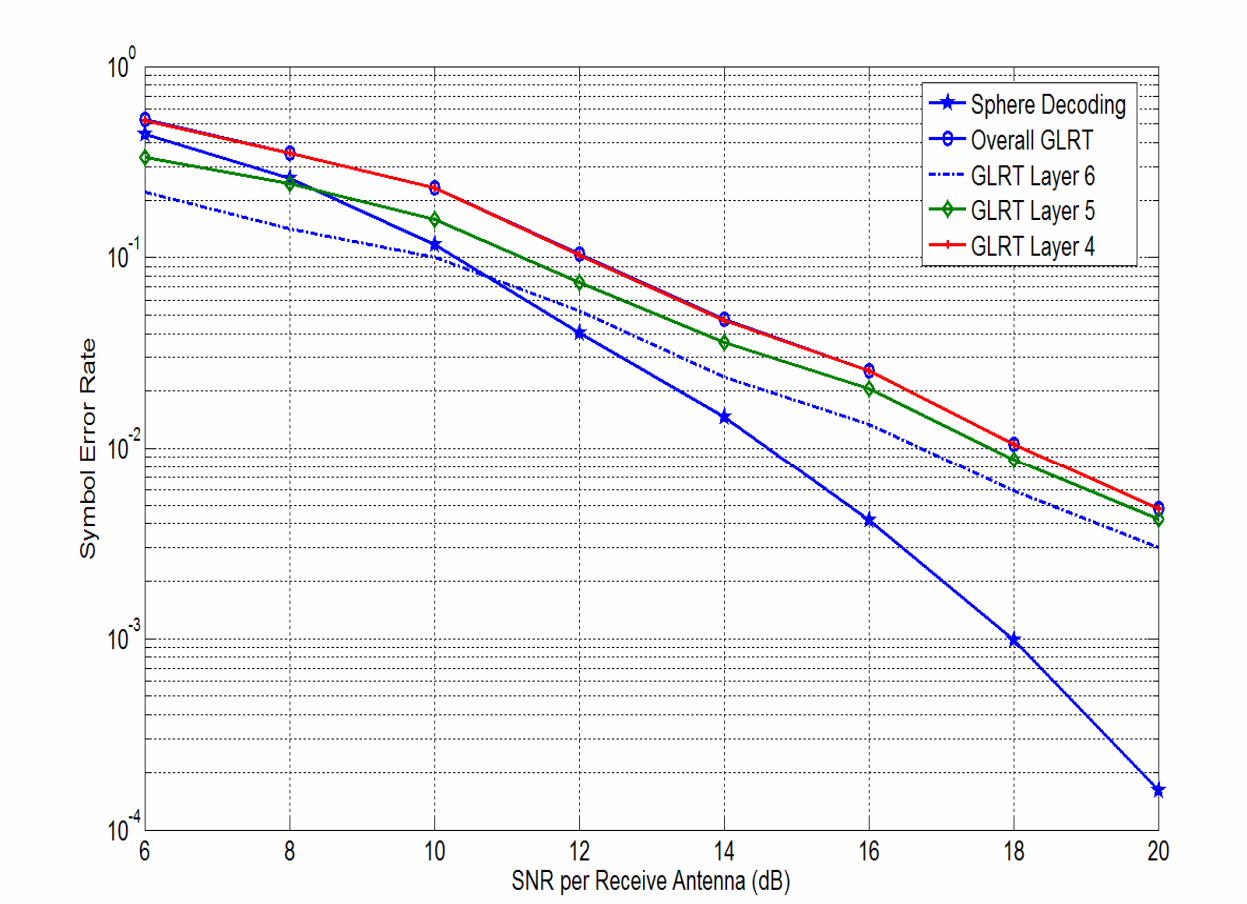

Figure 8: Performance of successive designed signal constellations $D = 6$, $M = 2$, $N = 2$

low complexity decoding algorithm. This coding theorem laid the information theoretic foundation of LUSTC.

We then gave a theoretical design guideline for LUSTC through the layered capacity analysis. The numerical values of the layered capacity were obtained via Monte-carlo simulation then used to design practical LUSTC codebook. Our simulation indicated that more signals should be put in inner layer rather than outer layers. We applied a successive algorithm to effectively design LUSTC signal constellations. The design example given in this chapter was a numerical optimization based on Given's angle parametrization of unitary matrices. In case of designing high spectral efficiency signal constellations with large dimensionality, a heuristic design approach was also provided.

Three different layered signal decoding algorithms namely multistage decoding, GLRT decoding and sphere decoding algorithms were also given in this chapter. The successive decoding, though theoretically simpler, is more complex than the ML decoding of the unitary signals when all the layered signals are constrained in a finite set. The performances of two promising practical decoding algorithms – GLRT and sphere detector were shown in the simulation.

3.0 LIMITED FEEDBACK MIMO SYSTEM

3.1 INTRODUCTION

A well known result of information theory establishes that feedback (of the received symbol at channel output) does not improve the capacity of a discrete memoryless channel [27]. On the other hand, for the cases where the channel is selective in either time, frequency or space, feedback of the channel state to the transmitter can bring substantial benefits to the forward communications system in terms of either capacity, performance or complexity.

The theoretical study of capacity and coding with channel state information at the transmitter (CSIT) can be traced back as early as to Shannon [33] and Dobrushin [34]. For practical applications, researchers have long utilized the channel information for antenna array beamforming which has been known as an effective way to combat fading in wireless environment [35]. More recently, information-theoretic capacity on channels with both perfect [36, 37, 38] and imperfect [39] CSIT and practical coding schemes using CSIT [40] [41] have been extensively studied.

With the advent of Multiple Input and Multiple Output (MIMO) antenna systems, investigation on the potential benefits of CSIT for MIMO systems has been intensified and design of a practical scheme to achieve the potential benefits as closely as possible has become very important. The channel estimation done at the receiver needs to be sent back to the transmitter to obtain the potential CSIT benefit. Unlike for a scalar channel case, for a MIMO system, the number of channel coefficients which need to be fed back to the transmitter is large and a naive feedback method (feedback the scaler quantization of each and every channel coefficients) would require a large

capacity in the feedback channel. Thus, the study of MIMO system with a limited rate feedback is of practical interests. In the past, various options in MIMO transmit beamforming with limited rate feedback have been considered in [42, 43, 44, 45, 46]. In the beamforming setting, however, one notes that the capacity loss is usually large, as compared to the optimal waterfilling (WF) solution [3], especially in high SNR region when multiple receive antennas are used. To remedy this shortcoming, in [1, 2], the problem is approached from the perspective of designing a codebook which contains a finite number of pre-determined matrices. The optimal WF solution is a mixture of optimal antenna phase rotation and power adaptation, which changes subject to a particular realization of the channel. The proposed codebook design methodology in [1] includes both the phase rotation and power allocation matrix. Whereas in [2], the authors propose a codebook design based only on the phase rotation which consequently has some performance degradation as compared to the method in [1].

In this chapter, we attempt to strike a balance between the sub-optimality and the signal processing complexity of the CSI feedback mechanism and propose a new CSI feedback methodology for MIMO systems which maximizes the forward channel capacity at a given feedback rate. Our design is based on a suboptimal waterfilling (SWF) scheme. Under this scheme, either a fixed level of constant power or no power at all is assigned to each subchannel based on a threshold test on the (eigen) channel gain. Making use of upperbound analysis on the capacity difference, we determine that the most important factor for the SWF scheme is the identification of those subchannels to which nonzero transmit power should be allocated. The transmit power is equally divided among these channels, and thus the SWF is simple for implementation. In addition, it incurs small capacity loss.

Inspired by the small capacity loss of the SWF as compared to the WF, we then propose a SWF based transmission scheme due to the smaller feedback rate requirement and implementation simplicity. As the feedback channel subjects to a rate constraint, the practicality of the scheme depends on whether a good limited rate feedback codebook can be designed. We use the generalized Lloyd algorithm [47] and design the codebook which is a finite set of unitary (orthonormal column) ma-

trices. The codebook for a given rate is constructed by a gradient search method applied on Given's angle parameterizations of the unitary matrix. The proposed algorithm, although it is based on a sub-optimal SWF scheme rather than on the optimal WF as in [1], surprisingly outperforms the algorithm in [1] as the numerical results indicate. We believe the gain in performance is largely owing to the optimal computation of the centroid. On the other hand, our proposed algorithm also presents superior performance over the multimode precoding scheme in [2], where the codebook is also a set of unitary matrices. The gain over [2] comes mainly from the fact that the cardinality distribution of the designed codebook employing our proposed algorithm is much closer to optimal (if not exactly optimal) as compared to the precalculated one in [2].

3.2 SYSTEM OVERVIEW

3.2.1 System Model

We consider a MIMO system with M inputs and N outputs. The $[N \times 1]$ output signal vector is modeled as

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{3.1}$$

where $\mathbf{H}$ is a $[N \times M]$ channel matrix with circularly symmetric complex Gaussian entries of zero mean and unit variance; $\mathbf{s}$ is a $[M \times 1]$ transmitted column vector with total power constraint P_t, such as $\mathbb{E}[\mathbf{s}^*\mathbf{s}] \leq P_t$; $\mathbf{n}$ is a $[N \times 1]$ column vector representing the additive white Gaussian noise present at the receiver, with $\mathbb{E}[\mathbf{n}\mathbf{n}^*] = \sigma^2\mathbf{I}_{N\times N}$. Without loss of generality, we assume $\sigma^2 = 1$ throughout the chapter.

3.2.2 Optimal and Sub-optimal Waterfilling

Assume the channel matrix is perfectly known to the receiver. By the singular value decomposition, the channel matrix $\mathbf{H}$ can be decomposed as $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^*$, where both $\mathbf{U}$ and $\mathbf{V}$ are the unitary matrices and $\mathbf{\Lambda}$ is a diagonal matrix with the singular values of $\mathbf{H}$ on its diagonal. The

capacity achieving power allocation solution in such a case is the well known Water Filling (WF), with the analogy of pouring water over a surface. The instaneous WF capacity as given in [3] is:

$$C_w = \sum_{i=1}^{m^*} \log(\mu\lambda_i).^{1} \tag{3.2}$$

μ is the water level and λ_i is the corresponding i-th sub-channel gain (the square of the i-th diagonal element in $\mathbf{\Lambda}$) and $\log$ is the logarithmic with base 2. The parameter m^* denotes the number of sub-channels whose channel gain is greater than the inverse water level, namely $m^* = \sum_i I(\lambda_i > \mu^{-1})$, where $I(\cdot)$ denotes the indicator function. The water level relates to P_t by satisfying the total power constraint

$$P_t \;\; = \;\; \sum_{i=1}^{m^*} (\mu - \lambda_i^{-1}). \tag{3.3}$$

Dividing both sides in (3.3) by m^* and defining $S_0 := \frac{P_t}{m^*}$, we have

$$S_0 := \frac{P_t}{m^*} = \mu - \frac{1}{m^*} \sum_{i=1}^{m^*} \lambda_i^{-1} = \mu - \bar{g}, \tag{3.4}$$

where we define the sub-channel losses as $g_i := 1/\lambda_i$ and the sample-mean of the sub-channel losses as $\bar{g} := \frac{1}{m^*} \sum_{i=1}^{m^*} g_i$. For convenience, we define the set of *good* sub-channels as $\mathcal{E} = \{g_i : \mu - g_i > 0, i = 1, 2, \ldots, M\}$.

Under the SWF scheme, a fixed identical transmit power is allocated to every *good* sub-channel whose channel gain λ_i is greater than μ^{-1}, i.e., the sub-channels which belong to the set $\mathcal{E}$. A typical allocation of power in SWF is depicted in Fig.9. The instaneous channel capacity of the SWF scheme is

$$C_e = \sum_{i=1}^{m^*} \log(1 + P_t\lambda_i/m^*). \tag{3.5}$$

[1] WLOG, we assume the eigenvalues are ordered decreasingly, i.e., $\lambda_j \geq \lambda_k$, for $j < k$.

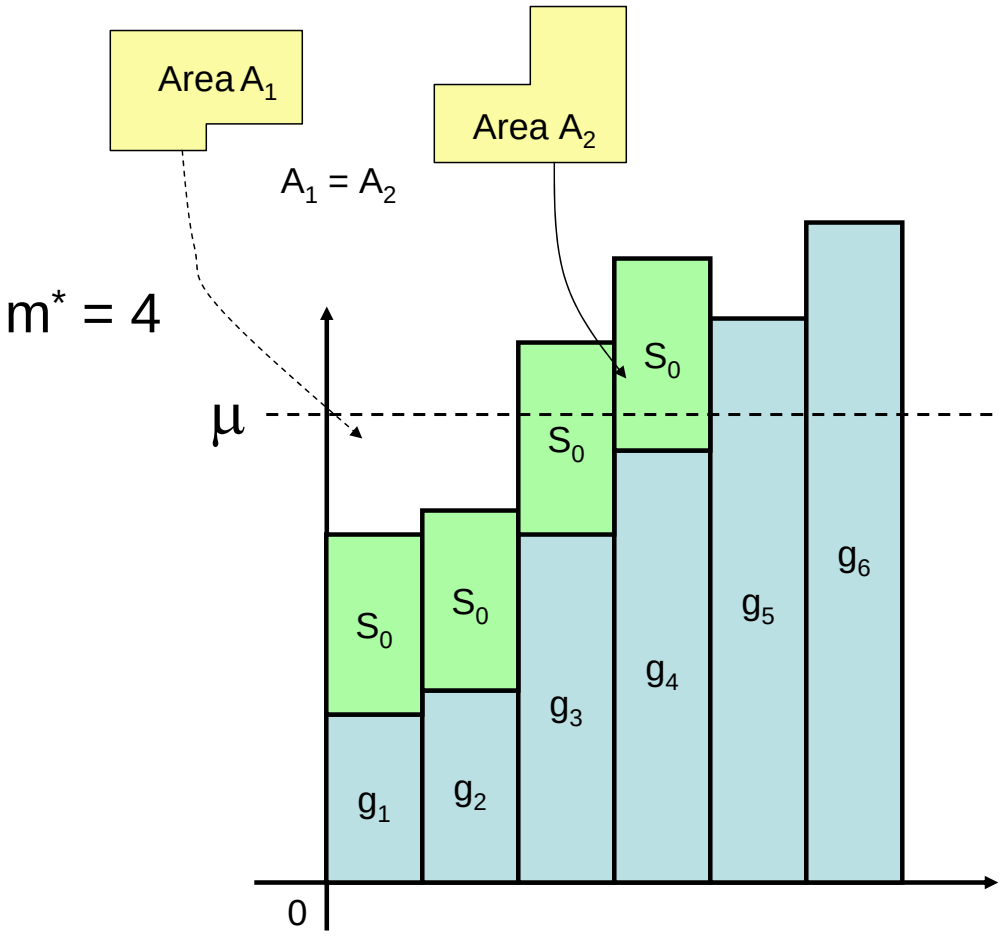

Figure 9: Optimal and Sub-optimal Power Allocation

3.3 TRANSMISSION SCHEMES WITH FINITE RATE FEEDBACK

Under the SWF scheme, only partial information – the first m^* columns of the optimal rotation matrix $\mathbf{V}$ is needed to be fed back to the transmitter; whereas for the WF, the matrix $\mathbf{\Lambda}$ needs to be transmitted back as well. This will decrease the amount of feedback traffic. In this section, we will focus on the design of practical transmission schemes under the SWF due to the lower feedback rate requirement and implementation simplicity of the scheme. However, before we proceed, the capacity loss of the SWF scheme will be examined first.

3.3.1 Capacity Difference between WF and SWF

The instaneous capacity difference between WF and SWF, using (3.5) and (3.2) can be written as

$$\begin{aligned} \Delta_C : &= C_w - C_e = -\sum_{i=1}^{m^*} \log \frac{1 + P_t\lambda_i/m^*}{\mu\lambda_i} \\ &= -\sum_{i=1}^{m^*} \log(1 + \frac{g_i - \bar{g}}{\mu}), \end{aligned} \tag{3.6}$$

where the last step follows from (3.4).

The capacity difference (3.6) depends on the optimal power level μ. Thus, it can only be evaluated numerically as no analytical solution to μ exists. To offer an insight into the capacity difference, we expand the summand in (3.6) by the Taylor series and obtain a series of upper- and lower-bounds. In the following theorem, we have a couple of upper bounds.

Theorem 3 *The capacity difference between the WF and the SWF scheme is upper bounded by*

$$\Delta_C \leq \frac{1}{\ln 2} m^* \frac{\sigma^2(g)}{\mu^2}, \tag{3.7}$$

where $\sigma^2(g)$ denotes the sample-variance *of the sub-channel losses (inverse sub-channel gains) in the set of good sub-channels $\mathcal{E}$. Meanwhile, by using the absolute value of the first order term, we have*

$$\Delta_C \leq \frac{1}{\ln 2} \sum_{i=1}^{m^*} \frac{|g_i - \bar{g}|}{\mu}. \tag{3.8}$$

Proof See Appendix B.

These bounds, though easy to obtain, are tight and provide useful tools to make inference on the loss of a suboptimal solution. It shall be noted that the SWF scheme subsumes the beamforming scheme in low SNR regime and the equal power allocation scheme to all transmit antennas in high SNR regime (when $M \leq N$). Therefore, making use of these upper bounds, we can explain why beamforming and the equal power allocation are asymptotically optimal in the low and the high SNR regime respectively. In the high SNR regime, i.e., $P_t \gg 1$ and $\mu \gg 1$, the upper bound (3.7) becomes zero, and thus the equal power allocation is optimal; in the low SNR regime, i.e., $P_t \ll 1$, $m^* = 1$ and $\sigma^2(g) = 0$, the upper bound also goes to zero, thus beamforming is optimal in this case.

In [48] [49], the authors obtained an upper bound using the duality concept in convex programming. They first set up a constraint optimization problem and solved it partly by using the Lagrangian multiplier method. However, they did not solve for all the optimal values of the Lagrangian multiplier problem; then consequentially, obtained a duality gap. When the power allocation scheme is constrained to be the so called SWF in this chapter, the gap is enlarged further. For the purpose of comparison of tightness, we include here one of the main results of the paper – equation (24) in [49], which is given as:

$$\Delta_C \leq \frac{1}{\ln 2} \sum_{k=1}^{m^*} p_k \left(\frac{\sigma^2/\lambda_k}{S_0 + \sigma^2/\lambda_k}\right), \tag{3.9}$$

where the parameters p_k and σ^2 need to be set. This upper bound (3.9) is a general result which can be applied to a variety of channels with different fading statistics. For the system given in this chapter, the parameters needs to be set as: $p_k = 1$ for all k (This setting makes the problem formulated in [49] become comparable to ours) and $\sigma^2 = 1$. Then, (3.9) can be written as

$\Delta_C \leq \frac{1}{\ln 2} \sum_{i=1}^{m^*} \frac{g_i}{S_0+g_i} =: U_w$ where the right hand side is defined as U_w and S_0 is shown in (3.4). Comparing U_w with the first order upper bound (3.8), we note that the summand in U_w is equivalent to $\frac{g_i}{\mu}$. That is, the denominator $S_0 + g_i$ can be replaced with μ as an approximation since the mean value of $S_0 + g_i$ is μ, as shown in (3.4). Now, let us make a comparison of $\frac{g_i}{\mu}$ with $\frac{|g_i-\bar{g}_i|}{\mu}$ the summand of the first order upper bound. We note that $\frac{g_i}{\mu} \gg \frac{|g_i-\bar{g}|}{\mu}$. This provides us with an approximate idea why even the coarse first order upper bound (3.8) provides tighter bound than U_w does.

Based on the ensemble of randomly generated channel matrix $\mathbf{H}$, the cumulative distribution function (CDF) on the capacity difference between SWF and WF is plotted in Fig.10. The true difference is compared with the lower and upper bounds derived. Selected system parameters are $M = N = 16$ and $\mathrm{SNR} = 10$ dB. The Yu-Cioffi bound is also provided for the purpose of comparison. We confirm that even the first order upper bound (3.8) is much tighter than the Yu-Cioffi bound. More importantly, the capacity difference between WF and SWF is indeed small as reflected in this figure. This suggests the low complexity SWF scheme can be used at the cost of negligible capacity loss.

3.3.2 SWF Based Transmission Scheme with Limited Rate Feedback

The small capacity difference between WF and SWF indicates that feedback of the first m^* columns of the rotation matrix $\mathbf{V}$ is sufficient when a negligible throughput loss is acceptable. However, even feedback of the first m^* columns of $\mathbf{V}$ itself may not be feasible due to the feedback channel rate constraint. Thus, one important issue in SWF implementation is the design of the limited rate feedback codebook. In practice, a designed codebook $\mathcal{C}$ can be stored at both the transmitter and the receiver. Each time when the channel $\mathbf{H}$ changes, the receiver selects a codeword $\mathbf{V}_{k^*}$ according to the selection criteria used in codebook design. For example, $k^* = \arg\max C_p(\mathbf{H}, \mathbf{V}_k)$, which will be discussed in the codebook design in section 3.4. The receiver will feedback the index k^* to the transmitter. At the transmitter side, the codeword $\mathbf{V}_{k^*}$ will be used as a precoding matrix. Under the SWF, the feedback codeword $\mathbf{V}_{k^*}$ is a $M \times m$

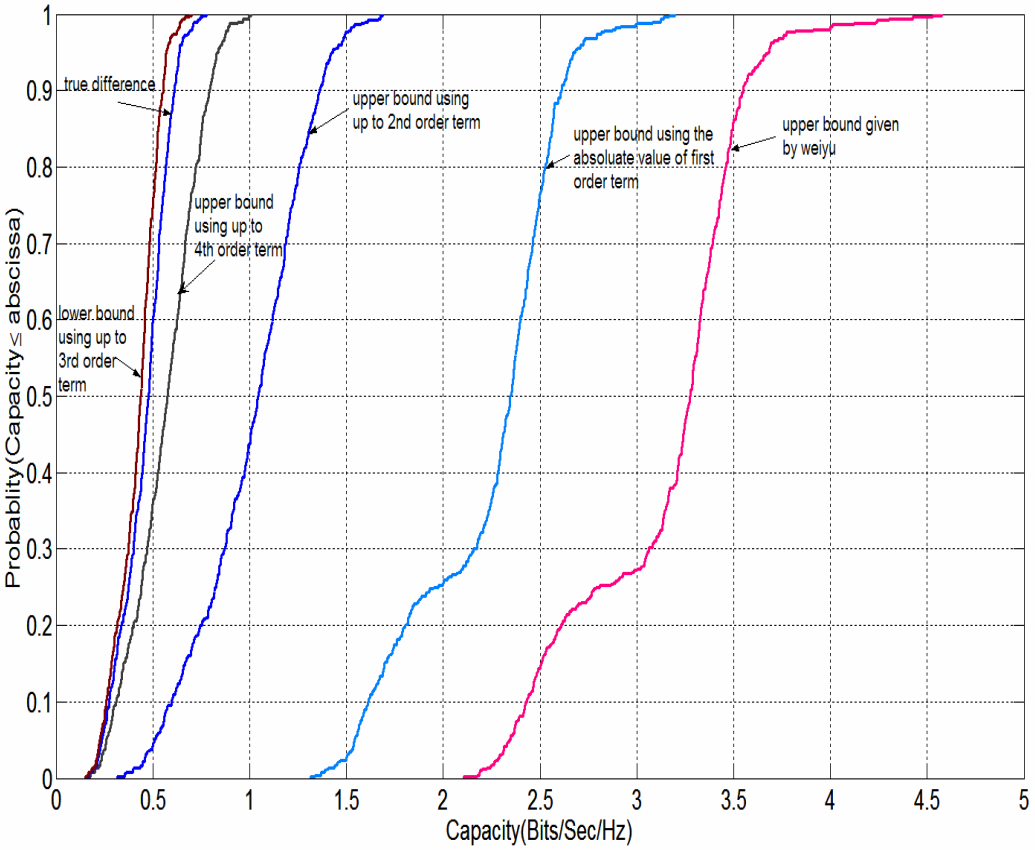

Figure 10: Cumulative Distribution Function of Capacity Difference

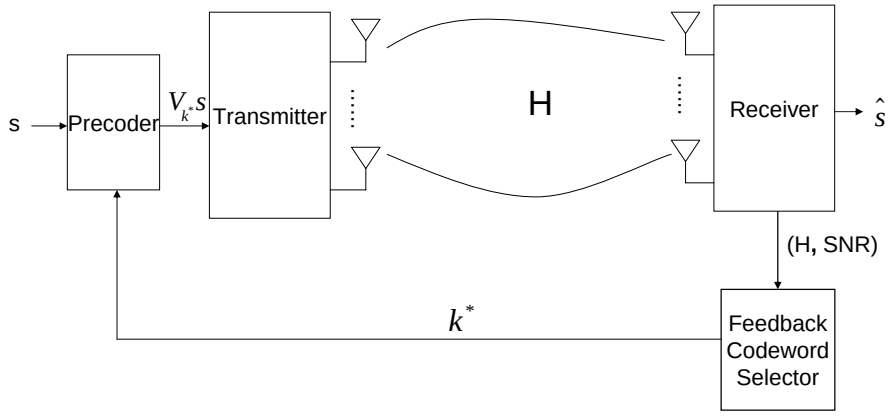

Figure 11: SWF based transmission scheme with limited rate feedback

($1 \leq m \leq M$) unitary matrix and the total transmit power P_t is equally divided among the m eigenmodes. Equal power allocation reduces the implementation complexity, compared to WF solution, which requires different power allocation to different eigenmode. The block diagram of MIMO system with limited rate feedback is shown in Fig.11.

The practicality of the SWF based transmission scheme depends largely upon whether a good limited rate feedback codebook can be designed as it is key to the actual performance of the scheme. The feedback codebook design is a lossy source coding problem and in principal the conventional rate distortion theory can be applied to find the ultimate limit of rate distortion curve. However, a couple of reasons prohibit the use of the rate distortion theory for designing the feedback codebook. First, the rate distortion theory is nonconstructive, which gives the ultimate limit but not a constructive design method. Second, the optimal rate distortion curve is available only in very special cases, such as compression of the Gaussian source with the mean square error dis-

tortion measure or use of Blahut's algorithm [50] to numerically derive the optimal rate distortion curve for discrete sources.

In this chapter, we restrict our codebook design specifically as a vector quantization problem without caring about the information theoretic limit. Under the SWF scheme, the problem of codebook design is to find the optimal quantizer, a deterministic mapping, from the space $\mathcal{H}$ of channel realization to the space $\mathcal{U}$ of unitary (orthonormal columns) matrices with the codebook cardinality constraint. Due to the randomness of the channel matrix $\mathbf{H}$ and consequently the number of good sub-channels m^*, the space of unitary matrices should have the form $\mathcal{U} = \bigcup_{m=1}^{M} \mathcal{U}_m$ with $\mathcal{U}_m$ being the set of complex $M \times m$ unitary matrix.

To solve the problem, one direct way as shown in [2] is to find the cardinality distribution and then design the codewords within each set $\mathcal{U}_m$. However, finding the optimal cardinality distribution itself is not easy as the distribution changes with the given total cardinality constraint. Moreover, intuitively the optimal cardinality distribution has to be dependent upon the channel statistics and the SNR value, which further complicates the problem. Even though the authors in [2] indeed found the cardinality distribution by expressing the system performance, such as capacity and probability of error in terms of the cardinality distribution, the distribution derived there is not exact and does not adapt to the channel statistics and the SNR value.

In Section 3.4, we circumvent the problem of finding the cardinality distribution by using the generalized Lloyd algorithm [47] to directly design the codebook. Utilizing the Lloyd algorithm, the codewords are searched over the whole space $\mathcal{U}$ with the total cardinality constraint rather than over each and every space $\mathcal{U}_m$ and thus the derivation of the cardinality distribution becomes unnecessary. The codebook designed this way naturally has codewords with different dimensionality and the cardinality distribution is adaptive to the channel statics and SNR value.

3.4 FEEDBACK CODEBOOK DESIGN

We now consider the problem of designing the feedback codebook $\mathcal{C}$ which is a finite set of orthonormal column matrices, i.e., $\mathcal{C} \subset \mathcal{U}$, under the the cardinality constraint $|\mathcal{C}| = N_{tot}$. Assuming the feedback channel is error free, a feedback codeword $\mathbf{V}_k \in \mathcal{C}$ ($k = 1, 2, ..., |\mathcal{C}|$) from the receiver serves as the precoding matrix at the transmitter. This then yields a modified input output relationship

$$\mathbf{r} = \mathbf{H}\mathbf{V}_k\mathbf{s} + \mathbf{n}. \tag{3.10}$$

As discussed in section 3.3.2, the problem of codebook design can be formulated as a vector quantization problem and hence the conventional generalized Lloyd algorithm will be applied to find a codebook that optimizes an overall distortion measure. Depending upon the interests of system performance, various criteria, such as probability of error [2, 51], capacity [2, 1] and the error exponent [52] can be utilized to design the codebook. In this section, our design goal is to find the codebook $\mathcal{C}$ using the Lloyd algorithm which maximizes the forward channel (3.10) capacity. It is equivalently the codebook that minimizes capacity loss compared to the optimal WF. The capacity loss is thus employed as our distortion measure, i.e.,

$$d(\mathbf{H}, \mathbf{V}_k) = C_w(\mathbf{H}) - C_p(\mathbf{H}, \mathbf{V}_k), \tag{3.11}$$

where C_w is the water-filling capacity defined in (3.2) and C_p is the forward channel capacity with $\mathbf{V}_k$ as the precoder

$$C_p(\mathbf{H}, \mathbf{V}_k) = \log\det(\mathbf{I}_N + \frac{P_t}{m}\mathbf{H}\mathbf{V}_k\mathbf{V}_k^*\mathbf{H}^*), \tag{3.12}$$

where $\mathbf{V}_k$ has the dimensionality $M \times m$.

According to the generalized Lloyd algorithm, a set of channel matrix $\mathbf{H}$ will be generated randomly according to a given channel statistics as the training sequence. We also randomly generate N_{tot} number of orthonormal column matrices as the initial codebook. The key to the generalized

Lloyd algorithm is to iteratively convert a given codebook to a new and improved one in the sense that the distortion measure with respect to the updated one is decreased.

In each iteration of updating the codebook, we first find the optimal partition of the training sequence according to the given codebook, i.e.,

$$\begin{aligned} \mathbf{V}_{k^*} &= \underbrace{\arg\min_{\mathbf{V}_k \in \mathcal{C}} \left\{ C_w(\mathbf{H}) - C_p(\mathbf{H}, \mathbf{V}_k) \right\}}_{T(\mathbf{H})} \\ &= \arg\max_{\mathbf{V}_k \in \mathcal{C}} C_p(\mathbf{V}_k). \end{aligned} \tag{3.13}$$

Define the k-th cluster as $\mathcal{R}_k = \left\{ \mathbf{H} : T(\mathbf{H}) = \mathbf{V}_k \right\}$ and the k-th partial distortion can be written as

$$D(\mathbf{V}_k) = \mathbb{E}\left[d(\mathbf{H}, \mathbf{V}_k) | \mathbf{H} \in \mathcal{R}_k \right]. \tag{3.14}$$

To update the codebook, we need to recompute the optimal centroid (new codeword) within each cluster $\mathcal{R}_k$ such that the distortion in the cluster is minimized, i.e.,

$$\mathbf{V}_k = \arg\max_{\mathbf{V}_k \in \mathcal{U}} D(\mathbf{V}_k),$$

where different from (3.13), the maximization is taken over the space of all unitary matrices $\mathcal{U}$ rather than a finite set of unitary matrices.

Once the new codebook is generated, we can repeat the previous process until the overall distortion $D = \sum_{k=1}^{|\mathcal{C}|} D(\mathbf{V}_k) p_k$ has changed little since the last iteration, where p_k is the probability that a specific channel matrix $\mathbf{H}$ falls into the cluster $\mathcal{R}_k$.

The detailed design algorithm is summarized as follows:

step 0. Specify the total available transmit power P_t.

step 1. Randomly generate the channel matrix $\mathbf{H}$ as the training sequence according to a given distribution.

step 2. Randomly generate an initial codebook $\mathcal{C}$ with N_{tot} codewords.

step 3. Repeat the following steps until the distortion has changed only by a small enough amount since the last iteration.

(a). *Given a codebook $\mathcal{C}$ redistribute each channel matrix $\mathbf{H}$ into one of the clusters in $\mathcal{C}$ by selecting the one whose centroid is closer to* $\mathbf{H}$, i.e.,

$$\mathbf{H} \in \mathcal{R}_{k^*} \iff d(\mathbf{H}, \mathbf{V}_k^*) \leq d(\mathbf{H}, \mathbf{V}_k) \text{ for } k^* \neq k.$$

(b). *Recompute the centroid for each cluster $\mathcal{R}_k$ just created, i.e., $\mathbf{V}_k = \arg\min_{\mathbf{V}_k \in \mathcal{U}} D(\mathbf{V}_k)$ to obtain a new codebook $\mathcal{C}$.*

(c). *Compute the overall distortion D for the new generated codebook $\mathcal{C}$.*

(d). *If an empty cluster was generated in (a), randomly generate an alternative codeword $\mathbf{V}_k$.*

The generalized Lloyd algorithm finds the codebook in iterations. A few issues of the algorithm need to be further addressed.

3.4.1 Random Generation of Initial Codebook

In the second step of the algorithm, we randomly generate N_{tot} unitary matrices as the initial codebook. However, noting from (3.12), the capacity of SWF applying any unitary precoding matrix $\mathbf{V}_k$ with dimensionality $M \times M$ is $C_p(\mathbf{H}, \mathbf{V}_k) = \log\det\left(\mathbf{I}_N + \frac{P_t}{M}\mathbf{H}\mathbf{V}_k(\mathbf{H}\mathbf{V}_k)^*\right) = C_p(\mathbf{H}, \mathbf{I}_M)$. This means that when more than one $M \times M$ codewords are presented, they essentially act as a single one. Thus, any codewords with dimensionality $M \times M$ can be replaced by $\mathbf{I}_M$ for practical design purpose. In our design algorithm, we start the initial codebook with at most one $M \times M$ unitary matrix. This is to ensure that the final codebook $\mathcal{C}$ meets the cardinality constraint, i.e., $|\mathcal{C}| = N_{tot}$; otherwise, the final codebook generated using the algorithm will have less than N_{tot} number of codewords.

3.4.2 Computation of the Centroid

One difficult part of the Lloyd algorithm is the computation of the centroid (3.14). In conventional scaler or vector quantization problems with the Euclidian distance measure [53], an explicit expression of computing the centroid can be obtained. However, this is mathematically intractable in our problem as the distortion measure (3.11) employed is non-linear. In [1], the authors

employed a similar Lloyd algorithm with a difference such that their design combines optimal antenna phase rotation and power adaptation, as compared to phase rotation only in the SWF scheme. They used a heuristic approximation to compute the centroid rather than an exact derivation. Their approximation enjoyed a closed form expression, but it inevitably has a performance degradation.

In this chapter, we setup an optimization problem to compute the centroid (3.14). In our SWF scheme, $\mathbf{V}_k$ is constrained to be unitary. To facilitate the solution of the optimization problem, we parameterize the unitary matrix using the Given's angle rotation introduced in Chapter 2.

By this parametrization, any $M \times M$ unitary matrix $\mathbf{U}$ can be written as a product of Given's rotation matrices and a diagonal matrix, i.e.,

$$\mathbf{U} = \mathbf{U}_{\mathbf{\Lambda}} \prod_{p=1}^{M-1} \prod_{q=M}^{p+1} \mathbf{U}^{p,q}(\phi_{p,q}, \sigma_{p,q}), \tag{3.15}$$

where $\mathbf{U}_{\mathbf{\Lambda}} = \mathrm{diag}\big(\exp(i\delta_1), \cdots, \exp(i\delta_M)\big)$. We choose to parameterize any $M \times m$ unitary matrix $\mathbf{U}_m$ in a similar manner as

$$\mathbf{U}_m = \mathbf{U} \begin{bmatrix} \mathbf{I}_m \\ \mathbf{0}_{(M-m)\times m} \end{bmatrix}.$$

Let Θ be the collection of M^2 parameters (σ, ϕ and δ). By this parametrization, the computation of centroid problem (3.14) becomes an unconstraint optimization over the parameter set Θ. To obtain the optimal solution, we randomly generate the initial values of the parameters and then update them along the direction of the gradient. The gradient with respect to the parameter set Θ is derived in Appendix C.

The optimization problem is solved for each and every m with $1 \leq m < M$. For $m = M$, we simply set $\mathbf{V}_k = \mathbf{I}_M$ as discussed in Section 3.4.1. We then choose from those M optimized unitary matrices the one that gives the minimum $D(\mathbf{V}_k)$ as the k-th centroid.

3.4.3 Adaptive Codebook Design

The proposed codebook design algorithm works not only for uncorrelated fading, it can also be applied to generate codebooks for correlated fading environments. In the case when the fading has a certain correlation, the only required modification is to randomly generate a training sequence (channel matrices) according to the correlated distribution (Step-1 of the algorithm). We can then utilize the algorithm to obtain the codebook for the particular distribution.

On the other hand, the designed codebook should vary with the SNR of the system as discussed in section 3.3.2. However, codebook for each SNR is burdensome in design overhead as well as in adaptive feedback effort. It would be desirable in some cases to use a codebook which can work for a range of SNR. In practice, we can partition the whole SNR region of the system into smaller regions and design codebooks for each one. As long as the operating SNR of the system does not change very fast, we can designate a codebook to use with little overhead incurred.

The goal of designing the codebook that works for a range of SNR can be achieved with slight modifications of the proposed algorithm. Specifically, Step-0 of the algorithm needs to be modified. We assume the instaneous operating SNR is a random variable taking values in a range according to a given SNR distribution. Then, in Step-0 of the algorithm, other than specifying a single value for power P_t, we randomly generate P_t according to the given distribution for each and every training channel matrix. On the other hand, as the distortion measure (3.11) is a function of P_t, we need to modify it as well to strike a design balance among different SNR values. The distortion measure employed in this case is the normalized capacity difference, i.e., $d'(\mathbf{H}, \mathbf{V}_k) = 1 - \frac{C_p(\mathbf{H},\mathbf{V}_k)}{C_w(\mathbf{H})}$.

3.5 NUMERICAL RESULTS

The empirical cumulative distribution function of the instaneous channel capacity with feedback is shown in Fig.12 for a 3×3 MIMO system. Our codebook was designed according to the generalized Lloyd algorithm. In the design, 1000 channel matrices $\mathbf{H}$ were randomly generated as

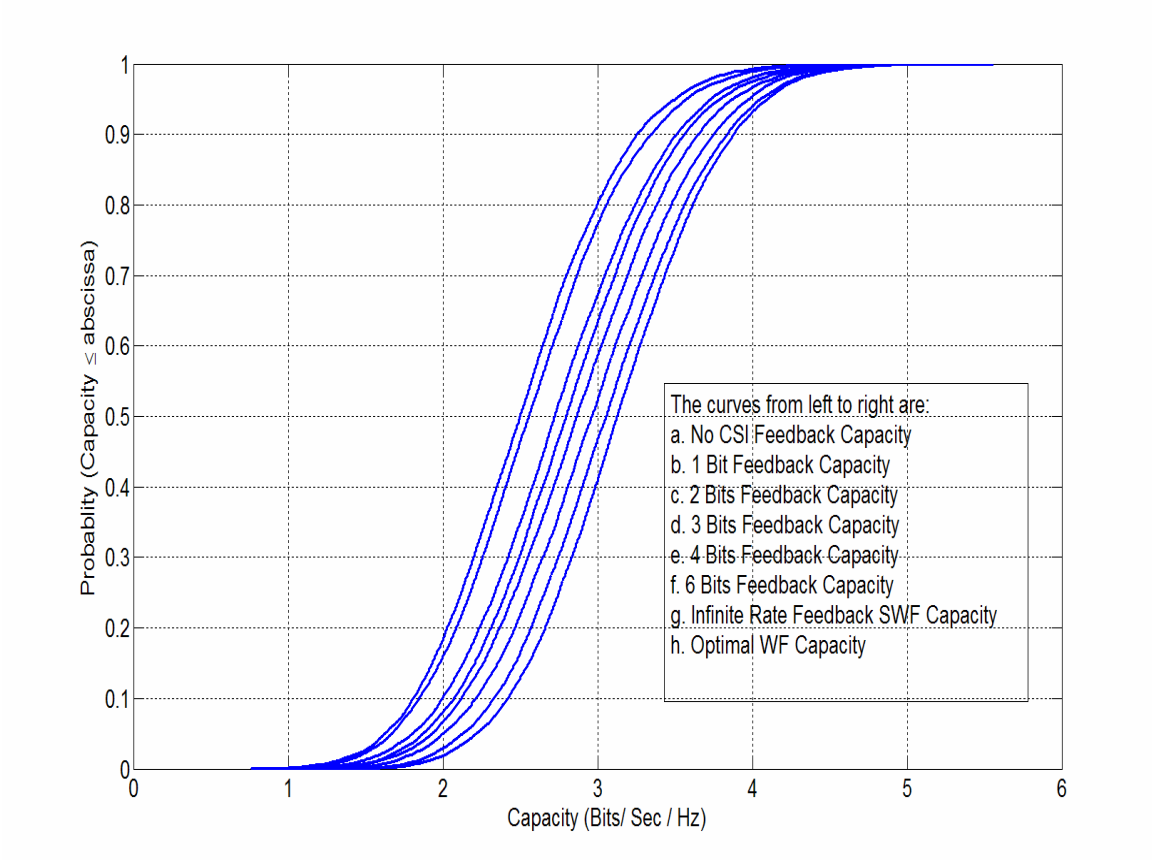

Figure 12: Cumulative Distribution Function of Capacity with Limited Rate Feedback for $3\times$ 3 MIMO System

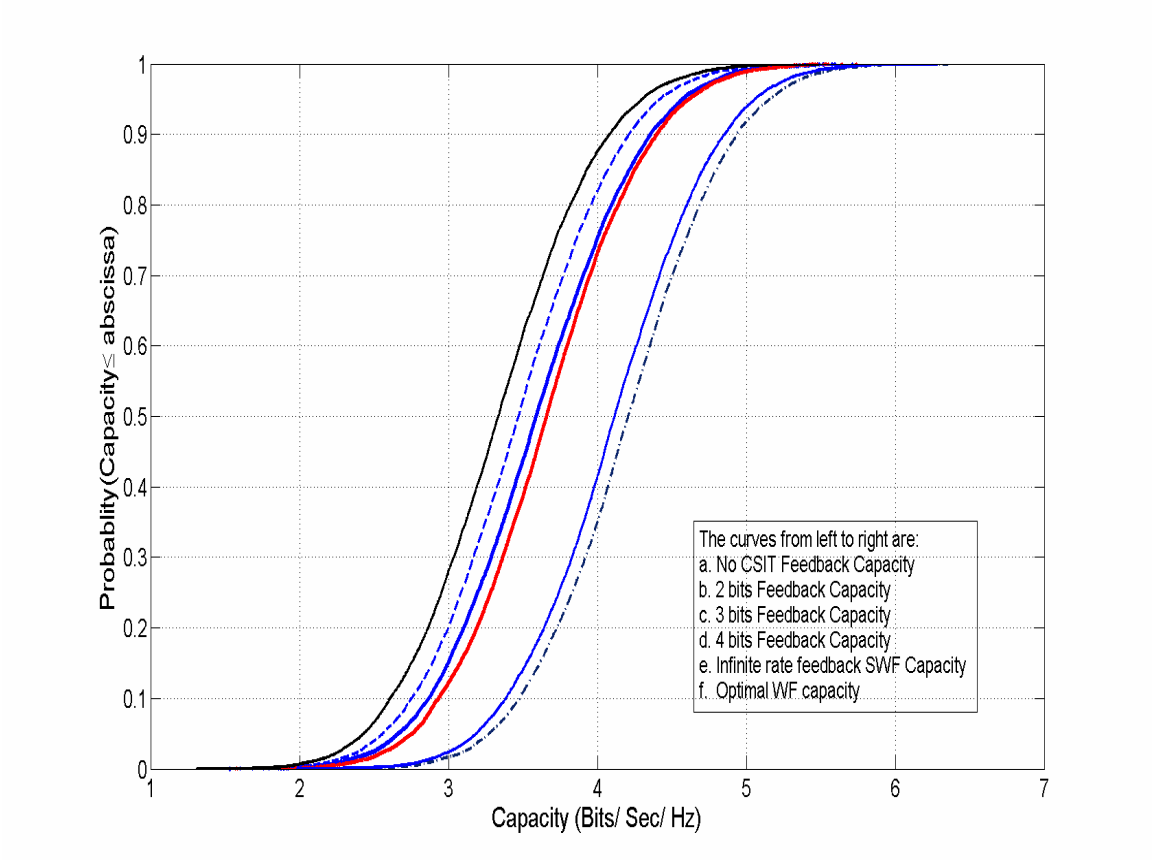

Figure 13: Cumulative Distribution Function of Capacity with Limited Rate Feedback for $4\times$ 4 MIMO System

the training sequence.[2] The targeted operating SNR is 0 dB. It is worthwhile to point out that as the codeword distribution in the feedback codebook is not necessarily uniform, the feedback rates (2, 3, 4 and 6 bit) listed in the figure are essentially upperbounds on the rate. We then randomly generated 10000 samples of $\mathbf{H}$ and calculated the forward channel capacity with the designed codebook at $\mathrm{SNR} = 0$ dB. The numerical results indicate that larger forward channel capacity will be achieved with more feedback rate. However, the benefit obtained from increasing the feedback rate from 1 bit to 2 bit is much larger than the benefit obtained by increasing the rate from 2 bit to 3 bit. As shown in the figure, the capacity with 6 bit feedback is very close to the infinite rate feedback SWF capacity. Similar result is also shown in Fig.13.

In Fig.14, the normalized (with respect to the optimal WF capacity) average channel capacity with feedback is shown for 4×4 MIMO system. The scheme by Lau, Liu and Chen (covariance feedback scheme) in [1] and the multimode precoding scheme in [2] are also depicted for comparison purpose. In multimode precoding, we use the codebook distribution derived from the capacity allocation criterion in [2] for fair comparison. For example, for 3 bit feedback case, we have $\mathrm{card}(\mathcal{U}_1) = 4$, $\mathrm{card}(\mathcal{U}_2) = 3$, $\mathrm{card}(\mathcal{U}_4) = 1$. The codebook in multimode precoding scheme is generated to minimize the Fubini-study distance of each mode as in [2]. The codebook in covariance feedback scheme is designed according to [1], which is a combination of power allocation and antenna phase rotation matrices.

From the simulation results, we see our SWF scheme outperforms both the covariance feedback and the multimode precoding schemes throughout all SNR. The gain of our SWF scheme over multimode precoding scheme comes from the fact that the codebook distribution according to the generalized Lloyd algorithm is closer to optimal as compared to the precalculated one in the multimode precoding scheme. On the other hand, our designed codebook based on a sub-optimal SWF scheme surprisingly outperforms the optimal WF based covariance feedback scheme. We believe that the gain comes from the more exact computation of the centroid.

[2]Based on the numerical results, changing the number of training sequence from 1000 to 10000 or even larger did not bring much benefit.

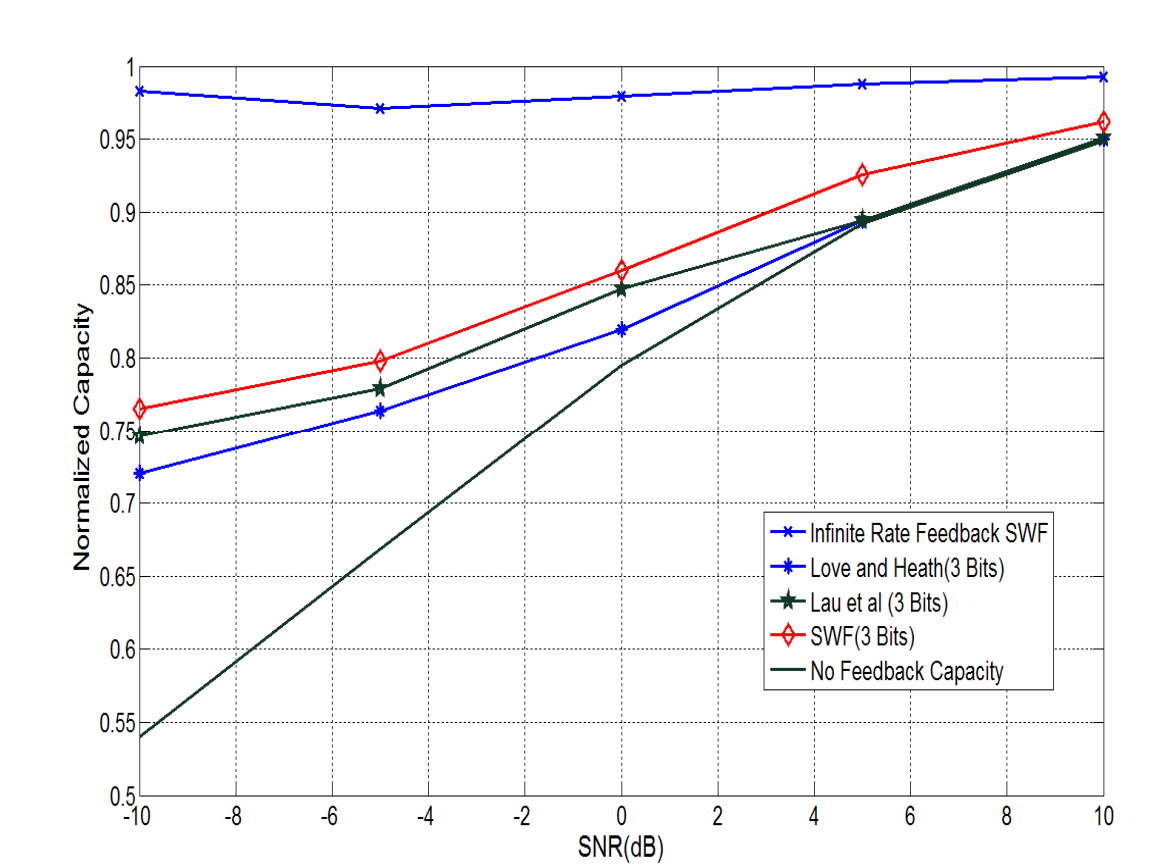

Figure 14: Capacity comparison of SWF limited rate feedback capacity with covariance feedback [1] and multimode precoding [2] schemes for 4×4 MIMO system

In Fig.15, the normalized average channel capacity with feedback is shown for 4×4 MIMO system with correlated channel matrix $\mathbf{H}$. The spatially correlated channel matrix $\mathbf{H}$ is generated by $\mathbf{H} = \mathbf{\Psi}_t^{1/2}\mathbf{H}_w(\mathbf{\Psi}_r^{1/2})^*$ [54], where $\mathbf{H}_w$ is uncorrelated channel matrix. In the simulation, we set $\mathbf{\Psi}_r = \mathbf{I}_M$ and $\mathbf{\Psi}_t$ is a Teoplitz matrix with the first row as $(1, 0.9048, 0.8187, 0.7408)$. Similar to the results in Fig.14, our scheme outperforms the multimode precoding scheme throughout all SNR with the same feedback rate. In some SNR region, even a lower rate feedback rate SWF scheme surpasses the performance of a higher rate feedback multimode scheme. For example, our 3 bit feedback SWF scheme performs better in low SNR region and almost the same in high SNR region as compared to the 4 bit feedback multimode scheme. Moreover, the superiority of our scheme to the multimode precoding scheme is more noticeable than that in Fig.14. This is due to the fact that the codebook distribution in the multimode scheme is derived from uncorrelated fading statistics. Ideally the gap would be smaller if the exact codebook distribution based on the specific fading statistics can be derived. However, deriving such an exact codebook distribution would be very difficult in the multimode precoding scheme if not impossible.

In Fig.16, we depict the normalized forward channel capacity of our SWF codebook design in the case when the operating SNR of the system varies. As pointed out in [2], it is assumed the codebook has to be redesigned when SNR changes in covariance feedback scheme. Thus, we only compare our SWF scheme with the multimode precoding scheme for fair comparison.

The results given in Fig.16 are for the uncorrelated 4×4 MIMO channel. In our adaptive codebook design, the SNR of the system is assumed to be uniformly distributed from -3 to 3 dB. Different from the previous setting, the training sequence used to design the codebook has 10000 channel matrices in our SWF scheme. Seen from Fig.16, our adaptive SWF scheme again outperforms the multimode precoding scheme for the same feedback rate, although the difference is smaller as compared to that in Fig.14. In some SNR region, even the performance of a lower rate feedback adaptive codebook surpasses that of a higher rate multimode scheme. For example, our 2 bit feedback SWF scheme outperforms 3 bit feedback multimode precoding scheme at SNR from -1 to 3 dB.

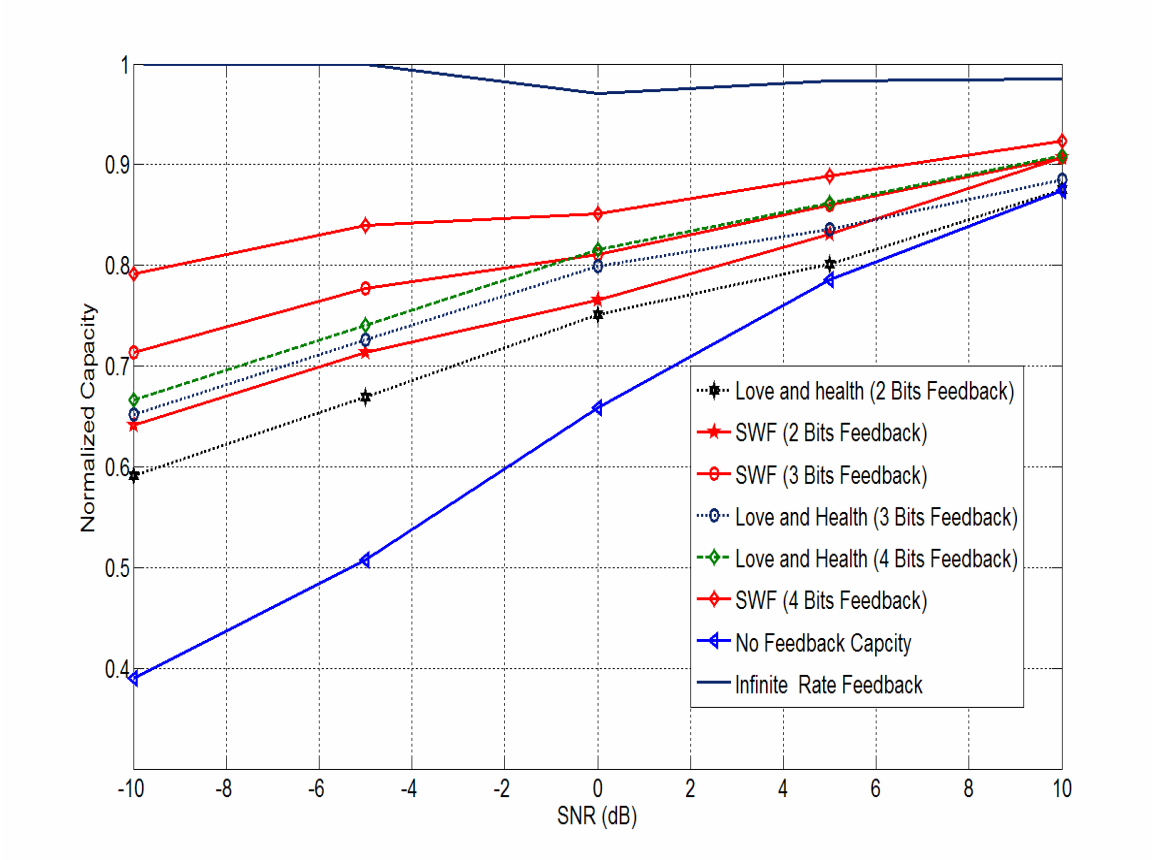

Figure 15: Capacity comparison of SWF limited rate feedback capacity with multimode precoding [2] for 4×4 spatially correlated MIMO system

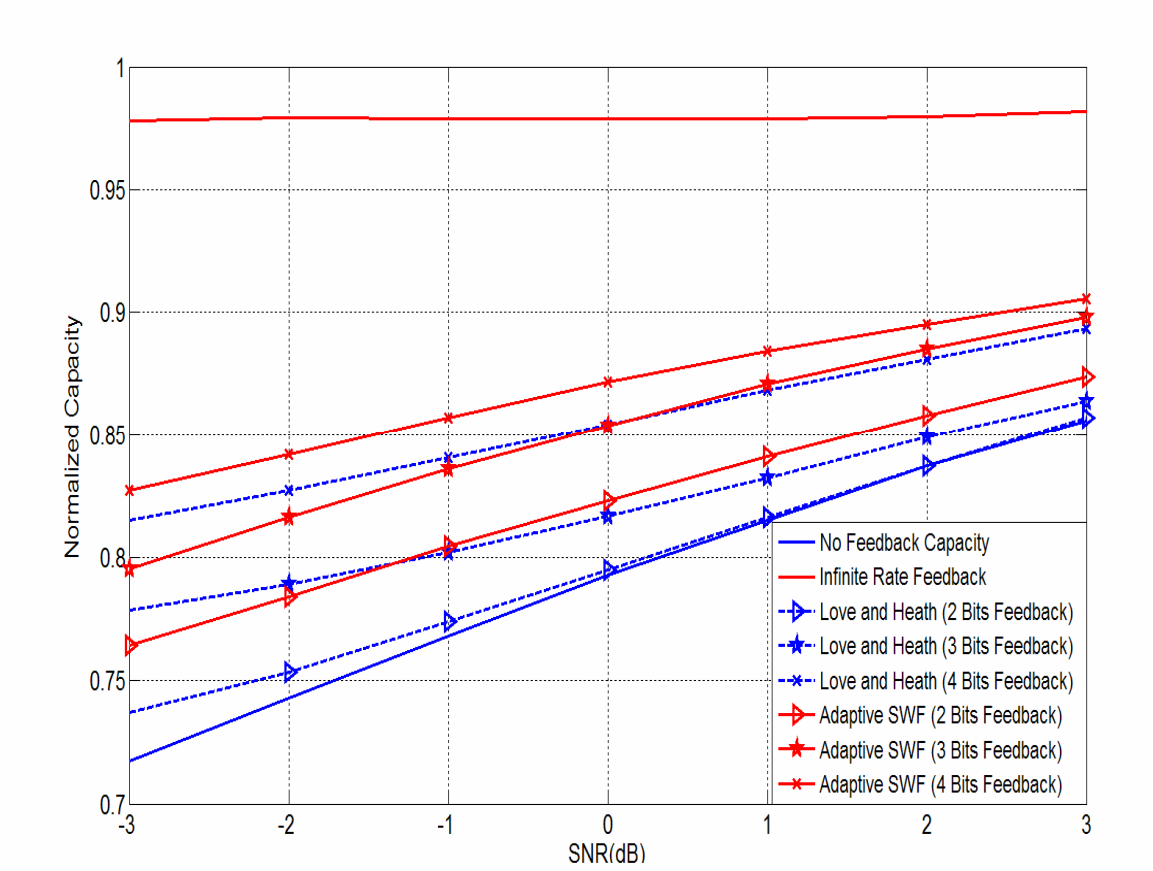

Figure 16: Capacity comparison of adaptive SWF limited rate feedback capacity with multimode precoding [2] for 4×4 MIMO system

3.6 CONCLUSION

In this chapter, we examined two different power allocation schemes over multiple input multiple output (MIMO) antenna system. They are namely the optimal water filling (WF) and sub-optimal water-filling (SWF) schemes. For SWF, equal power is allocated to every sub-channel whose gain is above a threshold level. A number of bounds on the capacity difference between WF and SWF are obtained in this chapter. These bounds explain why SWF is close to optimal WF in throughput.

Utilizing this SWF scheme, a limited feedback codebook design methodology is proposed which is to find a mapping from the space of channel realization to a finite set of unitary matrices. We used the generalized Lloyd algorithm to design the optimal codebook with respect to a distortion measure – the loss of forward channel capacity of limited feedback scheme compared to the optimal WF. One novelty of our design methodology lies in the computation of the centroid, which we set it as an optimization problem by the Given's angle parametrization of unitary matrix. The proposed algorithm is adaptive to the channel statistics and the SNR value. Numerical results showed that the proposed algorithm outperforms comparable algorithms reported in the literature.

4.0 CODED USER COOPERATION IN WIRELESS NETWORKS

4.1 INTRODUCTION

Inspired by the great capacity and diversity improvement of multiple input multiple-output (MIMO) systems using arrays of antennas, lots of researchers are now trying to improve the performance of the wireless networks by utilizing multiple transmit and receive antennas. However, it may not be feasible to deploy multiple antennas in a single wireless node due to the space limitation. Thus, user cooperation among different wireless nodes has been widely studied to increase the communication capability and information transmission reliability of the wireless nodes and thus to improve the overall performance of the networks.

The study of user cooperation was first initiated in the seventies by Van Der Meulen in his pioneering work "Three-terminal communication channels" [55]. Later on, Cover and El Gamal considered the information theoretic aspect of the relay channel and came up with an achievable coding scheme [56]. Recently, inspired by the huge diversity advantage of Space - Time coding, Laneman and Wornell proposed a distributed Space-Time Coded protocols to exploit the cooperation diversity in wireless networks [57, 58]. In their protocol, two or more active nodes jointly transmit their messages towards their destinations. Each wireless transmission can be overheard by neighboring nodes, which then process and retransmit the received signal to the final destination to provide additional reliability. The partnering nodes act as a virtual antenna and help to improve the reliability of the wireless link by providing a diversity transmission. Since then, wireless user cooperation has become one of the most popular topics in wireless networks.

Hunter and Nosratinia [59, 60] proposed a new paradigm for user cooperation - coded user cooperation. Different from space-time cooperation, coded cooperation integrates user cooperation with channel coding. Instead of repeating the received information in some form, the user decodes the partner's transmission and transmits additional parity symbols (e.g. incremental redundancy) according to a coding scheme. Bao and Li [61] further generalized the idea of coded cooperation from two users to of many (one thousand) users by utilizing long block code - Low Density Generator matrix codes (LDGM code). In their scheme, if users send their own information simultaneously to a common destination node, e.g., a base station in cellular system or a wireless access point in wireless LANs, they will cooperate together by randomly forming a LDGM code. The cooperation scheme has two phases. In the first phase, the transmitting nodes broadcast their messages to both their neighbors and the final destination node. In the second phase, each node selects some messages that were received correctly from their neighbor nodes to from a parity check sum and transmit it to the final destination node. It was shown numerically that by utilizing LDGM code, huge gain of some 20 - 40 dB can be achieved over conventional repetition schemes.

However, the assumption made in [61] that the cooperating nodes can detect whether the information received from other nodes is correct and thus only select the correct messages to form a parity check bit is unrealistic. With their signaling scheme, the information transmitted to the cooperating nodes in the first phase are un-coded, which results in very high chance of the transmitted bits being corrupted. In this chapter, on the contrary, the inter-user channel (the channel from the transmitting nodes to their partners) is modeled to be noisy, and no knowledge on whether a received message is correct is assumed at the partnering node.

The main contribution of the chapter is as follows: Under the realistic inter-user channel model, we modified the conventional sum-product message passing decoding algorithm for Low Density Parity Check (LDPC) code and LDGM code to incorporate the presence of the inter-user channel errors in decoding. The modified algorithm differs from the conventional sum-product algorithm only in the initialization part of the decoding process. Applying this modified algorithm, empirically we show the performance gain of coded user cooperation scheme over non-cooperative

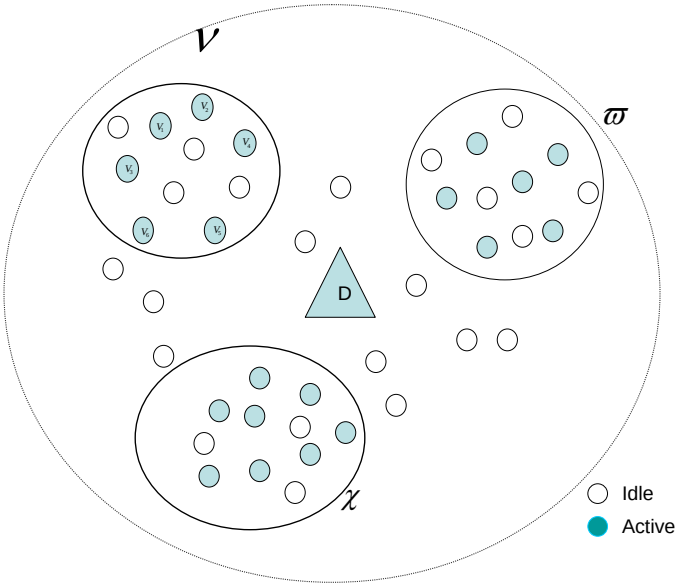

Figure 17: Cooperative network model. The active wireless nodes are clustered as different cooperation groups by their geographic location.

scheme. We also utilize a number of analysis tools (union bounds and density evolution methods) and analyzed the performance of the system.

4.2 SYSTEM MODEL

4.2.1 System Description

Fig.17 depicts the model considered throughout the chapter. The model of interests comprises a set of wireless nodes transmitting messages towards a single common destination D. The model is applicable to a wide range of wireless networks, such as cellular and wireless LANs. For example, in cellular system, the wireless nodes are mobile stations and the destination D is the base station. The transmission from wireless nodes to the destination is then uplink or reverse link transmission, i.e., from the mobile stations to the base station.

The links from the transmitting nodes to the destination suffer from signal strength fading due to multipath propagation in wireless environment. If each of the wireless nodes transmits information independently, the transmitted signal can be severely attenuated due to fading and it results in high bit error probability. To improve the link reliability, various forms of diversity, e.g. temporal diversity, frequency diversity and spatial diversity, need to be exploited. On the other hand, the information sent to the destination will be overheard by other neighboring wireless nodes due to the broadcast nature of wireless transmission. In such a case, independently of whether other forms of diversity can be exploited, the neighboring nodes can choose to cooperate in transmitting the information. The benefit of diversity which can be obtained from cooperation has been named as cooperative diversity.

Motivated by the success of MIMO system, several cooperative protocols appeared in the literature. Some utilize the cooperating nodes as virtual antennas. In these schemes [57, 58], the cooperating node either decode or amplify the received message and then retransmit it towards the destination node. However, rather than single forwarding of the information separately for each neighboring node, each cooperating node can form a parity check sum and transmit the parity check bit. This will maximize the diversity and the coding benefit.

4.2.2 Coded Cooperation Protocol

As done in many current wireless networks, we divide the available bandwidth into orthogonal channels and allocate these channels to the transmitting nodes. Under this assumption, all the wireless nodes can transmit information simultaneously to the common destination D without causing interference with each other. We note that even though this orthogonality constraint incurs throughput loss, it can significantly simplify signal processing complexity at the destination D as the multi-user detection problem is reduced to a number of independent parallel reception problems.

We will further assume all wireless transmissions are perfectly synchronous through some central authority, i.e, all the nodes start and stop to transmit simultaneously. Exactly how this syn-

chronization is achieved, and the effects of small synchronization errors on performance, however, is beyond the scope of the chapter.

As illustrated in Fig.17, the active wireless nodes (the nodes that are sending information to the destination) are clustered into different cooperation groups by their geographic locations at a particular time. The cooperation transmission is done within each cooperation group and we only consider a particular cooperation group in the sequel. Within a group, say $\mathcal{V} = \{V_i\}_{i=1}^{N}$, the wireless nodes V_i's are physically close to each other and the destination node is relatively far away from the group.

The channels from each node V_i to the destination D is modeled as Rayleigh fading and the fading coefficient α_i is assumed to be fixed for a sufficiently long period of time so that no temporal diversity can be exploited. As the the group of wireless nodes are collocated and the destination node is relatively far away, we will assume the average fading coefficients α_i's have the same average magnitude, which is determined by the path loss from the transmitting nodes to the destination D. As the transmitting wireless nodes are geographically separated, we further assume the fading channels from different wireless nodes to the destination are independent. Thus, the fading coefficients α_i's from wireless nodes V_i's to the destination node D are identically distributed independent (i.i.d.) Rayleigh random variables. On the other hand, the channels from any transmitting wireless node V_i to a node V_j within the group $\mathcal{V}$ are also modeled as Rayleigh fading channels with fading coefficient $\beta_{i,j}$. To simplify our analysis, we further assume the relative distance among those wireless nodes are more or less the same. Under this assumption, $\beta_{i,j}$'s are also i.i.d. Rayleigh random variables with the same average magnitude determined by the path loss among them.

The cooperation scheme consists of two phases of transmission, where each phase corresponds to a different time slot.

4.2.2.1 Phase-I Transmission In the first time slot, each of the wireless nodes V_i broadcasts information to the destination D and it's neighboring nodes. Assuming the modulation schemes is

Binary Phase Shift Keying(BPSK), the received signal from node V_i at the destination node D in the first time slot is

$$r_{1,i} = \alpha_i \sqrt{E_{c1}}(2v_i - 1) + n,$$

where E_{c1} is the transmitted power in Phase-I and n is the additive white Gaussian noise with power spectral density N_0. The ratio between the average magnitude square of the fading coefficients α_i and N_0 is defined as γ, i.e.,

$$\gamma = \mathbb{E}[\frac{\alpha_i^2}{N_0}],$$

and thus $E_{c1}\gamma$ is the SNR from the transmitting nodes to the destination D in phase-I.

Due to the broadcast nature of wireless transmission, the information sent out at node V_i will be overheard by node V_j ($j \neq i$) in the first time slot. The received signal at node V_j in phase-I is then

$$y_i = \beta_{i,j} \sqrt{E_{c1}}(2v_i - 1) + n.$$

Here without loss of generality, we assume power spectral density of the noise is also N_0.[1] The ratio between the average magnitude square of the fading coefficient $\beta_{i,j}$ and N_0 is defined as γ_0, i.e.,

$$\gamma_0 = \mathbb{E}[\frac{\beta_{i,j}^2}{N_0}]$$

and $E_{c1}\gamma_0$ is the SNR for transmissions within the group $\mathcal{V}$, which we will name as inter-user channel SNR in the sequel.

4.2.2.2 Phase-II Transmission Any node V_i will receive a number of messages from different transmitting nodes in the first phase. It will then decode a *select* subset $\mathcal{N}_i$ of the received messages and form a parity check sum

$$u_i = \bigoplus_{j \in \mathcal{N}_i} \hat{v}_j, \tag{4.1}$$

where $\bigoplus$ is a modulo-2 operation and $\hat{v}_j$ is the detected bit transmitted from node V_j . To distinguish the first and second phase transmission, we name node V_i as U_i in phase-II.

[1] This is generally not true as the noise variance at the cooperating node and destination node may not be the same.

During the second time slot, the parity check bit u_i is transmitted again from node U_i towards the final destination D. The received signal at the destination in phase-II is then

$$r_{2,i} = \alpha_i \sqrt{E_{c2}}(2u_i - 1) + n, \tag{4.2}$$

where E_{c2} is the transmitted power in the second phase and the SNR at the destination node in phase-II is then $E_{c2}\gamma$.

4.3 CODED USER COOPERATION

4.3.1 Coding Perspective of the Cooperative Transmission Scheme

The protocol described in Section 4.2.2 defines a coded user cooperation scheme. In phase-I, N wireless nodes send out independent information through orthogonal channels. The information transmitted in those N nodes in phase-I can be represented in a vector as

$$\mathbf{v} = [v_1 \; v_2 \; ... \; v_N]. \tag{4.3}$$

The transmitted messages in phase-II at each node as described in Section 4.2.2 are the parity check bits. We can also represent them in a vector as

$$\mathbf{u} = [u_1 \; u_2 \; ... \; u_N]. \tag{4.4}$$

The messages in phase-II $\mathbf{u}$ are redundant information of $\mathbf{v}$ in phase-I. The vector $\mathbf{v}$ and $\mathbf{u}$ can be regraded respectively as the information bits and the parity check bits of an encoder. Thus, the encoder is *systematic*. An illustrative example of the coded cooperation scheme is shown in Fig.18. Each u_j is the parity check bit of some v_j's as shown by the bipartite graph in the figure. For example, the information $u_1 = v_2 \oplus v_3$, assuming the v_2 and v_3 are perfectly received in phase-I, i.e., the inter-user channel is error free.

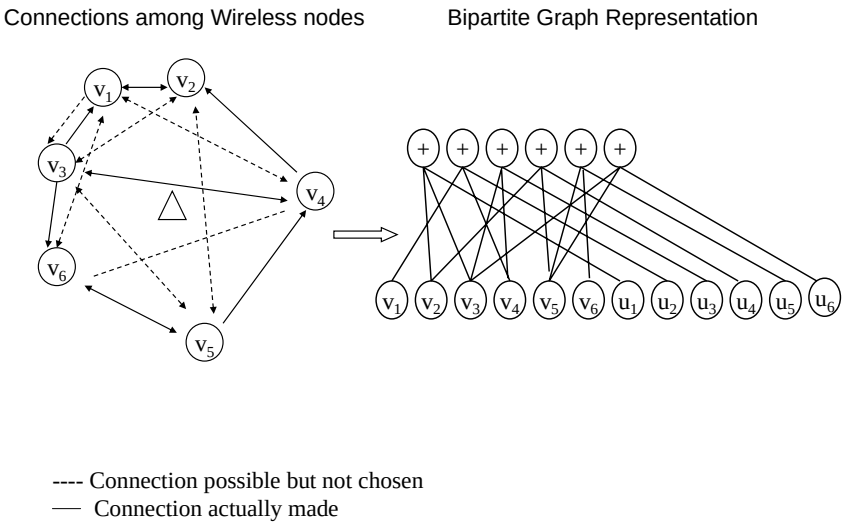

Figure 18: Coded user cooperation scheme and the corresponding bipartite graph representation

We can also write the relationship between the two vectors $\mathbf{v}$ and $\mathbf{u}$ using the generator matrix representation as follows

$$\mathbf{u} = \mathbf{v}\mathbf{G}, \tag{4.5}$$

where $\mathbf{G}$ is an $N \times N$ generator matrix with elements as either 0 or 1. $\mathbf{G}$ is determined by the connections of the cooperation scheme. For the cooperative scheme presented in Fig.18, $\mathbf{G}$ is

$$\mathbf{G} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

4.3.2 Random Cooperation vs. Selection Cooperation

One drawback in [61] is that the inter-user channel was assumed to be perfect, which can not be the case in real communications system, especially when the information bits transmitted in the first phase is un-coded. In this chapter, we consider imperfect inter-user channel condition and try to analyze the performance of the system incorporating the inter-user channel error.

The probability of making an erroneous decision using coherent detection on the transmitted message V_i at a receiving node V_j is

$$p(\hat{v_i} \neq v_i) = Q\left(\sqrt{\frac{2E_{c1}\beta_{i,j}^2}{N_0}}\right), \tag{4.6}$$

We note that one underlining assumption made here is that the receiving node can perfectly estimate the channel such that the coherent detection is possible. Throughout the chapter, we assume the channel state information is perfectly known to the receiver. In practice, however, the transmitter needs to send a pilot signal in order for the receiver to estimate the channel.

We consider two different cooperating schemes in the chapter, namely random cooperation and selection cooperation. In the random cooperation scheme, each cooperating node randomly (uniformly) selects a number of received messages to decode and compute the parity check bit. As the transmission within the cooperative group is uncoded, the message is prone to error due to the channel impairment. This will possibly make the cooperation scheme not performing well as a result of higher inter-user channel errors.

To improve the inter-user channel reliability, we consider another cooperation scheme in the chapter - selection cooperation. In the selection cooperation scheme, each cooperating node selects to decode a number of neighboring nodes' information based on the channel quality. The cooperating node will select to decode d number from all received messages with best channel qualities and compute the parity check bit. The selection cooperation makes use of the multi-user diversity and lowers the inter-user channel error as compared to the random selection scheme.

If each of the cooperating nodes chooses d of it's neighboring nodes' information to form a parity check sum, then according to (4.1), the transmitted bit u_i at node U_i in the second phase will be $\left(\bigoplus_{j=1}^{d} v_j\right) \oplus e$, where $e = 0$ with probability $1 - p_e$ and $e = 1$ with probability p_e.

The inter-user channel error probability p_e is different for different cooperation scheme. For a random selection scheme, probabilistically each of the links has equal channel condition, thus p_e can be computed as

$$p_e = \sum_{k=1, k \text{ is odd}}^{d} \left(\begin{array}{c} d \\ k \end{array} \right) \bar{p}^k (1-\bar{p})^{d-k} = \frac{1-(1-2\bar{p})^d}{2}. \tag{4.7}$$

$\bar{p}$ is the average error probability for a single link, i.e.,

$$\bar{p} = \frac{1}{2}(1 - \sqrt{\frac{E_{c1}\gamma}{E_{c1}\gamma + 1}}),$$

by averaging the instaneous error probability (4.6) over the fading coefficient.

The inter-user channel error p_e for selection cooperation scheme does not have an as easy closed form expression as (4.7) due to the fact that channel condition for each link is not the same. Numerically, p_e can be obtained by considering that there are odd number of errors from total d best channel qualities links.

4.3.3 Decoding Algorithm

Described in Section 4.2.2, the coded cooperation is equivalent to a conventional channel coding problem. However, unlike the conventional channel coding, the encoding is done distributively in each and every cooperating node within the cooperative group. The distributive manner of the encoding process introduces errors in the encoding part, which does not appear at all in conventional channel coding. To facilitate the analysis, however, we will view the error introduced in the encoding part as due to corruptions in a "virtual" channel.

We will treat the the transmitted messages in the first phase as information bits, the input to a systematic encoder. The output of the encoder has two parts. The first output is the information

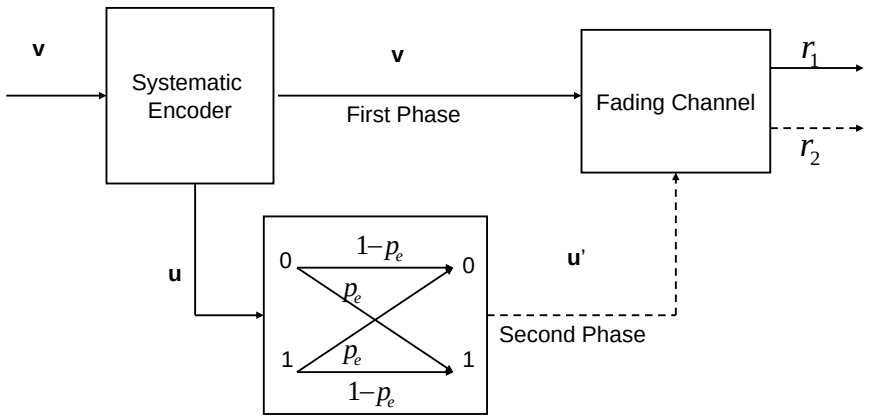

Figure 19: View of coded user cooperation scheme as channel coding

bits themselves, which will be sent out over the fading channel to the destination node D in the first phase. The second output is the parity check bits. The parity check bits can be regarded as first being sent out in the second phase to a binary symmetric channel (BSC) channel with crossover probability p_e defined in (4.7) and then sent over the air through the wireless channel to the destination D. This view of the coded user cooperation scheme is shown in Fig.19.

Based upon the argument given above, we have a coded codeword

$$\mathbf{c} = [\mathbf{v}; \mathbf{u}],$$

to be transmitted over the fading channel, where $\mathbf{v}$ and $\mathbf{u}$ are respectively the information bits and the parity check bits of the code. In conventional channel coding, all the coded bits are passed through a channel and received at the receiver side. However, in the case of our coded cooperation

scheme, the parity check bits are first passed through a BSC channel with crossover probability p_e. Thus, the actual transmitted codeword over the fading channel is

$$\mathbf{c}' = [\mathbf{v}; \mathbf{u}'],$$

where each element u_i in $\mathbf{u}'$ is $u_i' = u_i \oplus e$ and e a binary error introduced by the "virtual" BSC channel.

At the receiving side, the access node receives the information bits in the first phase and the parity check bits in the second phase. Since the cooperation scheme described essentially forms a channel code, conventional channel decoding can also be applied at the destination node to realize the coded cooperation gain. As the number of cooperating nodes is large and consequently the block length of the code is large, we focus on practical sum-product decoding algorithms in the chapter rather than on the exponential complexity maximum likelihood decoding algorithms. In order for the sum-product algorithm to work, however, one requirement is that the parity check matrix needs to be sparse. As the cooperative code is essentially a systematic code, this in turn requires that the generator matrix $\mathbf{G}$ needs to be sparse. In the chapter, we restrict in our cooperative scheme the column weight of the generator matrix d to be much smaller than the number of cooperating nodes N. Under this assumption, the cooperative coding scheme is naturally a Low Density Generator Matrix (LDGM) code.

A sum-product message passing algorithm can be used to decode the LDGM and LDPC code. It is an iterative decoding algorithm. The sum-product algorithm takes the likelihood of the received signal conditioned on transmitted bit as input and output the estimation of the transmitted codeword. The details of implementation of the algorithm are shown in [62].

Unlike conventional LDGM code, in the coded user cooperation scheme the parity check bits go through two serially concatenated channel (BSC channel + Rayleigh fading channel). To incorporate inter-user channel error in the decoding of the LDGM code, we need to modify the sum-product algorithm. The modification is only done in the initialization part of the algorithm,

i.e., computation of the likelihood function of the received signal given the transmitted bit u_i. If Log Likelihood Decoding Algorithm is used, the initialization part for the parity check bit is then

$$\begin{aligned} L_c(u_i) &= \log \frac{p(r_{2,i}|u_i=0)}{p(r_{2,i}|u_i=1)} \\ &= \log \frac{p_e p(r_{2i}|e=1,u_i=1)+(1-p_e)p(r_{2,i}|e=0,u_i=1)}{p_e p(r_{2,i}|e=1,u_i=0)+(1-p_e)p(r_{2,i}|e=0,u_i=0)} \end{aligned} \tag{4.8}$$

As for the information bits only go through one channel (Rayleigh fading channel), the initialization part for the parity check bit is the same as the conventional LDGM code.

4.4 PERFORMANCE ANALYSIS

Even though it was shown by simulation in [61] that a huge gain of coded cooperation over conventional repetition based cooperation scheme can be achieved, no analysis was given. In this section, we will analyze the performance of the coded user cooperation scheme using a number of performance bounds.

The first bounding technique we make use of is Gallager's 1968 error exponent bound [63]. Consider a discrete memoryless channel with K inputs and J outputs and a transaction probability P_{jk}, the block error probability can be bounded as

$$p_b \leq \exp\left[-n\left(-\rho R + E_0(\rho, p)\right)\right], \tag{4.9}$$

where R is the information rate in bits if the logarithmic is base 2, and

$$E_0(\rho, \mathbf{p}) = -\log \sum_{j=1}^{J} \left(\sum_{k=1}^{K} p_k P_{jk}^{1/(1+\rho)} \right)^{1+\rho}, \tag{4.10}$$

where ρ is an arbitrary number between 0 and 1 and $\mathbf{p} = [p_1, p_2, ..., p_k]$ is the probability vector of the input alphabet. Both ρ and $\mathbf{p}$ need to be optimized to get the tightest bound.

In the case when the output is unquantized ($J \to \infty$), the first summation becomes an integral, consequently

$$E_0(\rho, \mathbf{p}) = -\log \int_{-\infty}^{\infty} dy \left(\sum_{k=1}^{K} p_k P_{yk}^{1/(1+\rho)} \right)^{1+\rho} \tag{4.11}$$

As discussed in Section 4.2, the coded user cooperation scheme under the considered cooperation protocol can be regarded as a channel code. Therefore, Gallager's error exponent can be utilized to analyze the performance with a modification.

However, the error exponent bound assumes random code rather than the code ensemble under the cooperative protocol constraint. Since the minimum distance of the code ensemble under the cooperative protocol differs significantly from that of a random code, the error exponent bound does not reflect the actual network code and is not tight in general. In the next, we will try to develop a bound for the specific coded cooperation scheme.

4.4.1 Error Performance Analysis Based on the Actual Network Code Setup

In the protocol described in Section 4.2.2, each cooperating nodes choose exactly d different users' message (possible message from itself as well) to form a parity check bit and transmit it in the second phase. Under the assumption that the average receiving SNR at each node from all nodes within the cooperative group is the same, we can further assume that each cooperating node will choose d of it's neighbor's messages uniformly to form a parity check bit regardless of the cooperative scheme. The $N \times N$ generator matrix $\mathbf{G}$ in such a case will have the following properties:

- The column weight of matrix $\mathbf{G}$ is exactly d.
- The location of 1's in each column of the generator $\mathbf{G}$ is uniformly distributed.

4.4.2 Distance Spectrum Analysis

The performance a channel code under maximum likelihood decoding is uniquely determined by its' weight distribution. The weight distribution of a particular code is however hard to obtain as the number of codewords grows exponentially with the block length. Thus, to analyze the performance of the coded cooperation scheme, we will derive the weight distribution of the code ensemble defined by the properties of the generator matrix $\mathbf{G}$.

In the protocol, the number of cooperating nodes is exactly equal to the number of transmitting nodes. However, we will relax this requirement a little bit to derive a more general weight distribution of the ensemble. We assume in the first phase, n_i independent nodes are transmitting messages to D, and in the second phase n_p nodes will participate in the cooperation. This relaxation translates to a generator matrix $\mathbf{G}$ with dimensionality $n_i \times n_p$.

To obtain the weight distribution of the ensemble, i.e., the probability of the number of codewords with weight w, we want to get the probability of the parity check weight being m given a k information bit weight first. Shown in (4.5), the parity check vector $\mathbf{u}$ is $\mathbf{vG}$. Defined by the protocol, there is no constraint on the weight in each row of $\mathbf{G}$ and probabilistically each bit in $\mathbf{u}$ being 1 is identical. Thus, we only consider the bit u_1 WLOG. u_1 can be expressed as $u_1 = \bigoplus_{i=1}^{n_i} v_i \mathbf{G}_{i,1} = \bigoplus_{i \in \mathcal{N}_1} v_i \mathbf{G}_{i,1}$, where $\mathcal{N}_1 := \{i : v_i = 1, \mathbf{G}_{i,1} = 1\}$. The cardinality of $\mathcal{N}_1$, $|\mathcal{N}_1|$, is the number of locations where both vectors v and $\mathbf{G}_1$ are 1, i.e., the number of overlaps in the location of 1's of the two vectors. Given k 1's in the information vector $\mathbf{v}$ and d 1's in $\mathbf{G}_1$, the probability that $|\mathcal{N}_1| = j$ is a hypergeometric distribution. The probability of $\mathcal{N}_1 = j$ is then

$$p_{j|k} = \frac{\begin{pmatrix} d \\ j \end{pmatrix} \begin{pmatrix} n_i - d \\ k - j \end{pmatrix}}{\begin{pmatrix} n_i \\ k \end{pmatrix}},$$

where $\max(0, k + d - n_i) \leq j \leq \min(d, k)$.

According to modulo-2 operation, the parity check bit is 1 when the cardinality of $\mathcal{N}_1$ is odd. Thus, the probability of $u_1 = 1$ given the weight k of the input information vector $\mathbf{v}$ is

$$p = \Pr[u_1 = 1] = \sum_{j=\max(0,k+d-n_i),j \text{ is odd}}^{\min(d,k)} p_{j|k}$$

The output parity check weight m is then a binomial distribution conditioned on the given input information weight k, and

$$p_k^m = \begin{pmatrix} n_p \\ m \end{pmatrix} p^m (1-p)^{n_p - m}$$

Therefore, the total number of codewords with input weight (information bits weight) k and output weight (parity check bits weight) m is

$$N(k,m) = \begin{pmatrix} n_i \\ k \end{pmatrix} \begin{pmatrix} n_p \\ m \end{pmatrix} p^m (1-p)^{n_p - m},$$

and the total number of codewords with weight w can then be readily calculated as

$$N(w) = \sum_{(k,m) \in \{(i,j): i+j=w\}} N(k,m)$$

To illustrate the weight distribution of the coded cooperation scheme, the weight distribution of the LDGM code ensemble defined by the code cooperation protocol with chosen parameters $n_i = n_p = 200$ and $d = 7$ is depicted in Fig.20.

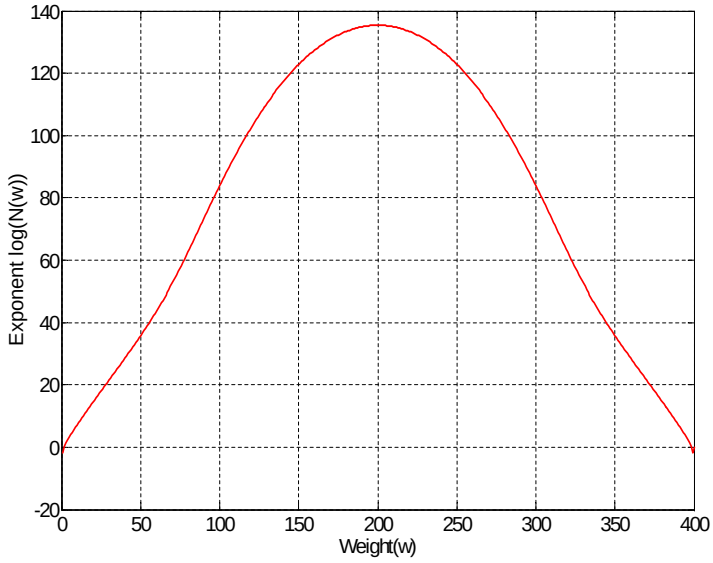

Figure 20: Weight distribution of coded user cooperation protocol defined LDGM code with $n_i = n_p = 200$ and $d = 7$

4.4.3 Average Performance of the Network Code

Based on the weight distribution derived, we are able to provide the union bound by summing up all the pair-wise errors. For the case of AWGN channel, one trivial union bound (bit error probability) can be readily derived as

$$p_b \leq \sum_{k=1}^{n_i} \sum_{m=0}^{n_p} N(k,m) \frac{k}{n_i} Q\left(\frac{k\sqrt{E_{c1}} + m\sqrt{E_{c2}}}{\sqrt{\frac{(k+m)N_0}{2}}} \right),$$

where $Q(x)$ is the tail integral of the standard normal Gaussian density function, i.e.,

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp(-\frac{x^2}{2}) dx.$$

For the case of Rayleigh fading channel, the instaneous union bound (bit error probability) can also be written as

$$p_b \leq \sum_{k=1}^{n_i} \sum_{m=0}^{n_p} N(k,m) \frac{k}{n_i} Q \left(\frac{\sum_{i=1}^{k} \alpha_i^2 \sqrt{E_{c1}} + \sum_{i'=1}^{m} \beta_{i'}^2 \sqrt{E_{c2}}}{\sqrt{\sum_{i=1}^{k} \alpha_i^2 \frac{N_0}{2} + \sum_{i'=1}^{m} \beta_{i'}^2 \frac{N_0}{2}}} \right).$$

Averaging over the fading coefficients α's and β's, we can obtain the average probability of error.

4.4.4 Performance Analysis Incorporating the Inter-user Channel Error

By the cooperative protocol, the code formed in the coded user cooperation scheme will suffer from inter-user channel error. To better evaluate the coded user cooperation scheme, we will try to derive the performance bound that incorporates the imperfect inter-user channel condition.

WLOG, we assume all zero codeword $C_0 = \{0\}^N$ is transmitted over the channel, and which is modulated to $S_0 = [-\sqrt{E_c} \cdots - \sqrt{E_c}]$ using BPSK modulation. The received channel output probability density function for AWGN channel conditioned on the transmitted signal S_0 can then be written as

$$p(r|S_0) = \prod_{i=1}^{N} p(r_i|S_{0i}) = \frac{1}{(\pi N_0)^{N/2}} \exp \left[\frac{\sum_{i=1}^{N} (r_i - S_{0i})^2}{N_0} \right].$$

The maximum likelihood detector is then

$$\begin{aligned} \hat{j} &= \arg_j \min \sum_{i=1}^{N} (r_i - S_{ji})^2 \\ &= \arg_j \max \sum_{i=1}^{N} r_i S_{ji}. \end{aligned} \tag{4.12}$$

Now we only consider a pair of codewords S_0 and S_k. Given S_0 is transmitted over the channel, the detector will choose S_k as transmitted over S_0 when

$$\sum_{i=1}^{d_{0,k}}(-\sqrt{E_c}+n_i)(-\sqrt{E_c}) < \sum_{i=1}^{d_{0,k}}(-\sqrt{E_c}+n_i)\sqrt{E_c},$$

i.e.

$$\sum_{i=1}^{d_{0,k}} n_i > d_{0,k}\sqrt{E_c},$$

where $d_{0,k}$ is the hamming distance between the two codewords S_0 and S_k. This probability can be derived readily as

$$p_{0\to k} = Q(d_{0,k}\sqrt{\frac{2E_c}{N_0}})$$

Next, we consider the analysis of incorporating the inter-user channel error. Suppose S_0 was transmitted over the channel and due to the inter-user channel error totally l bits got flipped. Let us further assume that among those l bits j of them are in the same positions where the codeword S_k also has 1.

In such a case, given S_0 is transmitted over the channel, the detector will choose S_k as transmitted over S_0 when

$$\sum_{i=1}^{d_{0,k}-j}(-\sqrt{E_c}+n_i)(-\sqrt{E_c}) + \sum_{i=d_{0,k}-j+1}^{d_{0,k}}(\sqrt{E_c}+n_i)(-\sqrt{E_c}) <$$
$$\sum_{i=1}^{d_{0,k}-j}(\sqrt{E_c}+n_i)(-\sqrt{E_c}) + \sum_{i=d_{0,k}-j+1}^{d_{0,k}}(\sqrt{E_c}+n_i)(\sqrt{E_c}), \tag{4.13}$$

i.e.

$$\sum_{i=1}^{d_{0,k}} n_i > (d_{0,k}-2j)\sqrt{E_c},$$

This probability can also be derived readily as

$$p_{0\to k} = Q\left((d_{0,k}-2j)\sqrt{\frac{2E_c}{N_0}}\right).$$

In the coded cooperation scheme, only the parity check bits suffer from the inter-user channel error, thus we need to differentiate the parity check bits and information bits and deal with them

separately. However, the probability of decision error can be derived in a similar manner as previous cases. The probability of choosing $S_{(k,m)}$ (a codeword with k information bits and m parity check bits) over the transmitted signal S_0 can be written as

$$p_{0\to k} = Q\left(\frac{k\sqrt{E_{c1}} + (m-2j)\sqrt{E_{c2}}}{\sqrt{\frac{(k+m)N_0}{2}}}\right)$$

Therefore, an upper bound on the probability of bit error can be derived as summing up all the pairwise probabilities, i.e.,

$$\begin{aligned} p_b \leq \quad & \sum_{k=1}^{n_i}\sum_{m=0}^{n_p} N(k,m)\frac{k}{n_i}\sum_{l=0}^{n_p}\sum_{j=\max(0,m+l-n_p)}^{\min(m,l)} \binom{m}{j}\binom{n_p-m}{l-j} \cdot \\ & p_e^l(1-p_e)^{n_p-l}Q\left(\frac{k\sqrt{E_{c1}} + (m-2j)\sqrt{E_{c2}}}{\sqrt{\frac{(k+m)N_0}{2}}}\right) \end{aligned} \tag{4.14}$$

For the case Rayleigh fading channel, the instaneous probability of error is thus

$$\begin{aligned} p_b \leq \quad & \sum_{k=1}^{n_i}\sum_{m=0}^{n_p} N(k,m)\frac{k}{n_i}\sum_{l=0}^{n_p}\sum_{j=\max(0,m+l-n_p)}^{\min(m,l)} \binom{m}{j}\binom{n_p-m}{l-j} \cdot \\ & p_e^l(1-p_e)^{n_p-l}Q\left(\frac{\sum_{i=1}^{k}\alpha_i^2\sqrt{E_{c1}} + \sum_{i'=1}^{m-2j}\beta_{i'}^2\sqrt{E_{c2}}}{\sqrt{\sum_{i=1}^{k}\alpha_i^2\frac{N_0}{2} + \sum_{i'=1}^{m}\beta_{i'}^2\frac{N_0}{2}}},\right) \end{aligned} \tag{4.15}$$

and the average error probability can be obtained by averaging out all the fading coefficients.

4.5 DENSITY EVOLUTION ANALYSIS OF CODED USER COOPERATION SCHEME

4.5.1 Density Evolution

Density evolution is one of the most powerful analytical technique which has been used to understand limits of performance of Low Density Parity Check (LDPC) code under the practical sum-product decoding algorithm. It also provides a tool which can be used in the design of families of LDPC codes, since their performance can be predicted using density evolution much more rapidly than the performance can be simulated.

In the sum-product message-passing algorithm, messages as the Log-Likelihood ratio were sent back and forth along the edges of bipartite graph. Variable nodes and check nodes receive messages from each edge that they are connected to, calculate them, and then send new messages to each edge. Consequently, messages are iteratively refreshed.

Since the code is linear and the channel is output symmetric, it suffices to assume that the all-zero codeword is sent. Let $\nu^{[l]}$ be the LLR message from a degree d_v variable nodes to a check node in the l-th iteration, and $\mu^{[l]}$ be the LLR message from a degree d_c check node to a variable node in the l-th iteration. For a regular LDPC code with degree (d_v, d_c), the messages transmitted from the variable node to the check node at iteration l is

$$\nu^{[l]} = \mu_0 + \sum_{i=1}^{d_v-1} \mu_i^{[l-1]}$$

The message transmitted from the check node to the variable node at iteration l is

$$\tanh \frac{\mu^{[l]}}{2} = \prod_{j=1}^{d_c-1} \tanh \frac{\nu_j^{[l]}}{2}$$

According to these two update equations and the initial probability density function μ_0 we can update the PDF of μ and ν iteratively until it goes to a delta function located at plus infinity or it converges to a probability density function that has certain fractions below zero.

4.5.2 Density Evolution for Coded User Cooperation Scheme

To analyze the performance of the coded user cooperation scheme using density evolution, we need to modify the density evolution analysis for regular LDPC code according to the cooperative protocol. Described by the coded user cooperation scheme, the generator matrix described in Section 4.4.2 has column weight d. Thus, the check node degree is exactly $d + 1$.

To get the variable node degree distribution, we need to treat the information bit variable nodes and parity check bit variable nodes separately. As shown in Fig.18, only one edge is associated with each parity check variable node, thus the degree of the parity check variable node is fixed to be 1. On the other hand, for the information bit variable node, the degree is a Poisson distributed random variable with average weight d due to the fact that each column of $\mathbf{G}$ has weight d and the location of 1's is uniformly distributed. The probability that a particular information bit variable node has degree i is thus

$$\chi_i = \frac{e^{-d} d^i}{i!}.$$

The distribution of the variable node weights can thus be represented as

$$\chi(\nu) = \sum_{i=1}^{\infty} \rho_i \nu^{i-1}, \tag{4.16}$$

where

$$\rho_i = \frac{i\chi_i}{\sum_{k=1}^{\infty} k\chi_k}.$$

The variable node weight distribution is plotted in Fig.21. Even though the information bit variable node degree can go up to infinity, noted in Fig.21, the probability that an information bit variable node has degree i drops to 0 very fast as i grows. For our practical purpose, we can truncate the weight greater than 30 in the sum in (4.16) for $d < 10$.

Once the degree distribution is obtained, we can modify the density evolution analysis for regular LDPC code. For the proposed coded user cooperation scheme, the messages transmitted from the variable node (information bit part) to the check node at iteration l is

$$\nu^{[l]} = \mu_0 + \sum_{k=1}^{30} \left(\rho_k \sum_{i=1}^{k-1} \mu_i^{[l-1]} \right),$$

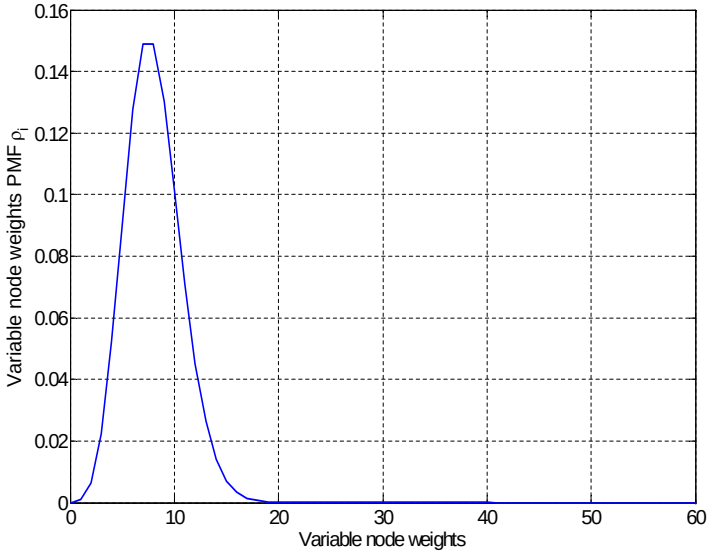

Figure 21: Probability Mass Distribution of Variable Node Weights

where $\mu^{[0]}$ is the initial LLR for the information bits from the channel. In the Gaussian channel, the initial likelihood ratio is

$$\mu_0 = \frac{2\sqrt{E_c}}{\sigma^2}(\sqrt{E_c} + n),$$

which is Gaussian distributed with mean and variance as

$$m = E[\mu_0] = \frac{2E_{c1}}{\sigma^2}$$

$$\mathrm{var}(\mu_0) = \frac{4E_{c1}}{\sigma^2}$$

The message transmitted from the check node to the variable node at iteration l is

$$\tanh\frac{\mu^{[l]}}{2} = \tanh\frac{\nu_0}{2}\prod_{j=1}^{d-1}\tanh\frac{\nu_j^{[l]}}{2}$$

where ν_0 is the initial LLR for the parity check bits from the channel. As there is only one edge connected to the parity check bits variable node, thus the information from them to the check node

stays the same as the iterations go on. As the parity check bits will first go through a BSC channel, thus the initial likelihood ratio, for the case of Gaussian channel, is

$$\nu_0 = (1-p_e)\frac{2\sqrt{E_{c2}}}{\sigma^2}(\sqrt{E_{c2}}+n) + p_e\frac{2\sqrt{E_{c2}}}{\sigma^2}(-\sqrt{E_{c2}}+n),$$

Given the two input PDF for μ_0 and ν_0, we can update the densities $\mu^{[l]}$ and $\nu^{[l]}$ in each iteration. Exactly how the density can be computed iteratively from these two initial densities is shown in [64] using a discrete density evolution method.

Density evolution can predict the water-fall region of the LDPC code. It introduces the idea of a channel threshold, above which the code performs well and below which the probability of error is non-negligible. This provides a single parameter characterizing code performance which may be used to gauge the performance compared to the ultimate limit of the channel capacity.

Unlike the LDPC code, due to the minimum distance of the code formed in the coded user cooperation scheme does not scale with the block length, the code will exhibit an error floor at relative high SNR. Thus, no single vaue parameter - the threshold can be determined from the density evolution. However, we can still make use of the density evolution as an powerful analytical tool to predict the performance of the coded cooperation scheme. .

4.6 SIMULATION RESULTS

4.6.1 Performance Bounds vs. Simulation Results

In Fig.22, the simulation result of LDGM code is shown and compared with the union bound and density evolution analysis. The selected parameter for the LDGM code is $N = 1000$ and $d = 7$. The simulation result shows that for the same bit error rate both the union bound and the density evolution analysis result are within 1 dB difference from the simulation result. This indicates the effectiveness of both analytical tools in predicting the performance of the LDGM code and then the coded user cooperation scheme.

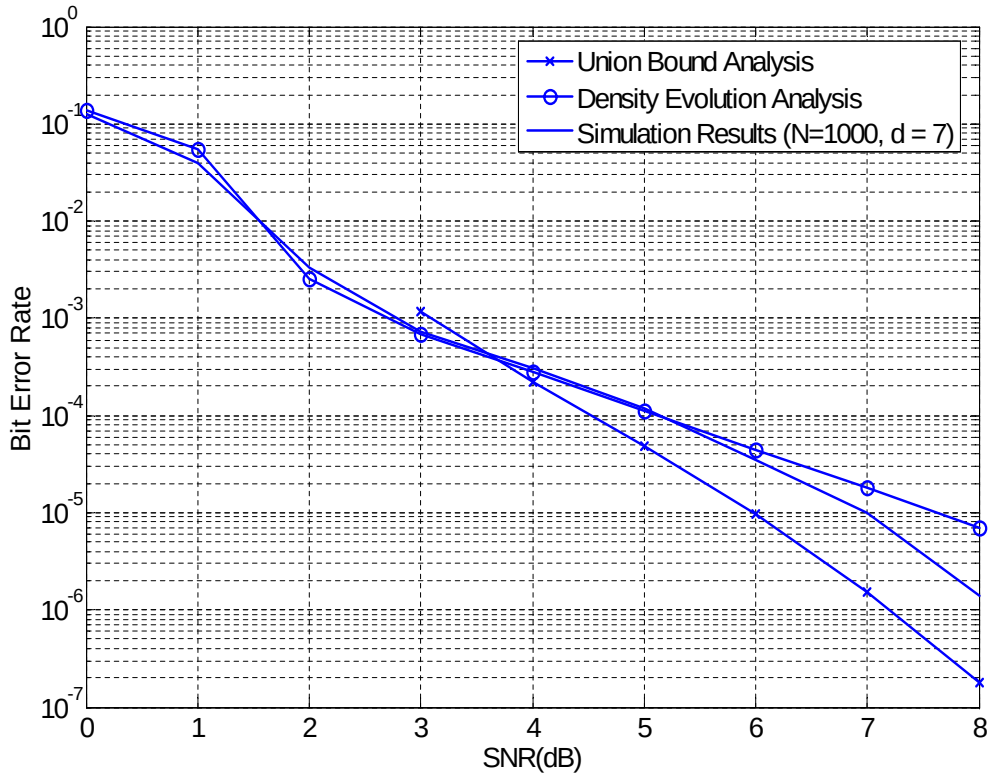

Figure 22: Comparison of simulation results with union bound and density evolution analysis for AWGN channel

4.6.2 Random Cooperation vs. Selection Cooperation

In Fig.23, the numerical results of random cooperation scheme and selection cooperation scheme for different inter-user channel conditions are shown. We see that under the same inter-user channel condition, selection user cooperation indeed performs better than the random user cooperation scheme. For 25 dB inter-user channel SNR, the gain of selection cooperation scheme is around 2dB at the bit error rate of 10^{-5}. For 15 dB inter-user channel SNR, the gain of selection cooperation scheme is greater than 6dB at the bit error rate of 10^{-5}.

4.6.3 Coded User Cooperation Scheme v.s. Non-cooperative Scheme

The performance of coded cooperation scheme versus non-cooperative scheme is plotted in Fig.24. In the coded user cooperation scheme, the number of wireless nodes in the cooperating group is $N = 200$, and each node select (based on the estimated channel quality) $d = 7$ users' messages to form a parity check sum and transmit it in phase-II. From the simulation results, we see that even under our realistic inter-user channel assumption, the performance of coded user cooperation scheme still have a huge gain over non-cooperative scheme. For example, at bit error rate 10^{-4}, the gain of coded user cooperation over non-cooperation scheme is over 20 dB. The performance bound for coded user cooperation scheme is also plotted in the figure. The bound even though is around 5 dB off the simulation result, it has the same trend with the simulation curve and can be unutilized in practice to predict the performance of the coded user cooperation scheme quickly.

4.7 CONCLUSION

In this chapter, we considered the problem of coded user cooperation in the wireless networks. A similar coded user cooperation scheme as in [61] was considered to realize both the diversity and coding gain. Differerent from [61], the inter-user channel was modeled to be noisy rather than

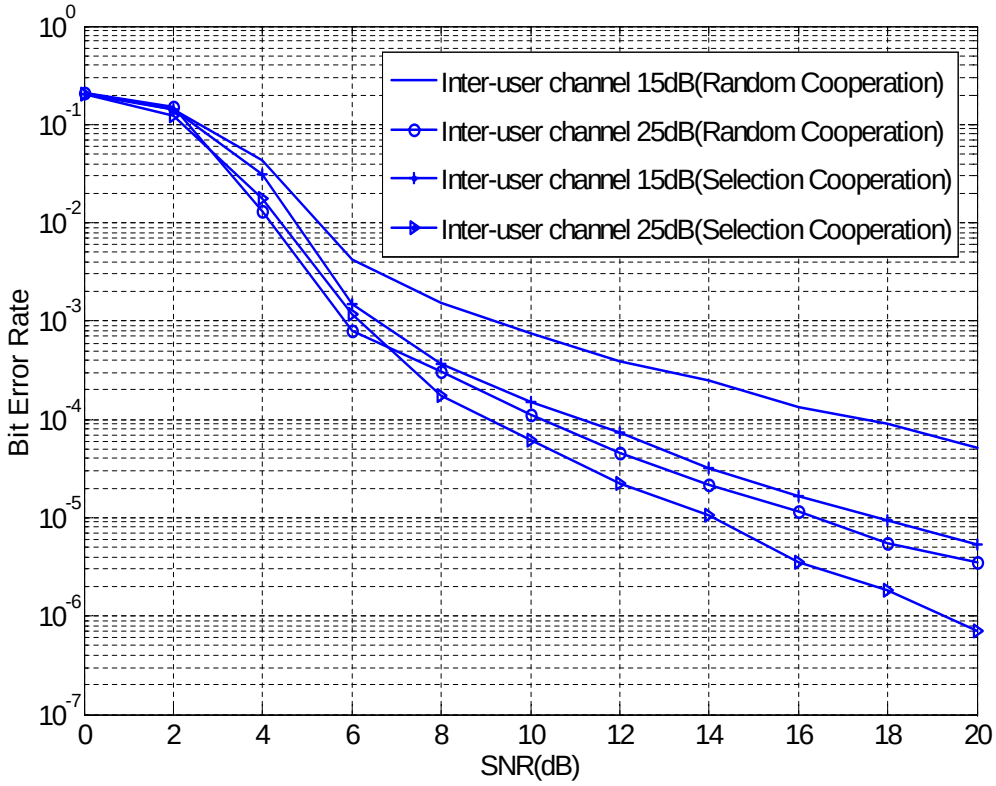

Figure 23: Random cooperation scheme vs. selection cooperation scheme

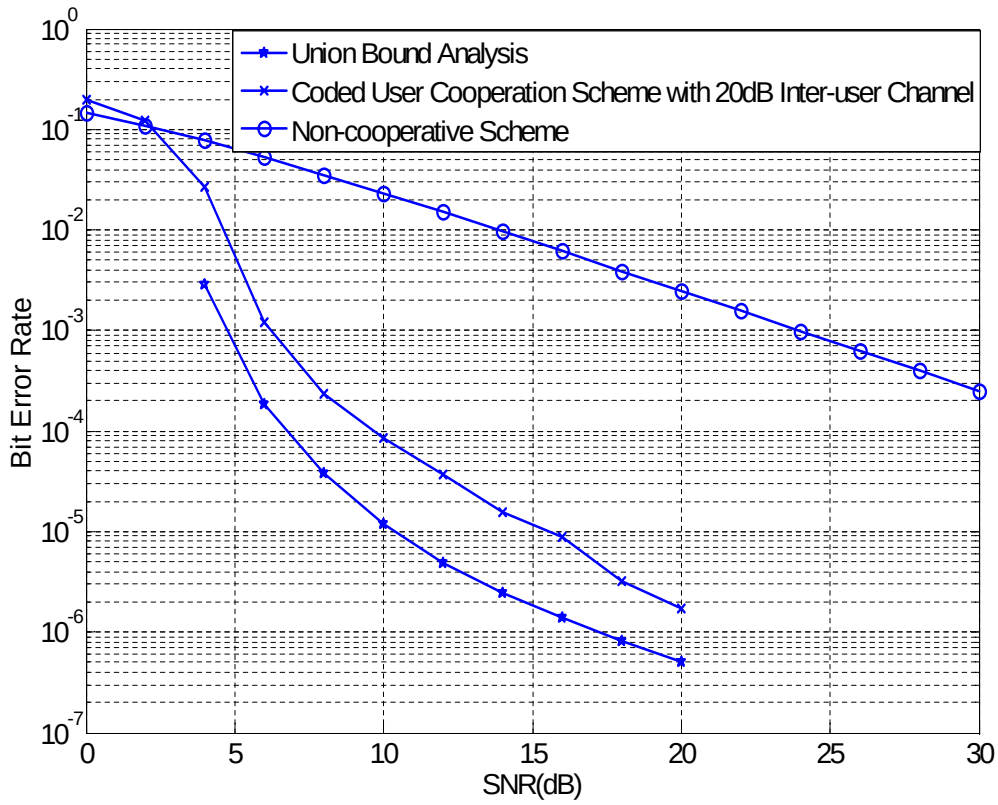

Figure 24: Comparison of coded user cooperation scheme with non-cooperative scheme for $N = 200$ and $d = 7$

perfect. To incorporate the practical inter-user channel condition, the conventional sum-product decoding algorithm was modified for better decoding.

Two different cooperation schemes were considered in the chapter, namely random cooperation and selection cooperation. By doing selection cooperation, the multiuser diversity can be employed to lower the inter-user channel error. A number of analytical tools (union bounds and density evolution analysis) were also developed to analyze the performance the coded user cooperation scheme. Simulation results indicate great performance advantage of this coded user cooperation scheme over non-cooperative scheme even in the case the inter-user channel is not perfect.

5.0 FUTURE DIRECTIONS

In this concluding chapter, we discuss future research directions in light of the results derived in this dissertation.

One direction to extend the work in Chapter 2 is to design the signal constellations that can optimize the performance under the Generalize Likelihood Ratio Test (GLRT) detection rule. This study of signal design under GLRT detector is of practical interests for its' simplicity.

Another possible direction is to come up with a design rule using the error exponent analysis in non-coherent MIMO system. While the design rule applied in the dissertation is layered capacity, the error exponent design rule would be extremely appealing to design a system with a targeted rate lower than the capacity. The computational complexity analysis of the various detectors can also be a direction of further research.

In Chapter 3, the feedback codebook design was considered primarily in a single user Multiple Input and Multiple Output (MIMO) antennas system. However, there are a rich set of problems in a multi-user MIMO setting, e.g. MIMO multi-access or MIMO broadcasting, to which the codebook design methodology developed in this dissertation can be generalized.

Finally, an important future direction is to utilize the analytical tools provided in Chapter 4 in the design of better coded user cooperation scheme.

APPENDIX A

PROOF OF LEMMA 2

We first show the following Lemma to prove Lemma 2.

Lemma 4 *If both* $\mathbf{A}$ *and* $\mathbf{B}$ *are* $i \times j$ *and* $j \times k$ *i.d. unitary matrices respectively, then their product* $\mathbf{C} = \mathbf{AB}$ *is also i.d.* $i \times k$ *unitary matrix.*

Proof For any $i \times i$ unitary matrix $\mathbf{U}$, we have

$$p(\mathbf{UC}) = p(\mathbf{UAB}) = p(\mathbf{AB}) = p(\mathbf{C}),$$

where the second equality follows from the definition of i.d. unitary matrix, i.e., $p(\mathbf{A}) = p(\mathbf{UA})$ for any deterministic matrix $\mathbf{U}$.

By this Lemma, Lemma 2 can be readily proved using induction.

APPENDIX B

PROOF OF THEOREM 3

For all $g_i \in \mathcal{E}$ we have $g_i < \mu$, i.e., for all $i \leq m^*$, $|\ \frac{g_i - \bar{g}}{\mu}\ | < 1$. Applying the Taylor series expansion to (3.6), we have:

$$-\ln(1 + \frac{g_i - \bar{g}}{\mu}) = -\sum_{i=1}^{\infty} \frac{(g_i - \bar{g})^i}{i \cdot \mu^i}. \tag{B.1}$$

In this series, the absolute value of i-th order term is greater than the summation from the $(i+1)$-th order term up (We omit the detailed proof here and which will be available upon request). By this property, both the second order and first order upperbound in Theorem 3 can be obtained trivially.

On the other hand, as all even power terms are negative, a tight lower-bound on the capacity difference can be obtained by truncating the series at an odd term (an upper bound for the Taylor series; but a lower bound on the capacity difference).

APPENDIX C

COMPUTATION OF THE GRADIENT

As the partial derivative is a linear operator, we can exchange the order of expectation and partial derivative to compute the gradient of the partial distortion $D(\mathbf{V}_k)$ with respect to Θ, i.e.,

$$\begin{aligned}\nabla_\Theta D(\mathbf{V}_k) &= \nabla_\Theta\Big\{\mathbb{E}_{\mathbf{H}\in\mathcal{R}_k}\big[C_w - C_p\big(\mathbf{H}, \mathbf{U}_m(\Theta)\big)\big]\Big\} \\ &= -\mathbb{E}_{\mathbf{H}\in\mathcal{R}_k}\Big\{\nabla_\Theta C_p(\mathbf{H}, \mathbf{U}_m(\Theta))\Big\}\end{aligned}$$

If $\Theta = \delta_k$ with $1 \le k \le M$, to compute $\nabla_\Theta C_p(\mathbf{H}, \mathbf{U}_m(\Theta))$, we rewrite the Given's parametrization (3.15) as $\mathbf{U}_m = \mathbf{U}_\Lambda \mathbf{U}_{+1}$, where

$$\mathbf{U}_{+1} = \prod_{p=1}^{M-1}\prod_{q=M}^{p+1} \mathbf{U}^{p,q}(\phi_{p,q}, \sigma_{p,q}) \begin{bmatrix} \mathbf{I}_m \\ \mathbf{0}_{(M-m)\times m} \end{bmatrix}.$$

As $C_p\big(\mathbf{H}, \mathbf{U}_m(\Theta)\big)$ depends on δ_k only through $\mathbf{U}_\Lambda$, using the results (Theorem 2 and Lemma 3) in [65], we have

$$\nabla_\Theta C_p(\mathbf{H}, \mathbf{U}_m(\Theta)) = 2\mathrm{ReTr}\Big\{\mathbf{H}^*\mathbf{H}\mathbf{U}_m\mathbf{E}\mathbf{U}_{+1}^* \times \nabla_\Theta \mathbf{U}_\Lambda^*\Big\},$$

where $\mathbf{E} = (\frac{m}{M}\mathbf{I}_m + \mathbf{U}_m^*\mathbf{H}^*\mathbf{H}\mathbf{U}_m)^{-1}$ is the MMSE matrix defined in [65] and

$$\nabla_\Theta \mathbf{U}_{\mathbf{\Lambda}}^*|_{\Theta=\delta_k} = -i\exp(-i\delta_k)\mathbf{e}_k\mathbf{e}_k^*,$$

with $\mathbf{e}_k$ being a unit norm column vector of length M with the k-th element being 1.

To compute the gradient $\nabla_\Theta C_p(\mathbf{H}, \mathbf{U}_m(\Theta))$ when $\Theta = \phi_{p,q}$ or $\Theta = \sigma_{p,q}$, we rewrite $\mathbf{U}_m$ as

$$\begin{aligned}\mathbf{U}_m &= \mathbf{U}_{\mathbf{\Lambda}} \prod_{p=1}^{M-1} \prod_{q=M}^{p+1} \mathbf{U}^{p,q}(\phi_{p,q}, \sigma_{p,q}) \begin{bmatrix} \mathbf{I}_m \\ \mathbf{0}_{(M-m)\times m} \end{bmatrix} \\ &= \underbrace{\mathbf{U}_{\mathbf{\Lambda}}\mathbf{U}^{M-1,M}(\phi_{M-1,M}, \sigma_{M-1,M}) \cdots}_{\mathbf{U}_{-1}} \underbrace{\mathbf{U}^{p,q}(\phi_{p,q}, \sigma_{p,q})}_{\mathbf{U}_0} \underbrace{\cdots \mathbf{U}^{1,M}(\phi_{1,M}, \sigma_{1,M}) \begin{bmatrix} \mathbf{I}_m \\ \mathbf{0}_{(M-m)\times m} \end{bmatrix}}_{\mathbf{U}_{+1}}\end{aligned}$$

Then, the input output relationship with this precoding marix $\mathbf{U}_m$ can be written as

$$\mathbf{r} = \mathbf{H}\mathbf{U}_{-1}\mathbf{U}_0\mathbf{U}_{+1}\mathbf{s} + \mathbf{n}$$

Again use the results in [65], we obtain

$$\nabla_\Theta C_p(\mathbf{H}, \mathbf{U}_m(\Theta)) = 2\mathrm{ReTr}\Big\{\mathbf{U}_{-1}^*\mathbf{H}^*\mathbf{H}\mathbf{U}_m\mathbf{E}\mathbf{U}_{+1}^* \times \nabla_\Theta \mathbf{U}_0^*\Big\}.$$

If $\Theta = \phi_{p,q}$,

$$\nabla_{\phi_{p,q}}\mathbf{U}_0^* = \begin{cases} -sin(\phi_{p,q}) & \text{if } j = k \text{ and } j = p, q \\ cos(\phi_{p,q})\exp(-i\sigma_{p,q}) & \text{if } j = p \text{ and } k = q \\ -cos(\phi_{p,q})\exp(i\sigma_{p,q}) & \text{if } j = q \text{ and } k = p \\ 0 & \text{otherwise.} \end{cases}$$

If $\Theta = \sigma_{p,q}$,

$$\nabla_{\sigma_{p,q}}\mathbf{U}_0^* = \begin{cases} -isin(\phi_{p,q})\exp(-i\sigma_{p,q}) & \text{if } j = p \text{ and } k = q \\ -isin(\phi_{p,q})\exp(i\sigma_{p,q}) & \text{if } j = q \text{ and } k = p \\ 0 & \text{otherwise.} \end{cases}$$

Another possible way of calculating the gradient is through the method introduced in [14]. By considering perturbation of $\mathbf{U}_m$ by $\mathbf{U}(\Theta)$ with $\Theta \approx 0$, the gradient can also be efficiently derived.

BIBLIOGRAPHY

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

[2] D. J. Love and J. Robert W. Heath, "Multimode precoding for MIMO wireless systems," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3674–3687, Oct. 2005.

[3] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans.Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov/Dec 1999.

[4] G.Foschini and M. Gans, "On limits of wireless communications in fading enviroment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–335, 1998.

[5] L. Zheng and D. N. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1073–96, 2003.

[6] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in rayleigh flat fading," *IEEE Trans. Inf. Theory*, Jan. 1999.

[7] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in rayleigh flat fading," *IEEE Trans. Inf. Theory*, Mar. 2000.

[8] M. O. Damen, K. Abed-Meraim, and J.-C. Belfiore, "Generalised sphere decoder for asymmetrical space - time communication architecture," *IEE Electronics Letters*, vol. 36, p. 166167, 2000.

[9] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[10] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, p. 1998, Mar. 1998.

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

[12] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 11, pp. 1841–1852, Nov. 1999.

[13] B. M. Hochwald, T. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Inf. Theory*, vol. 46, pp. 1962–1973, Mar. 2000.

[14] D. Agrawal, T. J. Richardson, and R. L. Urbanke, "Multiple-antenna signal constellations for fading channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2618–2626, Sep. 2001.

[15] I. Abou-Faycal and B. M. Hochwald, "Coding requirements for multiple-antenna channels with unknown rayleigh fading," *IEEE J. Sel. Areas Commun.*, 1999.

[16] Y. Jing and B. Hassibi, "Cayley codes for unitary space-time modulation," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2891–2904, Nov. 2003.

[17] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Math. Comput.*, vol. 44, pp. 463–471, Apr. 1985.

[18] M. L. McCloud, "Layered signals for use on the noncoherent MIMO block fading channel," in *Proc. 42th Annual Allerton Conference on Comm., Control and Comput.* Monticello, IL, Oct. 2004.

[19] ——, "Layered coded modulation for use on the noncoherent MIMO block fading channel," in *Proc. 39th Annual Conference on Information Sciences and Systems*. Baltimore, MD, Mar. 2005.

[20] ——, "Design and efficient detection of layered signals for use on the coded and uncoded noncoherent mimo block fading channel," *Submitted to IEEE Trans. Inf. Theory*, Jun. 2004.

[21] L. Zheng and D. N. Tse, "Communications on the grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, 2002.

[22] A. Barg and D. Y. Nogin, "Bounds on packings of spheres in the grassmann manifold," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2450–2454, 2002.

[23] B. Hughes, "Optimal space-time constellations from groups," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 401–410, 2003.

[24] ——, "Differential space-time modulation," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2567 –2578, 2000.

[25] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2041–2052, Dec. 2000.

[26] A. Shokrollahi, B. Hassibi, B. M. Hochwald, and W. Sweldens, "Representation theory for high-rate multiple-antenna code design," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2335–2367, Sep. 2001.

[27] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[28] B. Hassibi and T. Marzetta, "Multiple-antennas and isotropically random unitary inputs: the received signal density in closed form," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1473–84, Jun. 2002.

[29] H. Imai and S. Hirakawa, "Multilevel codes and multistage decoding for unequal error protection," *IEEE Trans. Inf. Theory*, vol. 23, pp. 371–377, 1977.

[30] M. Brehler and M. K. Varanasi, "Asymptotic error probability analysis of quadratic receivers in rayleigh-fading channels with applications to a unified analysis of coherent and noncoherent space-time receivers," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2383–2399, 2001.

[31] V. Tarokh and I.-M. Kim, "Existence and construction of noncoherent unitary space-time codes," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 3112 – 3117, 2002.

[32] M. L. McCloud, M. Brehler, and M. K. Varanasi, "Signal design and convolutional coding for space-time communication on the rayleigh fading channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1186–1194, 2002.

[33] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, pp. 289–293, Oct. 1958.

[34] R. L. Dobrushin, "Information transmission in a channel with feedback," *Theory Probab. Appl.*, vol. 3, pp. 395–412, Dec. 1958.

[35] W. C. Jakes, *Microwave Mobile Communications*. John Wiley & Sons, 1974.

[36] A. J. Goldsmith and P. Varaiya, "Capacity, mutual information, and coding for finite-state markov channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 868–886, May 1996.

[37] ——, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, pp. 1986–1992, Nov. 1997.

[38] G. Caire and S. S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2007–2019, Sep. 1999.

[39] H. Viswanathan, "Capacity of markov channels with receiver CSI and delayed feedback," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 761–771, Mar. 1999.

[40] A. J. Goldsmith and S.-G.Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 595–602, May 1998.

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

[42] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1423–1436, Oct. 1998.

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

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

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

[46] J. C. Roh and B. D. Rao, "Transmit beamforming in multiple-antenna systems with finite rate feedback: A VQ-based approach," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1101–1112, Mar. 2006.

[47] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, pp. 129–137, Mar. 1982.

[48] W. Yu and J. Cioffi, "On constant power water-filling," in *Proc. IEEE International Conference on Comunications*, Jun. 2001.

[49] ——, "Constant-power waterfilling: Performance bound and low-complexity implementation," *IEEE Trans. Commun*, vol. 54, pp. 23–28, Jan. 2006.

[50] R. E. Blahut, "Computation of channel capacity and rate-distortion function," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 460–473, Jul. 1972.

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

[52] V. K. N. Lau, Y. Liu, and T.-A. Chen, "Capacity of memoryless channels and block-fading channels with designable cardinality-constrained channel state feedback," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 2038–2049, Sep. 2004.

[53] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*. Kluwer Academic Press, 1992.

[54] D. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multi-element antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 502–513, Mar. 2000.

[55] E. C. V. D. Meulen, "Three-terminal communication channels," *Advances in Applied Probability*, vol. 3, no. 1, pp. 120–154, 1971.

[56] T. M. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.

[57] J.N.Laneman and G.W.Wornell, "Distributed space-time-coded protocols for exploiting cooperaative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415–2425, oct 2003.

[58] 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. 50, no. 12, pp. 3062–3080, 2004.

[59] T. Hunter and A. Nosratinia, "Diversity through coded cooperation," *IEEE Trans. Wireless Comm.*, vol. 5, pp. 283–289, Feb. 2006.

[60] T. Hunter, S. Sanayei, and A. Nosratinia, "Outage analysis of coded cooperation," *IEEE Trans. Inf. Theory*, vol. 42, pp. 375–391, 2006.

[61] X. Bao and J. Li(Tiffany), "Matching code-on-graph with network-on-graph: Adaptive network coding for wireless relay networks," in *Proc. 43rd Annual Allerton Conference on Comm., Control and Comput.*, 2005.

[62] T. K. Moon, *Error Correction Coding: Mathematical Methods and Algorithms*. John wiley & Sons, 2005.

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

[64] S.-Y. Chung, J. G. David Forney, T. J. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 db of the shannon limit," *IEEE Comm. Letter*, vol. 5, pp. 58–60, 2001.

[65] D. Palomar and S. Verdu, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141– 154, Jan. 2006.